European **Observatory** on Health Systems and Policies Series

Paying for Performance in Health Care

Implications for health system performance and accountability

Edited by
Cheryl Cashin
Y-Ling Chi
Peter Smith
Michael Borowitz
and Sarah Thomson

Paying for Performance in Health Care

Implications for health system performance and accountability



The European Observatory on Health Systems and Policies supports and promotes evidence-based health policy-making through comprehensive and rigorous analysis of health systems in Europe. It brings together a wide range of policy-makers, academics and practitioners to analyse trends in health reform, drawing on experience from across Europe to illuminate policy issues.

The European Observatory on Health Systems and Policies is a partnership between the World Health Organization Regional Office for Europe, the Governments of Austria, Belgium, Finland, Ireland, Norway, Slovenia, Sweden, the United Kingdom, and the Veneto Region of Italy, the European Commission, the World Bank, UNCAM (French National Union of Health Insurance Funds), the London School of Economics and Political Science, and the London School of Hygiene & Tropical Medicine.







Paying for Performance in Health Care

Implications for health system performance and accountability

Edited by

Cheryl Cashin, Y-Ling Chi, Peter C. Smith, Michael Borowitz and Sarah Thomson

Mc Graw Hill Education

Open University Press

Open University Press
McGraw-Hill Education
McGraw-Hill House
Shoppenhangers Road
Maidenhead
Berkshire
England
SL6 2QL

email: enquiries@openup.co.uk
world wide web: http://www.openup.co.uk

and Two Penn Plaza, New York, NY 10121-2289, USA

First published 2014

Copyright © World Health Organization 2014 (acting as the host organization for, and secretariat of, the European Observatory on Health Systems and Policies)

The views expressed by authors or editors do not necessarily represent the decisions or the stated policies of the European Observatory on Health Systems and Policies or any of its partners. The designations employed and the presentation of the material in this publication do not imply the expression of any opinion whatsoever on the part of the European Observatory on Health Systems and Policies or any of its partners concerning the legal status of any country, territory, city or area or of its authorities, or concerning the delimitations of its frontiers or boundaries. Where the designation 'country or area' appears in the headings of tables, it covers countries, territories, cities, or areas. Dotted lines on maps represent approximate border lines for which there may not yet be full agreement.

The mention of specific companies or of certain manufacturers' products does not imply that they are endorsed or recommended by the European Observatory on Health Systems and Policies in preference to others of a similar nature that are not mentioned. Errors and omissions excepted, the names of proprietary products are distinguished by initial capital letters. The European Observatory on Health Systems and Policies does not warrant that the information contained in this publication is complete and correct and shall not be liable for any damages incurred as a result of its use. The opinions expressed and arguments employed herein do not necessarily reflect the official views of the OECD or of the governments of its member countries.

Rights to translate into German, Spanish, French and Russian should be sought from WHO at WHO Regional Office for Europe, Marmorvej 51 DK-2100 Copenhagen, Denmark or by email at translationrights@euro.who.int. Rights to translate into all other world languages should be sought from Open University Press.

All rights reserved. Except for the quotation of short passages for the purpose of criticism and review, no part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of the publisher or a licence from the Copyright Licensing Agency Limited. Details of such licences (for reprographic reproduction) may be obtained from the Copyright Licensing Agency Ltd of Saffron House, 6-10 Kirby Street, London EC1N 8TS.

A catalogue record of this book is available from the British Library

ISBN-13: 978-0-33-526438-4 (pb)
ISBN-10: 0-33-526438-7
eISBN: 978-0-33-526439-1

Library of Congress Cataloging-in-Publication Data
CIP data applied for

Typesetting and e-book compilations by
RefineCatch Limited, Bungay, Suffolk

Fictitious names of companies, products, people, characters and/or data that may be used herein (in case studies or in examples) are not intended to represent any real individual, company, product or event.

European Observatory on Health Systems and Policies Series

The European Observatory on Health Systems and Policies is a unique project that builds on the commitment of all its partners to improving health systems:

- World Health Organization Regional Office for Europe
- Government of Austria
- Government of Belgium
- Government of Finland
- Government of Ireland
- Government of Norway
- Government of Slovenia
- Government of Sweden
- Government of the United Kingdom
- Veneto Region of Italy
- European Commission
- World Bank
- UNCAM
- London School of Economics and Political Science
- London School of Hygiene & Tropical Medicine

The series

The volumes in this series focus on key issues for health policy-making in Europe. Each study explores the conceptual background, outcomes and lessons learned about the development of more equitable, more efficient and more effective health systems in Europe. With this focus, the series seeks to contribute to the evolution of a more evidence based approach to policy formulation in the health sector.

These studies will be important to all those involved in formulating or evaluating national health policies and, in particular, will be of use to health policy-makers and advisers, who are under increasing pressure to rationalize the structure and funding of their health system. Academics and students in the field of health policy will also find this series valuable in seeking to understand better the complex choices that confront the health systems of Europe.

The Observatory supports and promotes evidence-based health policy-making through comprehensive and rigorous analysis of the dynamics of health care systems in Europe.

Series Editors

Josep Figueras is the Director of the European Observatory on Health Systems and Policies, and Head of the European Centre for Health Policy, World Health Organization Regional Office for Europe.

Martin McKee is Director of Research Policy and Head of the London Hub of the European Observatory on Health Systems and Policies. He is Professor of European Public Health at the London School of Hygiene & Tropical Medicine as well as a co-director of the School's European Centre on Health of Societies in Transition.

Elias Mossialos is the Co-director of the European Observatory on Health Systems and Policies. He is Brian Abel-Smith Professor in Health Policy, Department of Social Policy, London School of Economics and Political Science and Director of LSE Health.

Richard B. Saltman is Associate Head of Research Policy and Head of the Atlanta Hub of the European Observatory on Health Systems and Policies. He is Professor of Health Policy and Management at the Rollins School of Public Health, Emory University in Atlanta, Georgia.

Reinhard Busse is Associate Head of Research Policy and Head of the Berlin Hub of the European Observatory on Health Systems and Policies. He is Professor of Health Care Management at the Berlin University of Technology.

European Observatory on Health Systems and Policies Series

Series Editors: Josep Figueras, Martin McKee, Elias Mossialos, Richard B. Saltman and Reinhard Busse

Published titles

Regulating entrepreneurial behaviour in European health care systems
Richard B. Saltman, Reinhard Busse and Elias Mossialos (eds)

Hospitals in a changing Europe
Martin McKee and Judith Healy (eds)

Health care in central Asia
Martin McKee, Judith Healy and Jane Falkingham (eds)

Funding health care: options for Europe
Elias Mossialos, Anna Dixon, Josep Figueras and Joe Kutzin (eds)

Health policy and European Union enlargement
Martin McKee, Laura MacLehose and Ellen Nolte (eds)

Regulating pharmaceuticals in Europe: striving for efficiency, equity and quality
Elias Mossialos, Monique Mrazek and Tom Walley (eds)

Social health insurance systems in western Europe
Richard B. Saltman, Reinhard Busse and Josep Figueras (eds)

Purchasing to improve health systems performance
Josep Figueras, Ray Robinson and Elke Jakubowski (eds)

Human resources for health in Europe
Carl-Ardy Dubois, Martin McKee and Ellen Nolte (eds)

Primary care in the driver's seat
Richard B. Saltman, Ana Rico and Wienke Boerma (eds)

Mental health policy and practice across Europe: the future direction of mental health care
Martin Knapp, David McDaid, Elias Mossialos and Graham Thornicroft (eds)

Decentralization in health care
Richard B. Saltman, Vaida Bankauskaite and Karsten Vrangbæk (eds)

Health systems and the challenge of communicable diseases: experiences from Europe and Latin America
Richard Coker, Rifat Atun and Martin McKee (eds)

Caring for people with chronic conditions: a health system perspective
Ellen Nolte and Martin McKee (eds)

Nordic health care systems: recent reforms and current policy challenges
Jon Magnussen, Karsten Vrangbæk and Richard B. Saltman (eds)

Diagnosis-related groups in Europe: moving towards transparency, efficiency and quality in hospitals
Reinhard Busse, Alexander Geissler, Wilm Quentin and Miriam Wiley (eds)

Migration and health in the European Union
Bernd Rechel, Philipa Mladovsky, Walter Devillé, Barbara Rijks, Roumyana Petrova-Benedict and Martin McKee (eds)

Success and failures of health policy in Europe: four decades of divergent trends and converging challenges
Johan P. Mackenbach and Martin McKee (eds)

European child health services and systems: lessons without borders
Ingrid Wolfe and Martin McKee (eds)

Contents

List of tables, figures and boxes ix
List of contributors xiii
Foreword from the OECD xv
Foreword from CNAMTS xvii
Acknowledgements xix
List of abbreviations xxi

Part I ***An overview of health provider P4P in OECD countries***

one ***Health provider P4P and strategic health purchasing*** ***3***
Cheryl Cashin, Y-Ling Chi, Peter C. Smith, Michael Borowitz and Sarah Thomson

two ***P4P programme design*** ***23***
Cheryl Cashin

three ***Strengthening health system governance through P4P implementation*** ***41***
Cheryl Cashin and Michael Borowitz

four ***Evaluating P4P programmes*** ***63***
Y-Ling Chi and Matt Sutton

five ***Lessons from the case study P4P programmes*** ***83***
Cheryl Cashin, Y-Ling Chi and Michael Borowitz

Part II	***Case studies of P4P programmes in OECD countries***	
	Primary care	
six	***Australia: Practice incentives programme*** *Cheryl Cashin and Y-Ling Chi*	***109***
seven	***Estonia: Primary health care quality bonus system*** *Triin Habicht*	***127***
eight	***France: Payment for public health objectives*** *Frédéric Bousquet, Raphaëlle Bisiaux and Y-Ling Chi*	***141***
nine	***Germany: Disease management programmes*** *Y-Ling Chi*	***157***
ten	***New Zealand: Primary health organization performance programme*** *Cheryl Cashin*	***173***
eleven	***Turkey: Family medicine performance based contracting scheme*** *Rekha Menon, Son Nam Nguyen, Aneesa Arur, Ahmet Levent Yener and Iryna Postolovska*	***189***
twelve	***United Kingdom: Quality and outcomes framework*** *Cheryl Cashin*	***205***
thirteen	***United States: California integrated healthcare association physician incentive programme*** *Meredith Rosenthal*	***223***
	Inpatient care	
fourteen	***Brazil: São Paulo: Social organizations in health*** *Y-Ling Chi and Emily Hewlett*	***239***
fifteen	***Republic of Korea: Value incentive programme*** *Raphaëlle Bisiaux and Y-Ling Chi*	***251***
sixteen	***United States: Maryland hospital acquired conditions programme*** *Robert Murray*	***265***
seventeen	***United States: Hospital quality incentive demonstration*** *Cheryl Cashin*	***287***
	Index	301

List of tables, figures and boxes

Tables

1.1	Summary of case study P4P programmes	5
1.2	P4P definitions	7
1.3	P4P programmes and measures in OECD countries in 2012	11
1.4	Summary of objectives for P4P programmes in primary care	12
1.5	Summary of objectives for P4P programmes in hospitals	12
2.1	Summary of case study P4P programme design	25
2.2	Criteria for exception reporting under the UK QOF	35
3.1	Summary of governance structures and data sources and flows in the case study P4P programmes	44
4.1	Summary of case study P4P programme impact evaluation studies	67
5.1	Summary of the results of the case study P4P programmes	90
6.1	Incentives in the Australia PIP, 2010	112
7.1	Performance domains of the Estonia QBS	130
8.1	Performance indicators in the France ROSP	145
8.2	Sample of results of the France ROSP in December 2011–12 against objectives set by the National Health Insurance Fund	151
10.1	Priorities in New Zealand's 2000 Health Strategy	175
10.2	Current funded New Zealand PHO Performance Programme indicators, 2012	178
10.3	New Zealand PHO Performance Programme indicators collected for information only, 2012	179

11.1 Performance indicators and warning points in the Turkey FM PBC 194
12.1 Examples of indicators in the four performance domains of the UK QOF, 2011–12 209
13.1 Approved measurement set in the California IHA, 2012 227
13.2 Annual payouts in the California IHA, 2004–10 230
13.3 Average clinical quality achievement rates in the California IHA, 2006–09 232
14.1 Performance indicators of the Brazil OSS, 2002–04 242
15.1 Calculation of composite quality scores in the Korea VIP 256
15.2 Indicators for acute stroke care and use of prophylactic antibiotics collected by HIRA 257
16.1 Maryland HAC categories 270
16.2 Example of Maryland HAC categories and estimated resource use (charges) by category, fiscal year 2009 274
16.3 Example of the revenue-neutral reallocation (or 'scaling') of rewards/penalties in the Maryland HAC programme 276
16.4 State-wide trends in hospital acquired conditions in Maryland, fiscal years 2009–11 278
17.1 Performance indicators in the US HQID 291

Figures

2.1 Common elements of P4P programmes 24
2.2 UK QOF measurements along the continuum of clinical quality (coronary heart disease indicators) 30
3.1 Health sector governance and the performance improvement cycle 42
3.2 Information flows in the UK QOF 57
6.1 Distribution incentive payments in the Australia PIP, 2008–09 118
7.1 Mix of different payment methods for family physicians in Estonia, 2011 128
7.2 Distribution of points across domains (I and II) for bonus payments in the Estonia QBS 131
7.3 Participation of family physicians in the Estonia QBS, 2006–10 134
7.4 The share of family physicians participating in the Estonia QBS and receiving bonus payments, 2006–10 135
7.5 Participation of family physicians in QBS by county, 2010 135
8.1 Achievement rates for quality indicators for CAPI signers and non-signers in France between March 2009 and 2012 150
9.1 Health expenditure as a percentage of gross domestic product in Germany, 1998 158
9.2 Estimated number of complications avoided yearly by the implementation of the Diabetes DMP in Germany 160
9.3 Summary of the DMP institutional framework in Germany 161
9.4 Changes in payment mechanism for sickness funds in Germany 164

9.5 Number of DMPs developed by sickness funds by clinical condition in Germany 165
9.6 Number of DMP enrollees for each clinical condition in Germany 165
9.7 Comparison of mortality rates between diabetes patients enrolled and not enrolled in a DMP in Germany, 2010 168
10.1 Design of the New Zealand PHO Performance Programme 180
11.1 Design of the Turkey FM PBC 192
12.1 Structure of the primary care system in the NHS in England in 2004 207
12.2 Distribution of points across performance domains in the UK QOF, 2011–12 211
12.3 Share of total points achieved by GP practices across UK QOF performance domains, 2008–12 215
13.1 Survey of physician group responses to the California IHA, 2007 232
13.2 Survey of physician group perceptions of the California IHA, 2007 234
14.1 Data Envelopment Analysis (DEA) efficiency scores for hospitals following the implementation of the Brazil OSS, 2002 246
14.2 Bed turnover rate and average length of stay for OSS and non-OSS hospitals in Brazil, 2002 247
15.1 Annual average growth rate in health expenditure per capita in real terms in the Republic of Korea, 2000–09 (or nearest year) 252
15.2 In-hospital case-fatality rates within 30 days of admission for AMI and C-section in OECD countries, 2009 254
15.3 Ranking of hospitals by performance in the Korea VIP 258
15.4 Bonus payments disbursed under the Korea VIP by type of hospital, 2011 258
15.5 Improvement in process indicators in the Korea VIP, 2007–09 260
15.6 Decrease in variance in process indicators among hospitals in the Korea VIP, 2009 260
15.7 AMI composite quality score (high) and C-section (low) in the Korea VIP, 2007–09 261
17.1 Change in the average composite quality score in the US HQID between 2003 and 2008 295

Boxes

9.1 Examples of DMPs: Rhineland and Hamburg regions 162
14.1 Example of OSS contracting terms with Pirajussara Hospital, 2011 243

List of contributors

Aneesa Arur
World Bank.

Raphaëlle Bisiaux
Coffey International.

Michael Borowitz
Head of Strategic Investment & Partnerships, Global Fund to Fight AIDS, Tuberculosis and Malaria.

Frédéric Bousquet
CNAMTS.

Cheryl Cashin
Senior Fellow, Results for Development Institute.

Y-Ling Chi
DPhil Candidate, University of Oxford.

Triin Habicht
Estonia Health Insurance Fund.

Emily Hewlett
Organisation for Economic Co-operation and Development.

Rekha Menon
World Bank.

Robert Murray
Global Health Payment LLC.

Son Nam Nguyen
World Bank.

Iryna Postolovska
World Bank.

Meredith Rosenthal
Harvard School of Public Health.

Peter C. Smith
Emeritus Professor of Health Policy at Imperial College London and European Observatory on Health Systems and Policies.

Matt Sutton
University of Manchester.

Sarah Thomson
WHO Regional Office for Europe, London School of Economics and Political Science and European Observatory on Health Systems and Policies.

Ahmet Levent Yener
World Bank.

Foreword from the OECD

In 2010, the OECD published 'Health System Priorities when Money is Tight'[1] in response to the observation that health spending continues to grow faster than the economy in many OECD countries. Given the harsh fiscal realities of the recent economic downturn and the fact that most health spending comes from public budgets, countries are looking for ways to improve the efficiency of their health systems. The OECD report identified pay for performance (P4P) as one of the approaches countries are turning to in order to get better value for money.

The problem is that not enough is known about whether and how P4P actually increases value for money in health systems. The evidence that P4P improves health outcomes, or even the quality of processes of care, is limited at best. In fact, the OECD report coincided with the publication of a high-profile study that calls into question the effectiveness of P4P in improving quality of care and health outcomes.[2] Does the evidence suggest that P4P is intrinsically flawed, or are the relatively disappointing results rooted in problems with the design and implementation of P4P programmes, or limitations in the way in which programmes are evaluated?

This volume aims to shed light on these questions by analysing P4P programmes in their entirety within the health policy context of each country at the time the programme was introduced. The volume analyses the experience of P4P programmes in 10 OECD countries, selected to reflect the wide range of health system contexts and challenges across the OECD. Case study programmes are drawn from some of the highest health spending OECD countries (such as the United States), and some of the lowest (such as Turkey).

Programmes were selected to represent both a hospital focus and a primary care focus. Some of the programmes are implemented on a national or regional scale, and some are pilots. The case study authors systematically describe the design decisions and implementation arrangements for each programme. They critically assess the results against the objectives the programmes were designed to achieve, as well as their 'net' effects on health system objectives, which takes into account positive spillover effects, any unintended consequences, and programme net costs. The intent was to delve more deeply into the realities of P4P programme design and implementation, considering stakeholder roles and reactions, data constraints, and the evolution of governance structures to improve understanding of how financial incentives can be leveraged to achieve better quality of care and other health system objectives.

The findings of the volume in many ways mirror the findings of the few rigorous systematic reviews of P4P programmes, and the opinions of many leading commentators. Pay for performance does not lead to 'breakthrough' quality improvements, and performance measures and other key building blocks of P4P programmes remain highly inadequate. But the findings of the study also suggest that P4P has a broader role to play as an instrument for improving health system governance and strategic health purchasing. Several of the programmes that showed the most modest results also claim the most powerful impact on the relationship between purchasers and providers, in some cases opening the door to discussion of provider payment reform, quality measurement, and accountability for outcomes.

This volume will not provide answers to questions such as whether or not P4P works, which performance measures are most appropriate, or what is the right level of financial incentive to get results. Instead – and more importantly for real health financing policy in complicated contexts – are the insights about how P4P might be used to strengthen health system governance and strategic health purchasing to continue the shift taking place in many countries from paying for performance to paying for value.

Mark Pearson, Head of the Health Division,
Directorate of Employment, Labour and Social Affairs,
Organisation for Economic Co-operation and Development

Notes

1 OECD (2010) *Value for money in health spending*, OECD Health Policy Studies. Paris: OECD Publishing (doi: *10.1787/9789264088818-en*).

2 Serumaga, B., et al. (2011) Effect of pay for performance on the management and outcomes of hypertension in the United Kingdom: interrupted time series study, *BMJ* 342: d108.

Foreword from CNAMTS

Increasing value for money in health spending is a common challenge in all countries; we know that the quality and efficiency of care delivered can improve, and that there are significant gaps between actual care delivered by practitioners and best practices as they are defined by widely accepted standards and guidelines.

Among other levers, building on the development of information systems and performance measurement to design new payment mechanisms for health professionals, in order to align financial incentives with quality and efficiency goals, has emerged as a promising strategy. Indeed no traditional method of payment, whether it is fee for service, capitation or bundled payment, explicitly rewards the achievement of quality objectives.

Based on these considerations, a number of countries have developed P4P programs in the last decade. It is important to capitalise on lessons learned from these experiences: this is why this joint work of the OECD and of the European Observatory on Health Systems and Policies is extremely useful, to understand what P4P programmes have achieved and highlight key challenges encountered in the implementation of these programmes.

Even if the scientific evidence on the impact of P4P remains fragmented and incomplete, as pointed out in Chapter 1, the general perception is that the direct results of most programmes are modest. There is progress in some areas, but not in all, and the pace of change is rather slow. This is not specific to only P4P programmes: most of interventions in the health-care sector (e.g. guidelines dissemination, disease management programmes, public information, to name a few) do not lead to an immediate 'breakthrough' in quality as changing

practices and behaviours of health care providers and patients is not an easy task.

We can draw a few lessons from the evidence gathered in this volume and also from our own experience in France, where ROSP, a P4P programme in primary care, has been expanded since the first pilot in 2009:

P4P is one lever among others, and is probably more effective when implemented alongside other policies. We have a recent example in France with the combination of a P4P intervention targeting pharmacists and a financial incentive for patients, which has proved successful in increasing the substitution rate of branded drugs for generic drugs. In the field of prevention, informing and incentivising patients seems to be as important as incentivising physicians to prescribe screening or immunization. More generally, changing medical practices probably requires a variety of tools (financial incentives, information feedback, training, peer groups, development of new software or information technology tools …) and the involvement of different stakeholders, including physicians but also other caregivers and patients. We still have to find the right mix and there is room here for experimenting with new policies and cross-country learning.

Beyond its direct results, P4P may have positive collateral impacts: the development of a culture of performance measurement and monitoring among health professionals, the strengthening of a public health approach enhancing population-based outcomes. In France, P4P also gave strong impetus to the development of electronic medical records in primary care practice. These dynamics may contribute to quality improvement outside of the range of traditional 'performance indicators', but little is known about them.

Finally, P4P is a lever to develop strategic purchasing and to enrich the dialogue between purchasers of care and the medical profession, and in that sense it may be an element of a strategy for change.

Frédéric Van Roekeghem
Director General, Caisse nationale d'assurance maladie des travailleurs salariés

Acknowledgements

This volume is part of a series produced by the European Observatory on Health Systems and Policies. We are grateful for the efforts of Josep Figueras (Director of the European Observatory) to make this book part of the Observatory series. The production of the book would not have been possible without the support of the World Bank and the OECD both of which dedicated substantial time and resources to completing the case studies.

The idea for this book came from several heated debates about pay for performance and why it is viewed by some in extremes, while to others it just seems to be the latest distraction or fad in provider payment policy. A practical contribution to this debate seemed to be needed, with a closer look at how the programmes work in practice. Darren Dorkin, then leading the World Bank's Health Results Innovation Trust Fund, commissioned the initial set of six case studies of P4P programmes in OECD countries, and Anthony Meacham (consultant to the World Bank) helped conceptualize the analysis. Richard Scheffler and Brent Fulton of the University of California, Berkeley's Nicholas C. Petris Center on Health Care Markets and Consumer Welfare contributed an early background paper.

We wish to thank the many colleagues who shared their insights and time to contribute to and review the case studies: Stephen Campbell (University of Manchester), Sarbani Chakraborty (Ministry of Health, Turkey), Dr Cho (Korea Health Insurance Review and Assessment Service), Jackie Cummings (Victoria University of Wellington), Serena Curtis-Lemuelu (PHO Performance Programme), Janice Donaldson (National Services NZ), Neil Ellis (Australian Department of Health and Ageing), Valerie Frey (OECD), Dirk Göpffarth

(German Federal Insurance Agency), Andrea Harris (US Department of Health and Human Services), Julian Inch (DHBNZ), Ruth Kalda (University of Tartu), Vangool Kees (OECD), SunMin Kim (Health Insurance Review and Assessment Service), Jerry LaForgia (World Bank), Taavi Lai (Ministry of Social Affairs, Estonia), Magnus Lindelow (World Bank), Ruth McDonald (University of Nottingham), Zeynep Or (Institut de recherche en économie de la santé), Brett Paradine (PHO Performance Programme Governance Group), Dr Park (Korea Health Insurance Review and Assessment Service), Dominique Polton (CNAMTS), Tony Scott (Melbourne Institute of Applied Economic and Social Research), Steven Speldewinde (Australian Department of Health and Ageing), Rosanna Tarricone (Bocconi University), Mehtap Tatar (Hacettepe University), Stefan Thewissen (Leiden University), Jocelyn Tracey (PHO Performance Programme Governance Group Advisory Group), Evert Jan Van Lente (AOK Bundesverband), and Petra Vergeer (World Bank).

We would like to extend special thanks to Robert Berenson (Urban Institute), Joseph Kutzin (WHO), Meredith Rosenthal (Harvard School of Public Health), Dominique Polton and Frédéric Bousquet (CNAMTS), all of whom provided valuable insights and comments on the conceptual foundations of the book, multiple chapters, and case studies. We also would like to thank the members of the Interagency Working Group on Results-Based Financing, and in particular Monique Vledder (World Bank), who provided valuable feedback on our early results and conclusions.

Finally, we are most grateful to all of the case study authors who gave so generously of their time and expertise to contribute the valuable content for this book.

List of abbreviations

ACEI	ACE inhibitor (or angiotensin-converting-enzyme inhibitor)
ACO	US Medicare accountable care organization
AHA	American Hospital Association
AHRQ	Agency for Healthcare Research and Quality
AIC	Independent Assessment Commission
ALOS	Average Length of Stay
AMI	acute myocardial infarction
AMSs	Aboriginal Medical Services
ANAO	Australian National Audit Office
APR-DRG	All Patient Refined Diagnosis-Related Groups
ARB	Angiotensin Receptor Blocker
BBA	Balanced Budget Act
BCG	Bacille Calmette-Guerin
CABG	Coronary Artery Bypass Surgery
CAPI	Contracts for Improved Individual Practice
CDAC	Clinical Data Abstract Center
CHC	Community Health Centres
CMS	Centers for Medicare & Medicaid Services
CNAMTS	Caisse Nationale d'Assurance Maladie des Travailleurs Salariés (French National Health Insurance Fund)
CRMS	Core Health Resource Management System
C-section	Caesarean Section
CVD	Cardio-Vascular Diseases
DEA	Data Envelopment Analysis

DHB	District Health Boards
DHBSS	District Health Boards Shared Services
DMP	Disease Management Programs
DoHA	Department of Health and Ageing
DPT3	Diphtheria, Pertussis and Tetanus
DRG	Diagnosis-Related Group
EHIF	Estonia Health Insurance Fund
EHR	Electronic Health Records
ENMR	Expérimentation de Nouveaux Modes de Rémunération
EPP	Evaluation of professional practice
FM PBC	Family Medicine Performance Based Contracting Scheme
FMIS	Family Medicine Information System
FMP	Family Medicine Practitioners
FTE	Full Time Equivalent
GDP	Gross Domestic Product
GP	General Practitioner
GPFG	General Practice Financing Group
GPII	General Practice Immunization Incentive
HAS	National Authority for Health
HbA1c	measure of Glycosylated Hemoglobin
HepB3	Third Dose Hepatitis B
Hib3	Haemophilus Influenzae Type B
HIRA	Health Insurance Review Agency
HKR	Hip and Knee Replacement
HQA	Hospital Quality Alliance
HQID	Hospital Quality Incentive Demonstration
HSCRC	Health Services Cost Review Commission
HTP	Health Transformation Programme
ICD-9-CM	International Classification of Diseases, Ninth Revision-Clinical Modification
IGES	Institut für Gesundheits- und Sozialforschung
IHA	Integrated Healthcare Association Physician Incentive Program
IMIT	Information Management /Information Technology
IOM	Institute of Medicine
IPA	Independent Practitioner Association
JAMA	Journal of the American Medical Association
LDL	Low-density lipoprotein
LVSD	Left Ventricular Systolic Dysfunction
MBS	Medicare Benefits Schedule
MCH	Maternal Child Health
MHAC	Maryland Hospital Acquired Conditions Program
MMA	Medicare Prescription Drug, Improvement, and Modernization Act
MoH	Ministry of Health
MoSA	Ministry of Social Affairs
NCQA	National Committee for Quality Assurance
NHIC	National Health Insurance Corporation
NHIF	National Health Insurance Fund
NHS	National Health Service

NICE	National Institute for Health and Clinical Excellence
NQF	National Quality Forum
OECD	Organisation for Economic Co-operation and Development
OSS	Social Organizations in Health
P4P	Pay for performance
PCOs	Primary Care Organizations
PCTs	Primary Care Trusts
PHC	Primary Health Care
PHD	Provincial Health Directorates
PHO	Primary Health Organization
PIP	Practice Incentives Program
Pol3	Third Dose Polio
PPCs	Potentially Preventable Complications
PPI	Proton-Pump Inhibitor
QBR	Quality Based Reimbursement
QBS	Quality Bonus System
QMAS	Quality Management and Analysis System
QOF	Quality and Outcomes Framework
RACGP	Royal Australian College of General Practitioners
RBF	Results-Based Financing
RCS	Risk Compensation Structure
RCT	Randomized Control Trial
ROSP	Payment for Public Health Objectives (formerly CAPI)
RRMA	Rural Remote and Metropolitan Area
SES	State Secretariat of Health
SHA	Strategic Health Authority
SoI	Severity of Illness
SUS	Unified Health System
SWPE	Standardized Whole Patient Equivalent
UK NAO	UK National Audit Office
USAID	United States Agency for International Development
VBP	Value Based Puchasing
VIP	Value Incentive Programme
WHO	World Health Organization

Part I

An overview of health provider P4P in OECD countries

chapter one

Health provider P4P and strategic health purchasing

Cheryl Cashin, Y-Ling Chi, Peter C. Smith, Michael Borowitz and Sarah Thomson

Introduction

It is widely acknowledged that health systems of all types suffer from gaps between best practices supported by evidence and the actual delivery of health services. Many of these quality gaps are readily amenable to improvement (Institute of Medicine, 2001), yet they persist in spite of increased levels of health expenditure and numerous other reforms in health care financing, regulation, and service delivery. The quality gaps take many forms, including failure to implement evidence-based clinical practices, fragmentation of services, slow and incomplete responses to adverse indications, and lack of attention to appropriate preventive measures.

A series of studies worldwide has exposed specific examples of variations in the quality of care, even in the most widely acclaimed health systems (Institute of Medicine, 2001). Many of these countries fail to offer their populations consistently well-coordinated, high quality, cost-effective health care. Furthermore, the ageing of populations and the rising prevalence of complex chronic conditions has put increasing demands on the health care system and is changing the kinds of services needed. Chronic conditions often require coordinated preventive, curative, and disease management services, provided in a variety of settings, personalized to the specific circumstances of the individual patient.

Decision makers have sought to pull many types of policy levers to address gaps in quality, including publication of treatment guidelines, promoting competition and choice, professional exhortation, public reporting of quality and various forms of quality accreditation. In general, such approaches have had some success in certain settings, but any gains have on the whole been modest. It is therefore not surprising to find that over the last ten years policymakers have turned their attention to one of the most powerful instruments for altering provider behaviour – the provider payment echanism.

The way in which providers are paid is known to have a profound impact on the volume and quality of health services delivered (Dudley et al., 1998; Conrad & Christianson, 2004). However, traditional ways of paying health care providers – such as salary, fee-for-service, bundled payments, and capitation – do not explicitly reward providers for delivering better quality care. Any impact on quality of these payment methods is indirect and often incidental. For example, fee-for-service payment creates incentives for high levels of service provision, and thus might indirectly lead to higher levels of quality. That impact, however, is an accidental consequence of the incentives inherent in fee-for-service, and of course it also may contradict another key objective of health systems – the pursuit of efficiency. In contrast, traditional capitation payment might secure expenditure control, but it offers little direct incentive to promote high quality care and may instead create incentives for skimping on necessary services.

A growing number of new provider payment models are therefore emerging that explicitly seek to align payment incentives with health system objectives related to quality, care coordination, health improvement, and efficiency by rewarding achievement of targeted performance measures. These models are being tested in a wide range of countries: in OECD countries like the United States (US), United Kingdom (UK), and Germany; in middle-income countries like Brazil, China and India; and in low-income countries like Rwanda. They have become collectively known as 'pay for performance', or P4P for short.

The origin of the P4P movement in health care can be traced back to the private sector in the US in the late 1990s. In 1999 the Institute of Medicine (IOM) issued its now-famous report *To Err is Human: Building a Safer Health System* (Institute of Medicine, 1999). This watershed report made public the widespread preventable medical errors in hospitals that led to between 44,000 and 98,000 deaths each year. That report was followed by the IOM's (2001) *Crossing the Quality Chasm: A New Health System for the 21st Century*, which showed that health care in the US routinely deviated from clinical guidelines and best practices (Institute of Medicine, 2001). A key recommendation of that report was that payment incentives for providers needed to be realigned to support quality improvement. These reports coincided with a backlash against managed care efforts to contain costs, which were perceived as ignoring the consequences of managed care for quality (Robinson et al., 2009). There were a number of P4P programmes operational in the private sector in the US by 2002, but these initiatives remained mainly small and experimental. The first large-scale private sector P4P programme was initiated by the Integrated Healthcare Association in California in 2003 and is still ongoing (see Part II of this volume).

The P4P programmes implemented by strategic purchasers of health services in most countries have been used to augment and refine traditional payment systems. Although assuming a variety of forms, the common characteristic of P4P programmes is the deliberate adoption of explicit payment incentives associated with metrics for specific objectives, such as higher quality processes of care that follow evidence-based guidelines, increased coverage with preventive services, better management of chronic diseases, and better patient outcomes.

Objectives

The purpose of this book is to explore experience with P4P programmes through an assessment of existing literature and a series of case studies. The book looks specifically at the details of implementation of 12 P4P programmes and key contextual factors. In Part I we set out the principles of P4P in health care and draw on the set of case studies to illustrate how these principles are playing out in practice in current P4P programmes in OECD countries. Part II provides the detailed case studies from the 12 OECD P4P programmes (Table 1.1). Chapter 2 presents a general discussion of the

Table 1.1 Summary of case study P4P programmes

Programme focus	*Country*	*Programme*		*Year programme began*
Primary care	**Australia**	**PIP**	Practice Incentives Programme	1998
	Estonia	**PHC QBS**	Primary Health Care Quality Bonus System	2005
	France	**ROSP***	Payment for Public Health Objectives (formerly CAPI)	2009
	Germany	**DMP**	Disease Management Programmes	2002
	New Zealand	**PHO Performance Programme**	Primary Health Organization Performance Programme	2006
	Turkey	**FM PBC**	Family Medicine Performance Based Contracting Scheme	2003
	UK	**QOF**	Quality and Outcomes Framework	2004
	US–California	**IHA***	Integrated Healthcare Association Physician Incentive Programme	2002
Hospitals	**Brazil–Sao Paolo**	**OSS****	Social Organizations in Health	1998
	Korea, Rep. of	**VIP**	Value Incentive Programme	2007
	US–Maryland	**MHAC**	Maryland Hospital Acquired Conditions Programme	2010
	US–National	**HQID**	Hospital Quality Incentive Demonstration	2004

* Programme includes specialists providing outpatient services.
** Programme includes outpatient services delivered by the hospitals.

elements of any P4P programme, including the objectives, the domains of performance assessed, the metrics adopted, the basis for and nature of financial rewards and penalties, and the data needs. Chapter 3 then goes on to assess the system requirements to implement P4P programmes, including the governance structures and health information systems, and how the programmes can in turn strengthen these aspects of the health system. Chapter 4 discusses in more detail the issue of monitoring and evaluation of P4P programmes. Given the great uncertainty associated with the overall effect of P4P, all implemented programmes should be properly monitored, both for intended and unintended consequences, and should be capable of being rigorously evaluated. Chapter 5 draws together the lessons from the case studies, assesses the reasons for successes and failures, and summarizes the key messages for policy.

Underlying the analysis is the widespread perception that many P4P programmes have produced disappointing or only modest results and failed to improve provider performance in the intended fashion (Mullen, Frank & Rosenthal, 2009; Scott et al., 2009; Flodgren, 2011). There have been notable successes, however, and evidence on the role of P4P in improving quality of care, health outcomes and other health system objectives is at present fragmented and incomplete, in part because so few programmes have been rigorously evaluated. In the remainder of this chapter we define P4P in health care and its theoretical underpinnings, and summarize the current evidence about the impact of P4P on provider performance and health outcomes.

Defining P4P

There is no accepted international definition of pay for performance. The term often is used interchangeably with other closely associated terms, such as "performance-based funding", "paying for results", or "results-based financing" (RBF). Table 1.2 presents some of the more common definitions of pay for performance used to date. The first three definitions are from a US perspective, reflecting the origins of the P4P movement in the US: (1) Agency for Healthcare Research and Quality (AHRQ), (2) Centers for Medicare & Medicaid Services (CMS), and (3) RAND Corporation. These all focus on quality improvement, although each is defined somewhat differently. The RAND Corporation also includes efficiency as a measure. The latter three definitions take a broader approach and are more concerned with developing countries: (4) World Bank, (5) United States Agency for International Development (USAID), and (6) Center for Global Development. The World Bank, USAID, and Centre for Global Development definitions include both incentives on the supply side to providers and also demand-side incentives to patients like conditional cash transfers, although demand-side incentives are beyond the scope of this study.

In examining these definitions, it is important to note that payment mechanisms of all sorts offer implicit incentives that may promote (or inhibit) the achievement of health system objectives, including quality improvement.

Table 1.2 P4P definitions

Organization	*P4P definition*
AHRQ	Paying more for good performance on ***quality metrics*** (*Source*: AHRQ, undated).
CMS	The use of payment methods and other incentives to encourage ***quality improvement*** and patient focused high value care (*Source*: Centers for Medicare and Medicaid Services, 2006).
RAND	The general strategy of promoting ***quality improvement*** by rewarding providers (physicians, clinics or hospitals) who meet certain performance expectations with respect to health care quality or efficiency (*Source*: RAND Corporation, undated).
World Bank	A range of mechanisms designed to enhance the ***performance of the health system*** through incentive-based payments (*Source*: World Bank, 2008).
USAID	P4P introduces incentives (generally financial) to reward attainment of ***positive health results*** (*Source*: Eichler & De, 2008).
Center for Global Development	Transfer of money or material goods conditional on taking a ***measurable action or achieving a pre-determined performance target*** (*Source*: Oxman & Fretheim, 2008).

Source: OECD, 2010.

Furthermore, traditional payment systems can be adapted, without specific quality metrics, to create stronger incentives for quality.

Capitation payment offers a fixed payment to a provider to care for a specified population over a specified period, in effect offering a block contract. Of course there may be moderating influences and complicating factors, but the principal immediate incentives of capitation will be to discourage use of health services and secure expenditure control. Although quality ultimately may be improved by reducing unnecessary services and focusing more on prevention, there is no immediate incentive to promote quality under capitation.

In contrast, fee-for-service, under which providers are paid individually for each service delivered, creates a clear incentive to provide increased access to health services, with potential benefits for some aspects of quality, albeit in an indirect fashion. There are, however, likely negative consequences for expenditure control and overuse of inappropriate services. Bundled payments, or case payments, under which defined episodes of care are paid at a fixed rate, often independent of the services provided, create intermediate incentives. In particular these payment systems might exert downward pressure on the unit costs of delivering an episode of care, but they also create incentives to increase the number of cases treated. Although this may indirectly improve access to necessary care, the number of unnecessary services and overall costs also may increase.

Any health system will use some blend of such mechanisms to pay providers. For example, national health services traditionally have used fixed prospective payment for providers (e.g. capitation), augmented with a certain element of

retrospective payment (fee-for-service). Systems of social health insurance have in general moved from fee-for-service to bundled payments over the last 20 years, although retaining elements of retrospective payment, as well as block contracts. Many purchasers also are modifying underlying payment systems to address quality gaps and other aspect of performance. For example, Germany and the UK have modified their diagnosis-related group (DRG) hospital payment systems to refuse payment for cases that are readmissions within a certain period of time (Busse et al., 2011). Some may argue that any of these modifications to the base payment system to have an impact on improving quality could be labelled "pay for performance". The focus of this book, however, is on the more narrow set of mechanisms that blend or augment base payment systems with specific incentives and metrics explicitly to promote quality and other performance improvements. In light of this discussion we adopt the following definition of P4P:

> 'The adaptation of provider payment methods to include specific incentives and metrics explicitly to promote the pursuit of quality and other health system performance objectives.'

To address this topic, it is necessary to have a clear idea of what is meant by 'quality' and 'performance'. A wide variety of possibilities exists, and each of the case studies described in Part II of the book identifies the particular aspects of performance being addressed by the P4P programme. Most of the programmes identify quality, access to priority services, and efficiency as key performance domains. Quality poses particular challenges for measurement. In general, it is useful to consider two broad dimensions of quality, in the form of health outcomes and health system responsiveness. Outcomes are readily conceptualized, in the form of improvements in the length and quality of life created by health services. Responsiveness is a less well developed concept, but reflects a broad range of characteristics having an influence on patient experience and user satisfaction that are not immediately related to health outcomes, such as waiting times, other barriers to access and the way in which users, their caregivers or potential users are treated by the services.

Each of these concepts can be measured using indicators of the structure, process or outcomes of care. For example, ideally the health outcomes of care should be measured using indicators of improvement in future quality of life gained as a result of treatment. However, such measurement often is infeasible, largely out of the control of providers, and not helpful if it involves a long delay in securing results. So in practice measurement typically relies on measures of the structure of care (for example, the presence of certain elements of service infrastructure such as a dedicated stroke unit) or the processes of care (such as adherence to clinical guidelines). Given the current limitations of performance measures, recourse to structure and process indicators is often inevitable, but to use them as measures of quality is valid only if they are known – from research evidence – to lead to improvements in health outcomes.

The key elements of any P4P programme typically include: a statement of the quality objectives it seeks to promote; definition of quality metrics that will

influence payment; formulation of the associated rules for payment that make some element conditional on measured levels of attainment; rules for providers regarding provision of information and other standards; and governance arrangements to ensure that the system is working as intended. The elements of P4P programmes are considered in more detail in Chapters 2 and 3.

The theory underlying P4P

The theory underlying many P4P programmes can be traced to the economic principal/agent literature (Robinson, 2001; Christianson, Knutson & Mazze, 2006). A principal (such as a patient, or more often in health care, a strategic purchaser) wishes to structure the contractual relationship with the agent (either an individual practitioner or an organization, such as a hospital) to secure high quality health services at the lowest cost. It is assumed that increasing quality or efficiency requires 'effort' on the part of the agent, who must therefore be compensated with a financial reward (or face the threat of penalty) if improvements are to be secured. The agent will then assess how much effort to exert by comparing the expected financial benefits to the effort required. In the simplest form of this model, the principal then sets the financial rewards (or penalties) for the agent knowing how the agent will respond to the incentives, in terms of exerting increased effort, and thereby delivering improved performance. In setting the incentive regime, the principal must of course balance the expected costs of the rewards against the expected improvements in quality.

There are several elements in this model that bear more detailed scrutiny. First, measurement plays a key role. Effort usually cannot be observed and measured, so instead there must be some way of explicitly measuring the performance attained. Performance metrics therefore are central to any P4P programme. Ideally these should be accurate and timely indicators of the desired performance criterion, sensitive to variations in provider effort, and resistant to manipulation or fraud. In examining the programmes described in this book, it is important to assess the strengths and limitations of the metrics being used.

Second, design of the financial reward (or penalty) mechanism requires numerous judgements, such as the magnitude of the incentive, how it increases with increased quality, whether or not the rewards are based on performance relative to other providers, whether rewards are based on individual aspects of performance or an aggregate measure of organizational attainment, and whether they are based on absolute levels of attainment or on improvements from previous levels. These design considerations are a central concern of all the programmes described here, and are likely to play a crucial role in their effectiveness.

Third, the effect of any P4P programme depends crucially on the intrinsic motivation of the professionals and organizations at whom the programme is directed. If the desired improvements in quality are aligned with professional objectives, and the programme serves to offer focus and encourage professionals or the organizations in which they work to secure such improvements, then it

may indeed contribute to the desired outcomes. On the other hand, if the P4P programme contradicts or undermines professional motivation, it may prove ineffective or even lead to adverse outcomes (Woolhandler & Ariely, 2011; Cassel & Jain, 2012).

More generally, it is likely that contextual factors play a key role in the success or otherwise of P4P programmes. It may be, for example, that some aspects of health services are more amenable to P4P than others, that such programmes work better in market-based rather than centrally planned health systems, or that providers require a long-term commitment from payers to P4P before committing resources to quality improvement efforts. Furthermore, a persistent theme found throughout this book and the broader P4P literature is that effective governance arrangements are an essential prerequisite for the success of any P4P programme. Financial instruments create powerful incentives. As well as inducing the desired effects, they may also inadvertently create unintended, perhaps adverse, incentives. For example, if only certain aspects of quality are tied to the incentive, it may be the case that unrewarded (but nevertheless valued) aspects of quality will be ignored. Or if the performance metrics are inadequate, their use might stimulate adverse provider behaviour, such as excluding certain types of patients from treatment, even though those patients could benefit from care. Any full evaluation of a P4P programme should assess the nature and importance of any such side effects.

It can furthermore be argued that explicit incentives may be unnecessary to secure the desired quality standards. For example, if publicly available information sources are good, and payers and patients are able to select providers on the basis of reported performance, then the associated competition might lead to the optimal level of quality. However, the necessary information requirements are demanding, and experience with public reporting alone as a mechanism for stimulating improvement has been mixed at best (Mannion & Goddard, 2003; AHRQ, 2012). Other mechanisms such as professional regulation, central planning and democratic governance also have a role to play in performance improvement. A theme that emerges in this book is that the impact of P4P programmes on performance improvement is enhanced when the financial incentives are used in combination with and to reinforce these other actions for improving quality and provider performance.

Experience to date

Compared to many commentators, this book uses a quite restrictive definition of P4P that focuses on supply-side interventions (i.e. payments to providers, not to patients) that include some measure of quality of care. Such programmes are common within many OECD countries, and Tables 1.3, 1.4 and 1.5 report the P4P programme results from the OECD Survey on Health System Characteristics 2012. Pay for performance programmes were reported to exist in 15 OECD countries in the following categories: primary care physicians (15), specialists (8), and hospitals (8). For primary care physicians and specialists, most give bonuses for reaching performance targets such as preventive care,

Table 1.3 P4P programmes and measures in OECD countries in 2012

Country	*Primary care*	*Specialist care*	*Hospitals*
Australia	X		X
Austria			
Belgium	X		
Canada			
Chile	X	X	
Czech Republic	X		
Denmark			
Estonia			
Finland			
France	X	X	X
Germany	X		
Greece			
Iceland			
Ireland			
Israel			
Italy			
Japan			
Korea, Rep. of	X	X	X
Luxembourg	X		
Mexico	X		
Netherlands	X	X	X
New Zealand	X		
Norway			
Poland	X		
Portugal	X		X
Slovakia			
Slovenia			
Spain	X	X	X
Sweden	X		X
Switzerland			
Turkey	X		X
UK	X	X	X
US	X	X	X

Source: OECD work on health systems characteristics 2012 and authors' estimates, unpublished.

efficiency of care, patient satisfaction and management of chronic diseases. For hospitals, there are programmes that include bonuses or penalties, mostly for processes of care, but some also for clinical outcomes and patient satisfaction. As might be expected, there is significant variation amongst countries. Some such as Belgium, Japan, Turkey, United Kingdom, and United States report P4P in all three sectors (primary care, specialists, and hospitals). In contrast, Austria, Denmark, Finland, Greece, Iceland, Norway, and Switzerland do not report having any P4P programme, possibly due to underreporting.

The proportion of physicians and hospitals participating in a P4P programme was only reported for a few countries. The proportions for each sector were as follows: primary care: Belgium (90 per cent), Poland (80 per cent), and

Table 1.4 Summary of objectives for P4P programmes in primary care

Country	*Preventive care*	*Management of chronic diseases*	*Efficiency*	*Patient satisfaction*	*Uptake of IT services*	*Others*
Australia	X	X			X	X
Chile	X	X	X	X		X
Czech Republic	X					X
France	X	X	X			
Korea, Rep. of			X			
Mexico	X	X	X	X		X
New Zealand	X	X				
Portugal	X	X	X	X		
Spain	X	X	X			
Sweden	X	X	X	X	X	
UK	X	X	X	X		
US	X	X	X	X	X	X

Source: OECD work on health systems characteristics 2012 and authors' estimates, unpublished.

Table 1.5 Summary of objectives for P4P programmes in hospitals

Country	*Clinical outcomes of care*	*Use of appropriate processes*	*Patient satisfaction*	*Patient experience*
Australia				X
Korea, Rep. of	X	X		
Portugal	X	X	X	X
Spain	X	X		X
Sweden	X	X		X
UK	X	X	X	X
US	X	X	X	X

Source: OECD work on health systems characteristics 2012 and authors' estimates, unpublished.

United Kingdom (99 per cent); specialty care: Poland (5 per cent) and United Kingdom (68 per cent); and hospitals: Luxembourg (9 per cent). The share of the physician and hospital earnings represented by the bonus payment was only reported for a few countries, and they were generally five per cent or less, except for the United Kingdom. The bonus shares for each sector were as follows: primary care: Belgium (2 per cent), Poland (5 per cent) and United Kingdom (15 per cent); specialty care: Poland (5 per cent); and hospitals: Belgium (0.5 per cent) and Luxembourg (1.4 per cent). These data offer glimpses of current efforts, but clearly, more data are needed in order to understand better the attributes of these P4P programmes.

Summary of findings from the literature

Although many programmes have been implemented throughout the world over the past two decades, the evidence remains very fragmented about whether and how they are an effective way to improve quality of care and achieve other health system objectives. Most programmes have been implemented without provisions for adequate monitoring and evaluation, and the methods available to rigorously evaluate the programmes have been limited. Published studies tend to focus on narrow aspects of performance without placing programmes in the wider context within which they were conceived and implemented.

The evidence that does exist about the effect of P4P on improving health service delivery and patient outcomes remains mixed, but in general fails to show any 'breakthrough' quality improvements (Christianson, Leatherman & Sutherland, 2007; Damberg et al., 2009; Guthrie, Auerback & Binman, 2010). A review of the only five randomized control trials involving P4P programmes (defined as bonus payments tied to performance) found two programmes led to improved measures of quality, while three had no significant effect (Frolich et al., 2007). Most P4P programmes do show that performance measures that are tied to incentives tend to improve, but these improvements are often marginal. For example, a recent review of 128 P4P evaluation studies showed that P4P programmes led to a five per cent improvement effect on average, but there was a lot of variation across programmes and performance areas (Van Herck et al., 2010). One recent study did, however, find clinically significant effects on in-hospital mortality for conditions covered by a P4P programme in hospitals in one region of England (Sutton et al., 2012). The previously cited review examined 28 studies that analysed impacts on equity, mainly for the UK QOF, and showed that P4P programmes do not negatively affect equity and access to care, and in some cases managed to narrow equity gaps (Van Herck et al., 2010). Because equity effects are likely to hinge on key design and contextual factors, however, these results should be generalized with caution. Many questions remain about the degree of real improvement in quality of care and outcomes, and whether unintended consequences such as shifting away from activities and services that are not tied to incentives are significant.

The published and unpublished literature on P4P sheds even less light on aspects of design and implementation of the programmes that may be associated with their effectiveness, and no analyses have addressed the question of whether the programmes are a cost-effective way to achieve their various objectives.[1] A recent study suggests some aspects of P4P programme design and implementation that may be important for their success, including: (1) defining performance broadly rather than narrowly; (2) attention to limiting patient selection and health-reducing substitution; (3) including risk adjustment for outcome and resource use measures; (4) involving providers in programme design; (5) favouring group incentives over individual incentives; (6) using either rewards or penalties depending on the context; (7) more frequent, lower-powered incentives; (8) absolute targets preferred over relative targets; (9) multiple targets preferred over single targets; and (10) P4P as a permanent element of overall provider payment systems (Eijkenaar, 2011).

Several reviews conclude that P4P programmes in their entirety may be more powerful than the sum of their parts (Damberg et al., 2009; Van Herck et al., 2010). The most important effects of P4P programmes may be their reinforcing effect on broader performance improvement initiatives, and their "spillover effects", or other health system strengthening that occurs as a by-product of the incentive programmes. Some programmes report that the improved collection and use of data for performance improvement, faster uptake of IT, more quality improvement tools (e.g. guideline-based decision aids), sharper focus on priorities, and better overall governance and accountability are more important outcomes of the P4P programmes than improvements in the targeted performance indicators (Campbell, MacDonald & Lester, 2008; Martin, Jenkins & Associates Limited, 2008; Damberg et al., 2009).

Case study approach: examining the 'net effect' of P4P programmes on health system performance

Pay for performance programmes are based on the premise that if health care providers are paid more for certain behaviours, processes, and outcomes, then more of these will be delivered. Although this premise is not disputed, the actual power of P4P programmes and the incentives they create to improve provider performance, and ultimately health outcomes, can be altered by many institutional, behavioural, and system factors. Furthermore, the governance structures and information systems that may be created or strengthened to implement the programmes may have effects on provider performance and quality that are independent of the financial incentive. On the other hand, the programmes may lead to unintended consequences that detract from health system objectives. In practice, the net effect (the combination of performance improvement and unintended consequences) of P4P programmes on health system performance ultimately will be determined by the interplay of the financial incentives created by the P4P programme, the provider responses to those incentives, and implementation arrangements and contextual factors.

The approach taken for the case studies in Part II, therefore, aimed to describe the key design and implementation features of the P4P programmes in light of health policy objectives and contextual factors. The case study authors used a detailed matrix of 55 parameters for describing P4P programme design, implementation and results in a standardized way, which was developed by the editors specifically for this review (Appendix 1.1). The authors analysed the results of the programmes from the perspective that the overall effect on objectives such as health outcomes, clinical quality of care, and efficiency could be positive or negative, depending on the interplay of three sets of effects:

Net substitution effects

Net substitution effects take account of whether providers substitute more valuable activities for less valuable activities. Financial incentives will direct providers toward the rewarded activity and possibly away from other

activities. If the rewarded activity is more valuable than the foregone activities, the net substitution effect will be positive. If the providers substitute away from activities that have greater value in achieving quality or other objectives (such as, for example, time communicating with patients), the net substitution effect on health system objectives could be negative.

Net spillover effects

P4P programmes may have positive spillover effects that are not direct objectives of the programme, or negative unintended consequences. For example, P4P implementation may improve the governance of provider organizations and increase knowledge of providers on the latest clinical practice guidelines. The programme may improve decision making through, for example, the analysis of the data generated by the P4P programme. P4P programmes also may change provider culture, for example, giving more of a voice to nurses in improving organizational performance and being more open to trying new policies and approaches (Vina et al., 2009). Negative externalities are also possible, however. For example, the P4P programme may reduce intrinsic motivation or cause provider staff to become less team-oriented, because they are competing with each other for bonuses.

Net costs

P4P programmes typically, although not always, add costs to the health system to pay for the incentive itself, as well as the data collection, analysis and verification systems, and other governance and administrative functions. There also are costs to providers in terms of complying with reporting systems or other conditions of the programme. In some cases the improved processes of care and other outcomes of the programme lead to efficiency gains and cost savings. The net costs or savings of the programme either decrease or increase resources available for other health system improvement efforts.

Summary of key findings

In common with many other authors, we too find that P4P has not produced the direct significant change in performance that many advocates hoped for in the 12 case study programmes. This result is likely to be the consequence of numerous factors, such as weak programme design, inadequate incentives, deficient metrics, perceived lack of long-term commitment to the programme, countervailing incentives, or weaknesses in evaluation methodology. We nevertheless find that important system benefits have arisen from the implementation of these early experiments in P4P, such as clarification of the goals of providers, improved processes for purchasing health services, improved measurement of provider activity and performance, and a more informed dialogue between purchasers and providers.

In short, P4P appears to be having a beneficial effect on the strategic purchasing role in health systems. Hitherto, this has been a very weak and neglected function in most systems (Figueras et al., 2005). Organizations charged with strategic purchasing, such as local governments, social health insurers and other local health agencies, have tended to act as passive reimbursers, with scant regard for the nature or quality of the services purchased. We believe that the case studies indicate that an interest in P4P is forcing strategic purchasers to pay proper attention to the fundamental building blocks of their functions, such as setting coherent strategic objectives, putting in place appropriate information and reporting systems, ensuring proper auditing and governance arrangements are in place, and paying attention to the incentives under which providers operate. P4P is creating heightened awareness of the strategic purchasing function and its proper alignment with health system objectives.

We therefore believe that – far from being intrinsically flawed – P4P will progressively become a central element of the strategic purchasing function, which includes but is not limited to provider payment methods. It is inconceivable that health systems should reject the opportunity to use improved information and evidence to secure better value for money from their services. However, it has become evident that the design of P4P programmes is a complex undertaking that must balance the competing interests of different stakeholders, and it is important to view P4P within the context of the underlying payment methods and the broader health system. If P4P is implemented in isolation, without ensuring that other policy levers are aligned with its intentions, then it is likely to disappoint. Rather P4P should be used as a basis for creating a clear focus on the chosen goals of the health system, and better aligning incentives to steer the system towards those goals.

Appendix 1.1 Parameter matrix used for data collection and analysis of case studies

Programme component	*Parameters*	
Summary of P4P programme	Name.	
	Date programme was initiated.	
Policy objectives	What were the health system problems identified that the P4P programme was designed to address?	
Base payment system	What type of underlying payment system is used to pay providers participating in the P4P programme (e.g. capitation, fee-for-service, case-based payment)?	
Stakeholder involvement	Which stakeholders are involved in developing targets and indicators?	Government agencies
		Purchasers (public or private)
		Providers/provider associations
		Other independent associations
		Patients/ advocacy groups

Provider participation	Is participation mandatory or voluntary?		
	Is the programme implemented nationally or only in some regions, or by some purchasers?		
	Which providers participate?	All (public and private)	
		Only public	
		Only private	
		Some public and some private	
	What is the number (and share) of providers who participate?	Hospitals	
		Provider groups	
		Physicians	
		Nurses	
	Participation in multiple programmes	Do providers receive revenue from multiple payers? If yes, what share comes from the payer that sponsors this programme?	
		Do some providers participate in multiple programmes run by different purchasers?	
		What is the average share of provider bonus revenue from this programme?	
		Are performance measures coordinated across multiple programmes?	
Population covered	How many people are served by providers/interventions covered by the programme?		
Dimensions of performance linked to payment	What are the domains of performance that are rewarded?		
	Quality	Structure	Investment
			Data systems
			Others
		Process	Compliance with clinical guidelines
			Coordinated care/ disease management
			Coverage of priority services
			Others
	Outcomes	Clinical outcomes	
		Morbidity	
		Mortality	
	Patient satisfaction		
	Efficiency and cost savings		
	Other requirements for participation in the programme		

(*continued*)

Appendix 1.1 (*continued*)

Programme component	*Parameters*	
Performance measures	# of indicators	
	Frequency of reporting	
	How often are indicators and targets revised?	
	How are data reported to calculate performance measures?	
	Are some measures rewarded at a higher rate than others? If yes, which ones?	
Reward/penalty	Financial	Flat rate or % of total payment?
		Is the reward/penalty capped?
		What is the (absolute) average and maximum size of the reward/penalty?
		What is the average reward/penalty as a % of total payment to provider?
	Non-financial	
	Combination	
	What share of providers receives the reward/penalty?	
	Do providers compete for the reward? Are there winners and losers?	
	How often is the reward payment made?	
	Are there any restrictions on how the reward can be used?	
Basis for reward/penalty	Absolute level of measure	Are there targets?
	Change in measure	Is there a threshold level of change that is required?
	Relative ranking	What share of top performers is rewarded?
	How is the reward calculated?	
Assessment	Who assesses indicators? Purchaser, independent agency, other?	
	How are indicators assessed?	
	Is risk adjustment used? If yes, what is the methodology or adjustment parameters?	
	Do providers have the opportunity to validate/contest the results? If yes, how?	
	Is the assessment made public?	
Payment	Made to provider organization	
	Made to a team of providers	Which providers are included on the team?
	Made to individual provider	Physicians
		Nurses

	If payment is made to a provider organization or team, what are the criteria used for distributing bonuses among individuals?	What is the distribution: e.g. what is the ratio of the highest individual payment to the lowest? And/or what portion of individuals receive no bonus payment?
Other disincentives for non-performance	Are there any disincentives other than financial penalties for non-performance (under-performance)?	
Measures taken against unintended consequences	Has there been analysis of unintended consequences or steps taken to mitigate them?	
Evaluation	Has the programme been evaluated? If yes, what was the research design (e.g. randomized controlled trial, quasi-experimental design, pre- and post-measures without a control group)?	
Results	What are the overall results of the P4P programme?	
	Trends in performance measures	
	Organizational or other changes made by providers in response to the P4P programme	
	Cost of implementing the programme	
	Savings that resulted from implementing the programme	
	Spillover effects on other quality measures (positive)	
	Unintended consequences	
	Gaming	
	Facilitating factors for the P4P programme	
	Inhibiting factors for the P4P programme	

Note

1 One study by the Centre for Health Economics of the University of York examined the cost-effectiveness of a subset of specific indicators under the UK's Quality and Outcomes Framework (QOF) that have direct clinical impact, Anne Mason, Simon Walker, Karl Claxton, Richad Cookson, Elizabeth Fenwick and Mark Sculpher, *The GMS Quality and Outcomes Framework: Are the Quality And Outcomes Framework (QOF) Indicators a Cost-Effective Use of NHS Resources?* (York: Centre for Health Economics, University of York, 2008). The study compared the cost-effectiveness of the incentive for each indicator compared to no incentive, so it was not possible to assess whether the incentives are also cost-effective relative to other ways of achieving improvements in the indicators.

References

AHRQ (2012) *Public reporting as a quality improvement strategy*. Washington, DC: Agency for Healthcare Research and Quality.

Busse, R. et al. (2011) *Diagnosis-related groups in Europe*. (EO Series, Ed.) Maidenhead: Open University Press.

Campbell, S., MacDonald, R. and Lester, H. (2008) The experience of pay for performance in English family practice: a qualitative study, *Annals of Family Medicine*, 6(3): 228–34.

Cassel, C. and Jain, S. (2012) Assessing individual physician performance: does measurement suppress motivation, *Journal of the American Medical Association*, 307(24): 2595–6.

Christianson, J., Knutson, D. and Mazze, R. (2006) Physician pay-for-performance: implemenation and research issues, *Journal of General Internal Medicine*, 21(S2): S9–13.

Christianson, J., Leatherman, S. and Sutherland, K. (2007) *Paying for quality: understanding and assessing physician pay-for-performance initiatives*. Princeton, NJ: Robert Wood Johnson Foundation.

Conrad, D. and Christianson, J. (2004) Penetrating the 'black box': financial incentives for enhancing the quality of physician services, *Medical Care Research and Review*, 61(3 suppl): 37S–68S.

Damberg, C. et al. (2009) Taking stock of pay-for-performance: a candid assessment from the front lines, *Health Affairs*, 28(2): 517–25.

Dudley, R. A. et al. (1998) The impact of financial incentives on quality of health care, *Milbank Quarterly*, 76(4): 649–86.

Eijkenaar, F. (2011) Key issues in the design of pay for performance programs, *European Journal of Health Economics*, 14(1): 117–31.

Figueras, J., Robinson, R. and Jakubowski, E. (2005) *Purchasing to improve health systems performance*. Maidenhead: Open University Press.

Flodgren G. E. M. (2011) An overview of reviews evaluating the effectiveness of financial incentives in changing healthcare professional behaviours and patient outcomes, *Cochrane Database of Systematic Reviews*, 7.

Frolich, A. et al. (2007) A behavioral model of clinician responses to incentives to improve quality. *Health Policy*, 80(1): 179–93.

Guthrie, B., Auerback, G. and Binman, A. (2010) Competition for Medicaid enrolles based on performance does not improve quality of care, *Health Affairs*, 29(8): 1507–16.

Institute of Medicine (1999) *To err is human: building a safer health system*. Washington DC: National Academics Press.

Institute of Medicine (2001) *Crossing the quality chasm: a new health system for the 21st century*. Washington, DC: National Academies Press.

Mannion, R. and Goddard, M. (2003) Public disclosure of comparative clinical performance data: lessons from the Scottish experience, *Journal of Evaluation of Clinical Practice*, 9(2): 277–86.

Martin, Jenkins and Associates Limited (2008) *Evaluation of the PHO peformance programme: final report*. Auckland: Martin, Jenkins and Associates Limited.

Mason, A. et al. (2008) *The GMS quality and outcomes framework: are the quality and outcomes framework (QOF) indicators a cost-effective use of NHS resources?* York: Centre for Health Economics, University of York.

Mullen, K., Frank, R. and Rosenthal, M. (2009) *Can you get what you pay for? Pay-for-performance and the quality of healthcare providers*. Cambridge, MA: National Bureau of Economic Research.

Robinson, J. (2001) Theory and practice in the design of physician payment incentives, *Milbank Quarterly III*, 79: 149–77.

Robinson, J. et al. (2009) Financial incentives, quality improvement programs, and the adoption of clinical information technology, *Medical Care*, 27(4): 411–17.

Scott, A. et al. (2009) The effect of financial incentives on the quality of health care provided by primary care physicians, *The Cochrane Library*, 9.

Sutton, M. et al. (2012) Reduced mortality with hospital pay for performance in England, *New England Journal of Medicine*, 367: 1821–8.

Van Herck, P. et al. (2010) Systematic review: effects, design choices, and context of pay-for-performance in health care, *BMC Health Services Research*, 10: 247–60.

Vina, E. et al. (2009) Relationship between organizational factors and performance among pay-for-performance hospitals, *Journal of General Internal Medicine*, 24(7): 833–40.

Woolhandler, S. and Ariely, D. (2011) Will pay for performance backfire: insights from behavioral economics, *Health Affairs Blog*, 11 October (http://healthaffairs.org/blog/2012/10/11/will-pay-for-performance-backfire-insights-from-behavioral-economics/, accessed 22 December 2011).

chapter two

P4P programme design

Cheryl Cashin

Introduction

Pay for performance (P4P) programmes are intended to achieve a wide range of stated objectives, from improving clinical quality of care and coverage of priority preventive services to counteracting the adverse incentives of fee-for-service payment through better care coordination and integration, reducing health disparities, or improving the use of data and information technology. Within P4P programmes, a distinction can be made between those targeted at primary care providers or specialist physicians and those targeted at hospitals. While the programmes have the same conceptual basis, their objectives and scope often are quite different because of the performance problems they are trying to solve, the way care is organized, and data availability. In primary care, the objectives are typically broad based, focusing on covering a larger share of the population with evidence-based services delivered according to clinical guidelines. Hospital P4P programmes often are more narrowly defined to solve particular quality problems, such as reducing avoidable complications or adherence to clinical guidelines in specific clinical areas.

All P4P programmes include a common set of four basic elements, with a wide variety of choices made within those elements to meet different objectives. As shown in Figure 2.1, the common elements include: (1) performance domains and measures; (2) basis for reward or penalty; (3) nature of the reward or penalty; (4) data reporting and verification. The chapter addresses each element in turn, drawing on the 12 case study P4P programmes for illustration. Table 2.1 summarizes key design features.

Performance domains and measures

The first component of a P4P programme is the definition of the aspects (domains) of performance tied to the incentive and the metrics, or indicators.

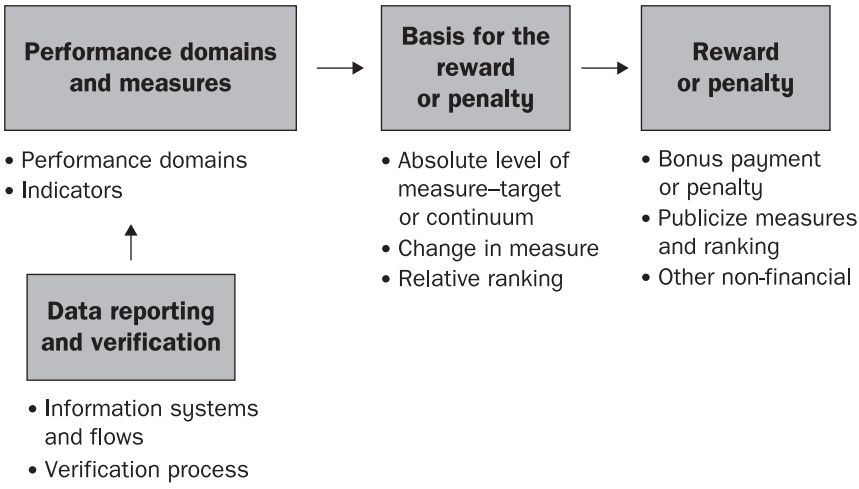


Figure 2.1 Common elements of P4P programmes

Source: Adapted from Scheffler, 2008.

Depending on the objectives of the P4P programme, different aspects of provider performance are selected for reward or penalty. Programmes tend to select performance measures relating to specific conditions that are widespread and contribute significantly to the overall burden of disease (such as cardiovascular disease), and where a particular problem has been identified (such as low coverage of vaccinations or inconsistent compliance with clinical guidelines).

The most common performance domain found in P4P programmes is **clinical quality**. The quality domains and measures follow the well-known paradigm of structure, process, and outcomes (Donabedian, 1966). Structure refers to the health care setting, including the facility, equipment, supplies, pharmaceuticals, information technology, and human resources. In the Australia PIP, for example, GP practices are rewarded for investing in infrastructure, such as computerization, or expanding services, such as providing after hours care or care in residential facilities. Process, broadly defined, is the set of procedures used to provide health care services, including how practice guidelines and disease management protocols are used. P4P programmes often use process measures related to clinical guidelines, such as the percentage of registered diabetes patients who have received the recommended cycle of care (Australia PIP, Estonia QBS, France ROSP, Germany DMP, UK QOF).

Outcome measures are the most difficult to implement and rarely include mortality or morbidity. Outcome measures better reflect the results that patients, and purchasers, want to achieve, but there are many challenges with outcome measures. An individual patient outcome is determined by many factors beyond the effectiveness of medical care, so risk adjustment of outcome measures is necessary to avoid penalizing providers who treat higher risk

Table 2.1 Summary of case study P4P programme design

Country	*Programme*	*Performance domains*	*Number of indicators*	*Basis for reward or penalty*	*Nature of reward or penalty*		
					Bonus, penalty or withhold as share of income		
Australia	Practice Incentives Programme (PIP)	Quality Capacity Rural support	15	*Absolute* Flat rate for participation, targets, and per patient reached	4–7 per cent	of total primary care practice income	Bonus
Brazil – São Paulo	Social Organizations in Health performance-based contracting (OSS)	Information quality Efficiency Patient satisfaction	9	*Varies* Performance targets negotiated for each OSS	10 per cent	of hospital budget	Withhold
Estonia	PHC Quality Bonus System (QBS)	Disease prevention Chronic disease management Additional activities	62	*Absolute* Minimum target threshold (lower payment) Target met (higher payment)	up to 4.5 per cent	of family doctor income	Bonus
France	Payment for Public Health Objectives (ROSP)	Prevention Chronic disease management Cost-effective prescribing Practice organization	29	*Absolute and change over time* Achievement rate-progress toward target relative to baseline performance	up to 11 per cent	of annual GP net income	Bonus

(*continued*)

Table 2.1 Summary of case study P4P programme design (*continued*)

Country	*Programme*	*Performance domains*	*Number of indicators*	*Basis for reward or penalty*	*Nature of reward or penalty*		
					Bonus, penalty or withhold as share of income		
Germany	Disease Management Programmes (DMP) [varies by region and sickness fund]	Documentation Follow-up of patients Additional services Training	10	*Absolute* Flat rate for participation and per service	Flat rate Additional payments Incentives	per enrolled patient per indicator met/service provided for patients	Flat rate payments
Korea, Rep. of	Value Incentive Programme (VIP)	Quality	7	*Relative ranking*	2 per cent 1 per cent	of total insurance payment to hospitals (highest/lowest performers) next grade of performers	Bonus/ Penalty
New Zealand	Primary Health Organization (PHO) Performance Programme	Chronic disease screening Prevention of infectious diseases	10	*Absolute* Per cent attainment of target	<1 per cent	of government budget for PHOs	Bonus
Turkey	Family Medicine Performance-Based Contracting Scheme	Coverage of maternal and child health services Governance Service delivery Quality	8 35 (warning system)	*Absolute* Per cent attainment of target with sliding scale for different achievement rates	up to 20 per cent	of the base payment of the family medicine unit	Withhold

UK	Quality and Outcomes Framework (QOF)	Clinical care Organizational Patient experience Additional services	142	*Absolute* Per cent of target met after minimum threshold is reached	~25 per cent	of GP practice income	Bonus
US – California	Integrated Healthcare Association (IHA) Programme	Clinical quality Meaningful use of health information technology Patient experience Appropriate resource use	78	*Varies by insurer*	<2 per cent	of provider group's per capita payment	Bonus
US – Maryland	Maryland Hospital Acquired Complication Programme	Potentially preventable complications	49	*Relative ranking*	per cent determined by reallocation formula	of hospital's inflation adjustment	Reallocation from low performers to high performers
US – National	Hospital Quality Incentive Demonstration	Clinical quality	33	*Relative ranking* and *Achievement of median performance level*	2 per cent 1 per cent	of total insurance payment to hospitals (highest/lowest performers) next grade of performers	Bonus/penalty

patients. There also are challenges with the validity of measures, sample sizes, and 'surveillance bias', or detecting more negative outcomes that are more closely scrutinized (Berenson, Pronovost & Krumholz, 2013). The US HQID and Korea VIP programmes, both hospital P4P programmes, are among the very few programmes that include mortality measures, and they are specifically for inpatient mortality for acute myocardial infarction and coronary artery bypass graft, which are more directly attributable to hospital performance (Premier Inc., 2006). In general, outcomes measures in P4P programmes are confined to intermediate clinical outcomes, such as blood pressure control, blood sugar levels, and cholesterol levels (California IHA, France ROSP, and UK QOF), or avoidable complications (Maryland HAC).

The Maryland HAC programme was one of the first programmes to penalize substandard clinical quality. In this programme, hospitals that have higher rates of potentially preventable complications are penalized with a reduction in their annual inflation adjustment. These funds are reallocated to better performing hospitals in the form of an increase in their annual inflation adjustment. Poor performance is also penalized in the Brazil OSS programme, Korea VIP, Turkey FM PBC scheme, and the US HQID.

Other performance domains that are commonly found in P4P programmes include **coverage of priority services** (such as immunization and screening for cancer and other chronic diseases), and **efficiency**. This is where P4P programmes often differ between higher and lower income countries. In high income countries, particularly those that rely on fee-for-service as the base provider payment method, the efficiency problem often is to constrain the ever-increasing demand for more and higher technology health services, and inefficient and fragmented care, particularly for chronic diseases. For example, primary care providers may be rewarded for patients using a below-average number of specialist services, inpatient hospitalizations, or branded medications. This type of measure creates the incentive for the primary care provider to internalize a portion of the health care costs that it influences but does not directly provide. One example is Medicare's Physician Group Practice Demonstration, which rewards physician groups for achieving lower cost growth (Colla et al., 2012). This type of P4P is known as 'shared savings'. Several higher income countries, such as France and New Zealand, specifically target pharmaceutical expenditures in their efficiency domains. Less common, but a promising direction for future initiatives, are attempts to reward both efficiency and quality achieved by better continuity of care and chronic disease management. The P4P programmes in France, Germany, Estonia, and the UK are moving in this direction, as well as the US Medicare accountable care organization (ACO) programmes (McWilliams & Song, 2012).

In many low-income countries with largely public health provision, where health personnel are salaried civil servants, the efficiency problem is often related to low productivity and lack of coverage of key public health interventions like immunization and antenatal care (Eichler et al., 2009). The goal in these contexts is to increase utilization, particularly for high priority services at higher quality. A number of P4P programmes in low-income countries such as Afghanistan, Burundi, and Rwanda pay providers per-service payments, adjusted by a quality score, for delivering a list of priority services

(Basinga et al., 2011; Bonfrer et al., 2013; World Bank, 2013). Other examples include rewarding physicians to work in the public sector instead of the private sector (e.g. Turkey), or to diagnose patients with tuberculosis (e.g. China) (Scheffler, 2008).

Other domains of performance are now more commonly being rewarded in P4P programmes, including **patient experience** or **satisfaction** (e.g. California IHA, UK QOF) and improved **equity**, or the **reduction of health disparities**. In the New Zealand PHO Performance Programme, for example, some indicators are measured separately for 'high-need populations', which are rewarded at a higher rate.

The number and range of performance domains and related indicators is a key feature of the design of P4P programmes, but there is little guidance in the literature about the right balance. Finding the right balance in the number of indicators is particularly challenging given the large gaps in available measures (Berenson, Pronovost & Krumholz, 2013). Fewer performance domains and indicators make the programme simpler to administer and provide clearer incentives, but the programme is more likely to be distorting, risking an overemphasis by providers on rewarded services or aspects of performance. Many domains and indicators may provide a more balanced set of incentives, but these programmes are more complex to administer and may dilute the incentives to providers (Eijkenaar, 2011).

Selecting the number and range of indicators, as well as the particular service areas, therefore, needs to strike a balance between having enough indicators to capture important aspects of performance and limit distortions, and not making the system overly complex so that it is administratively burdensome and the incentives become unclear. Some argue that having a larger set of indicators may somewhat guard against the risk of 'teaching to the test', or providers focusing disproportionately on those areas of care tied to an incentive payment, by assessing care more comprehensively and driving improvements more broadly (Damberg et al., 2009). With the critical gaps in available measures, however, a large set of indicators may do little to reduce such distortions, while adding to the reporting burden on providers. The Maryland HAC programme chose indicators that reflect broad-based measures of quality with indicators related to 49 complications that affect care across nearly all product lines of a full service hospital. This approach is considered to have provided an incentive to implement systematic approaches to reducing complications across all diagnoses, as opposed to targeting or reallocating resources to certain quality metrics tied to the incentive (Murray, 2012).

Some programmes limit the set of indicators but include at least those areas of care with high prevalence or disease burden (Eijkenaar, 2011). Other P4P programmes use a small number of proxies to capture one or several dimensions of clinical quality, while some try to combine multiple indicators to capture several points along the care continuum. The Australia PIP, for example, provides a one-off bonus for primary care practices using a diabetes register and call reminder system (structure), and per-patient bonus payments for diabetes patients who complete the recommended cycle of care (process). The UK QOF uses 142 indicators in an attempt to capture the full quality continuum from prevention all the way to clinical outcomes. One clinical area alone

(coronary heart disease) includes eight indicators, from recording, to initial and ongoing management, to clinical outcomes (see Figure 2.2). Given the gaps in performance measures, however, even such a large set of process indicators misses key areas of quality, such as diagnostic errors, appropriateness of care, and care for complex patients with multiple conditions (Berenson, Pronovost & Krumholz, 2013).

The final number and set of indicators tend to be driven by both negotiations between stakeholders and the limitations of current data and information systems. Most programmes settle on between 10 and 30 indicators that come from existing reporting systems, although some P4P programmes have driven the development of new data sources and refined information systems. Exceptions among the programmes reviewed for this volume include the UK QOF, California IHA programme, Maryland MHAC, and Estonia QBS. The UK QOF rewards four performance domains and uses 142 indicators. The California IHA programme includes four performance domains and 78 indicators, and the Estonia QBS includes three performance domains and 62 performance indicators. The Maryland HAC includes only one performance domain (Potentially Preventable Complications) but measures complication rates for 49 complications.

Programmes often weight performance domains or indicators differently to reflect different priorities. The UK QOF places priority on the clinical

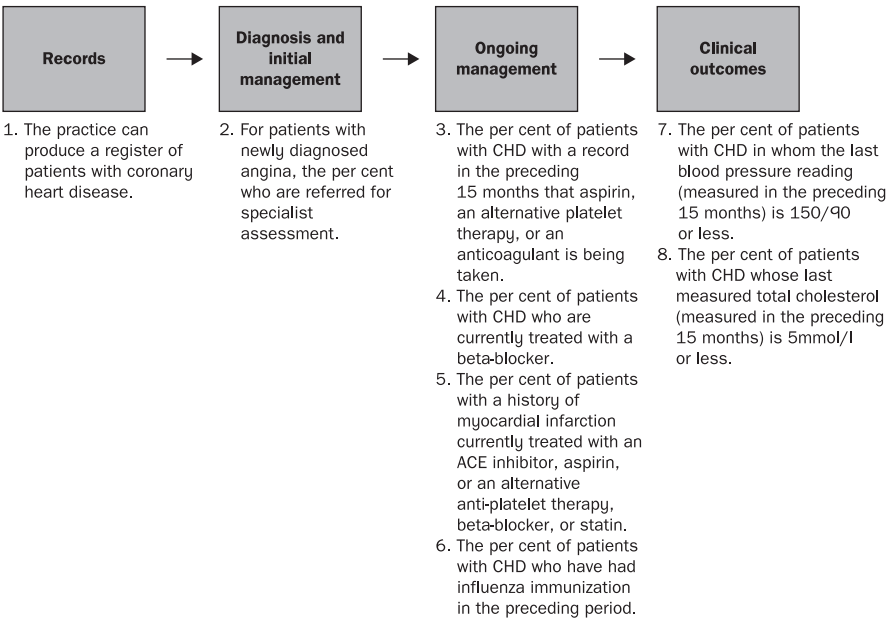


Figure 2.2 UK QOF measurements along the continuum of clinical quality (coronary heart disease indicators)

Source: Author's depiction from NHS Primary Care Commissioning, 2011.

care domain, which accounts for 70 per cent of possible points. In the France ROSP programme, emphasis is put on practice organization and efficiency (60 per cent) compared to chronic disease management and prevention (40 per cent). In the Estonia QBS programme, priority is placed on hypertension, which accounts for 41 per cent of possible points for the disease prevention and chronic disease management domains combined. The New Zealand PHO Performance Programme gives weight to reducing health disparities by scaling payments upwards for progress against targets for high needs populations.

Typically providers are required to participate in all of the performance domains, but the Australia PIP is an exception in which providers are able to selectively participate in incentive domains. The uptake and payment across incentive areas in the Australia PIP is highly skewed as a result, with relatively high-payment/low-effort incentive areas the most popular. Whereas computerization of GP practices ('eHealth') accounted for 33 per cent of all incentive payments (reflecting both higher uptake and relatively higher reward), all three priority service areas combined only account for 11 per cent of the total payout in 2008–09.

The appropriate definition of domains and indicators ultimately is context specific, and it should reflect the priorities of patients and purchasers. Performance measures are most credible when they reflect consensus among a wide range of stakeholders about what constitutes good performance and how it should be measured. Performance measures that are grounded in widely accepted clinical guidelines are more likely to be accepted by providers, but there also are challenges basing performance indicators on clinical guidelines (see Chapter 3). The performance measures, and the entire P4P programmes, are more readily accepted by providers when there is transparency and stakeholder participation in the development of the programmes and performance measures in particular (Martin, Jenkins & Associates Limited, 2008; Murray, 2012). A more detailed discussion of the relationship between performance domains and measures and health system objectives and governance is provided in Chapter 3.

Basis for reward or penalty

The second element of a P4P programme is the basis for reward or penalty, or how achievement against performance indicators is used to determine the level of the incentive payment earned by the provider. The most common options include: the absolute level of the measure (e.g. whether a target was achieved or the number of patients reached); the change in the performance measure (improvement), or how the provider performs against the measure relative to other providers (relative ranking). Most P4P programmes reward or penalize providers for each performance domain and indicator separately, but the Turkey FM PBC takes an 'all or nothing' approach, with providers only avoiding penalty if all members of the family medicine unit reach the targets for all of the indicators. Each of these approaches has some limitations, which are discussed below, and several programmes such as the California IHA programme, the France ROSP programme, and the US HQID use a combination.

Absolute measures

Absolute measures of achievement are the most common across P4P programmes, either paying based on targets or per patient reached. The use of absolute measures can create uncertainty for the purchaser about the amount of financial payment at risk. Of the programmes reviewed for the case studies, only the Australia PIP and the Germany DMP used the number of services provided as the basis for reward payment.

For targets, the reward is typically based on some combination of threshold targets being reached with additional payment possible up to an upper limit. For example, in the UK QOF, each indicator has a maximum point value, with a grand total of 1000 points possible. After reaching the minimum threshold for the indicator (e.g. 40 per cent of patients with coronary heart disease have blood pressure recorded within the past 15 months) providers are eligible for the minimum number of points. Providers then accumulate points up to a maximum threshold for the target (90 per cent of patients reached) and the point value for the indicator (17 possible points for recording blood pressure for patients with coronary heart disease). The use of targets for measuring and rewarding performance has been controversial. Targets can require elaborate risk adjustment mechanisms to account for different patient or population groups. Furthermore, targets do not provide incentives for providers who already have achieved upper limits, and they may encourage providers to focus on patients who are easier to reach, particularly if risk adjustment is inadequate. On the other hand, targets may help to focus performance priorities and make programmes more transparent and objective (Martin, Jenkins & Associates Limited, 2008). One approach, used by the New Zealand PHO Performance Programme, has been to set provider-specific targets that are adjusted each period as performance and priorities change. The France ROSP programme also takes provider baseline performance against national targets into account when computing achievement rates for bonus payment.

Improvement

The reward in a P4P programme also can be based on the change in a measure over time, or improvement. Using a provider's improvement as the basis for the reward often has more intuitive appeal for providers, but creating an individual basis for each provider's reward makes the system more complex and resource intensive. In the France ROSP programme, the reward is calculated using a formula that incorporates the individual provider's baseline value for the indicator, performance improvement, and national targets. In the New Zealand PHO Performance Programme, providers receive the full incentive payment if targets are reached, or partial payments if the target is not reached but improvement was achieved (Martin, Jenkins & Associates Limited, 2008). The standard methodology in the California IHA programme suggests that physician groups be scored on both attainment and improvement for each measure with the higher of the two used for each measure summed to the domain total, which is then weighted.

Use of improvement metrics as the basis for reward can encourage continual progress. Furthermore, it reduces the need for complex adjustment for case mix and other measurement challenges. Improvement measures can, however, favour organizations that were originally poor performers for which there is most scope for improvement, and therefore might be perceived to be rewarding previously poor performance. Furthermore, unless carefully designed, improvement measures can in some circumstances inhibit the search for improvements if the 'reward' for improvement is a tougher target in the future.

Relative ranking

The third option for the basis of the reward in P4P programmes is the performance of providers relative to others. The potential benefit of this approach is that it encourages greater effort among top performers because of the threat of being overtaken. Relative ranking also provides a means of filtering out common random shocks among providers that might affect performance, such as an epidemic or recession. In the US HQID programme hospitals in the top 20 per cent of achievement scores receive an incentive payment, with hospitals in the top 10 per cent receiving a higher reward. This competitive model is also used in the Korea VIP, which rewards the top performers in terms of quality improvement over time, and the Maryland HAC, which redistributes penalties from low performers to high performers as a bonus. The main concern raised with the relative ranking approach is that meaningful incentives may not operate on low performers, who may be the most urgent priorities for improvement and in greatest need of additional resources to improve performance (Damberg et al., 2009). Such 'tournaments' may therefore exacerbate inequalities, and penalize patients who already suffer from poor providers, especially in health systems where patients have little provider choice.

Calculation of achievement rates and risk adjustment

The way achievement rates are calculated also varies across P4P programmes. Most programmes rely on simple, transparent calculation methods, although some use more complicated formulae or composite measures. While these measurement approaches may allow more granularity and gradation in measuring quality differences, complicated measures that are not immediately clear to providers may risk diluting the incentive (Eijkenaar, 2011). In the US HQID programme, separate achievement scores were calculated for each clinical condition by 'rolling-up' individual process and outcome measures into an overall quality score (Premier Inc., 2006). The Korea VIP also uses a composite quality score for its two performance domains, quality of acute myocardial infarction (AMI) care and Caesarean section rate. For the AMI performance calculation, the programme uses a weighted average of five process indicators and one outcome indicator (risk-adjusted 30-day mortality

rate). Performance for the Caesarean section rate indicator is calculated as the difference between the observed rate and the expected rate, which is based on 15 clinical risk factors. The France ROSP programme has three different formulae for calculating achievement rates depending on the provider's baseline achievement rate relative to both a benchmark achievement rate (average achievement rate across providers) and a target achievement rate.

One concern about P4P programmes is that providers serving healthier people or those more likely to adhere to recommended care can demonstrate better performance with less effort. There may be an incentive to avoid sicker and more difficult patients (known as 'risk selection'), which would have a negative impact on equity. As a result, P4P programmes, and provider payment systems in general, often build in adjustments to compensate providers who serve a disproportionately higher risk (sicker and costlier) population so the incentive to avoid these patients is reduced. This is particularly important in P4P programmes that tie rewards or penalties to mortality, such as the US HQID and Korea VIP, both of which use commonly accepted risk adjustment measures for hospital mortality. The Maryland HAC calculates a hospital's achievement rate from the hospital's actual rate of preventable complications versus the expected rate given the severity of the hospital's patient case mix. Risk adjustment does not completely solve the problem of risk selection, however, and the models are complex and can lead to inaccurate results, particularly for diverse patient populations (Berenson, Pronovost & Krumholz, 2013; Wennberg et al., 2013).

The UK QOF, in addition to adjusting payments for practice size and disease prevalence relative to national average values, allows practices to 'exception-report' (exclude) certain patients from data collected to calculate achievement scores (Table 2.2). Patient exception reporting applies to those indicators in the clinical domain of the QOF where the level of achievement is determined by the percentage of patients receiving the designated level of care. Patients can be excluded from individual indicators if, for example, they do not attend appointments or where the recommended treatment is judged to be inappropriate by the GP (e.g. medication cannot be prescribed due to side effects). Of course a major concern with such mechanisms is the reliance on provider self-reporting of exceptions.

Nature of reward or penalty

The third common element of P4P programmes is the reward or penalty, which may be financial or non-financial, or a combination of both. Rewards are often a bonus or lump sum payment, or as in the case of the Korea VIP, Maryland HAC, and US HQID programmes, there can be increases (or decreases) in the rate of payment or rate of increase in payment. There are three main characteristics of the reward/penalty for P4P programmes: (1) the size of the reward/penalty; (2) the recipient (institutions or individuals); and (3) whether the financial reward/penalty is accompanied by non-financial incentives.

Table 2.2 Criteria for exception reporting under the UK QOF

1	Patients who have been recorded as refusing to attend review who have been invited on at least three occasions during the preceding 12 months.
2	Patients for whom it is not appropriate to review the chronic disease parameters due to particular circumstances, e.g. terminal illness, extreme frailty.
3	Patients newly diagnosed or who have recently registered with the practice who should have measurements made within three months and delivery of clinical standards within nine months, e.g. blood pressure or cholesterol measurements within target levels.
4	Patients who are on maximum tolerated doses of medication whose levels remain sub-optimal.
5	Patients for whom prescribing a medication is not clinically appropriate, e.g. those who have an allergy, contraindication or have experienced an adverse reaction.
6	Where a patient has not tolerated medication.
7	Where a patient does not agree to investigation or treatment (informed dissent) and this has been recorded in their medical records following a discussion with the patient.
9	Where the patient has a supervening condition which makes treatment of their condition inappropriate, e.g. cholesterol reduction where the patient has liver disease.
8	Where an investigation service or secondary care service is unavailable.

Source: NHS Primary Care Commissioning, 2011.

Size of the incentive

The appropriate size of the reward or penalty is a topic of much debate, but research and implementation experience provide few answers. The size of the incentive is important for creating a meaningful incentive to which providers will find it worthwhile to respond, but without distorting provider behaviour and leading to unintended consequences. The amount of the bonus or penalty that is 'meaningful' to a provider will be strongly influenced by its underlying revenues and margins. In the US, for example, hospitals operate on low margins, typically under ten per cent (AHA, 2013), so the relatively small bonus/penalty of one–two per cent of Medicare payment in the US HQID programme appears low in absolute terms, but it is meaningful for hospitals.

The P4P programmes reviewed for this volume provide relatively small rewards as a share of total provider income, typically less than five per cent. The exceptions are the UK QOF, in which about 25 per cent of GP practice income is tied to incentive payments, and the Turkey FM PBC, which ties up to 20 per cent of provider salaries to incentive payments. The France ROSP programme and the Brazil OSS programme are in the middle, with incentive payments at about ten per cent of GP and hospital income, respectively.

A number of studies have identified the small size of the incentive as a factor in the modest overall impacts of P4P on performance improvement (Damberg et al., 2009). On the other hand, larger incentives may exacerbate concerns

about unintended consequences, such as providers shifting excessive focus towards performance areas and services that are rewarded, or risk selection (Frey, 1997; Deci, Koestner & Ryan, 1999). Two reviews of published studies of P4P programmes found no consistent relationship between the size of the incentive and provider responses (Frolich et al., 2007; Van Herck et al., 2010).

In addition to the size of the incentive itself, it is important how much power a financial incentive can have to change performance, and whether the incentive payments represent new money in the system or simply a redistribution of existing funding. The power of the financial incentive to affect behaviour depends on many factors, including the reason for the underlying performance gaps, other incentives that are in place, the characteristics of the population served, and the flexibility and resources that providers have to make changes in response to the incentives (Dudley & Rosenthal, 2006). The incentive is also more powerful if it increases as performance improves. The Australia PIP, Estonia QBS, France ROSP, Germany DMP, New Zealand PHO Performance Programme, Turkey FM PBC, and the UK QOF all have higher payment rates for higher achievement levels, typically after a minimum threshold is reached.

In many systems the size of the performance-related incentive tends to be modest in relation to the incentives created by the underlying base payment system. In systems such as the US where providers receive revenue from multiple payers, the performance-related incentives are further weakened. Some programmes closely align and integrate the incentive payments with the underlying payment system as a 'blended payment system' to specifically strengthen or counteract the stronger incentives of the base payment system (e.g. Brazil OSS, Estonia QBS, Germany DMP, Maryland HAC, Turkey FM PBC, and UK QOF).

Whether the incentive payments represent new funds in the system or a redistribution of existing funds may be an important factor in both getting provider buy-in and bringing in new resources that may be needed to improve quality of care (Van Herck et al., 2010). The UK QOF and Turkey FM PBC scheme were initiated with large new infusions of funds in the primary care sector, which in the case of the UK QOF helped to overcome provider resistance initially. Programmes that are redistributive (Maryland HAC) and particularly those that involve the risk of a penalty (US HQID and Korea VIP) face potentially more resistance from providers and require careful stakeholder involvement and negotiation.

Payment to institutions or individuals

Whether the incentive payment is made to provider institutions or individuals may influence the extent to which the incentives will affect provider behaviour. Health care is increasingly provided by teams of individuals rather than solo physicians, so cooperative effort is likely needed to improve performance. On the other hand, incentives that do not reach front-line providers may have little power to change the individual behaviours that are most important for collective performance. A systematic review of published studies of P4P programmes showed that programmes that target incentives to individual providers or

teams show more positive results than those targeted to institutions (Van Herck et al., 2010). For many performance measures, however, it would not be valid to hold an individual health professional accountable for the results (Berenson, Pronovost & Krumholz, 2013). In any case, transmitting the objectives of the programme to the front-line providers is an important part of the overall impact of the programme.

Of the P4P programmes reviewed, only three give incentive payments directly to individual practitioners (Estonia QBS, France ROSP, and Turkey FM PBC). Of those, only the Turkey FM PBC links physician salary to incentive payments within the context of a larger provider organization, whereas in Estonia and France the physicians are typically solo practitioners so institutional payments are not an option. In the other nine P4P programmes reviewed, the incentive payments are made to provider institutions, which then have a large degree of freedom to determine how the payments are used. In most cases, provider institutions appear to use the flexibility of the incentive payments to make general improvements in service delivery (particularly related to performance measures), such as hiring more staff for screening or disease management, improving IT systems, or expanding outreach services. In some cases, however, the lack of guidance on the use of bonus payments has weakened the incentive or caused tensions. In the New Zealand PHO Performance Programme, for example, there are no guidelines for how PHOs should distribute bonus payments across individual providers, and the ambiguity led to some tensions and delays in using the funds in the past (Martin, Jenkins & Associates Limited, 2008). In the UK QOF, although GP incomes have increased significantly with the bonus payments, some tensions have arisen because nurse incomes have not been affected by the incentive payments and nurses are instrumental in the achievement of GP practices under the QOF (Audit Commission, 2011).

Non-financial rewards

A non-financial reward may be to publicize provider rankings based on different measures (Brazil OSS, Estonia QBS, California IHA programme, Korea VIP, Maryland HAC, New Zealand PHO Performance Programme, UK QOF, US HQID). Although public rankings are not directly financial, they can become financial if patients or insurers use the rankings to determine which provider to use. Public reporting of provider performance is not always possible, depending on the laws and norms in a country related to the privacy of health data. In the Australia PIP, for example, data on the performance of individual GP practices are not made publicly available because of concerns about patient privacy (Australian National Audit Office, 2010).

The literature on the impact of public reporting on provider behaviour and patient choices shows positive but small effects (Robinowitz & Dudley, 2006), but public reporting also serves a transparency and accountability function. In some cases, provider organizations have voiced opposition to public reporting. In the Maryland HAC, for example, the Maryland Hospital Association, which is largely an advocacy organization on behalf of the 46 acute care hospitals, was

not involved in the development of the web-based reports and indicated their opposition to public reporting on methodological grounds.

Data reporting and verification

Data availability is a key determinant of both the design of P4P programmes and their ability to drive performance improvement. P4P programmes often rely on claims data, which are typically the most readily available. Claims data can be a useful starting point, but they are not designed to measure performance and can provide an incomplete picture of provider activity. Most P4P programmes therefore eventually either move away from claims data or supplement claims data with other data sources (Eijkenaar, 2011). The UK QOF is one of the few P4P programmes that relies mainly on data that are extracted anonymously from electronic medical records.

Verification is the process through which the purchaser measures and validates the results that are being rewarded or penalized. Verification is a critical element in fiduciary processes and discharge of financial responsibilities in line with the contractual arrangement. It is of particular interest to governments, which are sensitive to the potential for 'overpayments' based on inflated reporting or other possible gaming. Verification is an important opportunity for a two-way dialogue between the purchaser and providers about current performance, barriers to improvement, and the joint efforts that may be necessary to make performance improvement for individual providers. The role of data and information systems, verification, and the feedback loop of information between purchasers and providers in P4P programmes are examined in depth in Chapter 3.

Conclusions

All P4P programmes include a common set of design decisions with a wide variety of options within each. The design decisions should be based on the objectives of the P4P programme – in particular the priorities of patients and purchasers. But options are constrained by system factors, particularly data availability. Also, the P4P programme design almost always evolves and bends through negotiations with providers and other stakeholders. Ultimately, P4P programme design and implementation arrangements reflect factors and objectives that sometimes conflict, and compromises that sometimes weaken the overall incentives but make the programmes feasible.

References

AHA (2013) *Chartbook: Trends affecting hospitals and health systems* (http://www.aha.org/research/reports/tw/chartbook/ch4.shtml, accessed June 2013).

Audit Commission (2011) *Paying GPs to improve quality*. London: Audit Commission.

Australian National Audit Office (2010) *Practice incentives program audit report No. 5 2010–2011*. Canberra: Australian National Audit Office.

Basinga, P. et al. (2011) Effect on maternal and child health services in Rwanda of payment to primary health-care providers for performance: an impact evaluation, *The Lancet*, 377(9775): 1421–8.

Berenson, B., Pronovost, P. and Krumholz, H. (2013) *Achieving the potential of health care performance measures*. Washington, DC: Urban Institute.

Bonfrer, I. et al. (2013) The effects of performance-based financing on the use and quality of health care in Burundi: an impact evaluation, *The Lancet*, 381: S19.

Colla, C. et al. (2012) Spending differences associated with the Medicare Physicians Group Practice Demonstration, *Journal of the American Medical Association*, 308(10): 1015–23.

Damberg, C. et al. (2009) Taking stock of pay-for-performance: a candid assessment from the front lines, *Health Affairs*, 28(2): 517–25.

Deci, E., Koestner, R. and Ryan, R. (1999) A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation, *Psychology Bulletin*, 125: 627–68.

Donabedian, A. (1966) Evaluating the quality of medical care, *Milbank Memorial Fund Quarterly*, 44(3): 166–206.

Dudley, R. A. and Rosenthal, M. (2006) *Pay for performance: a decision guide for purchasers*. Washington, DC: Department of Health and Human Services, Agency for Healthcare Research and Quality.

Eichler, R., Levine, R. and the Performance-Based Incentives Working Group (2009) *Performance incentives for global health: potential and pitfalls*. Washington, DC: Center for Global Development.

Eijkenaar, F. (2011) Key issues in the design of pay for performance programs, *European Journal of Health Economics*, 14(1): 117–31.

Frey, B. (1997) On the relationship between intrinsic and extrinsic work motivation, *International Journal of Industrial Organization*, 15(4): 427–39.

Frolich, A. et al. (2007) A behavioral model of clinician responses to incentives to improve quality, *Health Policy*, 80(1): 179–93.

Martin, Jenkins and Associates Limited (2008) *Evaluation of the PHO peformance programme: final report*. Auckland: Martin, Jenkins and Associates Limited.

McWilliams, J. and Song, Z. (2012) Implications for ACOs of variations in spending growth, *New England Journal of Medicine*, 366(19): e29.

Murray, R. (2012) *United States Maryland Hospital Acquired Complication Program*. Paris: Organization on Economic Co-operation and Development.

NHS (2003) *Practice QOF assessment visits*. London: UK National Health Service.

NHS Primary Care Commissioning (2011) *Quality and outcomes framework: guidance for GMS contract 2011/12*. UK National Health Service Primary Care.

Premier Inc. (2006) *Centers for Medicare and Medicaid Services (CMS)/Premier Hospital Quality Incentive Demonstration Project: project overview and findings from year one* (http://www.premierinc/hqi-whitepaper041306.pdf, accessed July 2010).

Robinowitz, D. and Dudley, R.A. (2006) Public reporting of provider performance: can its impact be made greater? *Annual Review of Public Health*, 27: 517–36.

Scheffler, R.M. (2008) *Is there a doctor in the house? Market signals and tomorrow's supply of doctors*. Stanford, CA: Stanford University Press.

Van Herck, P. et al. (2010) Systematic review: effects, design choices, and context of pay-for-performance in health care, *BMC Health Services Research*, 10: 247–60.

Wennberg, J. et al. (2013) Observational intensity bias associated with illness adjustment: cross sectional analysis of insurance claims, *British Medical Journal*, 346: F549.

World Bank (2013) *Using results-based financing to achieve maternal and child health – progress report 2013*. Washington, DC: World Bank.

chapter three

Strengthening health system governance through P4P implementation

Cheryl Cashin and Michael Borowitz

Introduction

Governance structures in the health sector should align incentives for health care providers with the organizational and service delivery strategies designed to improve the quality of care and health outcomes. These organizational and service delivery strategies, such as adhering to clinical guidelines or engaging in outreach and follow-up, may not be adopted by providers for a variety of reasons. Providers may not be aware of them, may not agree with them, or they may not have the resources to make the required investments. Furthermore, providers may not have sufficient information about what they are currently doing, or the financial incentives in the underlying payment system may direct providers away from these strategies. Better governance of the health system seeks to create the institutional arrangements and rules that influence provider behaviour to adopt these organizational and service delivery strategies and hold them accountable for results.

Governance is about the rules that guide the roles and responsibilities of different actors and how they relate to each other. Governance of the health system involves putting in place mechanisms that 'steer' the health system toward desired objectives. Mechanisms might include ensuring that strategic policy frameworks are in place that set clear system objectives and priorities, creating the right regulatory environment and incentives, and using appropriate performance monitoring instruments and accountability measures (World Health Organization, 2010; Smith et al., 2012). These governance structures function at the national, local and organizational levels, and might include various regulatory mechanisms, electoral processes, markets that promote patient choice, or professional oversight and accreditation.

Governance arrangements can range from hierarchical, to market based, to network driven. A hierarchical governance structure relies on top-down definition of rules and resource allocation, whereas market-based governance structures place more emphasis on purchasing, regulation, and incentives. Network-based

governance structures establish common values and knowledge, and manage accountability through professional norms and information sharing (Smith et al., 2012). A mix of these different governance arrangements exists at different levels in the health sector. Strategic purchasing and provider payment typically make up part of the market-based governance structures, even in countries such as the UK with a traditionally more hierarchical organization of the health system (Smith et al., 2012).

Good governance in the health system can create or strengthen a 'virtuous cycle' of performance improvement, in which performance objectives are clear, data and information shed light on strategies that are working well and where more effort is needed, providers are accountable for results, and performance of the overall system can continuously improve (Figure 3.1).

Provider payment systems can be a tool to improve health system governance by clarifying roles and relationships between purchasers and providers, and creating the right incentives to guide health provider behaviour toward reaching health system objectives. Most traditional provider payment systems do not by themselves contribute to strengthening the governance functions in the health sector, however, and in fact often work against them. For example, fee-for-service payment systems can create incentives that frustrate progress toward increased prevention, care coordination and chronic disease management. Also, traditional payment systems, particularly capitation, do not generate adequate data for performance monitoring, and therefore do not allow payers and patients to hold providers properly accountable for many aspects of performance.

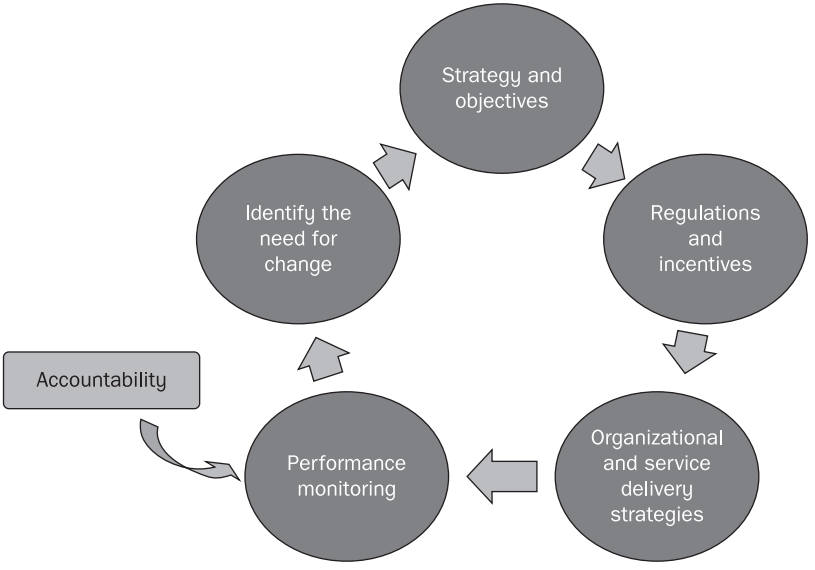


Figure 3.1 Health sector governance and the performance improvement cycle

Source: Author's adaptation of Performance Management Cycle (Public Health Foundation, 2013).

There is a movement in the health systems of some higher income countries toward reorienting service delivery toward more integrated, coordinated care and aligning payment systems with broader accountability for patient outcomes (Miller, 2011; Stock et al., 2011). One example is the experiments with bundled payment for integrated chronic disease care in the Netherlands (Struis & Baan, 2011). This approach would make better use of provider payment systems to reinforce governance structures and accountability in the health sector and could be considered a movement from 'pay for performance' to 'pay for value' (Berenson, 2010). This transition is in the early stages, however, and the evidence on different models is just beginning to emerge. In the meantime, P4P programmes can be used strategically to complement traditional payment systems to focus the attention and efforts of providers on objectives, create incentives for better generation and use of data, and provide a direct way to increase accountability for performance. If P4P works effectively, it may help create the foundation for a more fundamental shift in underlying provider payment systems that are aligned with improved governance structures and processes.

The remainder of this chapter discusses how the implementation of P4P programmes both requires and can strengthen health sector governance structures and processes based on the experience of the 12 case study programmes. Table 3.1 provides a summary of the key aspects of the case study programmes related to governance and accountability, including the underlying strategies, governance structures and stakeholder involvement in the programmes, data sources and flows, feedback mechanisms, and public reporting of performance results.

The role of P4P in strengthening governance and the performance improvement cycle

P4P programmes can play an important role in strengthening the health system governance cycle: sharpening the focus on strategic objectives; creating incentives to adopt evidence-based clinical guidelines and other service delivery approaches; better generation and use of information; creating or strengthening feedback loop so purchasers, providers, patients and policymakers can use information on performance to identify areas for further change and improvement. All of the case study P4P programmes are positioned within a larger health system strategy or legislation. Some programmes, such as the Brazil OSS and the UK QOF, are also tied to broader public sector reforms to improve accountability through performance-based contracting. Although the programmes typically are implemented by a public health purchaser (with the exception of the California IHA), the strategies and programme objectives are almost always developed by the government health department or ministry.

While governments are ultimately responsible for strategy and objectives, on behalf of patients and implemented by purchasers, aligning incentives to achieve better governance requires collaboration among many different stakeholder groups. Most of the case study P4P programmes have made an effort to involve stakeholders, in particular health provider associations, in the design

Table 3.1 Summary of governance structures and data sources and flows in the case study P4P programmes

Country	*Programme*	*Overarching strategy and objectives*	*Programme governance structures and stakeholder participation*	*Data sources*	*Feedback of information to providers*	*Public reporting*
Australia	**PIP**	• Department of Health and Ageing (DOHA) broader strategy to reform primary health care. • Main objective is to encourage continuing improvements in general practice through financial incentives to support quality care, and improve access and health outcomes.	• Administered by Medicare Australia on behalf of DoHA. • Payment formula developed in consultation with a negotiating body comprising the government and GP professional organizations. • DoHA regularly consults with GP professional organizations through an advisory group. • Periodic audits by national audit agency.	• Clinical indicators come from Medicare claims and other routine reporting, such as from the National Prescribing Service for the Quality Prescribing Incentive payment. • For non-clinical incentive streams, information is self-reported by providers.	No systematic feedback of performance information to providers.	No public reporting.
Brazil – São Paulo	**OSS**	• Part of the Law on Reform of Public Administration creating not-for-profit civil entities manage public organizations in a areas such as health, education, culture and research. • Objectives were to increase efficiency and civil participation to reduce deficits and limit waste.	• Public health department negotiates performance and volume targets with hospitals. • Independent Assessment Commission (AIC), made up of representatives of government and civil society, reviews performance indicators and calculates penalties. • State audit agency carries out audits.	• Standardized cost accounting and data collection system developed for the programme. • Other data (e.g. hospital-acquired infections) self-reported by hospitals directly.	Performance assessment results discussed with hospitals.	Quality indicators and reports on the government health system website.

Estonia	**PHC QBS**	• Health Services Organization Act, which established the regulatory framework for primary care and family medicine. • Objective is to encourage family doctors to widen their scope of services and focus on prevention.	• The Society of Family Doctors developed performance indicators, and the Estonia Health Insurance Fund (EHIF) provided recommendations for implementation arrangements. • Ongoing development of the QBS jointly by the EHIF and the Society on consensus basis.	• Claims data, with the exception of training of physicians and nurses, which is provided by professional associations. • Chronic disease status added to claims data as a result of the programme.	Physicians receive personal feedback on performance results electronically.	Performance results published on EHIF website.
France	**ROSP**	• Most of the performance measures selected based on objectives and criteria defined by the 2004 Public Health Law and recommendations of the National Authority for Health (HAS).	• Programme prepared by the National Health Insurance Fund (NHIF) as an amendment to the convention on setting tariffs and regulating the relations between providers and the NHIF. • NHIF is working with unions of physicians to review existing performance indicators and develop new ones for specialists.	Claims data mainly and self-reported measures for clinical outcome.	Providers can check indicator achievement rates on the NHIF website.	Results are not made public.

(*continued*)

Table 3.1 Summary of governance structures and data sources and flows in the case study P4P programmes (*continued*)

Country	*Programme*	*Overarching strategy and objectives*	*Programme governance structures and stakeholder participation*	*Data sources*	*Feedback of information to providers*	*Public reporting*
Germany	**DMP**	• Series of reforms to the SHI and experiments to improve care coordination in the area of chronic illness. • Specific objectives include to enhance access to treatment and care for patients with chronic conditions.	• MOH, Joint Federal Committee, and sickness funds jointly oversee DMPs. • Disease-specific committees of experts from universities and medical associations draft programme requirements based on evidence-based guidelines. • Sickness funds define the care packages for each condition, which are then approved by the Federal Insurance Agency. • The Institute for Quality and Efficiency in Health Care checks the recommendations against international norms.	Varies by programme, but systematic use of IT for documentation, evaluation, and quality assurance a key feature of all programmes.		Results are not made public.
Korea, Rep. of	**VIP**	• National Health Insurance Act of 2000, which integrated health insurance funds into a single payer, with a legal mandate for strategic health purchasing, including adjusting payment based on performance.	• Designed and implemented by Health Insurance Review Agency (HIRA) without involving hospitals and other key stakeholders.	• Claims data supplemented by medical records data through a web-based hospital quality data acquisition system. • Mortality data from the Ministry of Public Administration and Security.	Providers receive feedback on performance results.	Performance results published on HIRA website.

New Zealand	**PHO Performance Programme**	• Health Strategy of 2000 with 13 priorities for population health and three for reducing specific health inequalities. • Related PHC strategy that established PHOs. • Objectives of the Programme are to improve population health and reduce inequalities through clinical governance and continuous quality improvement.	• The Programme is administered by the national organization of district health boards (DHBSS). • Development of performance indicators is overseen by a governance group, which includes mandated representatives of practitioners, district health boards, and MOH. • Programme Advisory Group provides expert advice about the content of the Programme.	• Data for most indicators come from claims or data are self-reported. • Data for other indicators are retrieved by the programme from existing databases (e.g. breast cancer screening register).	Quarterly progress reports are provided to PHOs.	Results published on Programme website.
Turkey	**FM PBC**	• MOH reform agenda to improve access, efficiency and quality through the Health Transformation Program (HTP). • A key element of HTP was the introduction of family medicine within a performance-based contracting framework.	• Purchasing and contract management is the responsibility of Provincial Health Directorates (PHDs).	• Core Health Resource Management System, a MOH-wide financial tracking system, and the Family Medicine Information System (FMIS).	The FMIS allows providers to track their progress for indicators that are linked to payment deductions.	

(*continued*)

Table 3.1 Summary of governance structures and data sources and flows in the case study P4P programmes (*continued*)

Country	*Programme*	*Overarching strategy and objectives*	*Programme governance structures and stakeholder participation*	*Data sources*	*Feedback of information to providers*	*Public reporting*
UK	**QOF**	• In the 2000 NHS Plan for Reform and Investment, the UK government committed to large investment tied to expansion of quality-based contracts with GPs. • Specific objectives: increase productivity; redesign services around patients; improve skill mix; create governance structure to improve quality; extend range of services; improve recruitment, retention and morale.	• Implemented solely by the NHS, and PCOs (local purchasing arm of NHS) manage the contracts. • National Institute for Health and Clinical Excellence (NICE) reviews indicators and recommends additions or subtractions through a participatory process.	• GPs use electronic medical records, with QOF information extracted and fed into the Quality Management and Analysis System (QMAS) to calculate achievement scores and reward payments. • Data and evidence for most organizational indicators is submitted manually by GP practices.	• Performance reports made available to GP practices for their approval. • GP practices can access QMAS and run reports at any time.	QOF scores published on NHS Information Centre for health and social care (The NHS IC) maintains an online database.

US – California	**IHA**	• By early 2000 there was growing support for aligning the various health plan and purchaser performance measurement and incentive efforts in California. • Underlying objective was to address specific quality deficits and also the perception that purchasers and providers were excessively focused on cost control.	• Private, voluntary initiative with government involvement limited to the public reporting of results for consumer use. • Programme governance is provided by the IHA Board of Directors. • The programme is managed by IHA P4P Steering, Executive, and Technical Committees, with assistance from technical experts. • Within IHA, there is prominent representation of physicians, including the leadership of the major participating groups, insurers, other large purchasers, and consumer groups.	• IHA produces data collection and reporting guidelines. • Physician groups self-report or health insurers report on their behalf. • All data derived from standardized electronic sources subject to audit. • Data on use of IT collected by survey and validated by accrediting organization. • Patient experience surveys are conducted by a survey vendor. • A vendor generates performance scores.	IHA issues preliminary reports to both physician groups and insurers, which may be appealed.	Performance results reported publicly by the office of the Patient Advocate, a state government agency.

(*continued*)

Table 3.1 Summary of governance structures and data sources and flows in the case study P4P programmes (*continued*)

Country	*Programme*	*Overarching strategy and objectives*	*Programme governance structures and stakeholder participation*	*Data sources*	*Feedback of information to providers*	*Public reporting*
US – Maryland	**MHAC**	• In 2008 the US government purchaser (CMS) made plans to implement national hospital P4P programmes to promote the use of evidence-based process measures. • The State of Maryland used its all-payer system as the basis for developing its own versions of the programmes.	• The state Health Services Cost Review Commission (HSCRC) assembled work groups to design the programme, including staff of the HSCRC and representatives of hospitals and private and public insurers.	• Data come from information system created by the HSCRC to operate the all-payer system – uniform cost data from hospitals, and patient-level case mix.	• HSCRC provides state-wide performance data to each hospital.	HSCRC presents results on its website, with relative rankings.
US National	**HQID**	• Medicare Modernization Act of 2003 included financial incentives for hospitals for public reporting of quality information. • Premier Inc. submitted proposal to CMS to go further and link quality measures to payment.	• HQID designed jointly between CMS and Premier Inc., a nationwide organization of not-for-profit hospitals. • Premier submitted an unsolicited proposal to CMS for the demonstration programme and was selected as the sole programme partners.	To participate in the demonstration, hospitals were required to allow Premier Inc. to submit patient-level data and hospital-level quality data to the government purchaser (CMS) for all discharges from the five clinical areas.	The Premier alliance provided a communication mechanism to discuss performance with hospitals and disseminate best practices	Hospital quality measures reported on the CMS website 'Compare', a subset of which was used in HQID.

of P4P programmes and the governance structures that oversee their ongoing implementation. The programme governance itself can add structures that in some cases are effective beyond the programme. The tripartite governance group that oversees the New Zealand PHO Performance Programme, for example, includes mandated members representing practitioners, primary health organizations (PHOs), district health boards and the MOH. Overall governance of the primary care sector has become more participatory as a result, as multiple stakeholders have remained actively involved, and PHOs and providers have made ongoing investments in the governance structure (PHO Performance Programme, 2009).

The strategies, objectives and governance structures provide the overarching guidance for the design and implementation of the P4P programmes, which should work together to move the system toward better overall performance. In the sections that follow, each of the steps in the governance cycle is discussed in terms of the infrastructure needed to support P4P programmes and how, in turn, elements of P4P programmes can be leveraged to strengthen these governance cycle steps.

Strategy and objectives: the basis for performance domains and measures

P4P programmes can help focus providers and other health system actors more clearly on strategic objectives by explicitly linking those objectives to financial incentives. The domains and measures of the P4P programmes typically directly mirror the overarching health sector strategy and strategic objectives. In New Zealand, for example, the PHO Performance Programme was introduced in 2006 in part to sharpen the focus of PHOs on the priorities of the 2000 Health Strategy. The performance domains and measures map directly to the 13 population health priorities and three priorities for reducing health disparities identified in the strategy. In the Estonia QBS, the Health Services Organization Act provides the overarching strategy for the P4P programme and the regulatory framework for primary care and family medicine. The specific objective of the programme is to encourage family doctors to widen their scope of services and focus on prevention. The three performance domains in the QBS mirror the strategy and objectives: disease prevention, chronic diseases management, and additional activities. Other P4P programmes are designed to achieve specific health service delivery or quality objectives, such as better disease prevention and chronic disease management in Australia, France, California, Estonia, Germany and the UK.

Clinical performance domains

Health service delivery objectives are reflected in national standards of care or clinical guidelines (Campbell, Roland & Wilkin, 2001). The clinical performance domains and indicators of P4P programmes form the cornerstone of most programmes. It is widely believed that the barriers to wider adoption

of evidence-based guidelines by providers can be at least partially overcome by tying their use to financial reward (Institute of Medicine, 2001; Garber, 2005; Kenefick, Lee & Fleishman, 2008). The Germany DMP is an example of the potential success of this approach (Stock et al., 2011). There are a number of challenges, however, with using clinical guidelines as a basis for performance measurement and reward.

First, most guidelines require very detailed patient-level information to determine whether they have been followed. Second, the strength of the evidence underlying guidelines can vary, and the evidence may be generated by trials in very limited settings for patients with a narrow set of characteristics. Most guidelines therefore have been written to be flexible and allow a large degree of clinical judgment, making it difficult to assess whether a guideline was followed appropriately (Garber, 2005). Third, good clinical performance measures are related to conditions that are widespread and contribute significantly to the overall burden of disease, events related to the conditions should be common, there should be well-established evidence that relates the intervention to outcomes, and it should be feasible to collect reliable data related to the measure (Werner & Asch, 2007). The challenge with identifying clinical performance indicators is that the number of clinical situations that satisfy these conditions is likely to be limited. And finally, it is important that adherence to guidelines does not inhibit the search for innovative ways of delivering care, and linking a financial incentive to adherence to guidelines could possibly make this of even greater concern.

In primary care, it is often relatively easy to identify priority clinical areas with high burden of disease and well-established clinical guidelines, such as immunization and the management of common chronic conditions, including diabetes and cardiovascular disease. For example, the Australia PIP, California IHA, Estonia QBS, France ROSP, Germany DMP, New Zealand PHO Performance Programme, and the UK QOF all include performance measures derived from evidence-based guidelines for diabetes management. P4P programmes could potentially have a large impact if the incentives drive an increase in detection, better recording, and coverage with evidence-based services. The challenge is more difficult in hospitals, where the range of clinical conditions and services is so complex that only a fraction of them have potentially high impact with widely accepted guidelines that can be feasibly translated into valid performance measures.

To identify appropriate clinical domains and performance indicators that are grounded in evidence, several implementers of P4P programmes have delegated the responsibility to provider groups or other clinical governance institutions. In the Germany DMP, for example, disease-specific committees of experts from universities and medical associations draft programme requirements grounded in evidence-based guidelines. In Estonia, the QBS was developed jointly by the Society of Family Physicians and the Estonia Health Insurance Fund (EHIF). The provider organization was responsible for developing valid clinical indicators, and the EHIF was responsible for developing the details of implementation of the programme. Ongoing refinement of the programme has been undertaken by both organizations together on a consensus basis. On the other hand, if providers have disproportionate responsibility for developing

clinical performance domains and measures, the process may be open to capture by providers. Programme implementers had to contend with this issue both in the Maryland HAC and the UK QOF, but both programmes managed to find a balanced solution (Gillam & Siriwardena, 2010).

Another approach has been for the purchasing organization to carry out the initial analytical work and draft proposals, which are then discussed, revised, and validated with stakeholders. In the Maryland HAC programme, for example, the all-payer hospital rate-setting authority of the Health Services Cost Review Commission (HSCRC) convened working groups that included both clinical and financial staff of the HSCRC, representatives of hospitals, and public and private insurers. The HSCRC staff did the foundational analytic work and prepared draft recommendations for the P4P programme, then the working groups met over a nine-month period to discuss and amend the original recommendations. This process led to a near consensus on the final recommendations for the programme.

Stakeholder involvement in refining the UK QOF has evolved into a highly transparent and participatory process. In the UK QOF, clinical indicators were initially developed by a group of primary care academic experts in each QOF clinical area (Campbell & Lester, 2011). In 2009, the responsibility for indicator refinement was given to the National Institute for Health and Clinical Excellence (NICE), an independent organization that provides guidance on evidence-based health care services. NICE reviews current indicators, prioritizes areas for change, and develops and proposes new indicators for the QOF. After each revision, the proposed menu of indicators is reviewed through an open consultative process before the final selection is made (UK Department of Health, 2009).

Non-clinical performance domains

There is no dispute that evidence-based clinical guidelines should be the basis for clinical performance measures when possible, but there is no such gold standard for non-clinical performance measures, such as organizational measures and patient experience. Furthermore, these measures typically require additional data collection and report generation. With the exception of IT uptake (Hillestad et al., 2005), few of the non-clinical indicators used in the P4P programmes reviewed have been justified by a demonstrated impact on improved clinical quality of care.

Direct incentives related to improving the organization of service delivery are common in P4P programmes targeted at primary care. In Australia, providing incentives to improve the organization of service delivery is a large focus of the PIP. GP practices must be accredited or registered for accreditation, and two of the three performance domains relate to the organization of service delivery. The Capacity Stream gives additional resources to GP practices that invest in infrastructure, such as computerization, or to expand services, such as providing after hours care. The Rural Support Stream provides additional resources to GP practices in more rural and remote settings to bring services to these areas. Although the accreditation requirement has motivated a large number

of GP practices to undergo accreditation, a systematic review of 66 published studies failed to show a clear relationship between accreditation and improved clinical quality measures (Greenfield & Braithwaite, 2008), and the high cost of complying with the accreditation process has been a barrier to small practices in remote locations serving vulnerable communities (Australian National Audit Office, 2010). The UK QOF has 36 indicators in the organizational performance domain covering such aspects of GP practice organization as record keeping, information for patients, education and training of staff, practice management, and medicines management. The California IHA and France ROSP programmes both include organizational performance measures focused only on the use of IT in managing patient care.

In addition to direct incentives, pay for performance also can drive organizational changes and investments indirectly, as providers make organizational improvements to achieve clinical performance targets. In the UK, for example, GP practices have made internal changes to focus services more clearly on the targets set in the QOF, including increased employment of nurses for chronic disease management, and a more prominent role for IT (Roland, 2006). In California, the IHA initiative has spurred a variety of investments and policy changes, including increased patient outreach and use of data for internal quality improvement (Damberg et al., 2009). Larger providers with more resources may be more likely to make many of these organizational improvements, as has been the case in the California IHA programme and possibly the France ROSP programme, as primary care providers are mainly sole practitioners and small groups. In the California IHA, better performance achievement is found among large provider groups, which suggests that they are better able to make the necessary investments than smaller groups (Rosenthal et al., 2005). This may be intentional, if one objective of the programme is to secure consolidation amongst smaller groups, but it could lead to unintended consequences, such as negatively affecting the supply of services in rural areas (Rosenthal et al., 2001).

Performance monitoring: data and information flows

P4P programmes rely on valid, timely, and reliable data for performance indicators that can be generated easily by providers, and aggregated, analysed, and compared by purchasers. This requirement of P4P programmes has created a useful lever to motivate providers to make the leap from their current clinical information systems to more automated practices that can generate data for secondary uses. P4P programmes have contributed to the movement toward improved health information in several ways. In the US, the first steps toward measuring performance under the federally funded health insurance programmes began with 'pay-for-reporting' programmes, which laid the groundwork for the subsequent HQID pay for performance programme. Some programmes provide direct incentives for providers to invest in IT and electronic medical records (Australia PIP, California IHA, France ROSP). Other programmes have made reaching minimal IT standards a criterion for participation in the programme (UK QOF). Finally, some P4P programmes

contribute to better secondary use of clinical data by bearing the costs of developing the tools and applications that bring data together from multiple sources and analytical programmes that generate useful reports that may be costly for individual providers to generate themselves (California IHA, New Zealand PHO Performance Programme, UK QOF, US HQID).

Data for P4P programmes come from a variety of sources: administrative or claims data, medical records, self-reported data from providers, and patient surveys. Each source of data has strengths and limitations (Dudley & Rosenthal, 2006). In the case study P4P programmes, claims data are the main source of information for clinical indicators for most programmes, with the exception of the UK QOF that relies on electronic medical records. Data for non-clinical domains are almost always self-reported by providers.

Claims data

In eight of the case study P4P programmes, claims data are the primary source of information for calculating performance achievement rates. In the Estonia QBS, for example, all data for QBS come from the EHIF's electronic billing system. The Korea VIP programme also uses routinely collected data on hospital activity. In the California IHA programme, the majority of data are derived from encounter records (also known as 'shadow claims', because they mimic billing data but are not used for payment) and laboratory billing data. In the Maryland HAC programme, potentially preventable complications are identified through secondary diagnoses recorded in the hospital discharge abstracts submitted as part of Medicare claims. Claims data are the main source of performance information also in the Australia PIP, France ROSP, New Zealand PHO Performance Improvement Programme, and the US HQID.

Claims data can be useful for some performance measures, especially in the early stages of a programme, and using existing claims systems has the benefit of not placing additional reporting burdens on providers. Claims data often lack the clinical detail for meaningful performance measures, however, and may be particularly prone to error in identifying patients that are in the target groups for specific indicators, for example, the total number of patients diagnosed with asthma who should complete annual cycles of care (Berenson, Pronovost & Krumholz, 2013). This secondary use of claims data can in itself represent progress for governance, but ultimately it usually proves inadequate for measurement of provider performance.

Enhanced information systems and electronic medical records

In the New Zealand PHO Performance Programme, performance measures were initially chosen based on data already available through claims or other existing databases (e.g. the breast cancer screening register). As the programme evolved, however, there was a demand to link performance measures more closely with priority areas, which meant that the programme had to invest in the infrastructure required to generate new data directly from the GP practices,

particularly related to diabetes, hypertension, and smoking. The District Health Boards and MOH shared the cost of this infrastructure for the new data collection and made an effort to avoid placing an additional reporting burden on providers.

In the US HQID, data for performance measures come from discharge summaries of the Medicare claims submissions, but additional data must be submitted by hospitals from the patient records. There are two additional layers of data aggregation and analysis beyond the discharge summary data carried out by the Premier Quality Measures Web Tool. The hospitals are required to pay for their subscription to Premier's relatively expensive database tool to perform these aggregations as a condition for participation in HQID, and the cost may have reduced hospital participation in the programme (Grossbart, 2008).

The UK QOF is the most dramatic example of a P4P programme driving improved generation and use of data. Significant investments have been made to strengthen clinical information systems at the provider level and to build an application that can aggregate and analyse anonymized patient level data, the Quality Management and Analysis System (QMAS). Early in the QOF implementation, Primary care trusts (PCTs),[1] the local purchasing arm of the NHS, were expected to provide resources to upgrade the clinical systems of those GP practices that did not have compliant systems (UK Department of Health, 2003). The achievement calculation, verification and payment under the QOF are highly automated and use the electronic medical record in the GP clinical data system as its foundation for most indicators (Figure 3.2). Data relating to most of the organizational indicators cannot be automatically extracted from the QMAS, so practices enter organizational data manually on the QMAS website. QMAS can be accessed at any time by GP practices to get feedback on the number of services and the quality of care they are delivering, as well as their current performance against QOF achievement targets.

Self-reported data

Although most of the case study P4P programmes rely on administrative claims data or other standardized, audited data systems for performance measures, a large number of measures across the programmes are self-reported by providers. In particular, nearly all of the non-clinical indicators are self-reported. Even some clinical data, such as measures of hospital-acquired infections in the Brazil OSS, are self-reported by hospitals directly. The Turkey FM PBC has a new clinical information system that was introduced as part of the system-wide primary care reforms and introduction of family medicine, the Family Medicine Information System (FMIS). The routine information system relies on self-reported data input directly by family medicine teams rather than extracted from electronic medical records. The self-reported data are audited monthly by the Provincial Health Directorates for a ten per cent sample of family physicians.

Self-reported data not only raise obvious concerns about reliability (Anema et al., 2013), but they also place additional reporting burden on providers, which can be significant in some cases. The data required for most of the UK QOF non-clinical performance indicators, for example, are verified by separate

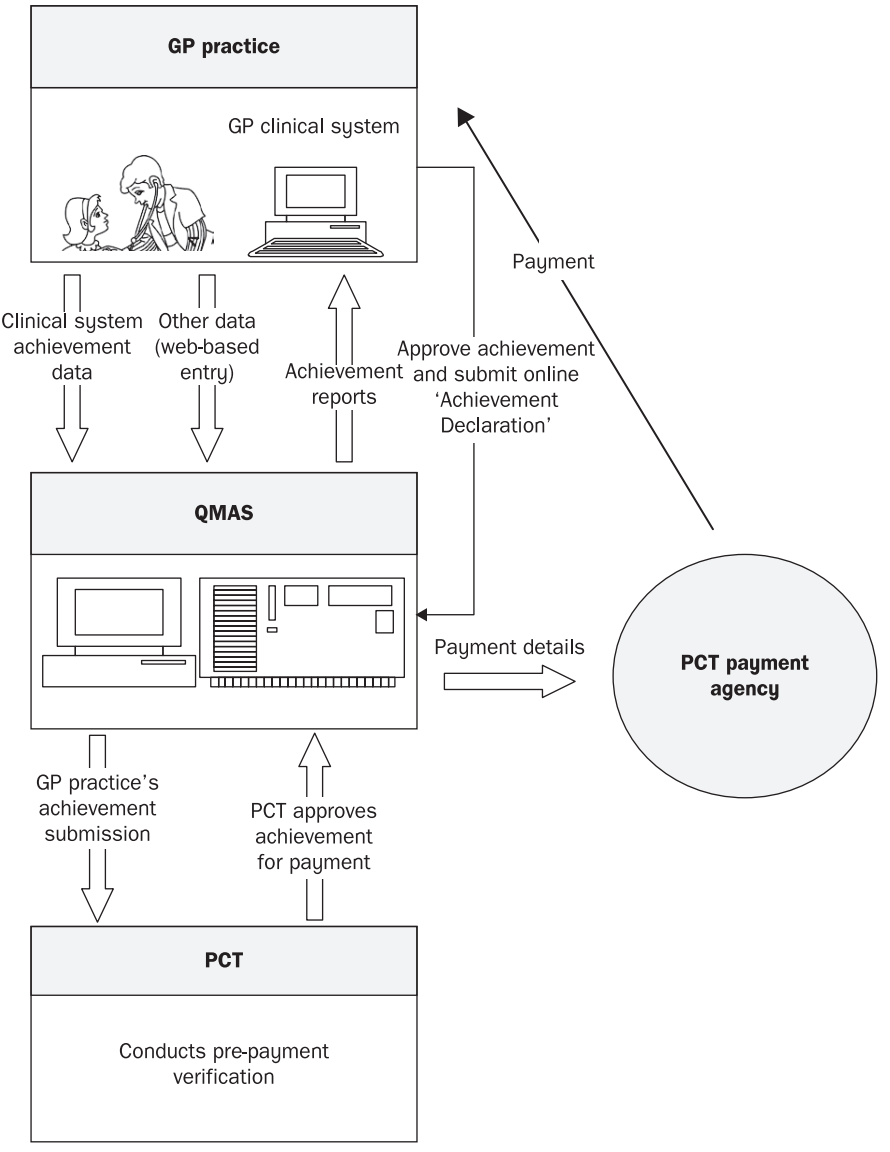


Figure 3.2 Information flows in the UK QOF

Source: Adapted from UK Department of Health, 2003.

reports or other sources of evidence supplied by the GP practices. The QOF guidance documents outline the types of evidence required for non-clinical indicators, which includes, for example, a 'report on the results of a survey of a minimum of 50 medical records of patients who have commenced a repeat medication', and a report of 'the results of a survey of the records of newly registered patients'. There are at least 15 such reports that are specified in

the guidance documents, with about half that need to be generated each QOF period and half that are one-off reports of policies and procedures which would not change every QOF period (NHS, 2010).

Verification

An important aspect of monitoring for clinical governance and for operating P4P programmes is verification of the accuracy and validity of data that are reported by providers. Verification serves three important functions. First, it makes the reporting, achievement calculation, and payment fair and transparent. Second, verification serves an audit function to guard against gaming and overpayment. Finally, the verification process can be used as an opportunity for dialogue between purchaser and providers.

In several P4P programmes reviewed, the verification process is one-directional and only serves the audit function. In the Australia PIP, for example, the Continuous Data Quality Improvement Program controls the quality of payments on a sampled basis, recording all sources and types of errors commonly found in the reporting of results. Medicare Australia also conducts random and targeted audits to ensure that practices meet the eligibility requirements. Other programmes, however, also use the verification process as an opportunity for dialogue with providers. The New Zealand PHO Performance Programme works closely with providers in the verification process. A number of measures are taken to validate the data submitted by PHOs. Every quarter, information from PHOs is run through logic algorithms that highlight unusual changes in indicators. No data are made publicly available until they have been validated and agreement has been reached with the PHOs.

The UK QOF has an intensive bi-directional verification process, which facilitates communication between providers and PCTs. PCTs oversee the automated assessment of performance and calculation of scores, and carry out a three-pronged verification process: (1) review visit of all GP practices at least once in three years; (2) pre-payment verification of achievement; and (3) post-payment audit of five per cent of practices randomly selected. The first prong of the verification process also has a supportive function and is focused on reviewing the practice's expected achievement, identifying barriers to improvement, and assessing data quality. The second prong of the verification process is meant to confirm the validity of the data and other evidence submitted for the QOF payment. The third prong of the verification process has solely an audit function as part of the anti-fraud system (Cashin & Vergeer, 2013).

Identifying the need for change – feedback loop for performance improvement

To start and sustain the virtuous cycle of health system governance and performance improvement, it is not enough to generate more and better data. The cycle is perpetuated when the information is used by purchasers and fed back to providers so they can identify and manage necessary changes, and so

effective practices can be identified and eventually incorporated into clinical guidelines and performance measures. Tremendous value is added when the data generated by individual providers are aggregated across the system, synthesized and analysed, then returned to purchasers and providers in actionable form.

Better generation and use of data also have been shown to directly support better quality of care and cost management (Hillestad et al., 2005), and is considered to be critical for the next generation of decision support and quality management in health care (Rhoads & Ferrara, 2012). Analysing data in populations over time can be used to identify ways to engage in health promotion and disease management; decision support tools can promote adherence to evidence-based practices; analytic tools can reveal patterns that correlate treatments with outcomes and identify which care practices are most effective (Teasdale et al., 2007; Rhoads & Ferrara, 2012). Better use of data and technology can also be used to avoid duplication, which affects both quality and costs.

Most health care providers generate a large amount of data – some form of a clinical information system, financial information systems, registers, and other data sources are available in most provider organizations. Most providers make very limited use of such data, however, and the movement to mobilize data to improve patient care, manage costs, and monitor performance is relatively new in many systems. Often providers, particularly at the primary care level, do not have the technical resources to integrate their data, run queries on indicators of interest, or generate reports (Teasdale et al., 2007). Lack of aggregated and analysed performance data in a useful format at the individual provider level also means a lack of useful performance data for purchasers and policymakers that are essential to the governance cycle.

Several of the case study P4P programmes identified improved generation, but more importantly *use*, of data as an important contribution of the programme to overall clinical governance and health sector management. In some cases, key data and analysis of their own performance became available and in the hands of providers for the first time. Other programmes identified the concrete nature of performance targets and achievement rates as a facilitator in the dialogue between providers and other players in the health system, including purchasers, regulators and patients. In the Brazil OSS programme, for example, the practice of routinely analysing hospital indicators has transcended the P4P programme objectives and is now part of routine hospital management (Radesca, 2010). In the Maryland HAC, the method for categorizing potentially preventable complications provided a useful communication tool that was essential to achieving reduced complications over time. Data showing each hospital its relative performance by category provided clinical and financial staff with the information they needed to systematically target specific problem areas to reduce the frequency of hospital acquired complications (Murray, 2012).

The PHO Performance Programme in New Zealand is an example of how the power of the feedback loop created by the P4P programme might exceed the power of the incentive to motivate behaviour change. Because of the low budget for the incentive, the programme had to find other ways to drive change and performance improvement. The programme provides PHOs with monthly reports for four of their indicators and raw data on a quarterly basis, with the information used to calculate their indicators. This information was

not previously available to providers. In the UK QOF, there are multiple opportunities for feedback on performance and dialogue between GPs and PCTs. The QMAS was developed to support the QOF, but its use extends beyond calculating achievement against performance measures. The QOF review visit as part of the verification process, for example, is meant to give both GP practices and PCTs 'early warning' of any issues related to data, reporting, or predicted performance achievement levels.

In the France ROSP, the data system developed for the programme can be accessed online, and individual physicians can track their scores over time and also benchmark them against national targets and regional and national averages. In the Australia PIP, although promoting uptake of IT among health care providers is a key focus of the PIP, the programme lacks a supporting information system and feedback loop. No reports are available showing trends in performance against the different indicators, and the possibility of monitoring trends is further diminished by the design of PIP, which allows PIP practices to move in and out of specific incentive schemes, making it difficult to monitor aggregate trends.

Accountability

P4P programmes inherently introduce more accountability of health care providers to purchasers and the populations they serve. The programmes themselves also include internal accountability mechanisms that ensure that the interests of purchasers, providers and patients are all represented in programme implementation. The governance structures and accountability mechanisms of P4P programmes are varied, but most include some multi-stakeholder oversight, involvement of professional associations, some form of external audit, and public reporting of results. Consumer groups have not been actively involved in the oversight of the programmes, with the exception of the California IHA programme, which includes consumer groups that are already active in IHA itself. Performance results are made public in all of the programmes, with the exception of the Australia PIP, France ROSP, Germany DMP, and Turkey FM PBC. No information is available, however, to assess whether the public is actually able to readily access and interpret the performance information, and whether and how the performance information is used by consumers.

Conclusions

P4P programmes can play an important role in strengthening overall health system governance when the incentive is used to strengthen one or more steps in the governance cycle: sharpening the focus on strategic objectives; creating incentives to adopt evidence-based clinical guidelines and other service delivery approaches; better generation and use of information for performance monitoring; strengthening the feedback loop so purchasers, providers, patients and policymakers can use information on performance to identify areas for further change and improvement. When P4P programmes align with health system objectives and the organizational objectives of providers,

the programmes can catalyse broader initiatives and approaches to improve service delivery and the health system as a whole. If the financial incentive provides a focus on particular objectives, clinical areas, more meaningful use of data and IT, or other aspects of governance, it can take on greater importance than simply a reward or penalty.

Note

1 Primary care trusts are now known as primary care organizations (PCOs).

References

Anema, H. et al. (2013) Influences of hospital information systems, indicator data collection and computation on reported Dutch hospital performance indicator scores, *BMC Health Services Research*, 13: 212–22.

Australian National Audit Office (2010) *Practice incentive, program audit report no. 5 2010–2011*. Canberra: Australian National Audit Office.

Berenson, B. (2010) *Moving payment from volume to value: what role for performance measurement?* Washington, DC: Urban Institute.

Berenson, B., Pronovost, P. and Krumholz, H. (2013) *Achieving the potential of health care performance measures*. Washington, DC: Urban Institute.

Campbell, S. and Lester, H. (2011) Developing indicators and the concept of QOFability, in S. Gilliam and A.N. Siriwardena, *The quality and outcomes framework: transforming general practice*. London: Radcliffe Publishing.

Campbell, S., Roland, M. and Wilkin, D. (2001) Improving the quality of care through clinical governance, *British Medical Journal*, 322(7302): 1580–2.

Cashin, C. and Vergeer, P. (2013) *Verification in results-based financing: the case of the United Kingdom*. Washington, DC: World Bank.

Clavreul, L. (2012) Les médecins se convertissent au paiement à la performance, *Le Monde*, 2 January.

Damberg, C. et al. (2009) Taking stock of pay-for-performance: a candid assessment from the front lines, *Health Affairs*, 28(2): 517–25.

Dudley, R. A. and Rosenthal, M. (2006) *Pay for performance: a decision guide for purchasers*. Washington, DC: Department of Health and Human Services, Agency for Healthcare Research and Quality.

Galvin, R. (2006) Pay-for-performance: too much of a good thing? A conversation with Martin Roland, *Health Affairs*, 25(5): 412–19.

Garber, A. (2005) Evidence-based guidelines as a foundation for performance-based incentives, *Health Affairs*, 24(1), 174–9.

Gillam, S. and Siriwardena, N. (2010) *The quality and outcomes framework–QOF–transforming general practice*. Milton Keynes: Radcliffe Publishing.

Greenfield, D. and Braithwaite, J. (2008) Health sector accreditation research: a systematic review, *International Journal for Quality in Health Care*, 20(3): 172–83.

Grossbart, S. (2008) Effectiveness of pay for performance as a quality improvement strategy, *Prescriptions for Excellence in Health Care*, 3: 1–3.

Hillestad, R. et al. (2005) Can electronic medical record systems transform health care? Potential health benefits, savings, and costs, *Health Affairs*, 24(5): 1103–17.

Institute of Medicine (2001) *Crossing the quality chasm: a new health system for the 21st century*. Washington, DC: National Academy Press.

Kenefick, H., Lee, J. and Fleishman, V. (2008) *Improving physician adherence to clinical practice guidelines: barriers and strategies for change*. Cambridge, MA: New England Healthcare Institute.

Miller, H. (2011) *Transitioning to accountable care: incremental payment reforms to support higher quality, more affordable health care*. Pittsburg: Center for Health Care Quality and Payment Reform.

Murray, R. (2012) *United States Maryland Hospital Acquired Complication Program*. Paris: Organization for Economic Co-operation and Development.

NHS (2010) *QOF management guide: volume 4*. London: UK National Health Service.

Or, Z. (2010) *P4P for generalists: first results* (http://www.hpm.org/survey/fr/a16/3, accessed November 2011).

PHO Performance Programme (2009) *PHO performance program annual report: 1 July 2008–30 June 2009*. Auckland: District Health Boards in New Zealand.

Public Health Foundation (2013) *From silos to systems: using performance management to improve the public's health. Performance management national excellence collaborative*. Washington, DC: Public Health Foundation.

Radesca, E. (2010) ENC: respostas solicitadas (R. Faria, interviewer; 23 July).

Rhoads, J. and Ferrara, L. (2012) *Transforming healthcare through better use of data*. Falls Church, VA: Global Institute for Emerging Health Care Practices.

Roland, M. (2006) Financial incentives to improve the quality of primary care in the UK: predicting the consequences of change, *Primary Care Research and Development*, 294: 18–26.

Rosenthal, M. et al. (2001) Scale and scope of capitated physician organizations in California, *Health Affairs*, 20: 109–19.

Rosenthal, M. et al. (2005) Early experience with pay for performance: from concept to practice, *Journal of the American Medical Association*, 294(14): 1788–93.

Smith, P. et al. (2012) Leadership and governance in seven developed health systems, *Health Policy*, 106: 37–49.

Stock, S. et al. (2011) *Disease-management programs can improve quality of care for the chronically ill, even in a weak primary care system: a case study from Germany*. Washington, DC: Commonwealth Fund.

Struis, J. and Baan, C. (2011) Integrating care through bundled payments – lessons from the Netherlands, *New England Journal of Medicine*, 364(11): 990–1.

Teasdale, S. et al. (2007) Secondary use of clinical data in primary care, *Informatics in Primary Care*, 15: 157–66.

UK Department of Health (2003) *Delivering investment in general practice: implementing the new GMS contract*. London: UK Department of Health.

UK Department of Health (2009) *Developing the quality and outcomes framework: proposal for a new, independent process*. London: UK Department of Health.

Werner, R. and Asch, D. (2007) Clinical concerns about clinical performance measurement, *Annals of Family Medicine*, 5: 159–63.

World Health Organization (2000) *World health report 2000: health systems: improving performance*. Geneva: World Health Organization.

World Health Organization (2010) *Everybody's business: strengthening health systems to improve health outcomes: WHO's framework for action*. Geneva: World Health Organization.

chapter four

Evaluating P4P programmes

Y-Ling Chi and Matt Sutton

Introduction

P4P programmes are becoming increasingly popular in spite of the lack of conclusive evidence that they improve health care quality or health outcomes. Does the lack of evidence suggest P4P is intrinsically flawed, or are disappointing results rooted in problems with the design and implementation of P4P programmes and limitations in the way in which programmes are evaluated?

Programme evaluation is at the heart of public demand for effective use of limited resources, providing evidence about impact that can drive improvement in policy design, transparency and accountability. For example, Mexico's conditional cash transfer programme Opportunidades (previously Progresa) was made famous for its rigorous evaluation and the valuable evidence this generated. Evaluation showing how the programme achieved important improvements in child health and school enrolment rates helped maintain funding for the intervention through electoral cycles, and the programme was successfully included in the national poverty reduction agenda (Skoufias & McClafferty, 2001). The success of the intervention also informed the design and implementation of similar interventions in other countries (e.g. Honduras and Columbia).

Progresa's case is unusual, however. In practice, evaluation is rarely planned in advance and often has to rely on opportunistic data and administrative arrangements that may limit the scope for convincing insight into a programme's impact. P4P programmes are no exception and present additional challenges for evaluators given the potential for spillovers and unintended consequences. Also, since P4P programmes are just one option for the use of health care resources, it is important to have a better understanding of their costs as well as their benefits. This chapter begins with a short overview of impact evaluation techniques for social interventions, then reviews the evaluation of P4P programmes in OECD countries and highlights key issues.

Overview of impact evaluation techniques for social interventions

The shift towards evidence-based policymaking has encouraged the development and refinement of methodological and technical instruments for measuring the effectiveness of policies. These increasingly advanced methods have been applied across countries and in diverse social interventions, from provision of school textbooks to vaccination campaigns. Different studies adopt numerous methodological approaches ranging from highly qualitative, in-depth assessments of perceptions of a programme's impact to advanced statistical and econometric approaches using massive data sets. In this chapter we make no judgement on the relative merits of these approaches. Each will be more or less appropriate depending on the focus of the evaluation, the resources available, and feasibility. If resources are sufficient, it is likely that mixed methods will yield the deepest understandings. However, of all methods, it is the quantitative analysis of inputs, processes and outcomes that has proved most persuasive amongst decision makers in assessing the generalizability of P4P programme results.

Impact evaluation studies aim to examine the causal relationship between changes in outcomes and implementation of a given intervention in a target population. Moreover, understanding the channels of impacts and the extent to which the results could be replicable to other contexts have been the focus of recent works on the refinement of technical tools (Jones et al., 2009).

Establishing a causal relationship between a programme (often referred to as 'treatment') and one or several outcomes (i.e. the endpoints where improvements are expected to occur) is the most challenging part of impact evaluation. Simply observing changes in outcomes between groups with different treatment statuses or over time usually fails to account for underlying trends and the issue of selection bias. This latter phenomenon refers to the fact that individuals or providers who choose to participate in a programme may differ systematically from those who do not participate. The differences between programme participants and non-participants may be observable (e.g. level of assets) or unobservable (e.g. willingness to take risks). Selection bias limits a researcher's ability to estimate the true causal effect of the policy intervention; the estimated effects of programme participation may be caused by participant traits rather than the policy itself. Successful outcomes, for instance, may be driven by the eagerness of volunteer programme participants. The task of a researcher, then, is to use intelligence on the programme planning process and data from the pre- and post-implementation period in order to separate the causal effect of some intervention from the behaviour of self-selected participants.

For these reasons, amongst all available techniques, randomized control trials (RCTs), which employ an experimental design, are often put forward as the 'gold standard' of evaluation. Despite their strong power, however, randomly assigned policies and interventions are unusual in OECD countries, and occasions for experimental evaluations have been rare, as evaluation was almost never thought as a part of programme implementation. This is partly because implementers have failed to recognize the need for monitoring and evaluation. In addition, there may also be profound political, technical and financial barriers to randomized designs. Analysts have therefore developed

several alternative non-experimental methods for successful evaluation that seek to mimic the randomization process.

Depending on the challenges posed by the design of the intervention, researchers have a range of techniques at their disposal, both qualitative and quantitative. Quantitative models can broadly be classified in three categories: experimental, quasi-experimental, and non-experimental (or observational) techniques. These evaluation techniques differ in rigour, complexity, feasibility and cost.

Qualitative techniques can offer a good alternative or complementary approach to quantitative analysis, and are often less costly and difficult to implement ex-post. They enable researchers and policymakers to investigate and understand different aspects of policy impact, but they are considered less exhaustive than quantitative techniques because they rely on smaller sample sizes. In health interventions, qualitative techniques could be integrated in quantitative impact evaluation to understand heterogeneous results and investigate complex socio-economic mechanisms (Glenton et al., 2011). Nonetheless the use of such methods remains marginal, even when combined with quantitative models (Glenton et al., 2011).

In an **experimental impact evaluation**, a policy 'treatment' is typically assigned to a randomly selected group of recipients, while a randomly assigned control group receives no treatment. This allows evaluators to estimate the effect of a policy treatment while avoiding selection bias. RCTs are derived from medical testing protocols, and aim to answer the following question: what would the outcome be if the policy had not been implemented? Experimental strategies answer this question by building a counter-factual and comparing a *control group* to a *treatment group*. If planned and implemented correctly, these types of evaluation produce unbiased and reliable results by overcoming problems of selection often encountered by other types of evaluation. By randomizing policy interventions, researchers can examine the causal link between programme implementation and programme impact.

Quasi-experimental methods refer to a broad range of techniques that mimic experimental design while using observational data. RCTs may present logistical, ethical, political, and other challenges. Consequently, when evaluating a programme it is often easier for researchers to compare groups that have equal probability of participating in the programme but who differ in whether or not they actually received a programme 'treatment'. Such comparison groups are often constructed by matching the participants on the basis of observed traits. Common quasi-experimental approaches include difference in differences estimation, interrupted time series, regression discontinuity, and propensity score matching, in which researchers build ex-ante a comparison and treatment group using matching methods (Jones & Rice, 2011).

Finally, **non-experimental (or observational) designs** usually compare the outcome of programme participants before and after the intervention (reflexive comparison), or compare the outcomes of programme participants to that of a comparison group (without matching or controlling for group differences). While this evaluation technique is the cheapest and most easily implemented, it is considered to have little internal and external validity due to potential selection biases. Because various factors influence programme

participation (and consequently affect programme outcomes), it is difficult to isolate the true effect of the policy.

Compared to experimental studies, non-experimental and quasi-experimental studies need to provide richer and stronger evidence that they have fully controlled for secular trends, selection biases, and confounding caused by other factors. A detailed description of the programme planning process and the mechanisms by which participants were selected for the programme is critical for constructing a plausible justification for the selected control groups.

Evaluation of P4P programmes in OECD countries

Impact evaluation is seldom conducted, due to a lack of funding and incentives and political, bureaucratic and administrative obstacles (Savedoff et al., 2006). In OECD countries, despite the large sums of money spent on P4P programmes, very few have designed impact evaluation into the programme, and evaluation results are usually largely disconnected from political choices to expand, scale-up, reform or withdraw the programmes. For instance, the HQID programme in the United States has been expanded even in the absence of conclusive evidence from the pilot phase of the programme (Lindenauer et al., 2007; Ryan, 2009; Jha et al., 2012). Moreover, programme evaluation is often included as a marginal and neglected part of programmes in OECD countries.

Flodgren et al. (2011) undertook a review of the systematic reviews that have evaluated the impact of financial incentives on health care professional behaviour and patient outcomes. They concluded that existing studies had serious methodological limitations and were very limited in their completeness and generalizability. In the same vein, Scott et al.'s (2011) review of the use of financial incentives in primary care concluded that poor study designs could lead to substantial risk of bias likely to misinform policymaking. This review was particularly concerned that none of the existing studies addressed issues of selection bias, caused by providers being able to select in or out of the incentive programme. Van Herck et al. (2010) identified more than a hundred studies assessing the impact of P4P on quality of care and showed that the prevalent evaluation method of P4P programmes in published peer-reviewed literature was cross-sectional design, i.e. non-experimental group comparison. Randomized control trials were applied in just nine out of 128 studies. In addition to the study design, it is of interest to understand what institution commissioned the evaluation of the programme, what aspects of impacts (clinical performance, providers' behaviour, patient health, cost) have been investigated and whether the results have been fed into a broader policy discussion or decision.

Table 4.1 summarizes evaluation of the case study P4P programmes. Most of the evaluations use group comparison or before/after comparison. Only a handful of programmes have used economic modelling or quasi-experimental techniques in external reviews (e.g. United Kingdom and California). It is also important to note that in almost none of the cases have impact evaluation results been used to inform decisions with regards to the evolution of the programmes. This underlines a key problem: the often inevitable delay in reporting evaluation

Table 4.1 Summary of case study P4P programme impact evaluation studies

Country/P4P project	*External evaluation*	*Internal evaluation*	*Type of evaluation (number of studies)*	*Results of the evaluation*	*References*
Germany, Disease Management Programme	X		• Pair-wise matching (1) • After/before comparison (2) • Comparison between participants and non-participants (2)	• Improvements in implementation of practice guidelines for diabetes. • Improvements in mortality rates for diabetic patients.	Eichenlaub et al. (2004) Altenhofen et al. (2004) Szecsenyi et al. (2008) Miksch et al. (2010) Schäfer et al. (2010)
US, Hospital Quality Incentive Demonstration		X	• Mostly before/after comparison and yearly time series • Premier Inc. was the main body in charge of evaluation but did not disclose details of methodology	• Quality Index scores raised by 18.6 per cent over the past six years. • Reported reductions in mortality (– 8500 patients deaths due to heart attacks in the past five years).	https://www.premierinc.com/quality-safety/tools-services/p4p/hqi/index.jsp
US, Hospital Quality Incentive Demonstration	X		• Pair-wise matching (hospitals) (1) • Comparison between participants and non-participants (1) • After/before (1) • Modelling (1)	• Moderate to insignificant improvements of patient health outcomes and quality indicators for acute myocardial infarction, heart failure, pneumonia, or CABG.	Lindenauer (2007) Grossbart (2008) Glickman et al. (2007) Ryan (2009)

(*continued*)

Table 4.1 Summary of case study P4P programme impact evaluation studies (*continued*)

Country/P4P project	*External evaluation*	*Internal evaluation*	*Type of evaluation (number of studies)*	*Results of the evaluation*	*References*
Estonia, Quality Bonus System	X		• Comparison between participants and non-participants (1)	• Participating physicians providing better care continuity than those non-participating (resulting in fewer hospitalizations and specialists consultations).	Västra (2010)
France, Payment for Public Health Objectives		X	• Comparison between participants and non-participants (1) • Before/after comparison	• Better performance amongst participants with regards to quality of care to diabetic patients. • Mixed results in other areas of care, especially prevention.	Caisse Nationale d'Assurance Maladie (2013)
UK, Quality and Outcomes Framework		X	• National Audit Office conducted a performance assessment: qualitative assessment and before/after comparison (1)	• Good improvements of retention and participation. • 'No progress' on redesign of patients and productivity improvements. • 'Some progress' on quality of care.	UK National Audit Office (2008)

UK, Quality and Outcomes Framework	X		• Interrupted time series (2) • Modelling (3) • Difference in difference analysis (quasi experimental) (1)	• No impact on the health outcomes of patients with diabetes, asthma and hypertension. • Positive impact of QOF on recording of smoking cessation and prevention. • Minor impact of QOF on mortality. • No impact on non-incentivized items, modest impact on incentivized items.	Serumaga et al. (2011) Campbell et al. (2007) Taggar et al. (2012) Fleetcroft et al. (2010) Doran et al. (2011) Sutton et al. (2012)
Australia, Practice Incentives Programme		X	• Australian National Audit Office carries out regular assessment of effectiveness	• Accreditation rates have gone up as a result of implementation of PIP, which in turn has improved the quality of care for patients.	Australian National Audit Office (2010)
Australia, Practice Incentives Programme	X		• Modelling (1)	• Diagnosis of patients with diabetes improved (use of HbA1c test increased).	Scott et al. (2009)
US, Maryland Acquired Hospital Infections				• No evaluation commissioned but analysis of the performance data shows decrease in hospital acquired infections.	See case study

(*continued*)

Table 4.1 Summary of case study P4P programme impact evaluation studies (*continued*)

Country/P4P project	*External evaluation*	*Internal evaluation*	*Type of evaluation (number of studies)*	*Results of the evaluation*	*References*
Brazil, Social Organizations in Health	X		• Comparison between participants and non-participants (2) • Pair-wise matching (2)	• C-section rates are higher in non-participating hospitals (lower quality of care). • Productivity is higher in participating hospitals, despite lower funding.	Barata et al. (2009) Barata and Mendes (2007) La Forgia and Couttolenc (2008) Harding (2011)
US, California, Integrated Healthcare Association Physician Incentive Programme		X	• Independent evaluation with publication of data yearly	• Improvements over the period of measure.	Integrated Healthcare Association (2012)
US, California, Integrated Healthcare Association Physician Incentive Programme	X		• Modelling (2) • Difference-in-difference analysis (quasi experimental) (1)	• Modest impact on cervical screening and mammography. • Improvements in take-up of tests for patients with diabetes. • Improvements were uneven across different indicators.	Rosenthal et al. (2005) Mullen et al. (2010) Coleman et al. (2007)

Korea, Value Incentive Programme	X	• Health Insurance Review and Assessment Service (HIRA) internally evaluated the programme using yearly publication of performance data	• Comparison shows improvement over the period of measure for acute myocardial infarction. • Reduction in C-section rate over the period of measure.	Health Insurance Review and Assessment Service (2010)
New Zealand, Primary Health Organizations	X	• Performance Management Programme – issues yearly reports • No formal external evaluation	• All indicators have improved, but in an uneven manner across indicators. • Improvements were usually modest.	Performance Management Program (2009)

results. With the exception of a few programmes (e.g. Quality and Outcomes Framework and Advancing Quality (Sutton et al., 2012) in the United Kingdom), dissemination of such evaluation studies has been limited to policymaking circles. In a number of countries, scholars and other third party institutions have undertaken rigorous impact evaluation (e.g. Germany or the United Kingdom), but it is unclear how these results are included in policy discussions. Finally, while ideally impact evaluation should inform the development of the programme throughout time, in reality, we observe that it is usually designed as a 'one-off' task aiming at a certain point in time, specifically for a certain range of measures.

Issues to consider in evaluating P4P programmes

Evaluation of P4P interventions can take numerous forms, depending on the final objectives of the evaluation process. Ex-post evaluation will typically aim to look at the impact of a P4P programme over a relatively long time range. Nonetheless, ex-ante evaluation (e.g. pilot phases) can inform policy decisions in a first learning phase of implementation. Ex-ante evaluation is becoming more popular and in some instances has become an integral part of informed policymaking in other areas of health care policy (e.g. health technology assessments). Ex-ante and ex-post evaluation can be as broad and complex as to understand the overall impact of a policy using extensive quantitative techniques; or more simple to rapidly inform policymaking by using systematic yearly comparisons. Therefore, the first step and issue to consider is to more precisely define the scope and purpose of the evaluation process.

What to evaluate? Choosing the right indicators

First, design of impact evaluation should address the question of 'what to evaluate'; therefore indicators used in programme evaluation will depend largely on the goals of the P4P programme itself and the impact evaluation. Ultimately, P4P programmes aim to improve patient outcomes by motivating changes in the way care is delivered. However, evaluation studies rarely intend to attribute changes in patient outcomes to programme implementation.

As discussed in Chapter 2, P4P indicators usually follow closely the Donabedian (1966) framework of structure, process and outcomes. Most evaluation studies look at the impact of the programme using the same paradigm. This evaluation framework can pose several challenges. For instance, some studies analyse treated patient outcomes (e.g. in hospital mortality rates of patients admitted with a diagnosis of acute myocardial infarction) in the group of participating and non-participating physicians. The risk of using outcome measures is the problem of attribution, i.e. whether the measures of changes in outcome can be linked only to the increased efforts of providers. Amongst other problems, evaluation using patient outcomes measures can be heavily influenced by differences in patient case mix, beyond physician's control. Evaluation design should seek to capture and control for these differences, especially when using

non-experimental and quasi-experimental design such as simple comparison between participating and non-participating physicians.

Alternatively, P4P programmes usually specify aspects of the production process that are believed to represent good quality care, which can be used in the context of evaluation. These quality indicators are often designed on the basis of expert opinion and clinical effectiveness evidence. However, there is surprisingly little clinical effectiveness evidence to support many of the aspects of care that are widely believed to represent good quality (Mason et al., 2008).

P4P also increases the importance of providers keeping good quality records. The introduction of financial incentives where the monitoring of performance relies on provider-generated data is therefore likely to result in changes in data recording as well as in the real quality of care. Evaluation of P4P should attempt to distinguish between these two effects, preferably using data that are not affected by changes in provider recording. A particular form of changes in data recording can occur if providers have control over which patients are judged to be 'eligible' for inclusion in the P4P programme. In the QOF, for example, providers can 'exception report' patients and a small number of providers have been found to exploit this to maximize their revenue (Gravelle et al., 2010). Future evaluations of P4P should therefore pay attention to changes in the size and composition of the eligible population as well as to achievement amongst the eligible population.

The choice of indicators used in impact evaluation also relies largely on the availability and quality of data. The case studies reviewed in this book show that evaluation was seldom planned ex-ante, which has consequences on the extent to which comprehensive impact evaluation is possible. With an ex-ante design, evaluation could rely on timely and convincing information produced as a natural by-product of implementation. Nevertheless, most studies use data routinely collected to assess performance and process payment, and therefore limit the scope of evaluation to collected indicators.

Identifying a suitable comparison group: dealing with selection bias

Having identified the scope of the evaluation, the most important consideration in evaluating a P4P programme is the adequate specification of the 'counter-factual', i.e. what would have happened if the P4P programme had not been introduced. Since most studies show that the quality of health care is improving, it is misleading to simply compare performance after the programme is introduced with the level of performance before the programme was introduced. Studies that compare two groups of providers randomized to participate or not in a trial offer the most plausible counter-factuals; as long as treatment is assigned and implemented randomly, a programme's aggregate or micro-level causal effects can be identified. Nevertheless, in the real world, selection in programme participation is almost never performed on a random basis, or programme applies to all providers in the target group. The use of RCTs is also complicated by administrative, political and ethical concerns.

Some quasi-experimental methods can help address the issue of selection bias. For instance, as long as trends in pre-programme behaviour are similar

across groups, a difference-in-difference design allows researchers to compare changes over time across participants and non-participants. Regression discontinuity methods compare groups that are very similar on a continuous variable but where the probability of participation differs substantially around a threshold value. Though a powerful analytical technique, the estimated programme effects relate only to those either barely eligible or barely ineligible for programme participation. Moreover, while such technique might provide consistent results for the group of interest, it is not clear that these can be generalized to all providers.

Clearly, in the absence of adequate randomization, selection bias presents an important challenge to impact evaluation. To compensate for selection bias, policymakers need to gain a detailed understanding of the provider participation process. In the majority of cases, provider participation is not necessarily controlled by the evaluator, and is either is voluntary or universal. The studies of Germany's DMP show better processes of care and outcomes for DMP enrolees, but the studies fail to account for the fact that individuals choosing to enrol may be more motivated to take control of their own treatment (Altenhofen et al., 2004; Miksch et al., 2010; Drabik et al., 2012). In Australia, participation in the programme is voluntary, and providers can also cherrypick the domains of performance on which they wish to be assessed. The US Premier Hospital Quality Incentive Demonstration (HQID), for example, allowed providers to choose whether to participate or to withdraw during the operation of the programme. Although this is the most widely evaluated programme (Mehrotra et al., 2009), no study has satisfactorily addressed the issue that only five per cent of potential providers participated in the programme.

One problem with self-selection is that the direction of bias is unknown. Providers may select into a programme if they believe that they are already high achievers or if they know that they can improve substantially once the programme is introduced. Participants may therefore be more likely to be high achievers or low achievers prior to the introduction of the programme. Scott et al. (2011) highlighted this problem in their review of financial incentive programmes for improving quality in primary care. In the absence of randomization, a detailed understanding of the participation process is required in order to identify a group of providers that could plausibly serve as a control group. This may, for example, be a group of providers that were not eligible to participate for reasons unrelated to their likely performance had they been eligible to participate in the programme.

Looking beyond targeted indicators: spillover effects and unintended consequences

Examination of only the process indicators incentivized by the P4P programme might be too restrictive for assessing the full impact of P4P interventions. Most impact evaluations are conducted based on the measures and indicators collected to calculate the programme performance scores. It is likely that providers respond to financial incentives with regard to indicators they know to be measured for payment, especially in cases where the size of the bonus

is large. Nonetheless, a major concern is that the providers shift their efforts and attention to the measured indicators, at the expense of other unmeasured aspects of quality of care. In this case, solely relying on the targeted indicators could overestimate the impact of programmes on quality of care. On the other hand, case study P4P programmes show improvements in data collection (especially patient records), transparency, accountability and governance arising from P4P programmes. These results could have important positive effects on quality of care, and might not be reflected in the measured indicators.

The issue of whether P4P is intended to increase provider efforts overall or to divert effort onto prioritized activities is important for the evaluation of their effects. 'Spillovers' of P4P onto non-incentivized elements need to be considered. These unintended effects may come in two forms (Sutton et al., 2010). There may be 'horizontal spillovers' for patients targeted by the programme; for example, if encouraging providers to improve certain aspects of the care of particular patient groups leads to general improvements in their treatment. There may also be 'vertical spillovers' for patients not targeted by the programme. These vertical spillovers may be positive or negative. Positive vertical spillovers may arise if providers begin to deliver certain aspects of care (e.g. more regular blood pressure monitoring) for all patients, regardless of whether such patients are in the groups targeted by the programme. Negative vertical spillovers may arise if providers focus their efforts on the patients targeted by the programme at the expense of patients not targeted by the programme. The effects of P4P on non-incentivized aspects of care are not well studied – Sutton et al. (2010) found substantial positive horizontal spillovers while Doran et al. (2011) found that improvements associated with financial incentives seem to have been achieved at the expense of small detrimental effects on aspects of care that were not incentivized.

The possibility of spillovers has a profound effect on the design of evaluations of P4P programmes. Studies that focus only on whether providers improved incentivized aspects of care risk omitting some important consequences. Sutton et al. (2010) found that inclusion of positive horizontal spillovers reduced the implicit unit costs of the QOF by a factor of two. To measure the spillover effects of P4P programmes, evaluations should examine changes in non-incentivized aspects of care, both for the targeted patients and the untargeted patients. This possibility also affects the choice of patient groups and activities that can serve as 'controls' for the evaluation. If all patient groups and activities can be affected by the introduction of P4P for a subset of patients and activities, then information on the counterfactual can only be obtained from providers not exposed to the financial incentives.

Focusing on equity: evaluations should take a closer look at the beneficiaries of P4P

P4P programmes are designed to change the way in which providers treat patients. It is unlikely that providers will start from a position of offering the incentivized aspects of care to none of their patients and to finish up delivering the incentivized aspects of care to all of their patients. There are therefore likely

to be distributional consequences, with some patient groups receiving good quality care regardless of the incentive programme, some patient groups not receiving good quality care regardless of the incentive programme, and some patients only receiving good quality care because of the incentive programme. There is relatively little evidence on the distributional consequences of incentive programmes (Alshamsan et al., 2010). A concern frequently expressed is that providers will 'cherrypick' the easiest patients for inclusion in the P4P programme. However, providers may already be electing to provide good quality care to the 'easiest' patients and P4P may force providers to focus on more costly patients. Future evaluations of P4P programmes should facilitate further understanding of their distributional consequences by estimating average treatment effects for different socio-economic and demographic groups. An equity focus requires the existence of disaggregated data, but is well suited to quasi-experimental methods if data permit examination of trends in different social groups.

When to evaluate? Short-term vs long-term effects

The timeframe over which P4P programmes are expected to achieve results, and therefore be evaluated, is important. P4P programmes are meant to trigger changes at different levels (e.g. *provider practice of care, patient outcomes*), which can operate on different time horizons. The timeline for change is also intrinsically determined by the way programmes are planned, designed and implemented (Sridharan et al., 2006). Furthermore, programmes may trigger a 'spike' effect in improvement early on in the programme, which may plateau or decline as the programme matures, or alternatively some effects may take time to realize when they are dependent on provider investment, organizational changes, or complex behaviour change.

The indicators that are used to monitor achievement on P4P programmes are often short term so as to reward providers quickly for their additional costs and efforts. However, the health gains may accrue over a longer period of time, meaning that the evaluation should in principle continue after the end of the monitoring period. In addition, providers may make quality improvements in the short term that cannot feasibly be sustained in the longer term. From an initially high baseline, providers cannot continue to make five per cent performance improvements year-on-year, making the impact of P4P programmes decline over time.

Finally, very little is known about how providers respond when financial incentives are removed. If quality improvement is an investment activity, i.e. providers 'learn' how to improve the quality of their production process, then higher quality may be sustained when the financial incentives are removed. Alternatively, if quality improvement is transitory, performance may reduce once the incentives are removed. There is remarkably little evidence on whether decision makers should continue P4P programmes in the longer term. One paper addressing this question found that performance dropped to levels below that which was delivered prior to the introduction of financial incentives, once the incentives were removed (Lester et al., 2010).

Notwithstanding the frequent need to take a longer evaluative perspective, policymakers clearly need timely feedback on the effectiveness of a P4P programme. Without such information they are unable to judge whether to expand, abandon or amend the programme. The tension between comprehensiveness and timeliness is a recurring theme in evaluative studies and in policy circles. One option is to build a timeline for change with the involvement and expertise of key stakeholders with expertise in programme implementation (Sridharan & Nakaima, 2011).

Focusing on programme costs as well as effectiveness

There are a variety of ways in which to understand P4P as an intervention. P4P programmes are frequently described as 'bonus' programmes that reward providers for making additional effort and improving the efficiency of their care delivery. It is also possible, however, that such programmes are a form of cost reimbursement, with the additional revenue acting as compensation for the providers for the costs they incur in improving the quality of their care delivery. P4P often involves an increase in the amount of resources that purchasers make available to providers as well as the change in the way that providers are paid. In evaluating P4P it is therefore important to be clear about what the comparator is. If the purpose is to evaluate P4P as a way of paying providers, the comparator should be an equivalent expected amount of resources paid in an alternative manner (e.g. block grant or increase in all per-case tariffs). Otherwise, the evaluation is of P4P as a way of increasing payments to providers in a particularly manner.

In a recent commentary, Maynard (2012) highlights the 'curious' focus of research to date on the effectiveness of P4P programmes, with a neglect of their costs, and therefore cost effectiveness. In some of the programmes documented in the book, information on the cost of the programme, average payment per physician/institution, and distribution of payments was not readily available. No programme attempts to measure the cost to providers of participating in the programme or meeting initial requirements, which appears to be significant in some cases. The literature on how to evaluate the cost effectiveness of P4P programmes remains underdeveloped (Meacock et al., 2012), with most studies focusing only on the costs of the incentives paid out (which only constitute a part of the total costs) and on the intended, direct consequences. More comprehensive assessments of the wider costs and consequences of P4P programmes are required.

Conclusions

Capturing the full impact of P4P programmes, controlling for underlying trends, and finding suitable cost and benefit measures and counter-factual groups for evaluation are important challenges in evaluating P4P programmes. The way in which P4P programmes are introduced will determine the choice of evaluation technique and, since randomization is rarely practical or politically

acceptable, innovative quasi-experimental techniques will be required. Use of these techniques places more onus on the evaluator to demonstrate that possible causes of bias and confounding have been satisfactorily addressed.

The choice of evaluation technique depends on the availability of data and P4P programme design and contextual factors: e.g. size, motivation and observed characteristics of participants or the use of supporting levers, including public reporting of results, the potential for patient choice and the facilities for shared learning (Van Herck et al., 2010). Evaluation methods also differ markedly in levels of rigour and costs. More attention should be paid to building evaluation into P4P programmes at the design stage, to ensure that relevant information can be collected in order to properly address evaluation questions.

Given large variations in design and context, it is questionable whether evaluations of specific P4P programmes as a whole will produce transferable results. It may be more useful for future evaluations to examine the effects of specific design decisions (e.g. whether to use bonuses or penalties, whether to reward achievement of targets or improvement) across different programmes in a similar context. If P4P influences provider behaviour, then these design aspects of P4P programmes – effectively the underlying intervention ingredients and causal mechanisms – should matter. Implementation of differently designed P4P programmes in similar contexts may be more feasible, and evaluations of these initiatives may offer more useful knowledge to purchasers considering new P4P programmes.

All studies of P4P programmes have identified unexpected effects, both positive and negative. It would therefore be most desirable to undertake evaluation alongside the implementation of a P4P programme and to agree with providers in advance that the P4P programme will evolve over time in response to the evaluation findings. This would allow purchasers the opportunity to identify performance measures that are most closely linked to outcomes, find meaningful levels for bonuses or penalties without overpaying, introduce approaches to protect areas suffering negative spillovers, and adjust implementation arrangements to ensure that programmes are fair and transparent.

References

Alshamsan, R. et al. (2010) Impact of pay for performance on inequalities in health care: systematic review, *Journal of Health Services Research and Policy*, 15(3): 178–84.

Altenhofen, L. et al. (2004) *Report on quality assurance – disease – management programmes in North-Rhine*. Düsseldorf: KV Nordrhein [in German] (http://www.kvno.de/importiert/qualbe_dmp04.pdf, accessed June 2013).

Australian National Audit Office (2010) *Practice incentives program audit report No 5 2010–2011*. Canberra: Australian Audit Office.

Barata, L. and Mendes, J. (2007) *Organizações Sociais de Saúde: a Experiência Exitosa de Gestão Pública de Saúde do Estado de São Paulo*. São Paulo: Secretaria de Estado da Saúde – Gabinete do Secretário.

Barata, L. et al. (2009) Comparação de grupos hospitalares no Estado de São Paulo, *Revista de Administração em Saúde*, 11(42): 8–15.

Caisse Nationale d'Assurance Maladie (2013) *Rémunération des objectifs de santé publique: une mobilisation des médecins et de l'assurance maladie en faveur de la*

qualité des soins. Press release 11 April (http://www.ameli.fr/fileadmin/user_upload/documents/DP_Bilan_ROSP_1_an_11042013_VDEF3.pdf, accessed June 2013).

Campbell, S. et al. (2007) Quality of primary care in England with the introduction of pay for performance, *New England Journal of Medicine*, 357(2): 181–90.

Coleman, K., Reiter, K.L. and Fulwiler, D. (2007) The impact of pay-for-performance on diabetes care in a large network of community health centers, *Journal of Health Care for the Poor and Underserved*, 18(4): 966–83.

Donabedian, A. (1966) Evaluating the quality of medical care, *Milbank Memorial Fund Quarterly*, 44(3): 166–206.

Doran, T. et al. (2011) Effect of financial incentives on incentivised and non-incentivised clinical activities: longitudinal analysis of data from the UK quality and outcomes framework, *British Medical Journal*, 342(7814).

Drabik, A. et al. (2012) Life prolonging of disease management programmes in patients with type 2 diabetes is cost-effective. Diabetes, *Diabetes Research and Clinical Practice*, 95(2): 194–200.

Eichenlaub, A. et al. (2004) Evaluation des Modellprogramms 'Diabetes Sachsen-Anhalt', Im Auftrag des AOK Bundesverbandes und des IKK Bundesverbandes, Köln.

Fleetcroft, R. et al. (2010) The UK pay-for-performance programme in primary care: estimation of population mortality reduction, *British Journal of General Practice*, 60(578): 345–52.

Flodgren, G. et al. (2011) An overview of reviews evaluating the effectiveness of financial incentives in changing healthcare professional behaviours and patient outcomes, *The Cochrane Library 2011*, Vol. 7.

Glenton, C., Lewin, S. and Scheel, I.B. (2011) Still too little qualitative research to shed light on results from reviews of effectiveness trials: a case study of a Cochrane review on the use of lay health workers, *Implementation Science*, 6(53).

Glickman, S. et al. (2007) Pay for performance, quality of care, and outcomes in acute myocardial infarction, *Journal of American Medical Association*, 297(21): 2373–80.

Gravelle, H., Sutton, M. and Ma, A. (2010) Doctor behaviour under a pay for performance contract: treating, cheating and case finding?, *The Economic Journal*, 120(542): 126–57.

Grossbart, S. (2008) Effectiveness of pay for performance as a quality improvement strategy, *Prescriptions for Excellence in Health Care*, 3: 1–3.

Harding A. (2011), Hospital performance: what is it? How do you get it? Presentation prepared for the health system performance flagship course, Tallinn, Estonia, 14–18 November.

Health Insurance Review and Assessment Service (2010) *Comprehensive quality report of national health insurance 2009*. Seoul: Health Insurance Review and Assessment Service.

Integrated Healthcare Association (2012) *IHA pay for performance report of results for measurement year 2010*. Oakland, CA: Integrated Healthcare Association.

Jha, A.K. et al. (2012) The long-term effect of premier pay for performance on patient outcomes, *New England Journal of Medicine*, 366(17).

Jones, N., Steer, L. and Datta, A. (2009) Improving impact evaluation production and use, *Overseas Development Institute*, Working Paper 300, March.

Jones, A.M. and Rice, N. (2011) Econometric evaluation of health policies, in S. Glied and P. Smith (eds), *The Oxford handbook of health economics*. Oxford: Oxford University Press.

La Forgia, G.M. and Couttolenc, B.F. (2008) *Hospital performance in Brazil – the search for excellence*. Washington, DC: World Bank.

Lester H. et al. (2010) The impact of removing financial incentives from clinical quality indicators: longitudinal analysis of four Kaiser Permanente indicators, *British Medical Journal*, 340: c1898.

Lindenauer, P.K. et al. (2007) Public reporting and pay for performance in hospital quality improvement, *The New England Journal of Medicine*, 365(5): 486–96

Mason, A. et al. (2008) The GMS quality and outcomes framework: are the quality and outcomes framework (QOF) indicators a cost-effective use of NHS resources?, in University of East Anglia and University of York *Joint Executive Summary: Reports to the Department of Health.*

Maynard, A. (2012) The powers and pitfalls of payment for performance, *Health Economics*, 21(1): 312.

Meacock, R., Kristensen, S.R. and Sutton, M. (2012) The cost-effectiveness of using financial incentives to improve provider quality: a framework and application. Paper presented to the Health Economists' Study Group, Oxford, 25–27 June.

Mehrotra, A. et al. (2009) Pay for performance in the hospital setting: what is the state of the evidence?, *American Journal of Medical Quality*, 24(1): 19–28.

Miksch, A. et al. (2010) Is there a survival benefit within a German primary care-based disease management program? *American Journal of Managed Care*, 16(1): 49–54.

Mullen, K., Frank, R., and Rosenthal, M. (2010). Can you get what you pay for? Pay for performance and the quality of healthcare providers, *RAND Journal of Economics*, 41(1): 64–91.

Performance Management Program (2009) *Annual Report: 1 July 2008 – 30 June 2009*. Auckland: Ministry of Health.

Rosenthal, M. et al. (2005) Early experience with pay for performance: from concept to practice, *Journal of the American Medical Association*, 294(14): 1788–93.

Ryan, A.M. (2009) Effects of the premier hospital quality incentive demonstration on Medicare patient mortality and cost, *Health Services Research*, 44(3): 821–42.

Savedoff, W.D. et al. (2006) When will we ever learn? Improving lives through impact evaluation: report of the Evaluation Gap Working Group. Washington, DC: Center for Global Development.

Schäfer, I. et al. (2010) The disease management program for type 2 diabetes in Germany enhances process quality of diabetes care – a follow-up survey of patient's experiences, *BMC Health Services Research*, 10(55).

Scott, A. et al. (2009) The effects of an incentive program on quality of care in diabetes management, *Health Economics*, 18(9): 1091–1108.

Scott, A. et al. (2011) The effect of financial incentives on the quality of health care provided by primary care physicians, *The Cochrane Library 2011*, Vol. 9.

Serumaga, B. et al. (2011) Effect of pay for performance on the management and outcomes of hypertension in the United Kingdom: interrupted time series study, *British Medical Journal*, 342: d108.

Skoufias, E. and McClafferty, B. (2001) Is Progresa Working? Summary of the results of an evaluation by IFRI, *FNCD Discussion Paper*, 118.

Sridharan, S., Campbell, B. and Zinzow, H. (2006) Developing a stakeholder-driven timeline of change for evaluations of social programmes, *American Journal of Evaluation*, 27(2): 1–15.

Sridharan, S. and Nakaima, A. (2011) Ten steps to making evaluation matter, *Evaluation and Program Planning*, 34(2): 136–45.

Sutton, M. et al. (2010) Record rewards: the effects of targeted quality incentives on the recording of risk factors by primary care providers, *Health Economics*, 19: 1–13.

Sutton, M. et al. (2012) Reduced mortality with hospital pay for performance in England, *New England Journal of Medicine*, 367: 1821–8.

Szecsenyi, J. et al. (2008) German diabetes disease management programmes are appropriate for restructuring care according to the chronic care model: an evaluation with the patient assessment of chronic illness care instrument, *Diabetes Care*, 31(6): 1150–4.

Taggar, J.S. et al. (2012) The impact of the quality and outcomes framework (qof) on the recording of smoking targets in primary care medical records: cross-sectional analyses from the health improvement network (thin) database, *BMC Public Health*, 12(1): 329.

UK National Audit Office (2008) *NHS pay modernisation: new contracts for general practice services in England*. London: National Audit Office.

Van Herck, P. et al. (2010) Systematic review: effects, design choices, and context of pay-for-performance in health care, *BMC Health Services Research*, 10: 247.

Västra, K. (2010) Assessing the impact of implementing primary care quality bonus system on follow up of patients with hypertension and type 2 diabetes based on Estonian health insurance fund claims registry data in 2005–2008. Unpublished Masters Thesis: University of Tartu.

chapter five

Lessons from the case study P4P programmes

Cheryl Cashin, Y-Ling Chi and Michael Borowitz

Introduction

Most OECD countries are implementing some form of pay for performance (P4P) in the health sector to better align incentives for health care providers with health system objectives, particularly improved quality of care. These efforts continue to expand in spite of the limited evidence that P4P leads to significant improvements in quality of care and health outcomes, and in the absence of guiding information on effective design and implementation of P4P programmes. In this study, we examined the objectives, design, implementation and results of 12 P4P programmes in OECD countries. The case studies qualitatively examined the 'net effect' of P4P programmes on health system objectives, which included not only the direct effects on quality, outcomes, equity and efficiency, but also the unintended consequences, both positive and negative. Ultimately, the net effect of the programmes is determined by the interplay of the financial incentives, the provider responses to those incentives, and implementation arrangements and contextual factors. Although we do not categorize individual case study programmes as more or less successful, we draw conclusions in the following sections by considering more effective programmes to be those that are likely to have a net positive effect on health system performance and objectives, as reflected by trends in performance indicators, published studies, and stakeholder perceptions.

The main finding from the case studies that follow is that P4P did not lead to 'breakthrough' performance improvements in any of the programmes. Most of the programmes did, however, contribute to a greater focus on health system objectives, better generation and use of information, more accountability, and in some cases a more productive dialogue between health purchasers and providers. This also can be described as more effective health sector governance and more strategic health purchasing.

The findings of this study are in line with several reviews that conclude that P4P programmes in their entirety may be more powerful than the sum of

their parts (Damberg et al., 2005; Campbell, MacDonald & Lester, 2008; Martin, Jenkins & Associates Limited, 2008; Van Herck et al., 2010). We also find that the most important contributions of P4P programmes may be their reinforcing effects on broader performance improvement initiatives, and their spillover effects, or other health system strengthening that occurs as a by-product of the incentive programmes. Several programmes report that the improved generation and use of data for performance improvement, faster uptake of IT, more quality improvement tools (e.g. guideline-based decision aids), sharper focus on priorities, and better overall governance and accountability are more important outcomes of the P4P programmes than improvements in performance indicators. In some cases, the programmes provided the opportunity for dialogue around performance measures and accountability, which previously had been topics too sensitive to raise directly.

In the sections that follow, we discuss the overall effect of P4P programmes on provider performance in the 12 case study programmes (summarized in Table 5.1) and any unintended consequences. We highlight key lessons about programme design and implementation and identify contextual factors that appear to enhance the effectiveness of programmes or detract from success.

Overall results of the case study P4P programmes

Quality of care

Improvements were achieved for coverage of preventive services in some programmes and for some conditions but not others. Two of the programmes that rewarded increased coverage of preventive services, the Estonia QBS and New Zealand PHO Performance Programme, showed large increases in coverage rates, particularly for childhood immunization, and screening for breast cancer, cervical cancer, and cardiovascular disease risk factors. Childhood vaccination rates increased 30 percentage points in New Zealand (from 60 to 90 per cent) over a six-year period, and cardiovascular disease screening increased 20 percentage points from 30 to 50 per cent of the target population. In Estonia cholesterol screening increased 20 percentage points between 2007 and 2010, while other cardiovascular disease prevention services such as rates for follow-up tests for high-risk patients actually decreased (possibly due to the increased case finding). No significant improvements in coverage of most preventive services have been found in the France ROSP programme, with the exception of increases in the prescribing of vasodilators and benzodiazepines for elderly patients. None of these results, however, control for underlying trends that may have been occurring independently of the programmes.

Results based on more rigorous evaluation that controlled for underlying trends are more mixed. In the UK QOF, for example, coverage with influenza immunization increased only 3.5 percentage points during the first three years of the programme (from 67.9 to 71.4 per cent), controlling for other factors. Larger increases were observed for populations with the lowest immunization rates, with increases up to 16 percentage points for individuals less than

65 years of age with a previous stroke (Norbury, Fawkes & Guthrie, 2011). In the California IHA programme, a rigorous study of changes in performance that could be attributed to the programme found that only cervical cancer screening improved differentially among the IHA participants, and improvement was modest at 3.5–6 percentage points. In Australia, one study found that the PIP was associated with an increase in the probability of diabetes testing of 20 percentage points (Scott et al., 2009). A more recent study, however, found that neither signing onto the PIP programme nor claiming incentive payments was associated with increased diabetes testing or cervical cancer screening (Greene, 2013).

Some programmes have shown modest-to-significant improvements in chronic disease management. The Germany DMP has been widely studied and demonstrated the most improved processes of care and better patient outcomes. Sickness funds received higher payments for DMP enrolees through the risk adjustment system, and payments continue to be made to physicians for care management services, such as documentation and patient education that were not previously reimbursed as separate services. Studies have found significantly better processes of care in general as a result of these incentives (Schäfer et al., 2010), including more time spent with a care coordinator and more patient education (Schoul & Gniostko, 2009). The DMPs are found to lead to more patient-centred care for diabetes and asthma, with patients reporting better understanding and control of their conditions (Schoul & Gniostko, 2009; Mehring et al., 2012).

In the Australia PIP, linking bonus payments to the completion of evidence-based cycles of care for asthma and diabetes led to a significant increase in the number of cycles completed for both conditions according to claims data analysed by the Australia National Audit Office (ANAO, 2010). The PIP's Practice Nurse Incentive also has been associated with improved management of chronic diseases through a general greater involvement of nurses in chronic care, leading to increased time spent with patients and reduced waiting times (ANAO, 2010). A study of the Estonia QBS found that family physicians participating in the programme and achieving a high enough performance score to receive a bonus perform better in providing continuous follow-up for patients with chronic conditions, and their patients tend to require specialist services and hospitalization less frequently (Västra, 2010).

The improvements in chronic disease management found in these programmes appear to be driven by better alignment of incentives with evidence-based processes of care rather than through targeted, indicator-based incentives. In the Australia and Germany programmes in particular, P4P payments have served as a way to pay providers for aspects of chronic disease management that are not typically reimbursed under fee-for-service payment systems and therefore have tended to be neglected. In the Australia PIP, part of the bonus being linked to the completion of a cycle of care rather than for each individual contact appeared to increase compliance with treatment guidelines. The France ROSP programme and the UK QOF rely on targeted indicator-based incentives, such as the percentage of diabetic patients receiving appropriate tests, and these programmes have shown more modest improvements in chronic disease management.

The programmes achieved very limited or no improvement in specific processes of care in hospital-based programmes. The two P4P programmes targeted to specific hospital processes of care showed only modest improvements in performance at best. The US HQID programme has shown very limited positive results. One study found that hospitals participating in HQID hospitals had slightly greater improvements in quality over a two-year period than comparable hospitals with public reporting alone (Lindenauer et al., 2007). Another study, however, found that the performance of HQID hospitals accelerated in year one of the programme, but that the scores converged with non-HQID hospitals over three years (Grossbart, 2008). A third study found that participation in the HQID was not associated with a significant improvement in quality of care processes or outcomes for acute myocardial infarction (Glickman et al., 2007). In the Korea VIP, the overall composite score for acute myocardial infarction (AMI) increased only 5.3 percentage points during the first three years of the programme, but the baseline level was high (92.1 per cent). The programme led to almost no reduction in the Caesarean section rate.

Health outcomes

Programmes generally fail to have an impact on health outcomes. The experience of the P4P programmes reviewed is consistent with the lack of evidence in the literature that selected process measures can be linked to improved outcomes (Bradley et al., 2006; Mattke et al., 2007; Morse et al., 2011; Pimouguet et al., 2011; Shahian et al., 2012). Even the highly stylized indicator framework and high achievement rates in the UK QOF have failed to show any impact on health outcomes. Only the Germany DMP was able to demonstrate an impact on health outcomes, and the results were modest. One study found that participation in a diabetes DMP was associated with a reduction in hospitalization rates and a reduction in the three-year mortality rate from 14.4 to 11.3 (Miksch et al., 2010). Another study found participation in a DMP was associated with an additional 60 days survival time over a three-year period (Drabik et al., 2012).

Equity

Programmes have mixed effects on equity, even when explicit steps are taken to favour underserved populations or geographic areas. The Australia PIP and New Zealand PHO Performance programme emphasized improving quality and accessibility of care for underserved populations or rural and remote areas through targeted incentives or higher payment rates. The Australia PIP aims to improve equity through higher overall payment rates for rural primary care practices, which represents an important source of revenue for some practices (ANAO, 2010). The additional resources available to rural practices have contributed to financial viability for some, possibly contributing to the reduction of rural–urban inequalities (ANAO, 2010). The

New Zealand PHO Performance Programme has a strong focus on the MOH priority of reducing health disparities. Some indicators are measured separately for high-needs populations, and payments are weighted more highly for achieving targets for high-needs populations. Progress on reducing health disparities has been modest, however, with only breast cancer screening rates improving disproportionately for the high-needs population (PHO Performance Programme, 2012).

In programmes without explicit steps to improve equity, the picture is also mixed. The UK QOF does not have a specific objective to achieve improvements in equity, but a number of studies have explored its effect, and some modest positive impacts have been found. Although QOF performance initially was slightly lower in deprived areas, there is evidence of some 'catch up' (Doran et al., 2008; UK National Audit Office, 2008). The difference in the mean QOF score between least deprived and most deprived quintiles fell from 64.5 points (2004/05) to 30.4 (2005/06) (Ashworth et al., 2007). A systematic review of the equity effects of the QOF found small but significant differences that favoured less deprived groups, but these differences were no longer observed after correcting for practice characteristics (Boeckxstaens et al., 2011).

In the California Integrated Healthcare Association (IHA), it appears that the P4P programme may not have distributed its benefits equally. First, while there has been some compression in the distribution of performance scores, physician groups that performed poorly on quality measures at the launch of the programme have not caught up with high performers and overall have received only a small share of payments (Damberg et al., 2005). Second, there is substantial geographic variation in performance, which may be associated with factors such as socio-economic status and local health care delivery system capacity (IHA, 2009). Finally, interviews with physician group leaders revealed some concerns that the P4P programme further entrenched existing health inequities and possibly has caused groups to avoid patients whose health or health behaviour would negatively affect the group's performance (Hood, 2007).

In the US HQID, there is some evidence that the programme helped close the performance gap between hospitals serving poorer and wealthier populations. Among hospitals caring for a high proportion of poor patients, those participating in HQID improved at a more rapid rate than those not participating in HQID (Jha, Orav & Epstein, 2010).

Patient experience

Patient experience is not a common performance domain, and no improvements have been shown in the programmes that include patient experience measures. Among the 12 programmes reviewed, patient experience was included as a performance domain only in the Brazil OSS, UK QOF and California IHA. In the UK QOF, measures of patient experience initially included three indicators related to time to get an appointment with a GP and length of the consultation. In spite of achievement rates consistently well over 90 per cent in the other three QOF performance domains, patient experience

showed results closer to 70 per cent, until all but one performance measure in the domain was eliminated, which made the results appear better. In California's IHA programme, patient experience measures have typically been over 80 per cent, with no significant improvement over the life of the programme. Measures of timely access to care and care coordination are lower at around 75 per cent, again with no change as a result of the programme. It is questionable whether performance measures for patient experience have been adequately validated for use in the P4P programmes, and whether enough is understood about the steps which providers can take to improve the perceptions and experience of their patients, and the investment that may require.

Efficiency and costs

Programmes that have achieved broad-based improvements in processes of care have also generated some efficiency gains and cost savings. Significant improvements in general processes of care have been found in the Germany DMP, Maryland HAC, and Estonia QBS. All of these programmes also report efficiency gains, and even direct cost savings in the case of the DMP and MHAC. DMP enrolees had a lower annual net cost per patient (€122 vs. €169) (Drabik et al., 2012). Germany's largest insurer AOK reports net cost savings ranging from 8–15 per cent of total annual costs of care for enrolees with chronic conditions (Stock et al., 2011). By reducing avoidable hospital complications by 15 per cent, the Maryland HAC programme has generated $110.9 million savings to the system (see Chapter 16). In Estonia, no direct cost savings have been reported as a result of the QBS, but lower referral rates to specialist providers and hospitalization led to net savings of the programme (Västra, 2010).

In the Brazil OSS, greater hospital autonomy combined with performance-based financial incentives has led to large efficiency gains and cost savings. Hospitals with performance-based contracts provided care of equal or better quality than non-contracted hospitals, with a 50 per cent lower cost per discharge. Other indicators of efficiency, such as hospital occupancy rate, bed turnover rate, and average length of stay also showed significantly better performance for contracted hospitals. It is difficult to disentangle, however, how much of the improvement can be attributed to the financial incentive and how much simply to greater autonomy in decision making and resource use (La Forgia & Couttolenc, 2008).

Generating efficiency gains and cost savings through targeted incentives has been less successful. In other programmes that attempt to generate efficiency gains through targeted incentives, the results have been disappointing. In the France ROSP programme, for example, the National Health Insurance Fund intended to make the programme cost neutral by offsetting the costs of the incentive payments and programme administration with savings generated by the replacement of branded medicine by generic prescribing. Results show, however, that prescribing practices have not changed significantly in response to this programme. In the UK QOF providers are rewarded for prescribing medicines that are cost effective, but higher

quality scores related to prescribing are not associated with lower spending on medicines (Fleetcroft et al., 2011). The California IHA programme recently added a set of 21 indicators related to more effective resource use, such as generic prescribing and emergency department visits. Providers will be rewarded by sharing any savings that are generated by better performance in these areas. This new domain will start payouts in 2013, so results are not yet available at the time of writing.

Costs to providers of participating in P4P programmes have not been measured but may be significant in some cases, and may have to be offset by the programme. Very little information is available about the cost to providers of participating in P4P programmes and complying with reporting and other programme requirements. Although most programmes rely mainly on existing claims data, they also typically have new data reporting requirements, particularly for non-clinical indicators. Putting the appropriate data systems in place or preparing new reports can be costly to providers. In the US HQID Premier, Inc. required that hospitals renew their subscription to the relatively expensive database tool as a condition for participation, and cost was seen as a limiting factor for expanded participation (Grossbart, 2008). The UK QOF requires sophisticated standardized clinical information systems, which has involved significant investment that has been shared between the NHS and GP practices. In 2004 alone 30 million GBP additional capital funding was made available to support the upgrading of clinical data systems and to provide systems for non-computerized practices (UK National Health Service, 2004).

A number of programmes have prerequisites for participation that may require investments to be made by providers. In the New Zealand PHO Performance Programme, for example, PHOs must fulfil eligibility criteria demonstrating that they have clinical governance structures in place to support the programme. In some cases the programme has offset the additional investment costs to lower the burden on providers and encourage participation. In the Australia PIP, accreditation is a prerequisite of participation, and the Department of Health has had to bear some of the costs, particularly for rural practices (see below).

In some programmes providers question whether the incentive payments are sufficient to cover the costs of participation in the programme and generate net financial gains. One review found that participation in Australia's PIP accounted for nearly 33 per cent of GP practice administrative costs (Productivity Commission, 2003). The issue was taken up again by the Regulation Task Force in 2006 (Commonwealth of Australia, 2009). In New Zealand, one large network of PHOs estimated that just under half of the funds it anticipated earning from the PHO Performance Programme would be needed to run the Programme (Buetow, 2008).

These claims of higher provider costs may ignore cost savings to providers from better processes, particularly in hospitals. Some leaders of hospitals participating in HQID, for example, claimed that the bonus money did not cover the administrative costs that the project imposes on their institutions (Hospitals and Health Networks, 2007). Premier Inc. on the other hand claimed that their analyses showed cost savings to hospitals related to the quality improvements driven by the programme.

Table 5.1 Summary of the results of the case study P4P programmes

Country	*Programme*	*Results*				
		Coverage of preventive services	*Processes of care*	*Outcomes*	*Equity*	*Efficiency*
Australia	**PIP**	Mixed results – one study showed no increase in diabetes testing or cervical cancer screening (Greene, 2013); one study showed probability of diabetes testing increased by 20 per cent points (Scott et al., 2009).	Practice Nurse Incentive has led to improved management of chronic diseases, increased time spent with patients, and reduced waiting times (ANAO, 2010). Some positive impact on the completion of treatment cycles for asthma and diabetes based on claims data – no control for underlying trends (ANAO, 2010).	Not available.	Higher payments for rural primary care practices and the rural loading factor has been important for financial viability in some cases (ANAO, 2010).	Not available.
Brazil – São Paulo	**OSS**	Not applicable.	Not available.	Not available.	Not available.	Cost per discharge 50 per cent lower; bed turnover and occupancy rates higher; ALOS 20 per cent shorter (La Forgia & Couttolenc, 2008).

Estonia	**PHC QBS**	Rate children's immunization and check-ups increased 1–10 per cent points Rate of cholesterol screening increased from 41 to 61 per cent Other cardiovascular disease prevention services decreased 3–12 per cent points All results (Estonian Health Insurance Fund, 2011) – no control for underlying trends	Family physicians achieving a high enough performance score to receive a bonus perform better in providing continuous follow-up for chronic patients (Västra, 2010).	Patients of family physicians achieving a high enough performance score to receive a bonus tend to require specialist services and hospitalization less frequently (Västra, 2010).	Not available.	Possibly net savings from lower rates of specialty referrals and hospitalization, but not analysed (Västra, 2010).
France	**ROSP**	Modest improvements.	Appropriate diabetes management and generic prescribing improved between 2011–2012 but still below NHIF–set targets (CNAM, 2013).	Not available.	Not available.	No significant changes.

(*continued*)

Table 5.1 Summary of the results of the case study P4P programmes (*continued*)

Country	*Programme*	*Results*				
		Coverage of preventive services	*Processes of care*	*Outcomes*	*Equity*	*Efficiency*
Germany	**DMP**	Not available.	Significantly better processes of care for DMP participants (Schäfer et al., 2010). Type 2 diabetic patients more likely to receive patient-structured and coordinated care (Szecsenyi et al., 2008). Asthma education increased from 4.4 to 23.4 per cent and utilization of an individual self-management plan increased from 40.3 to 69.3 per cent (Mehring et al., 2012). 99 per cent of patients reported spending more time with coordinating doctor; 97 per cent understand their disease and treatment much better; and 87 per cent feel they are in better control of their disease (Schoul & Gniostko, 2009).	Reduction in the share of diabetes patients with blood sugar levels outside the target range from 8.5 to 7.9 per cent within a six-month period (Altenhofen et al., 2004). Participation in a diabetes DMP was associated with a reduction in hospitalization rates and a lower three-year mortality rate (11.3 vs. 14.4 per cent) (Miksch et al., 2010). Participation in a DMP was associated with a modest increase in survival time over a three-year period (1045 vs. 985 days) (Drabik et al., 2012).	Not available.	Lower costs per patient (€122 vs. €169) including DMP administration and service costs (Drabik et al., 2012). Germany's largest insurer reports net cost savings ranging from 8–15 per cent of total annual costs of care for enrolees with chronic conditions (Stock et al., 2011).

Korea	**VIP**	Not available.	Overall composite score for acute myocardial infarction treatment increased from 92.1 to 97.4 per cent (HIRA, 2010). Caesarean section rate decreased by only 1.6 per cent points (HIRA, 2010).	Not available.	Not available.	Not available.
New Zealand	**PHO Performance Programme**	Breast cancer screening rates increased from 55 to 68 per cent for the total population (from 42 to 63 per cent for high-needs population) between 2006 and 2012. Cervical cancer screening increased from 66 to74 per cent for the total population (63 to 66 per cent for high-needs population). Cardiovascular disease screening rate increased from 30 to 50 per cent. Diabetes detection and follow-up rate increased from 46 to 72 per cent for the total population (50 to 70 per cent for high-needs population). Childhood vaccination rates increased from 60 to 90 per cent.	Not available.	Not available.	Only breast cancer screening rates increased disproportionately for high-needs population (PHO Performance Programme, 2010).	Not available.

(*continued*)

Table 5.1 Summary of the results of the case study P4P programmes (*continued*)

Country	*Programme*	*Results*				
		Coverage of preventive services	*Processes of care*	*Outcomes*	*Equity*	*Efficiency*
		No change in flu vaccination rate. All results (PHO Performance Programme, 2012) – no control for underlying trends.				
Turkey	**FM PBC**	Significant improvement in primary care service utilization, patient satisfaction, and health outcomes over the period of the programme, as well as reduced regional disparities, but cannot be directly attributed to the programme alone.				Not available.
UK	**QOF**	Influenza immunization rates increased from 67.9 to 71.4 per cent between 2003/04 and 2006/07. Rates of increase were higher for populations with previously low rates (e.g. up to 16 per cent point increase for individuals <65 years of age with previous stroke) (Norbury, Fawkes & Guthrie, 2011).	The high level of achievement of targets linked to only modest measurable improvements in quality asthma and diabetes (Campbell et al., 2007). Some improvements in the quality of diabetes care in the first year related to documentation of recommended aspects of clinical assessment, not patient management or outcomes of care. Improvements insubsequent years were more modest (Kontopantelis et al., 2013).	No impact on rates of heart attacks, kidney failure, stroke or death (Serumaga et al., 2011).	QOF performance is slightly lower in deprived areas (UK NAO, 2008), but evidence of some 'catch up' (Doran et al., 2008). The difference in mean QOF score between least and most deprived quintiles fell from 64.5 to 30.4 points between 2004 and 2006 (Ashworth et al., 2007). A systematic review found small but significant differences that favoured less deprived groups, but no longer observed after correcting for practice characteristics (Boeckxstaens et al., 2011).	Higher quality scores related to prescribing not associated with lower spending on medicines (Fleetcroft et al., 2011).

US – California	**IHA**	Only cervical cancer screening improved differentially among the IHA participants (3.5–6 percentage points) (Mullen, Frank & Rosenthal, 2009).	Appropriate treatment for upper respiratory infection increased by two percentage points – no control for underlying trends (IHA, 2009).	Cholesterol control for cardiovascular disease patients increased by 9.4 percentage points – no control for underlying trends (IHA, 2009).		
US – Maryland	**MHAC**	Not available.	Hospital avoidable complication rates declined by 15 per cent during the first two years of the programme (Health Services Cost Review Commission – HSCRC, 2012).		Not available.	$110.9 million savings for reduced complications (HSCRC, 2012).
US National	**HQID**	Not applicable.	Participation in HQID over the period 2003–2006 was not associated with a significant improvement in quality of care processes or outcomes for acute myocardial infarction (Glickman et al., 2007).		Among hospitals caring for a high proportion of poor patients, those participating in HQID improved at a more rapid rate than those not participating in HQID (Jha, Orav & Epstein, 2010).	An independent study failed to find any impact of HQID, either positive or negative, on Medicare's costs (Ryan, 2009).
			HQID hospitals had slightly greater improvements in quality over a two-year period than comparable hospitals with public reporting alone (Lindenauer et al., 2007). Another study found that the performance of HQID hospitals accelerated in year one of the programme, but that the scores converged with non-HQID hospitals over three years (Grossbart, 2008).	No evidence that HQID improved 30-day mortality rates for acute myocardial infarction, heart failure, pneumonia, or CABG (Ryan, 2009).		

Positive spillover effects

Nearly all of the P4P programmes have some documented or perceived positive spillover effects on individual provider activity and the health system as a whole. No programme, however, demonstrated significant positive spillover effects on specific quality measures that were not rewarded through the financial incentive, but this was only rarely measured. As discussed in detail in Chapter 3, the most important potential positive spillover effects of the P4P programmes reviewed are the general strengthening of health sector governance through better data systems and performance feedback loops. In several cases, P4P appears to have raised awareness, and possibly acceptance, of objective measurement of provider performance. This could represent a profound cultural shift in some cases – with increased accountability and transparency in clinical interactions becoming the norm.

Improved generation and use of data

Improved generation and use of data is possibly the most important positive spillover effect of the P4P programmes. There is evidence that several of the case study P4P programmes have leveraged new or improved data systems for quality improvement activities well beyond reporting of performance measures. In the UK QOF, for example, the upgrading of computer systems and increased role of IT in GP practices has been used to a large extent in the quality improvement process within practices, including decision support templates and patient reminder systems. The increased use of computerized templates to guide clinicians and to assist in collecting data during consultations also could have more general positive impacts on overall quality of care (Campbell et al., 2007). The Estonia QBS introduced a new chronic disease status variable into patient records, which has facilitated the overall clinical management of these conditions. The standardized cost accounting system introduced through Brazil's OSS performance-based contracting model has led to improved capacity among hospital managers in planning and monitoring hospital activities.

Better use of information as a result of P4P programmes has come about largely as a result of investment by the programme in infrastructure and general incentives of P4P programmes for more effective use of information. Programmes that give direct incentives to expand IT infrastructure have had mixed results. Reviews of the Australia PIP have found that despite the high take-up of the eHealth incentive, major improvements in quality of care related to better electronic information have lagged (Australian National Audit Office, 2010). The California IHA programme does not regularly report the results of its 'meaningful use of IT' indicators, so it is difficult to assess impact. One report shows a large increase in the use of IT for some care management activities, but improvement levelled off after the first three years of the programme (IHA, 2009). More information is needed to assess the value of direct incentives to upgrade IT, but it is clear that providers need to see that the cost of investing in IT will be offset by direct revenue benefits from the incentive, as well as benefits from improved management and patient care.

Improved communication between purchasers and providers

Feeding performance data back to providers facilitates performance improvement and is an opportunity for productive dialogue between purchasers and providers. Several P4P programmes appear to have facilitated this communication by providing concrete organizing platforms for such dialogue. In the California IHA programme, for example, although only modest improvements in provider performance have been achieved, observers have noted the importance of the initiative for establishing a basis for collaboration and trust among participants. An important feature of the Maryland HAC programme was that it created a specific tool for discussing, assessing and evaluating overall quality of care and the relative performance of individual providers. The use of a uniform method for categorizing complication rates provides a useful communication tool for all professionals (clinical, managerial, and coding personnel), which helped drive behaviour change over time.

In the France ROSP, physicians initially strongly opposed the idea of linking performance to payment. Over the course of implementation of first CAPI and then ROSP, however, the close dialogue between unions of physicians and the National Health Insurance Fund has led to support from unions for including a P4P pillar in the national agreement on tariffs. This is considered to have opened the door to further refinements of provider payment in the French national health insurance system, and perhaps future steps away from the entrenched fee-for-service payment system.

Unintended consequences

None of the programmes carefully assessed unintended consequences, but no serious effects have been reported. Several unintended consequences may result from P4P programmes, including shifting provider focus disproportionately towards rewarded activities resulting in neglect of non-rewarded areas that may also be important for improving patient care and outcomes. Concerns also have been raised that focusing too much on financial incentives may detract from the intrinsic motivation of providers and negatively affect the relationship between providers and patients. These consequences are difficult to measure, and no rigorous attempts have been made to examine them in any of the programmes.

In the Maryland HAC complication rates for included conditions declined by 18.6 per cent in two years, while complication rates for excluded conditions increased by 2.8 per cent. Although the increase in the complications for excluded conditions may reflect real changes in these complications or improvements in documentation and coding, the increase in the rate of hospital acquired complications for excluded conditions may be the result of hospitals shifting the focus of their quality efforts toward rewarded conditions. In the UK QOF, in a study of physician attitudes toward the QOF, physicians noted the emergence of potentially competing 'agendas' during office visits if patient concerns do not relate to activities that are tied to the incentive (Campbell, MacDonald & Lester, 2008). Another study found that 75 per cent of GPs believed that they

spend more time on areas which attract QOF points and significantly less time on areas which were less likely to be rewarded under QOF (UK National Audit Office, 2008).

Although these results suggest that P4P programmes may subtly divert resources and attention away from activities that are not rewarded, more analysis is needed to understand whether these changes have negative consequences for service delivery and health outcomes that outweigh any positive contributions of the programmes.

Important design, implementation and contextual factors

Few clear lessons emerge from the case studies about specific aspects of programme design that may contribute to, or detract from, the effectiveness of programmes. There is no 'right' number of performance measures or level of bonus or penalty, although payment rates that are too low and do not reach frontline providers have been blamed for weak programme results in some cases (e.g. Australia PIP, California IHA and New Zealand PHO Performance Programme). It is not clear whether bonuses get better results than penalties or withholds. Some lessons do emerge, however, about general design, implementation and contextual factors that may contribute to more effective programmes, and some programme design decisions to possibly avoid.

Factors contributing to the effectiveness of P4P programmes

Programmes are most effective when they are aligned with and reinforce overarching strategies, objectives and clinical guidelines that are accepted by stakeholders. In the Estonia QBS, New Zealand PHO Performance Programme, UK QOF, Maryland HAC, and Turkey FM PBC, the P4P programmes are used as instruments in support of more comprehensive strategies to improve quality and strengthen health service delivery. The Turkey FM PBC, for example, is a key element of the Ministry of Health's comprehensive Health Transformation Programme, which created a new primary care specialty and service delivery approach, brought family physician salaries on par with those of specialists, promoted the use of clinical guidelines, and implemented well-functioning health information and decision support systems.

The Estonia QBS, has been used as a key lever in support the country's strategy of strengthening primary care by raising awareness of the role of family physicians in providing the full scope of high quality services, particularly preventing and managing chronic diseases. In the Maryland HAC programme, the incentive to reduce hospital-acquired complications coincided with and reinforced other programmes, such national initiatives to eliminate specific hospital-acquired infections. This reinforcing effect, though important for the success of the programmes, makes it difficult to attribute performance improvements to the programme in general, or to the incentive specifically.

The programmes are more successful when the incentive is integrated into and complements the underlying payment system. In most of the P4P programmes the power of the performance-related incentive payments tends to be modest relative to the incentives created by the underlying base payment system. In systems such as the US where providers receive revenue from multiple payers, the performance-related incentives are further weakened. Incentive payments seem to have the most potential to change provider behaviour where the P4P system is closely aligned and integrated with the underlying payment system particularly in a way that counteracts adverse incentives of the underlying payment system (e.g. Brazil OSS, Estonia QBS, Germany DMP, Maryland HAC, Turkey FM PBC, and UK QOF).

The Germany DMP has demonstrated improved processes of care and better patient outcomes that are attributed not to a targeted financial incentive but to better alignment of the incentives of the underlying payment system with the evidence-based care processes for chronic conditions. The Maryland HAC programme carefully layered the incentive onto the underlying DRG payment system to counteract the incentive to reduce inputs per case and possibly skimp on quality. In the Brazil OSS programme, targeted financial incentives are integrated into the underlying payment system through the performance contracts to counteract the adverse incentives for low productivity under global budget payment. In the Australia PIP, on the other hand, higher volume practices have been disproportionately rewarded by PIP, which suggests that the P4P incentive payments have reinforced the adverse incentives of the underlying fee-for-service payment system.

Programmes are more effective when they focus on specific performance problems that require broad-based approaches for improvement. Some programmes have led to improved performance when they target specific performance problems and processes of care are targeted that can be addressed through broad-based approaches to quality improvement. The Maryland HAC programme, for example, focuses on avoidable hospital complications related to specific clinical areas, but the improvement process has required broad-based improvement in processes. The Korea VIP, on the other hand, targets some very specific care processes in hospitals related to acute myocardial infarction, which do not necessarily require broad-based improvement approaches, and one more general process, the Caesarean section rate, where the performance problem may be difficult to pinpoint. Only modest improvements at best have occurred in the VIP in both clinical areas.

The structure of service delivery is important for whether or not providers can and do respond to the incentives, and programmes tend to favour larger, more urban providers. At the primary care level, teams or group practices appear to have greater incentive and more opportunity to make the investments and organizational changes necessary to improve performance. In France, for example, primary care is mainly organized through solo practices, and the ROSP programme does not appear to be driving large changes in the organization of service delivery in response to the P4P programme. In the UK, on the other hand, where primary care is organized in GP group practices, changes in practice organization, such as employing

nurses for chronic disease management, and investment in quality management tools have been common responses to the QOF.

In the California IHA programme, better performance achievement is found among large provider groups, which suggests that they are better able to make the necessary investments than smaller groups. In the Australia PIP participation rate for solo practices (34 per cent) is half the overall participation rate (67 per cent) (ANAO, 2010). Although some rural primary care practices have benefited from higher payment rates in the Australia PIP, equity may have been negatively affected when the requirement of accreditation proved to be a more difficult barrier for GP practices in rural and remote areas serving more vulnerable populations. This has been addressed by Australia's Department of Health and Ageing through targeted support to those practices to make the required investments to achieve accreditation.

Autonomy for health facilities together with broad performance-based contracting based on penalties or withholds appears to be effective in some settings. This is particularly the case for health systems starting with public health service provision. In the Brazil OSS programme, such contracting arrangements led to greater efficiency and productivity of contracted hospitals, which was largely attributed to autonomy (World Bank, 2006; La Forgia & Couttolenc, 2008). Provider autonomy combined with performance-based contracting with the possibility of penalty has also yielded positive results in the Turkey FM PBC programme.

What to avoid: design and implementation features that weaken the incentive

Complex and non-transparent programme structure

The structure of the Australia PIP, for example, with 13 incentives with requirements that can change from year to year, does not allow for a coherent set of policy objectives with clear priorities, and the mix of different payment mechanisms within PIP (between target and key performance indicators, sign-on, take-up of the incentive, etc.) has made payments less transparent. In the France ROSP the achievement rate calculation is rather complex, incorporating the providers' baseline performance and calculated using a different formula depending on the level of achievement relative to national targets. It is not clear whether this has affected the ability of providers to understand and respond to the incentives. In the California IHA programme, one possible explanation for weak results has been the continued expansion of the measure set and the difficulties that physician organizations face in making investments in quality improvement when the targets are continuously moving.

Selective participation in programme domains

The Australia PIP allows providers to select those areas in which they have the greatest potential for reward. This has resulted in a high uptake of an incentive

that is relatively easy to achieve and that comes with a big reward (eHealth) and much lower uptake of the incentives related to service delivery for chronic conditions, which require much more effort on the part of the practices. The movement in and out of incentive streams also makes it difficult to monitor performance trends or provide meaningful aggregate analyses and other feedback to providers.

Specific incentives to improve the organization of service delivery

Several of the programmes for primary care include targeted incentives related to the organization of service delivery and infrastructure. These performance measures generally are not based on evidence and typically require separate self-reported documentation for indicator measurement. The California IHA programme, for example, includes 22 indicators on 'meaningful use of health IT'. Performance against these indicators is measured by a self-reported survey and signed attestation documents (NCQA, 2011). The UK QOF includes 36 indicators in the 'organizational' performance domain covering such aspects of GP practice organization as record keeping, information for patients, education and training of staff, practice management, and medicines management. Performance against these measures also requires separate self-reported documentation which includes at least seven to 15 reports generated by the GP practice.

Since evidence is lacking that links these organizational indicators to improved processes of care, it is questionable whether direct incentives to improve the organization of service delivery are valid and a cost-effective way to achieve the desired results, particularly given the high administrative burden on the providers to prove achievement of these indicators. While some success has been achieved through direct incentives for IT uptake, there is no clear benefit observed from the other organizational performance indicators in use. Alternative approaches may be more effective, such as direct support and investment to upgrade infrastructure. The Australia PIP, for example, now includes direct investment to help rural practices achieve accreditation. Furthermore, P4P programmes should be structured to indirectly drive organizational changes and investments, as providers make organizational improvements to achieve clinical performance targets.

Conclusions

The experience from the case study P4P programmes reviewed for this volume suggests that by itself a targeted financial incentive linked to specific performance metrics may be a costly way to achieve small improvements in coverage of priority services and processes of care. Little or no impact on health outcomes should be expected with the way programmes are currently designed and implemented. Putting all of the health system support structures in place to implement P4P programmes adds costs beyond the cost of the

incentive payments. None of the programmes reviewed has estimated these costs or additional administrative costs to providers. Typically new money is required in the system not only to pay for the incentive, but also to invest in support structures, particularly IT and verification/monitoring. We do not know whether the benefits to the system from implementing P4P programmes outweigh these costs, or if P4P programmes are crowding out other more cost-effective approaches to reaching health system objectives.

The experience of the case study P4P programmes also shows, however, that the incentives can have greater value if they are applied strategically to focus attention on high-impact performance problems, and to strengthen key elements of health purchasing and health sector governance. When P4P programmes contribute to aligning incentives and strengthening governance structures and processes, the spillover effects of the programmes may be more important than the incentive itself. The contribution of P4P programmes to strengthening governance and these wider spillover effects, however, typically are not captured in current studies and evaluations of P4P programmes.

The results of this study suggest that the emphasis in P4P programmes should be not on the performance measures and incentive payments alone, but rather on using comprehensive approaches in which the indicators and incentives play a supporting rather than a central role. Used in this way, P4P programmes may contribute to establishing or sustaining a cycle of performance improvement in the health system, yielding benefits beyond changes in performance measures. When P4P programmes do not contribute to strengthening key aspects of health system governance and health purchasing, the already modest impact on performance measures is even less significant, and the overall effectiveness and justification of the programmes can be questioned.

More importantly, if P4P programmes do work effectively to strengthen data systems and feedback loops and reinforce a culture of accountability, they may create the foundation for a more fundamental shift in underlying provider payment systems. P4P may be most useful as a 'stepping stone' to more sophisticated provider payment systems that improve contracts between purchasers and providers and better align incentives with outcomes. Better contracts define the output more clearly – specifying continuity of care, disease management and clinical guidelines and hold providers accountable not just for volume but also for processes and outcomes.

P4P programmes should contribute to building the experience base with different performance measures, their validity, feasibility and link to outcomes; to a move toward richer clinical information systems, electronic health records and platforms to aggregate, analyse and compare provider-level data; and to promoting more transparent and constructive communication between purchasers and providers to identify the sources of performance problems, whether they lie with providers or with the system, and to work together to solve them creatively. Viewed in this way, pay for performance is not an end in itself, but an instrument for achieving better underlying provider payment systems, more strategic health purchasing, and stronger health system governance.

References

Altenhofen, L. et al. (2004) *Report on quality assurance – disease – management programmes in North-Rhine* [in German]. Duesseldorf: KV Nordrhein (http://www.kvno.de/importiert/qualbe_dmp04.pdf, accessed July 2011).

Ashworth, M. et al. (2007) The relationship between social deprivation and the quality of primary care: a national survey using indicators from the UK Quality and Outcomes Framework, *British Journal of General Practice*, 57(539): 441–8.

Australian National Audit Office (2010) *Practice incentives program audit report no. 5 2010–2011*. Canberra: Australian National Audit Office.

Boeckxstaens, P. et al. (2011) The equity dimension in evaluations of the quality and outcomes framework: a systematic review, *BMC Health Services Research*, 11: 209–24.

Bradley, E.H. et al. (2006) Hospital quality for acute myocardial infarction: correlation among process measures and relationship with short-term mortality, *Journal of the American Medical Association*, 296: 72–8.

Buetow, S. (2008) Pay-for-performance in New Zealand primary health care, *Journal of Health Organization and Management*, 22(1): 36–47.

Caisse Nationale d'Assurance Maladie (2010) *Contrat d'Amélioration des Pratiques Individuelles (CAPI): une dynamique au bénéfice des patients*. Press release September 2010 (http://www.ameli.fr/fileadmin/user_upload/documents/Dp_capi_16_09_2010_vdef.pdf, accessed January 2011).

Campbell, S., MacDonald, R. and Lester, H. (2008) The experience of pay for performance in English family practice: a qualitative study, *Annals of Family Medicine*, 6(3): 228–34.

Campbell, S. et al. (2007) Quality of primary care in England with the introduction of pay for performance, *New England Journal of Medicine*, 357(2): 181–90.

Commonwealth of Australia (2009) *Government response to the Productivity Commission annual review of regulatory burdens on business*. Canberra: Department of Industry.

Damberg, C.L. et al. (2005) Paying for performance: implementing a statewide project in California, *Quality Management in Health Care*, 14(2): 66–79.

Doran, T. et al. (2008) Effect of financial incentives on inequalities in the delivery of primary clinical care in England: analysis of clinical activity indicators for the quality and outcomes framework, *The Lancet*, 372(9640): 728–36.

Drabik, A. et al. (2012) Life prolonging of disease management programmes in patients with type 2 diabetes is cost-effective. Diabetes, *Diabetes Research and Clinical Practice*, 95(2): 194–200.

Fleetcroft, R. et al. (2011) Correlation between prescribing quality and pharmaceutical costs in English primary care: national cross-sectional analysis, *British Journal of General Practice*, 61(590): e556–64.

Glickman, S. et al. (2007) Pay for performance, quality of care, and outcomes in acute myocardial infarction, *Journal of American Medical Association*, 297(21): 2373–80.

Greene, J. (2013) An examination of pay-for-performance in general practice in Australia, *Health Services Research*, 48(4): 1415–32.

Grossbart, S. (2008) Effectiveness of pay for performance as a quality improvement strategy, *Prescriptions for Excellence in Health Care*, 3(1): 3.

HIRA (2010) *Report on value incentives program 2010*. Seoul: HIRA.

Hood, R. (2007) Pay-for-performance – financial health disparities and the impact on healthcare disparities, *Journal of the National Medical Association*, 99(8): 953–8.

Hospitals and Health Networks (2007) Inside the premier/CMS pay-for-performance project: an H&HN exclusive: insights from the 5 top hospitals, *Hospitals and Health Networks*, March (http://www.finance.gov.au/publications/response-to-the-pc-manufacturing-and-distributive-trades-review/index.html, accessed June 2013).

Integrated Healthcare Association (2009) *Measurement year 2009 results report*. Seoul: Integrated Healthcare Association.

Jha, A., Orav, J. and Epstein, A. (2010) The effect of financial incentives on hospitals that serve poor patients, *Annals of Internal Medicine*, 153(5): 299–306.

Kontopantelis, E. et al. (2013) Recorded quality of primary care for patients with diabetes in England before and after the introduction of a financial incentive scheme: a longitudinal observational study, *BMJ Quality and Safety*, 22(1): 53–64.

La Forgia, G.M. and Couttolenc, B.F. (2008) *Hospital performance in Brazil – the search for excellence*. Washington: World Bank.

Lindenauer, P.K. et al. (2007) Public reporting and pay for performance in hospital quality improvement, *New England Journal of Medicine*, 356(5): 486–96.

Martin, Jenkins and Associates Limited (2008) *Evaluation of the PHO peformance programme: final report*. Auckland: Martin, Jenkins and Associates Limited.

Mattke, S., Seid, M. and Ma, S. (2007) Evidence for the effect of disease management: is $1 billion a year a good investment?, *American Journal of Managed Care*, 13: 670–6.

Mehring, M. et al. (2012) Disease management programmes for patients with asthma in Germany: a longitudinal population-based study, *Respiratory Care*, 58(7): 1170–77.

Miksch, A. et al. (2010) Is there a survival benefit within a German primary care-based disease management program?, *American Journal of Managed Care*, 16(1): 49–54.

Morse, R.B. et al. (2011) Hospital-level compliance with asthma care quality measures at children's hospitals and subsequent asthma-related outcomes, *Journal of the American Medical Association*, 306(13): 1454–60.

Mullen, K., Frank, R. and Rosenthal, M. (2009) Can you get what you pay for? Pay for performance and the quality of healthcare providers, *RAND Journal of Economics*, 41(1): 64–91.

NCQA (2011) *Integrated Healthcare Association California pay for performance program: measurement manual year 2011 P4P manual*. Washington, DC: National Committee for Quality Assurance.

Norbury, M., Fawkes, N. and Guthrie, B. (2011) Impact of the GP contract on inequalities associated with influenza immunisation: a retrospective population-database analysis, *British Journal of General Practice*, 61(588): e379–85.

PHO Performance Programme (2010) *National summary of PHO performance 1 July 2009–31 December 2009*. Auckland: PHO Performance Programme.

PHO Performance Programme (2012) *National summary of PHO performance 1 July 2012–31 December 2012*. Auckland: PHO Performance Programme.

Pimouguet, C. et al. (2011) Effectiveness of disease-management programmes for improving diabetes care: a meta-analysis, *Canadian Medical Association Journal*, 183(2): E115–27.

Productivity Commission (2003) *General practice administrative and compliance costs*. Canberra: Australian Government Productivity Commission.

Ryan, A. (2009) Effects of the premier hospital quality incentive demonstration on Medicare patient mortality and cost, *Health Services Research*, 44(3): 821–41.

Schäfer, I. et al. (2010) The disease management program for type 2 diabetes in Germany enhances process quality of diabetes care – a follow-up survey of patient's experiences, *BMC Health Services Research*, 10(55).

Schoul, A.K. and Gniostko, C. (2009) *AOK Bundesverband: evaluation AOK Curaplan COPD* (http://www.aok-gesundheitspartner.de/imperia/md/gpp/bund/dmp/evaluation/befragung/auswahl_copd_ergebnisse_2009.pdf, accessed May 2013).

Scott, A. et al. (2009) The effect of financial incentives on the quality of health care provided by primary care physicians, *The Cochrane Library 2011*, vol. 9.

Serumaga, B. et al. (2011) Effect of pay for performance on the management and outcomes of hypertension in the United Kingdom: interrupted time series study, *British Medical Journal*, 342: d108.

Shahian, D.M. et al. (2012) Association of National Hospital Quality Measure adherence with long-term mortality and readmissions, *British Medical Journal*, 21(4): 325–36.

Stock, S. et al. (2011) Disease-management programmes can improve quality of care for the chronically ill, even in a weak primary care system: a case study from Germany, *The Commonwealth Fund, Issues in International Health Policy*, November.

Szecsenyi, J. et al. (2008) German diabetes disease management programmes are appropriate for restructuring care according to the chronic care model: an evaluation with the patient assessment of chronic illness care instrument, *Diabetes Care*, 31(6): 1150–4.

UK National Audit Office (2008) *NHS pay modernisation: new contracts for general practice services in England*. London: The Stationery Office.

UK National Health Service (2004) *National programme for information technology: GMS IT-QMAS Bulletin*. London: NHS.

Västra, K. (2010) *Assessing the impact of implementing primary care quality bonus system on follow up of patients with hypertension and type 2 diabetes based on Estonian Health Insurance Fund claims registry data in 2005–2008*. Unpublished Masters Thesis: University of Tartu.

Van Herck, P. et al. (2010) Systematic review: effects, design choices, and context of pay-for-performance in health care, *BMC Health Services Research*, 10: 247–60.

World Bank (2006) *Brazil: Enhancing performance in Brazil's health sector, lessons from innovations in the state of Sao Paulo and the city of Curitiba*. Washington: World Bank (report no. 35691-BR).

Part II

Case studies of P4P programmes in OECD countries

Primary care

chapter six

Australia: Practice incentives programme*

Cheryl Cashin and Y-Ling Chi

Introduction

Australia's health care system is considered to be one of the best performing health systems overall, demonstrating success in controlling costs, while at the same time achieving high levels of health outcomes. Australia spends a little above the OECD average on health (USD PPP $3137 per capita compared to the OECD average of $2984) and has managed to contain the growth in health expenditure, unlike in other OECD countries where spending has increased steadily over the last ten years (OECD, 2009). Australia has achieved one of the highest life expectancies, ranking third after Japan and Switzerland in 2007 (OECD, 2009). In spite of these achievements, however, concerns have emerged in recent years about the quality and coordination of care and prevention. Chronic conditions such as diabetes are reaching epidemic proportions, and incidents involving quality and safety of hospital care have received attention. The fee structure of the Medicare Benefits Schedule (MBS) under Australia's national health insurance programme (Medicare) encourages a large number of short consultations and provides minimal incentives for quality or preventive activities (Australian Government Department of Health and Ageing, 2010).

Australia has experience in using pay for performance (P4P) programmes as a solution to some problems in health care delivery (Boxall, 2009). Two large, ongoing P4P programmes date back to the 1990s: the General Practice Immunization Incentive (GPII) programme to increase vaccination coverage among children, and the Practice Incentives Program (PIP) to encourage continuous improvements in primary health care. More recent P4P initiatives reward hospital quality achievement, including a programme run by the Australian Government Department of Veteran's Affairs introduced in 2006, and the Clinical Practice Improvement Payment system in the Australian state of Queensland, which was introduced in 2007. As part of the Pharmaceutical Benefits Programme reform initiated in 2008, community pharmacies receive

a small incentive payment for dispensing substitutable, premium-free brands, as well as an increase in pharmacy mark-ups and dispensing fees (Australian Government Department of Health and Ageing, 2009).

Faced with serious challenges in fragmented primary health care, brought about partially by Medicare – Australia's fee-for-service payment system – the Australian Government Department of Health and Ageing (DoHA) introduced the PIP in 1998 as part of a broader strategy to reform primary health care (Russell & Mitchell, 2002). These 'practice incentive payments' reward a number of areas of primary health care including comprehensive after-hours care, rural practices, teaching medical students, and use of electronic health records (eHealth). The PIP allows GP practices to participate once they have been accredited against the Royal Australian College of General Practitioners' (RACGP's) *Standards for General Practices*. Practices can choose among 13 incentive areas to participate. Incentive payments reached A$61,600[1] on average per practice in 2008-09, or A$19,700 per FTE GP (Australian National Audit Office, 2010). The programme is among the largest in the world, with some A$2.7 billion spent since its inception.

Health policy context

What were the issues that the programme was designed to address?

Australia's primary health care is delivered in large part by a network of private GP practices that are permitted to set their own fees. Patients receive a rebate from Medicare Australia for eligible services as determined by the Medicare Benefits Schedule (MBS). A large share of practices choose to direct bill Medicare (known as 'bulk billing'), which holds them to the MBS fee levels without being able to charge additional fees to patients (Russell & Mitchell, 2002). This fee-for-service payment system was considered to be at least partially responsible for increasingly fragmented primary health care and the shift away from prevention, and has contributed to the poor management of chronic diseases.

Reform efforts began in 1991, resulting in the 'General Practice Reform Strategy', which was designed to improve the integration, quality, and comprehensiveness of GP care (Australian Government Department of Health and Ageing, 2010). A key reform introduced in the early 1990s established about 120 'Divisions of General Practice', which are geographically based organizations that represent networks of approximately 150 GPs (ranging from 12 to 800). The Australian Government provides infrastructure funding to enable Divisions to engage in cooperative activities to address health needs at the local level (National Health Strategy, 1992). The PIP started in July 1998 in response to a series of recommendations made by the GP Strategy Review Group, a group of DoHA officials and general practice interests, appointed by the then Minister for Health and Family Services. The group recommended a programme that would move toward a 'blended payment' model, providing a

portion of funding to GP practices that was unrelated to the volume of fee-for-service payments (Australian National Audit Office, 2010). The programme aimed to create incentives for practices to provide longer visits and discourage a high volume of brief consultations.

The main objective of the PIP is to encourage continuing improvements in general practice through financial incentives to support quality care, and improve access and health outcomes for patients. Practices are required to be accredited or registered for accreditation to participate in the PIP. PIP practices may be eligible for a number of incentive payments, providing a more flexible payment model that can influence both short- and long-term changes in service delivery. Improving accountability, reporting and data collection on selected health issues were implicit, if not explicit, objectives, as shown by the introduction of the Information Management/Information Technology (later evolved to eHealth) Incentive, one of the largest payment components of the programme. The programme is under the umbrella of wider incentive initiatives in health carried on by DoHA, which also comprise the Rural Incentive Programme, Mental Health Nurse Incentive Programme, and the GPII Programme.

Stakeholder involvement

The PIP is administered by Medicare Australia on behalf of DoHA. DoHA has overall policy responsibility for the programme, while Medicare Australia is responsible for the day-to-day administration, including verifying compliance with programme and payment eligibility criteria, and calculating and making payments. Other stakeholders have participated in the design and governance of the programme. For example, the basis for the PIP payment formula was developed in consultation with the General Practice Financing Group (GPFG), which was a negotiating body comprising the Royal Australian College of General Practitioners, Australian Medical Association, Rural Doctors Association of Australia, Australian Divisions of General Practice, and the Australian Government (Medicare Australia, 2010). DoHA regularly consults with GP professional organizations through an advisory group.

Technical design

How does the programme work?

Performance domains and indicators

The programme was designed around 13 incentive areas organized in three main streams – quality of care, capacity, rural support (Table 6.1). Not all of the incentives are strictly related to performance, and some of them could be considered to be conditional cash transfers to practices upon implementation of certain services. Two incentive streams recently were discontinued, the Practice Nurse Incentive and the Domestic Violence Incentive, and the After

Hours Incentive is ending during 2013 (Australia Department of Human Services, 2013). The Quality Stream incentives pay for coverage of services that comply with evidence-based guidelines, which the programme treats as a proxy for outcomes. The Capacity Stream incentives give additional resources to GP practices that invest in infrastructure, such as computerization, or to expand services, such as providing after hours care or providing care in residential aged care facilities. The incentive related to Information Management/Information Technology (IMIT) has been particularly important in PIP. This stream has evolved over time as the IT capacity and needs of practices have changed and available technology has become more sophisticated (Australia Department of Human Services, 2013). The original IMIT Incentive was instrumental in driving computerization of GP practices. The incentive was updated in 2009 to become the eHealth Incentive, which aims to encourage general practices to keep up to date with the latest developments in eHealth.

The Rural Support stream incentives provide additional resources to GP practices in more rural and remote settings and compensate them for bringing services to these areas that otherwise would be difficult to access for these populations, such as some more specialized surgical and obstetric procedures.

Table 6.1 Incentives in the Australia PIP, 2010

Incentive	*Activity*	*Payment amount*
Quality stream		
Quality Prescribing	Practice participation in quality use of medicines programmes endorsed by the National Prescribing Service. Paid annually in May.	A$1 per SWPE[2]
Diabetes Incentive	*Sign-On Payment*: one-off payment to practices using a diabetes register and recall/reminder system.	A$1 per SWPE
	Outcomes Payment: payment to practices where at least 2% of practice patients are diagnosed with diabetes and GPs have completed a cycle of care for at least 20 per cent of them.	A$20 per diabetic SWPE/year
	Service Incentive Payment: payment to GPs for each patient completing an annual cycle of care	A$40 per patient/ year
Cervical Screening Incentive	*Sign-on Payment*: one-off payment to practices for engaging with the state/territory cervical screening registers.	A$0.25 per SWPE
	Outcomes Payment: payment to practices if at least 65 per cent of women aged 20–69 screened have been screened in the 30-month reference period.	A$3 per female SWPE aged 20–69

	Service Incentive Payment: payment to GPs for screening women aged 20–69 years who have not had a cervical smear within the last 4 years.	A$35 per patient/ year
Asthma Incentive	*Sign-on Payment*: one-off payment to practices that: • use a patient register, and a recall and reminder system; • agree to use the asthma cycle of care; and • agree to have their details forwarded to appropriate bodies.	A$0.25 per SWPE
	Service Incentive Payment: payment to GPs for each cycle of care completed for patients with moderate to severe asthma.	A$100 per patient/ year
Indigenous Health Incentive	*Sign-on Payment*: one-off payment to practices that agree to undertake specified activities to improve the provision of care to their Aboriginal and/or Torres Strait Islander patients with a chronic disease.	A$1000 per practice
	Patient Registration Payment: Payment to practices for each Aboriginal and/ or Torres Strait Islander patient aged 15 years and over, registered with the practice for chronic disease management.	A$250 per eligible patient/year
	Outcomes Payment Tier 1: Payment to practices for each registered patient for whom a target level of care is provided in a calendar year.	Tier 1: A$100 per patient/year
	Outcomes Payment Tier 2: Payment to practices for providing the majority of care for a registered patient in a calendar year.	Tier 2: A$150 per patient/year
Capacity stream		
eHealth Incentive	The PIP eHealth Incentive has three eligibility requirements. Practices must meet each of the eligibility requirements to qualify for payments.	A$6.50 per SWPE capped at A$12,500 per practice, per quarter.
Practice Nurse Incentive	**Practices in urban areas of workforce shortage (RRMA[3]s 1–2):** Payment to PIP practices that employ a practice nurse, Aboriginal health worker and/or allied health worker, for the minimum number of sessions per week over the payment quarter.	A$8 (RRMA 1–2) capped at A$40,000/ year.

(*continued*)

Table 6.1 Incentives in the Australia PIP, 2010 (*continued*)

Incentive	*Activity*	*Payment amount*
	Practices in rural and remote areas (RRMAs) 3–7: Payment to practices in rural and remote areas that employ a practice nurse and/or Aboriginal health worker for the minimum number of sessions per week over the payment quarter.	A$7 (RRMA 3–7) capped at A$35,000/ year.
After Hours Incentive	Tier 1 – Practice patients have access to 24-hour care, seven days a week through formal external arrangements.	A$2 per SWPE annually
	Tier 2 – Practice GPs must provide at least 10 or 15 hours per week of after hours cover depending on practice size. At all other times practice patients have access to after hours care through formal external arrangements.	A$2 per SWPE annually (+ payment for Tier 1)
	Tier 3 – Practice GPs provide their practice patients with 24-hour care, seven days a week.	A$2 per SWPE annually (+ Tiers 1&2)
Teaching Incentive	Teaching of undergraduate medical students. Maximum of two 3-hour teaching sessions per GP, per day.	A$100 per session
Aged Care Access Incentive	Tier 1 – GPs must provide at least 60 eligible services in residential aged care facilities (RACF) in the financial year.	A$1500 per year
	Tier 2 – GPs must reach the QSL 2 by providing at least 140 eligible services in RACF in the financial year.	A$3500 per year
Rural support stream		
Rural Loading	The practice's main location is outside metropolitan areas (increases with extent of remoteness) based on the RRMA 3–7 Classification. Rural loading is applied to the practice's total PIP payment.	0–50 per cent loading
Procedural GP Payment	Tier 1 – A GP in a rural or remote practice provides at least one procedural service (services typically provided in hospital setting), in the six-month reference period.	A$1000 per six-month reference period
	Tier 2 – A GP in a rural or remote practice meets the Tier 1 requirement and provides after hours procedural services.	A$2000 per six-month reference period
	Tier 3 – A GP in a rural or remote practice meets the Tier 2 requirements and provides 25 or more eligible surgical and/or anaesthetic and/or obstetric	A$5000 per six-month reference period

	services in the six-month reference period.	
	Tier 4 – A GP in a rural or remote practice meets the Tier 2 requirements and delivers 10 or more babies in the six-month reference period or meets the obstetric needs of the community.	A$8500 per six-month reference period
Domestic Violence Incentive	Payment to encourage practices in RRMA 3–7 to employ a qualified practice nurse or Aboriginal health worker that is available to act as a referral point for women experiencing domestic violence for the minimum number of sessions per week.	A$1 per SWPE capped at A$4000 per practice/year

Source: Medicare Australia, 2010.

Incentive payments

The way incentive payments are calculated and made in the PIP is complex. The recipient (whether the general practice or GPs working in PIP practices), basis for payment amount, payment determination (prospective or retrospective), and frequency of payment vary across incentives, and they can vary further for components or tiers within incentives. Payments for most of the indicators are made to the practices, but some of the quality incentives are paid directly to individual GPs for each priority service they deliver.

Most of the incentive payments are flat-rate rewards per Standardized Whole Patient Equivalent (SWPE), which is a measure of a practice's patient load independent of the number of services provided, or per service provided. The exception is rural loading, which is paid as a percentage of the total incentive payments made to the practice. The Quality Stream incentives, with exception of the Quality Prescribing Incentive, give one-off payments to practices that participate and meet specific criteria, such as participating in the cervical cancer screening register. Practices are then paid a per-patient bonus for achieving specified coverage rates for priority services, such as achieving a 50 per cent rate of cervical cancer screening for the target group, or 20 per cent of diabetic patients with a completed cycle of care.[4] Some incentives in the Quality Stream include a third element of payment, which is made directly to individual GPs for each priority service they provide. For example, individual GPs receive a payment for each of their patients with diabetes completing an annual cycle of care.

Payments are made on a quarterly basis for diabetes, asthma and cervical screening after a one-off payment for signing on to the incentive. To qualify for payments, practices must be participating in the PIP and meet the eligibility requirements of the incentives at the 'point in time' that corresponds to the last day of the month prior to the next quarterly payment month. There are no restrictions on how the practices can allocate their incentive payments.

The guidelines established by DoHA stipulate that 'payments are intended to support the practice to purchase new equipment, upgrade facilities, or increase remuneration for doctors working at the practice' (Australian Government Department of Health and Ageing, 2009). There are no reporting requirements for how the incentive payments are used.

Data sources and flows

Information on indicators related to the number of services delivered is collected through the Medicare claims processing system and other routine reporting, such as from the National Prescribing Service for the Quality Prescribing Incentive. For other incentive streams, information is submitted to the PIP database that documents the activity of the practitioner. An annual Confirmation Statement process was introduced in May 2010. Practices are required to check, complete and confirm whether the practice is continuing to meet the eligibility requirements of the incentives which the practice has applied for. A new online administrative system was introduced in October 2010 to allow practices to apply for new PIP incentives and review payment levels, and is aimed at reducing the administrative burden of practices (Medicare Australia, 2010).

Data are collected by Medicare Australia, which has the responsibility to assess the performance of the practice on some selected indicators, calculate the practices SWPEs, and decides on the total payment to practices and individual GPs. The Continuous Data Quality Improvement Programme controls the quality of payments on a sampled basis, recording all sources and types of errors commonly found in the reporting of results. Medicare Australia is also conducting random and targeted audits to ensure that practices meet the eligibility requirements.

Reach of the programme

Which providers participate and how many people are covered?

Participation in the PIP is voluntary and conditional on the GP practice being accredited or registered for accreditation against the Royal Australian College of General Practitioners *Standards for General Practices*. About 5000 GP practices throughout the country participate in PIP, which represents about two-thirds of all practices and about 21,000 Full-time Equivalent General Practitioners. It is estimated that 82 per cent of GP patient care was delivered through PIP practices in 2009 (Australian Government Department of Health and Ageing, 2009). After meeting the requirements to participate in the PIP, practices decide on enrolment in individual incentive areas within the general PIP framework, according to their eligibility for the different initiatives. This allows for flexibility and provides tailored incentives to each practice. Some practices also participate in other programmes, such as the General Practice Immunization Incentive Programme and the Mental Health Nurse Incentive Programme.

Practices receive quarterly payments following enrolment in the programme. The average payment to a practice in 2009–2010 was A$57,800, which is typically between 4 and 7 per cent of total practice income. There have been great disparities in payment, however. One practice alone received A$576,000, with FTE GPs receiving individually A$36,000, or 90 per cent more than the average.

Improvement process

How is the programme leveraged to achieve improvements in service delivery and outcomes?

Whether and how the PIP is driving performance improvement in Australia's GP practices is difficult to ascertain. There is very little information about how the incentive payments are used by the practices, or whether improvement processes have been put in place or strengthened. There is no structured dialogue between the programme administrators (DoHA and Medicare Australia) and the practices on the performance measures, and there is no systematic feedback of performance information to providers for their internal management purposes. DoHA does, however, regularly consult with GP professional organizations through an advisory group, where feedback from member GPs may be provided. Data on the performance of individual practices are not made publicly available because of privacy issues. Several of these weaknesses were highlighted by a recent review of the Australia National Audit Office (ANAO) released in 2010.

GP practices receive incentive payments for becoming accredited and providing certain priority services according to established guidelines. Whether that in fact leads to improved quality of care and better outcomes has not been verified. Furthermore, the uptake and payment across incentive areas is highly skewed. Whereas eHealth accounts for 33 per cent of all incentive payments (reflecting both higher uptake and relatively higher reward), all three priority service areas combined only account for 11 per cent of the total payout in 2008-09 (Figure 6.1). Only 17 per cent of practices eligible to participate in the Domestic Violence Incentive participated (Australian National Audit Office, 2010).

Both the choice of GP practices about which incentive streams to participate in and the way they use their incentive payments show that IT is the part of GP practice development and quality improvement that is supported most by PIP. Although GP practices can apply for as many of the incentives as they are eligible for, by far the largest payout is for the eHealth Incentive. Furthermore, although there is no good information on how GP practices use PIP incentive payments, it is generally believed that most practices distribute at least a portion of the funding to staff GPs and the rest into practice infrastructure, with most of the money going to IT (Ferguson, 2006). Whether and how upgraded IT supported by PIP is being used to improve service delivery and whether improved IT can be linked to improved quality of care and better outcomes are unknown.

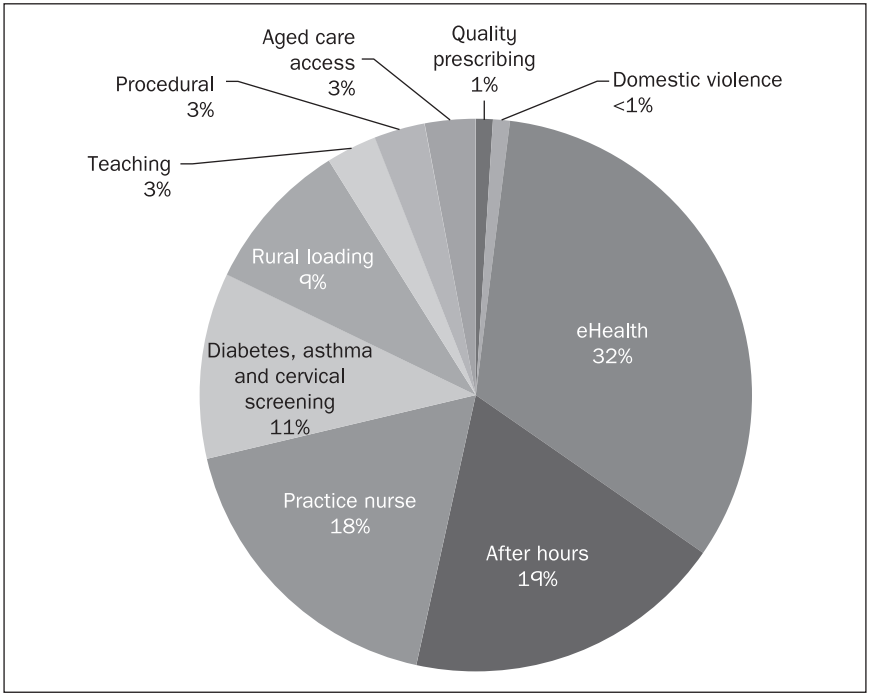


Figure 6.1 Distribution incentive payments in the Australia PIP, 2008–09

Source: ANAO, 2010.

Results of the programme

Has the programme had an impact on performance, and have there been any unintended consequences?

Programme monitoring and evaluation

In spite of the longevity of the programme there are no comprehensive, rigorous evaluations of PIP. The monitoring done by DoHA is related mainly to the uptake of the programme. DoHA tracks and reports on several programme coverage indicators: (1) number and share of practices participating in PIP; (2) the volume of payments made; (3) the percentage of care provided by practices participating in PIP; and (4) the proportion of Australian Government funding for general practice that is channelled through PIP. The lack of more in-depth monitoring and evaluation may be related to the main stated objective being to increase accreditation among primary care practices, which is easily observable and measurable. DoHA claims that the percentage of all primary care that is provided by PIP practices is a proxy for care provided in accredited practices, which reflects higher overall quality of care.

The primary accountability mechanism for PIP is regular reviews by ANAO (with five reports since the creation of the programme). Although

these reviews are not impact evaluations, they do provide some assessment of the effectiveness of programme implementation and the performance of PIP against some of its stated objectives. The latest report sought to address the question to which extent the programme met the new policy objectives set up in 2006. This report provides a mixed picture on the overall results of the programme, especially on the lack of reliable data to estimate the impact of the programme (Australian National Audit Office, 2010).

Overall, the latest ANAO report emphasizes the need to define adequate effectiveness measures to fully assess the overall impact of the programme. So far, the data for evaluation have mainly relied on the Key Performance Indicators (KPI), which also are those used in the definition of payment levels for individual practices. The report noted that evaluation indicators should be defined based on the objectives of the programme and should be different from the payment indicators used in the programme. Evidence on the effectiveness of the programme is thus limited, which has already been pointed out successively by the different audit reports. Data on the performance of practitioners outside the PIP programme should also be collected and analysed, for instance, from MBS claims. Comparisons between the participating and non-participating programmes could provide conclusive evidence about PIP's effectiveness.

The latest ANAO review also found that PIP has been successful at meeting its objective of increasing rates of accreditation among general practices. Accreditation has increased to 67 per cent of practices as a result of PIP. In their survey of GPs, 43 per cent of practices responded that the main reason they applied for accreditation is to have access to PIP (Australian National Audit Office, 2010). Nonetheless, the report fails to provide evidence on the actual ongoing efforts of participating practices in improving standards of care. This may be attributed to self-selection into the programme, with 'better-off' practices applying for PIP. In fact, the review found that accreditation and PIP participation rates have levelled off, because not all practices find it worthwhile to incur the fixed costs to become accredited.

The PIP has been successful at meeting the objective of adding a flexible component to the fee for service payment system. The programme has been a means of funding general practices and GPs for a diverse range of activities outside the fee-for-service arrangements through the Medicare Benefits Schedule (MBS). On the other hand, reviews of the programme have been pessimistic about the extent to which PIP encourages GPs to spend more time with their patients. The analysis provided in the most recent ANAO report using MBS claims shows that higher volume practices have been disproportionately rewarded by PIP, which suggests that the blended payment system under PIP has not drastically changed the incentives for GP practices.

Performance related to specific indicators

Several independent studies of individual incentives also provide little evidence on the effectiveness of PIP in driving quality improvement and better outcomes. A recent study found that there was a short-term increase in diabetes testing and cervical cancer screens after the PIP began, but that could not be attributed to the programme at the individual GP practice

level. Neither signing onto the programme nor claiming incentive payments was associated with increased diabetes testing or cervical cancer screening (Greene, 2013). Two earlier studies published in 2005 and 2009 on diabetes case detection indicate ambiguity related to the effectiveness of the incentives. The 2005 study performed by the Healthcare Management Advisors (2005) found that PIP did not create incentives for GPs to diagnose more cases of diabetes.

The ANOA report points to more promising results, including a study finding that the Diabetes Incentive increased the probability of an HbA1c test being ordered by 20 percentage points (Scott et al., 2009). The ANAO report also cited studies based on claims data suggesting that the number of completed cycles of care for diabetes and asthma have increased as a result of the incentive, although there is no control for underlying trends (Australian National Audit Office, 2010). Finally, the ANAO report suggests that the Practice Nurse Incentive has led to improved management of chronic diseases, increased time spent with patients, and reduced waiting time (Australian National Audit Office, 2010).

The ANAO report concluded that the After Hours Incentive and the Domestic Violence Incentive have not met their stated policy objectives, however, although DoHA disagreed with this conclusion (Australian National Audit Office, 2010). The benefits of the implementation of eHealth also could be better leveraged, as the evaluation showed that despite the high uptake of the incentive, major improvements in quality of care related to better electronic information have lagged. Electronic transmission of documents, electronic patient record transferred, etc. would require a more coordinated system between the different practices, especially those using eHealth techniques and those not participating in the programme (Australian National Audit Office, 2010).

Equity

The accreditation process can be a significant barrier to certain GP practices including Aboriginal Medical Services (AMS) and to smaller practices. As such, AMS and small practices servicing remote locations and non-English speaking communities have been underrepresented in PIP. The PIP participation rate for solo practices (34 per cent) is half the overall participation rate (67 per cent) (Australian National Audit Office, 2010). This disparity in PIP participation across smaller practices and those serving more disadvantaged populations may contribute to inequity in the programme. If there is a geographical or economic self-selection of practices into PIP, additional revenues for the participating programmes is likely to further exacerbate these gaps in quality of care.

On the other hand, PIP has had a positive effect on access and provision of care in rural areas, contributing to the reduction of rural–urban inequalities. For some rural practices, PIP represents an important source of revenue, and the rural loading payment is an important component of the financial viability of rural practices. Furthermore, under the Closing the Gap Measure, DoHA has provided additional funding to AMS to assist them to become accredited. The net impact of the programme on equity has not been adequately assessed or monitored.

Costs and savings

The cost of PIP is significant, reaching nearly A$300 million per year in 2008–09, with almost A$3 billion in cumulative expenditures since its inception. The cost of the programme increased 25 per cent over the six-year period from 2002–03 to 2008–09, although it has been declining as a share of all government expenditure on primary care in Australia, from 8 per cent in 2002–03 to 5.5 per cent in 2008–09 (Australian National Audit Office, 2010). The costs to GP practices of participation, including accreditation and administrative burden, have not been quantified.

Provider response

The response of providers to the PIP was less than enthusiastic in the early stages of implementation. A government review of the programme in 2000 found that GPs claimed to participate in the programme mainly to supplement their income and fund maintenance of equipment and facilities (Wendy Bloom & Associates, 2000). During the Productivity Commission's review of the administrative burden of PIP in 2002, the Australian Medical Association submission was critical of the programme overall and particularly opposed to the perceived level of administrative burden of the programme (Australian Medical Association, 2002), which has been an ongoing source of dissatisfaction since the programme began. In its 2002 Annual Review of Regulatory Burdens on Business, the government's Productivity Commission found that PIP participation accounted for nearly 33 per cent of GP practice administrative costs (Productivity Commission, 2003). The issue was taken up again by the Regulation Task Force in 2006 (Commonwealth of Australia, 2009).

DoHA and Medicare Australia have been responsive to many of the concerns of providers, particularly attempting to simplify the administrative burden. Over time the providers have acknowledged a more positive role for the programme. In a survey of GPs conducted as part of the latest ANAO review, 88 per cent of PIP practices responded that they consider that PIP provides at least some support to them for providing patients with quality care and improved access. Views are still mixed, however, with 27 per cent of providers responding that PIP gives significant benefit to their practice, 36 per cent responding that there is medium benefit, and 27 per cent responding that the benefit is minor (Australian National Audit Office, 2010). A recent published study on the impact of the PIP included in-depth interviews to understand the perceptions of GPs about the programme. GPs reported that the incentive did not influence their behaviour, largely due to the modest payment and the complexity of tracking patients and claiming payment (Greene, 2013).

Overall conclusions and lessons learned

Has the programme had enough of an impact on performance improvement to justify its cost?

The PIP appears to have gained gradual acceptance among GPs, even if they do not appear to regard it as important in day-to-day service delivery decision making. The supplemental payments to practices seem to have contributed to enhancing quality of care to some degree, especially for chronic conditions. The structure of PIP – the umbrella structure for 13 different incentives – has allowed DoHA to provide flexible and tailored responses to quality of care in different areas. The emphasis put on quality and accessibility of care in rural and remote areas (by the different incentives and also the calculation of payment) also has contributed to addressing the crucial issue of care gaps between rural and urban areas. There is recognition that accountability and reporting have been improved to a certain extent. The introduction of the new online system will also contribute to reducing the administrative burden associated with the implementation of the programme.

Although there are modest results observed on service delivery and quality of care, the PIP has not been fully leveraged to drive performance improvement in primary care. There are several aspects of the design of the programme that limit the ability of PIP to significantly impact service delivery and reward real improvements in quality and outcomes:

1. *Complex and non-transparent programme structure*. The structure of the programme (13 incentives with requirements that can change from year to year) does not allow for a coherent set of policy objectives with clear priorities. In the New Zealand primary care P4P programme, for example, clarifying policy objectives and establishing priorities are seen as major benefits of the programmes to improving overall system performance (Buetow, 2008). Moreover, the mix of different payment mechanisms within PIP (between target and key performance indicators, sign-on, take-up of the incentive, etc.) has rendered monitoring difficult and payments less transparent. The calculation of payment levels based on SWPE has also added further confusion to the actual link between performance of practices and payments. The strength of the incentives and accountability also could be further enhanced by the publication of payment levels and rankings on performance, but limitations due to privacy regulations prevent the publication of payment levels and rankings for individual practices.
2. *Selective participation in lower effort incentive streams*. The structure of the incentive programmes allows providers to select those areas in which they have the greatest potential for award. This has resulted in a high uptake of an incentive that is relatively easy to achieve and that comes with a big reward (eHealth) and much lower uptake of the incentives related to service delivery for chronic conditions, which require much more effort on the part of the practices. The relative contribution of the two incentive areas to overall quality of care and performance is not known, but it seems that

increasing screening and appropriate management of chronic diseases is an essential element of providing good quality of care.

3. *Inadequate use of performance data for improvement processes*. Although IT-related incentives show the highest uptake, the potential of improved data, reporting and performance monitoring has not been fully exploited by the Australia PIP. No reports are available showing trends in performance against the different indicators, which would provide valuable information both for policy purposes and management of service delivery at the provider level. The possibility of monitoring trends is further diminished by the design of PIP, which allows PIP practices to move in and out of specific incentive programmes, rendering aggregate trends in indicator performance meaningless. Again, improved health information reporting, availability and use is found to be one of the main potential benefits of P4P programmes in a range of countries (Galvin, 2006; Sutton & McLean, 2006).

The evidence that the PIP has had impacts on quality of care and outcomes that justify the costs of the programme is limited. Furthermore, there are concerns about the role of the programme in exacerbating inequity across large urban practices and smaller practices serving more disadvantaged populations, and the possible spillover effects of the programme into other areas of service delivery and performance have not been addressed. An important contribution of the PIP, however, has been to pay providers for aspects of chronic disease management that are not typically reimbursed under fee-for-service payment systems and therefore have tended to be neglected. Part of the incentive payment being linked to the completion of a cycle of care rather than for each individual contact appeared to increase compliance with treatment guidelines. Overall, evaluation of both the impacts and spillover effects of the programme, particularly on small practices and those located in disadvantaged areas, should be the priority of DoHA. In the absence of such evaluation, conclusive evidence on the overall effectiveness of the programme is limited.

Notes

* This case study is based on the 2011 report 'RBF in OECD Countries: Australia – The Practice Incentives Program (PIP)' prepared by Cheryl Cashin and Y-Ling Chi for the International Bank for Reconstruction and Development, the World Bank and the OECD.

1 1 A$ = 0.994 US$ in January 2011.

2 Standardized Whole Patient Equivalent (SWPE) is a measure of a practice's patient load independent of the number of services provided. It is based on an estimate of the share of total care provided for a patient by the GP practice and is estimated from Medicare Australia claims data and weighted by age and sex.

3 Rural Remote and Metropolitan Area (RRMA).

4 Coverage targets for several of the quality stream indicators will be increasing during 2013. Practices will need to screen at least 70 per cent of their eligible patients to receive the Cervical Screening Incentive outcomes payment, up from 65 per cent. Practices will need to complete a diabetes cycle of care for at least 50 per cent of their diabetic patients to receive the PIP Diabetes Incentive, up from 20 per cent.

References

Australia Department of Human Services (2013) *Practice incentives program. Medicare* (http://www.medicareaustralia.gov.au/provider/incentives/pip/updates.jsp, accessed March 2013).

Australian Government Department of Health and Ageing (2009) *2008–09 Annual Report*. Canberra: Department of Health and Ageing.

Australian Government Department of Health and Ageing (2009) *Practice incentives program: guidelines*. Canberra: Department of Health and Ageing.

Australian Government Department of Health and Ageing (2009) *Primary health care reform in Australia: report to support Australia's first national primary health care strategy*. Canberra: Department of Health and Ageing.

Australian Government Department of Health and Ageing. (2010) *Primary health care reform in Australia: report to support Australia's first national primary health care strategy* (http://www.yourhealth.gov.au/internet/yourhealth/publishing.nsf/Content/nphc-draftreportsupp-toc~nphc-draftreportsupp-ch3, accessed June 2010).

Australian Government Department of Health and Ageing (2011) *Practice incentives program. Indigenous health incentives guidelines*. Canberra: Department of Health and Ageing.

Australian Medical Association (2002) *AMA submission to the Productivity Commission study on administrative and compliance costs associated with Commonwealth programs that impact specifically on general practice*. Barton: Australian Medical Association.

Australian National Audit Office (2010) *Practice incentives program audit report No. 5 2010–2011*. Canberra: Australian National Audit Office.

Boxall, A.M. (2009) *Should we expand the use of pay-for-performance in health care?* Canberra: Parliamentary Library Information, Parliament of Australia.

Buetow, S. (2008) Pay-for-performance in New Zealand primary health care, *Journal of Health Organization and Management*, 22(1): 36–47.

Commonwealth of Australia (2009) *Government response to the Productivity Commission annual review of regulatory burdens on business*. Canberra: Department of Industry.

Department of Health and Family Services (1996) *General practice in Australia*. Canberra: General Practice Branch, Commonwealth Department of Health and Family Services.

Ferguson, H. (2006) Sharing PIP, *Australian Doctor*, 28 June.

Galvin, R. (2006) Pay-for-performance: too much of a good thing? A conversation with Martin Roland, *Health Affairs*, 25(5): 412–19.

Greene, J. (2013) An examination of pay-for-performance in general practice in Australia, *Health Services Research*, 48(4): 1415–32.

Medicare Australia (2010) *Practice incentives program* (http://www.medicareaustralia.gov.au/provider/incentives/pip/index.jsp, accessed March 2013).

National Health Strategy (1992) *The future of general practice*. Canberra: Australian Government Publishing Service.

OECD (2009) *OECD health data*. Paris: Organization for Economic Co-operation and Development.

Productivity Commission (2003) *General practice administrative and compliance costs*. Canberra: Productivity Commission.

Russell, G. and Mitchell, G. (2002) Primary care reform: view from Australia, *Canadian Family Physician*, 48: 440–3.

Scott, A. et al. (2009) The effect of financial incentives on the quality of health care provided by primary care physicians. *The Cochrane Library 2009*, Vol. 9.

Sivey, P., Scott, A. and Schurer, S. (2008) *The effect of a performance incentive scheme on the allocation of time across multiple tasks: evidence from physicians*. Melbourne: Melbourne Institute of Applied Economic and Social Research, The University of Melbourne.

Sutton, M. and McLean, G. (2006) Determinants of primary medical care quality measured under the new UK contract: cross sectional study, *British Medical Journal*, 332: 389–90.

Wendy Bloom and Associates (2000) *Review of the practice incentives program. Conducted for the Commonwealth of Australia Department of Health and Aged Care, Medicare Benefits Branch*. Canberra: Department of Health and Ageing.

chapter seven

Estonia: Primary health care quality bonus system

Triin Habicht

Introduction

Estonia inherited the soviet Semashko-style health care system, which was in place prior to independence in 1991. The system was characterized by a large network of secondary care provider institutions and a fragmented primary health care (PHC) system, with separate parallel systems of care for adult services, children's services, and reproductive health services, and specialized dispensaries. Primary care doctors acted as referral points, or 'dispatchers', to specialists rather than as gatekeepers and care managers. This fragmented system was unable to effectively address the major shift in disease burden toward chronic diseases in Estonia and throughout the world that began in the 1970s. After independence, Estonia embarked on a fundamental reform of its health care system around a family medicine-centred PHC model to better address the health needs of the population, and in particular chronic diseases. At the suggestion of the new Society of Family Doctors, the newly established Estonia Health Insurance Fund (EHIF) introduced a pay for performance (P4P) programme in 2006 known as the Quality Bonus System (QBS). The objective of the programme was to reinforce the new position of family doctors and create an incentive for them to strengthen their role in disease prevention and chronic diseases management.

The Estonian health system is now based on compulsory, solidarity-based health insurance with providers operating under private law. Stewardship and supervision, as well as health policy development are the responsibility of the Ministry of Social Affairs (MoSA) and its agencies. The financing of health care is the responsibility of the independent EHIF, and out-of-pocket payments by individuals make up less than 25 per cent of total financing. The main role of the EHIF is to serve as an active purchasing agency, and its responsibilities include contracting health care providers and paying for health care services, reimbursing pharmaceutical expenditure, and paying for temporary sick leave and maternity benefits.

Health care provision has been almost completely decentralized since the new Health Services Organization Act was passed in 2001. The Act defines four types of health care: primary care provided by family doctors, emergency medical care, specialized (secondary and tertiary) medical care and nursing care. The Act established the regulatory framework for primary care and family medicine, under which primary care is organized as the first level of contact with the health system and provided by private family medicine practices contracted by the EHIF and serving the population on the basis of a practice list (Koppel et al., 2008).

The way family physicians are paid through the EHIF is a carefully crafted combination of payment methods to achieve a complete set of incentives for family doctors to take more responsibility for diagnostic services and treatment, as well as to compensate them for the financial risks associated with caring for older patients and working in remote areas (Koppel et al., 2008). Family physicians under contract with the EHIF are paid through a combination of a fixed monthly allowance, a capitation payment per registered patient per month, some fee-for-service payments, and additional payments based on the distance to the nearest hospital and performance-related payment through the QBS (Figure 7.1). The QBS incentive serves as a complementary and reinforcing part of this overall payment system design.

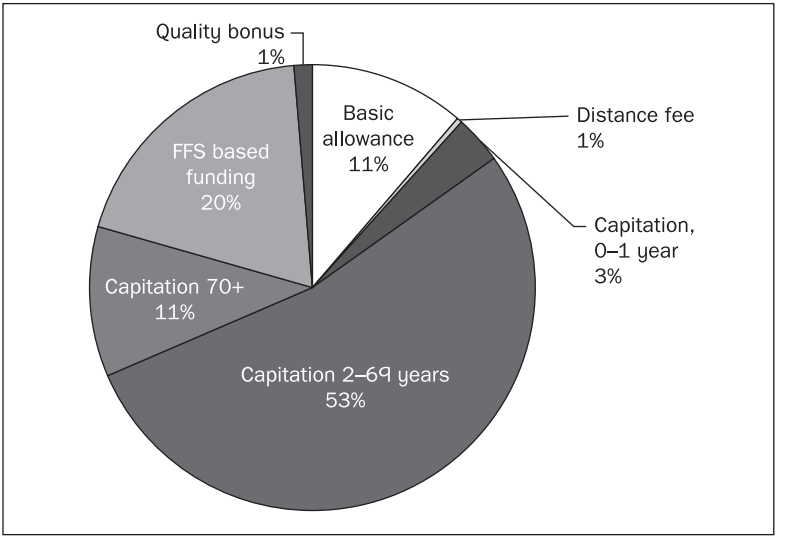


Figure 7.1 Mix of different payment methods for family physicians in Estonia, 2011

Source: Estonia Health Insurance Fund, 2011.

Health policy context

What were the issues that the programme was designed to address?

The QBS programme was launched in 2006 to highlight the importance of family physicians in disease prevention and chronic diseases management. By that time the primary health care reforms were complete, in the sense that the whole country was covered by family physicians, and all citizens were assigned to a family physician patient list. Even though family physicians were accepted as the first contact with the health system with responsibility for management of preventive work and chronic diseases management, the actual role of family doctors varied substantially based on different skills and motivation to take on new responsibilities. The QBS programme was seen as a tool to signal the importance of the role in chronic disease prevention and management and that it was clearly valued (also monetarily) by system. The objectives of the QBS were therefore defined as follows:

- To provide incentives to family physicians to focus on prevention to avoid high expenditures due to illness and incapacity to work in the future.
- To reduce morbidity from vaccine-preventable diseases and reduce hospitalization from chronic diseases.
- To improve the management of chronic diseases in PHC.
- To motivate FPs to widen the scope of provided services.

The initiative to develop the QBS was taken by the Society of Family Doctors, which started taking steps toward differentiated payment for providers based on performance already in 2001. As an initial step, the Society started the accreditation process of its members in 2002. The main goal was to give recognition to good professionals and create a basis for differentiation of payment for better performance. When the first 100 family physicians (out of approximately 800) passed the accreditation process, it emerged that the EHIF was not able to accept accreditation as a criterion for differentiated payment. The solution was to introduce a bonus payment as a 'new service' in the government-approved 'price list' (Aaviksoo, 2005).

Stakeholder involvement

In 2005 the Society of Family Doctors made a proposal to the EHIF to develop the QBS in collaboration. The Society developed the QBS, but it was done in a close collaboration with EHIF. The Society mainly took responsibility for the development of performance indicators, and the EHIF provided recommendations for implementation arrangements. Ongoing development of the QBS has been undertaken jointly by the EHIF and the Society of Family Doctors together on consensus basis. The joint development of the programme has ensured wider acceptance of the QBS by family physicians, as the system is not seen purely as initiative of financing organization.

Technical design

How does the programme work?

Performance domains and indicators

The QBS includes three domains: disease prevention, chronic diseases management, and additional activities. Each domain has several indicator groups with a total of 45 indicators and 600 possible points (Table 7.1). There are different total points available for each domain and for each indicator. Family physicians earn points for reaching performance targets for each indicator. The points are awarded on an 'all or nothing' basis; that is, if the physician reaches the target she or he is awarded all of the points. If the physician fails to reach the target, no points are awarded.

Table 7.1 Performance domains of the Estonia QBS

Indicators		*Maximum points*	*Points in total (maximum)*	*Minimum level of points to be eligible to bonus*
Domain I*: Disease prevention	Child vaccination (9 indicators)	90	200	160 (80 per cent of max)
	Children's preventive check-ups (5 indicators)	50		
	CVD prevention (4 indicators)	60		
Domain II: Chronic disease management	Diabetes, type II (6 indicators)	104	400	320 (80 per cent of max)
	Hypertension (14 indicators)	248		
	Myocardial infarction (2 indicators)	32		
	Hypothyreosis (1 indicator)	16		
Domain III: Additional activities*	FP and nurse training (1 indicator)	Coefficient 0.2	Coefficient 1.0	Coefficient 0.2
	Maternity care (1 indicator)	Coefficient 0.3		
	Gynaecological activities (1 indicator)	Coefficient 0.2		
	Surgical activities (1 indicator)	Coefficient 0.3		

Domain I, 'Disease prevention', includes three indicator groups: child vaccination, children's preventive check-ups, and cardiovascular disease (CVD) prevention. The target threshold for child vaccination and check-ups is 90 per cent of the target group covered. There is a procedure for 'exception reporting', so providers are not penalized for patient behaviour beyond their control. For example, children can be excluded from calculating vaccination rates when their parents refuse vaccinations through a written refusal, or if they have a medical condition that does not allow vaccination. Also, family physicians can apply to exclude those children who live abroad or were assigned to the family physician's practice list but have never visited the physician.

The target group for CVD prevention indicators is all adults age 40 to 60 without hypertension, type II diabetes or history of myocardial infarction. The target threshold for prevention is 80 to 90 per cent coverage, depending on the indicator. As the actual level of indicator values has been low, targets have been revised and set at the previous year's average achievement rate plus 10 per cent. So, if the actual average coverage rate is 45 per cent this year, next year's target will be 55 per cent coverage.

Domain II, 'Chronic disease management', includes indicators for four conditions: hypertension, type II diabetes, myocardial infarction and hypothyreosis. The indicators are directly linked to clinical guidelines and focus on key activities required of family physicians and nurses to manage these conditions. These are process indicators and do not include outcomes (e.g. target blood pressure) due to the lack of availability of necessary data. The indicators for hypertension are weighted most heavily, accounting for 40 per cent of the total potential points for Domains I and II (Figure 7.2). The

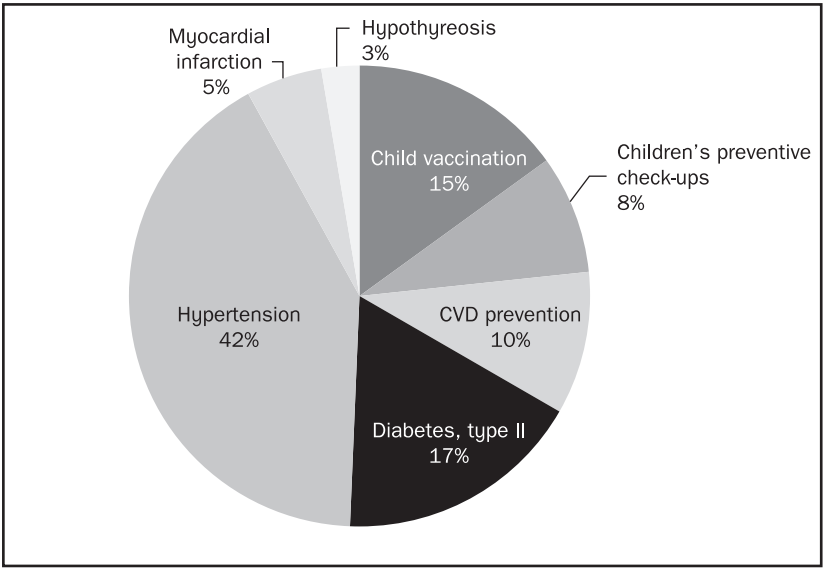


Figure 7.2 Distribution of points across domains (I and II) for bonus payments in the Estonia QBS

target thresholds for this domain are between 80 to 90 per cent coverage. Similar to the CVD prevention indicators, the actual targets used in practice follow a step-wise approach, with the current year's target based on the previous year's average achievement plus 10 per cent.

Domain III, 'Additional activities', includes indicators for four different areas: family physician and nurse recertification, maternity care, gynaecological activities, and surgical activities. In this domain each of the four indicators has a target level, and the family physician receives the respective coefficient when the target is achieved. The coefficient represents the share of the total possible award for Domain III. For example, if the family physician and nurse both have valid accreditation, the coefficient 0.2 (or 20 per cent of the maximum possible payment) is received. If the family physician has performed at least 40 surgical manipulations annually, the coefficient 0.3 is received. The maximum sum of coefficients is 1, which guarantees the physician is eligible for the full payment for Domain III.

Incentive payments

Domains I and II form the basic payment, which was a maximum of €3068 per year in 2011. Family physicians are eligible for bonus payments if they achieve at least 80 per cent of possible points. The bonus payment is paid to the family physician at 100 per cent (€3068) for Domains I and II if the physician achieves at least 560 points, and at 80 per cent (€2454) if the physician achieves at least 480 points. The first year of implementation of the QBS was an exception, as all family physicians who submitted their chronic patients lists received a bonus payment of at least 25 per cent regardless of achievement in order to send the message that all doctors interested in participating in the new system deserved a reward (EHIF, 2008).

Family physicians can earn an additional payment from Domain III extra activities, but only if they qualify for a bonus in Domains I–II at least at the 80 per cent level. The maximum payment for Domain III is €767 per year, but the amount paid to the physician depends on the coefficients achieved in each of the Domain III indicator areas.

The QBS bonus is paid directly to the family physician, who then decides whether and how the payment is shared among other staff such as nurses. If the family physician works in a group practice rather than as a solo practitioner, the bonus payment is still linked only to the individual physician's performance and not the practice as a whole. Initially bonus payments were made monthly, but since 2008 the payment is made annually for administrative simplicity.

In addition to the direct incentive of the QBS bonus payment, family physicians can earn additional revenue by participating in the programme through an expanded fee-for-service fund. Family physicians can earn up to 29 per cent of their income through fee-for-service payments in addition to the basic allowance and the capitation payments. If the physician participates in QBS, however, the fee-for-service fund increases to 34 per cent. If the physician participates in QBS and qualifies for a bonus payment, the fee-for-service fund increases to 37 per cent.

Data sources and flows

All necessary data to implement the QBS come from the EHIF's routine claims data. The EHIF has had an electronic billing system in place since 2001 for all providers in the country. This means that patient-level electronic information is available for all cases, including patient diagnosis, and all performed activities according to payment rules. Since the main payment method for family physicians is capitation, however, claims data are not used for payment and there is a separate system to code provider activity in the EHIF routine data system to get all necessary data input to the QBS without additional data collection. The only exception is recertification of physicians and nurses, which requires data to be provided by professional associations that oversee continuing medical education.

Before 2010 the lists of patients with chronic diseases covered by QBS was submitted to the EHIF separately. Since 2010, however, information on chronic disease status is available in the EHIF's billing data. The patient is categorized as a chronic disease patient if she or he has had at least one claim to the EHIF by the family physician in the last three years with that diagnosis. The list of chronic patients is presented to the family physicians by the EHIF. The family physician's confirmation of the list of patients with chronic diseases is considered as conformation of the family physician's participation in QBS.

Reach of the programme

Which providers participate and how many people are covered?

The Estonian QBS is a voluntary system used to reward well-performing family physicians who have a registered patient list. In 2006, the first year of implementation, 50 per cent of all family physicians participated in the system. Since that time, the share of participating physicians has been increasing, reaching 90 per cent in 2010, covering 90 per cent of insured people in Estonia (see Figure 7.3). The maximum QBS bonus payment across all three domains in 2011 was €3835, or 4.5 per cent of the total annual income for a family physician (€80,800). The total the cost of QBS in that year was €800,000, or about 1 per cent of the EHIF's total PHC budget.

Improvement process

How is the programme leveraged to achieve improvements in service delivery and outcomes?

The QBS is leveraged to drive improvement not only through the financial incentive, but also by providing feedback on performance. Every family physician receives personal feedback on her or his results electronically in the third quarter of the performance year, so there is time to improve results before the end of the year, and again at the end of the year with final results.

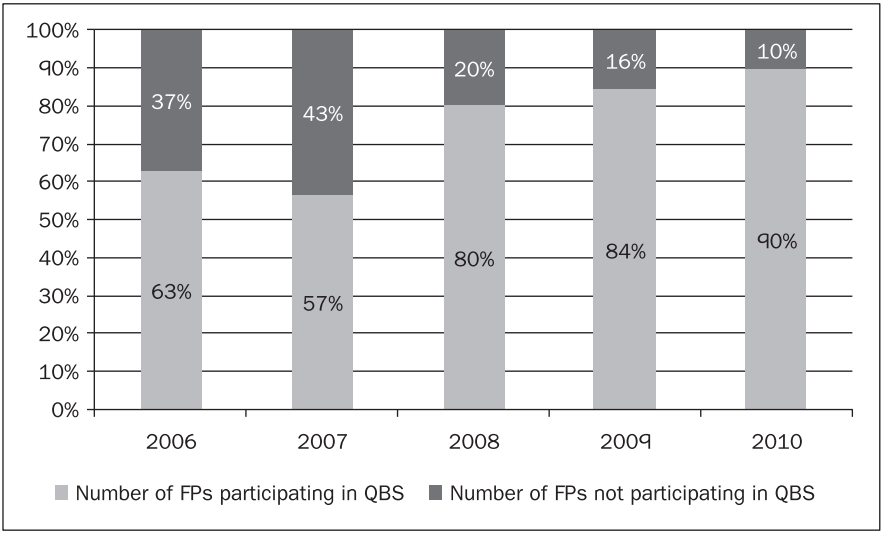


Figure 7.3 Participation of family physicians in the Estonia QBS, 2006–2010

In addition, the list of family physicians participating in the QBS is published annually on the EHIF website along with performance results. Public interest in the performance information, however, is only modest, possibly because the information presented may not be accessible and easy to interpret.

Results of the programme

Has the programme had an impact on performance, and have there been any unintended consequences?

The QBS system has been in place now for six years, and results are available for the first five years. Over that time, participation has increased and only 10 per cent of family physicians in 2010 did not participate in this voluntary system. About 25 per cent of family physicians received bonus payments at the maximum level for Domains I and II in 2010. Half of the family physicians participating in the QBS did not receive bonus payment (Figure 7.4). In 2010 there was a change in the system, as this was the first year when EHIF took the chronic patients lists from claims data, which may have increased the number of patients identified as chronic and therefore reduced coverage rates if nothing else changed.

There is wide variation across the counties in Estonia in both participation rates and the share of family physicians receiving bonus payments (Figure 7.5). For example, in Hiiu county all family physicians participated in the QBS in 2010, whereas in Viljandi and Põlva counties 22 per cent of family physicians did not participate in the system. Also, in Viljandi and Järva only 13 per cent

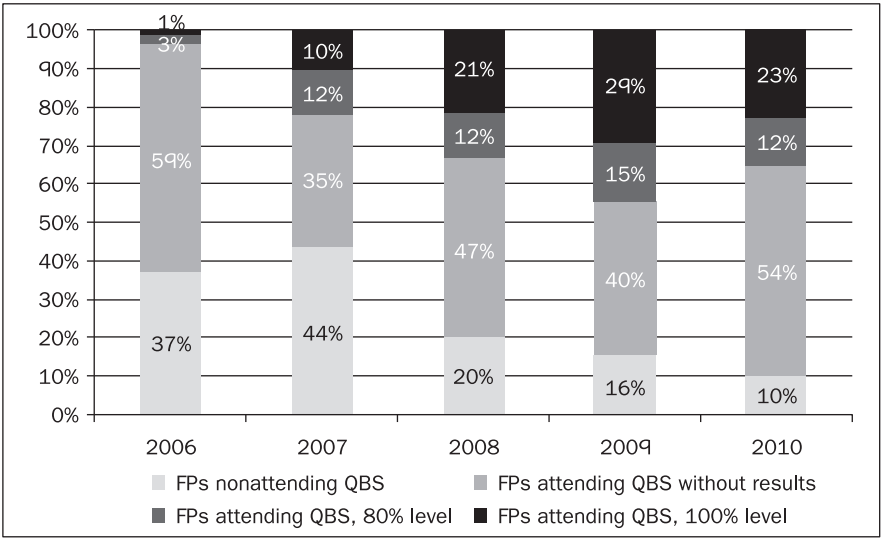


Figure 7.4 The share of family physicians participating in the Estonia QBS and receiving bonus payments, 2006–10

of family physicians received a bonus payment, while in Lääne, Pärnu and Hiiu counties 63 per cent of family physicians achieved a high enough performance score to receive a payment.

No formal evaluation has been done of the QBS. Several studies assessing the impact of QBS suggest that participation in the programme is linked to better chronic disease management and reduced hospitalization for chronic

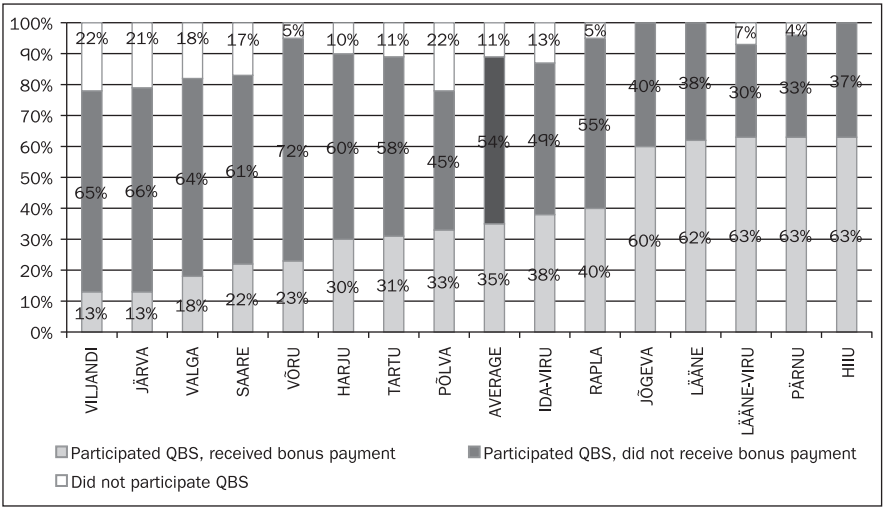


Figure 7.5 Participation of family physicians in QBS by county, 2010

conditions. Västra (2010), for example, analysed the impact of the QBS on management of hypertension and type II diabetes in 2005–08. That study found that family physicians participating in the QBS and achieving a high enough performance score to receive a bonus perform better in providing continuous follow-up for chronic patients, and their patients tend to require specialist services and hospitalization less frequently.

Overall conclusions and lessons learned

Has the programme had enough of an impact on performance improvement to justify its cost?

The most important impact of the QBS in Estonia has been raising awareness and understanding of the role of family physicians in providing the full scope of high quality services, particularly preventing and managing chronic diseases. The implementation of the QBS and the monitoring of performance results have highlighted the importance of clinical guidelines in performance monitoring at PHC level. The cost of the QBS is modest at only 1 per cent of the annual PHC budget.

The most important factor in implementing the QBS system successfully has been the electronic billing data collection system that covers all family physicians in Estonia. This detailed patient-level information makes it possible to assess performance measures without additional data collection, particularly now that even lists of patients with chronic diseases are extracted from the EHIF database. The limitation of billing data is that although it contains process-based information (i.e. which diagnostic tests have been done), it does not include outcome measures (i.e. values of blood pressure), and therefore QBS has been limited to including only process-based indicators.

So far the QBS has been the only initiative in the Estonian health care system to link provider performance with payment. The system is voluntary, but it has been widely accepted by family physicians. In fact, the initiative came from the family physicians themselves through the Society of Family Doctors, and the system was developed through a close collaboration between the Society and the EHIF. Nonetheless, only 35 per cent of family physicians consider QBS to be motivating for them (State Audit Office, 2011), which may be due to the relatively low bonus payment. There is ongoing discussion in Estonia about whether the bonus payment should be larger to increase the impact of the incentive.

Appendix 7.1 Domain I and II indicators with actual and goal coverage in QBS

Indicator	*2007*	*2008*	*2009*	*2010*	*Goal*
Domain I – Prevention					
Child vaccination					
Whooping cough 3 months	93%	96%	94%	94%	90%
Whooping cough 4.5 months	91%	95%	92%	93%	90%
Whooping cough 6 months	89%	92%	91%	91%	90%
Whooping cough 2 years	83%	86%	83%	86%	90%
Diphtheria 3 months	93%	96%	94%	94%	90%
Diphtheria 4.5 months	91%	95%	92%	93%	90%
Diphtheria 6 months	89%	93%	91%	91%	90%
Diphtheria 2 years	83%	86%	83%	86%	90%
Tetanus 3 months	93%	96%	94%	94%	90%
Tetanus 4.5 months	91%	95%	92%	93%	90%
Tetanus 6 months	89%	92%	91%	91%	90%
Tetanus 2 years	83%	86%	83%	86%	90%
Poliomyelitis 3 months	93%	96%	94%	94%	90%
Poliomyelitis 4.5 months	91%	95%	92%	93%	90%
Poliomyelitis 6 months	89%	92%	91%	91%	90%
Poliomyelitis 2 years	83%	86%	83%	86%	90%
Measles 1 year	87%	91%	89%	90%	90%
Mumps 1 year	87%	91%	89%	89%	90%
Rubella 1 year	87%	91%	88%	90%	90%
Hepatitis B3 5 days	95%	96%	96%	82%	90%
Hepatitis B 1 month	95%	96%	95%	95%	90%
Hepatitis B 6 months	89%	93%	91%	92%	90%
Haemophilus influenzae type b I 3 months	92%	96%	93%	94%	90%
Haemophilus influenzae type b I 4.5 months	90%	95%	92%	93%	90%
Haemophilus influenzae type b I 6 months	88%	92%	91%	91%	90%
Haemophilus influenzae type b I 2 years	75%	85%	82%	86%	90%
Children's preventive check-ups					
Children's check-up 1 month	83%	88%	90%	92%	90%
Children's check-up 3 months	81%	89%	91%	91%	90%
Children's check-up 1 year	78%	86%	88%	89%	90%
Children's check-up 2 years	75%	82%	81%	83%	90%

(*continued*)

Appendix 7.1 Domain I and II indicators with actual and goal coverage in QBS (*continued*)

Indicator	*2007*	*2008*	*2009*	*2010*	*Goal*
CVD prevention					
Total cholesterol measured for 40–60 year old person once in 5 years	41%	45%	54%	61%	80%
Glycose test for high CVD risk persons aged 40–60 once per year			79%	67%	90%
Fractions of cholesterol measured for high CVD risk persons aged 40–60 once per year			73%	61%	90%
Nurse counselling for high CVD risk persons aged 40–60 once per year			72%	69%	90%
Domain II – Chronic diseases management					
Type II diabetes					
Glycohemoglobin test done for patients with type II diabetes once per year	46%	61%	70%	66%	90%
Creatin test done for patients with type II diabetes once per year	49%	62%	69%	66%	90%
Total cholesterol test done for patients with type II diabetes once per year	54%	65%	71%	69%	90%
Fractions of cholesterol measured for patients with type II diabetes once per 3 years	42%	71%	86%	87%	90%
Albumin test done for patients with type II diabetes once per year	30%	54%	43%	42%	90%
Nurse counselling for type II diabetes patients	40%	53%	62%	64%	90%
Hypertension					
Glucoses test done for hypertension patients (low risk) once per 3 years	65%	80%	86%	79%	80%
Total cholesterol test done for hypertension patients (low risk) once per 3 years	64%	79%	87%	81%	80%
ECG done for hypertension patients (low risk) once per 3 years	54%	68%	79%	71%	80%
Nurse counselling for hypertension patients (low risk)	32%	47%	57%	49%	90%
Total cholesterol test done for hypertension patients (medium risk) once per year	49%	56%	62%	61%	90%
Fractions of cholesterol measured for patients with hypertension (medium risk) once per year	38%	47%	55%	54%	90%
Glucoses test done for hypertension patients (medium risk) once per year	52%	58%	61%	59%	90%
Creatin test done for patients with hypertension (medium risk) once per year	42%	51%	58%	57%	90%

ECG done for hypertension patients (medium risk) once per 3 years	56%	69%	78%	78%	80%
Albumin test done for patients with hypertension (medium risk) once per year	22%	44%	33%	32%	90%
Total cholesterol test done for hypertension patients (high risk) once per year	50%	59%	69%	67%	90%
Fractions of cholesterol measured for patients with hypertension (high risk) once per year	42%	50%	63%	61%	90%
Glucoses test done for hypertension patients (high risk) once per year	57%	62%	67%	65%	90%
Creatin test done for patients with hypertension (high risk) once per year	46%	55%	67%	65%	90%
Albumin test done for patients with hypertension (high risk) once per year	27%	45%	39%	37%	90%
Myocardial infarction					
Total cholesterol test done for patients with myocardial infarction once per year			71%	69%	90%
Glucoses test done for patients with myocardial infarction once per year			69%	68%	90%
Hypothyreosis					
TSH test done for patients with hypothyreosis once per year			63%	64%	90%

References

Aaviksoo, A. (2005) *Performance payment for family physicians*, *Health Policy Monitor*, 6.

Estonia Health Insurance Fund (2008) *Estonian Health Insurance Fund annual report* (http://www.haigekassa.ee/eng/ehif/annual, accessed June 2013).

Estonia Health Insurance Fund (2011) *Estonian Health Insurance Fund annual report* (http://www.haigekassa.ee/eng/ehif/annual, accessed June 2013).

Koppel, A. et al. (2008) Estonia: health system review. Health systems in transition, *WHO Regional Office for Europe on behalf of the European Observatory on Health Systems and Policies*, 10(1).

Västra, K. (2010) Assessing the impact of implementing primary care quality bonus system on follow up of patients with hypertension and type 2 diabetes based on Estonian Health Insurance Fund claims registry data in 2005–2008. Unpublished Masters thesis, University of Tartu.

State Audit Office (2011) *Family physicians web-based survey*. Tallinn: State Audit Office.

chapter eight

France: Payment for public health objectives

Frédéric Bousquet, Raphaëlle Bisiaux and Y-Ling Chi

Introduction

In 2000 the World Health Organization ranked France's health care system as the best performing in the world. The system produces good health outcomes and high levels of satisfaction among the French population, but threats to financial sustainability have been looming for two decades. The National Health Insurance Fund (Assurance Maladie) has been operating under a deficit, with the shortfall reaching 10 per cent of the insurance system's total budget in 2010. It has been difficult to control costs in a system characterized by fee-for-service payment and unlimited patient choice (Sandier, Paris & Polton, 2004). Successive reform plans introduced since 2004 were aimed largely at controlling France's unchecked health care demand, but also experimenting with new provider payment systems other than fee-for-service.

The French health care system is characterized by 'liberalism' and 'pluralism' (Rodwin, 2003), which translates into a high degree of freedom for physicians and choice for patients. The National Health Insurance Fund (NHIF) coexists with private medical practice under fee-for-service payment, with little control over the decisions of physicians or patients. Compared to the UK, where GPs work in teams with nurses and other health care providers, French primary care doctors generally work in solo private practice. French physicians enjoy a great deal of freedom in the practice of medicine and can enhance their incomes through a high volume of services. Thus, the fee-for-service system in France has not encouraged prevention or a coordinated approach to primary care, and general practitioners have not received any financial incentives for time-consuming activities such as managing chronic diseases (Degos et al., 2008). At the same time, patients have had virtually unlimited choice in utilization at all levels of care. Until recently, access to specialists in independent practice was not regulated in France. Patients also have high expectations about access to medicines, and physicians have no incentive to limit prescribing. Fewer than 10 per cent of consultations in France end without a prescription (Degos et al., 2008).

The high degree of independence and choice for both providers and patients has been a key driver of health care cost escalation in France, which has been accompanied by fragmented, uncoordinated care. A series of health reforms since 2004 has attempted to address these structural problems in the French health system, but progress was considered insufficient. In 2009 the NHIF introduced the pay for performance pilot programme Contracts for Improved Individual Practice (CAPI) for primary care physicians in an attempt to stimulate fundamental changes in the way health care is delivered in France. In 2012, CAPI was extended to all GPs and to some specialists for a set of specific indicators. At that time, CAPI was renamed Rémunération sur Objectifs de Santé Publique (ROSP; Payment for Public Health Objectives).

Under the new National Agreement on setting tariffs and regulating the relations between private medical practitioners and the NHIF in 2011,[1] private physicians are enrolled automatically in ROSP, but they remain free to opt out of the programme. Four domains of performance are rewarded based on a total of 29 indicators: prevention, chronic disease management, cost-effective prescribing, and the practice organization. The P4P programme aims to improve quality of clinical care and to encourage efficient practices and organization, but it does not alter the existing fee-for-service payment system (Or, 2010). ROSP is directed to both primary care physicians and a class of specialists for which the programme is still under development. For convenience, only primary care physicians' financial incentives will be discussed in this chapter, as payments to specialists are currently still limited.[2]

Health policy context

What were the issues that the programme was designed to address?

In spite of numerous reform initiatives over the past decade, the highly individualistic and pluralistic nature of health care in France combined with the deeply rooted fee-for-service payment system continues to promote fragmented and inconsistent care, inadequate focus on preventive services, and a high degree of heterogeneity in the quality of clinical practice. Reforms introduced through the Health Insurance Reform Act of 2004 have tried to regulate access to specialist care. Each French citizen now has to choose a 'gatekeeper general practitioner' (*médecin traitant*) who is responsible for all primary care and referrals to specialists. If patients do not follow the 'coordinated care pathway' and choose instead to self-refer to specialists, the rate of reimbursement by the health insurance fund is reduced from 70 per cent to 50 per cent.

The second pillar of the 2004 reform aimed to address quality of care. The Act established the National Authority for Health (HAS), which has the mandate to enhance quality throughout the French health system through a variety of mechanisms including health technology assessment, clinical

guidelines, and accreditation (Haute Autorité de Santé, n.d.). The 2004 reform included EPP – Evaluation of Professional Practice, which encourages primary care physicians to follow HAS recommendations. Since 2005, primary care physicians are expected to undergo a mandatory evaluation by HAS at least every five years. This reform has not been fully implemented for private physicians working in ambulatory setting, however. So far the HAS evaluations have mainly been limited to physicians in hospitals who collectively completed the evaluation process within hospital accreditation programmes. EPP was replaced in 2012 by an ongoing process mixing evaluation and trainings.

In spite of these measures to enhance quality, however, the gaps between HAS recommendations and actual service provision remain large, as demonstrated by several studies led by the NHIF. For instance, only 31 per cent of patients diagnosed with diabetes received all four recommended diabetes services in 2008 (Commonwealth Fund, 2008). In terms of prevention, only 61 per cent of women aged 50 years and above had a screening test for breast cancer during the previous two years (Aubert & Polton, 2009).

Two major laws introduced in 2008 and 2009 opened up the possibility to use new organizational models and payment methods, including pay for performance, to drive improvements in service delivery. The 2008 Finance Law for Social Security allowed experimentation with new payment systems other than fee-for-service for the next five years, with a compulsory annual evaluation to be sent to Parliament. The 2009 Hospital, Patients, Health and Territories Law opened the way for a new organization at the regional level, which is aimed to reinforce prevention, access to health care and modernization of hospitals. ROSP combines a number of elements from these reform initiatives of the past decade and reinforces them with financial incentives. The ROSP sets common objectives for health care professionals with respect to treatment, prescribing patterns and practice organization. In contrast with the previous years, however, the achievement of objectives is now assessed at the level of the individual physician.

Stakeholder involvement

In 2009, the contract model for CAPI was prepared by the NHIF as an amendment to the new National Agreement on setting tariffs and regulating the relations between medical practitioners and the NHIF. The NHIF developed the performance indicators based on the national public health objectives and recommendations of HAS. At that time, there was little direct involvement of providers in the design and implementation of the programme. The quality indicators were submitted to HAS, which validated them. Most of the performance measures had been selected based on objectives and criteria defined by the 2004 Public Health Law as well as different HAS guidelines. Not surprisingly, these indicators are consistent with those already validated and in use internationally in programmes such as the UK Quality and Outcomes Framework (QOF) and the US National Quality Forum.

An additional change with the transition from CAPI to ROSP was the

inclusion of medical professionals in the definition of quality indicators used for performance monitoring and payment. Since then, the NHIF has been working with unions of physicians to review existing performance indicators and develop new ones for specialist physicians. The measures are always based on applicable clinical recommendations. Experts from professional societies may be involved in the process, but the external validation process remains quite informal. Indicators are presented and adopted during meetings of the institution in charge of monitoring the National Agreement between the NHIF and private physicians.

Technical design

How does the programme work?

Following the adoption of the new National Agreement, all GPs have been automatically included in the programme. Nonetheless, participation remains non-mandatory, as physicians can notify the NHIF if they do not wish to take part.

Performance domains and indicators

Performance indicators used in ROSP include process, structure and outcome indicators in the four domains of performance: (i) prevention; (ii) chronic disease management (diabetes and hypertension); (iii) cost-effective prescribing; (iv) practice organization. Table 8.1 provides detailed information on the set of indicators used for performance assessment.

Incentive payments

Each indicator is associated with a number of points, and the achievement rate calculation takes into account the level of achievement and the progress made during the year on every measure, except for the practice organization domain. A baseline performance level is measured for each physician and two types of objectives are used to set payment:

- an intermediate objective that corresponds to the average score of physicians for the specific indicator, which would qualify the physician for half of the points that can be earned for that indicator;
- a target objective that is based on objectives defined by the Public Health Law, the National Health Authority guidelines, or international comparisons, which would qualify the physician for the maximum of points that can be earned for that indicator.

The performance calculation formula was developed in order to not penalize physicians whose baseline level is higher than the intermediate level, for which margin of improvement is smaller (i.e. good performing physicians). Thus the achievement rate is calculated differently if the providers' initial performance

Table 8.1 Performance indicators in the France ROSP

Domain	*Theme*	*Indicator*	*Minimum threshold*	*End objective*	*Intermediate objective*	*Number of points (total = 1300)*
Prevention and screening Total = 250 points	Flu immunization	1. Proportion of patients 65 and older with flu immunization.	20 patients	>= 75%	62%	20
		2. Proportion of patients aged 16–64 with chronic condition having flu immunization.	10 patients	>= 75%	62%	20
	Breast cancer screening	3. Proportion of women aged 50–74 who received a mammogram in the past two years.	20 patients	>= 80%	70%	35
	Cervical cancer screening	4. Proportion of women aged 25–65 with at least one pap smear in the last three programmes years.	20 patients	>= 80%	65%	35
	Prescribing vasodilators for elderly patients	5. Proportion of patients older than 65 treated with vasodilators during the year.	20 patients	<= 5%	7%	35
	Long half-life benzodiazepines prescribing	6. Proportion of patients older than 65 treated with long half-life benzodiazepines.	20 patients	<= 5%	11%	35
		7. Proportion of patients treated with long half-life benzodiazepines for 12 weeks or more.	5 patients	<= 12%	13%	35
	Antibiotic therapy	8. Number of antibiotic prescriptions for patients aged 16–65 per 100 patients.	20 patients	=<37	40	35
Quality of the management of chronic conditions Total = 250 points	Diabetes	9. Proportion of diabetes patients who had three or four tests of HbA1c during the last year.	10 patients	>= 65%	54%	30
		10. Proportion of diabetes patients who had an ophthalmological consultation or underwent an examination of the fundus in the last two years.	10 patients	>= 80%	68%	35
		11. Proportion of diabetes patients over 50 (men) or 60 (women) treated with antihypertensives who received statins.	10 patients	>= 75%	65%	35

(*continued*)

Table 8.1 Performance indicators in the France ROSP (*continued*)

Domain	*Theme*	*Indicator*	*Minimum threshold*	*End objective*	*Intermediate objective*	*Number of points (total = 1300)*
		12. Proportion of diabetes patients over 50 (men) or 60 (women) treated with antihypertensives and statins, who received low-dose aspirin or other anticoagulant.	10 patients	>= 65%	52%	35
		13. Proportion of type II diabetes patients whose HbA1c test <8,5%.	10 patients	>= 90%	80%	15
		14. Proportion of type II diabetic patients with HbA1c test <7,5%.	10 patients	>= 80%	60%	25
		15. Proportion of type II diabetic patients with LDL cholesterol test <1.5g/l.	10 patients	>= 90%	80%	10
		16. Proportion of type II diabetic patients with LDL cholesterol test <1.3g/l.	10 patients	>= 80%	65%	25
	Hypertension	17. Proportion of patients treated with antihypertensives with blood pressure <= 140/90 mmHg.	20 patients	>= 60%	50%	40
Efficiency of drug prescription Total = 400 points	Antibiotics	18. Generic antibiotics dispensed as % of total antibiotics dispensed (# of items).	40 packages	>= 90%	85%	60
	PPIs	19. Generic PPIs dispensed as % of total PPIs dispensed (# of items).	35 packages	>= 85%	68%	60
	Statins	20. Generic statins dispensed as % of total statins dispensed (# of items).	30 packages	>= 70%	46%	60
	Antihypertensives	21. Generic antihypertensives dispensed as % of total antihypertensives dispensed (# of items).	130 packages	>= 65%	64%	55
	Antidepressants	22. Generic antidepressants dispensed as % of total antidepressants dispensed (# of items).	30 packages	>= 80%	75%	55

	ACEI/ARBs	23. ACEI as % of ACEIs + ARBs dispensed (# of items).	50 packages	>= 65%	45%	55
	Aspirin	24. % of patients treated with antiplatelet drugs also treated with aspirin.	10 patients	>= 85%	84%	55
			Minimum requirement			
Practice organization Total = 400 points	GPs and specialists	25. Electronic medical patient file system and update of the clinical data for individual follow-up of patients.	The physician is equipped with up-to-date software able to transmit administrative data and at least two-thirds of administrative related to visits are actually transmitted.			75
	GPs and specialists	26. Certified software to assist prescription.				50
	GPs and specialists	27. Computer equipment and software allowing doctors to send information via internet and provide online consultation services.				75
	GPs and specialists	28. Provision of information on the opening hours and organizational features of the practice (e.g. equipment to improve access of patients with disability) in the practice and on website of the NHIF.				50
	GPs only	29. Yearly assessment of individual electronic patient files, and a synthesis report given to the patient.				150

Indicators 1 to 25 and 29 apply to GPs only while indicators 25 to 28 apply for GPs and specialists

Source: CNAMTS, 2012 (http://www.ameli.fr/professionnels-de-sante/medecins/votre-convention/nouvelle-convention-medicale.php).

level is below the intermediate objective, or between the intermediate objective and the target. The details are explained below:

- Current level below the intermediate objective

$$achievement\ rate = 50\% \times \frac{current\ level - initial\ level}{intermediate\ objective - initial\ level}$$

- Current level is between the intermediate objective and the target objective

$$achievement\ rate = 50\% + 50\% \times \frac{current\ level - intermediate\ objective}{target\ objective - intermediate\ objective}$$

The monetary value per point is negotiated within the National Agreement and is currently set at 7€ per point. The total payment per indicator is the point value multiplied by the achievement rate, adjusted by the number of patients who have chosen the physician as their attending physician (with the exception of indicators related to practice organization). If a GP does not agree with the assessment of achievement, he or she can request a meeting with the local representation of the NHIF for a second assessment.

Data sources and flows

Performance indicators are calculated using mainly insurance claims data. Since 2005, reimbursement claims processed by all French public Health Insurance Funds are centralized in a data warehouse with the identification of all professionals and hospitals and details of all items of care for each individual patient. These data are compiled by the NHIF and serve as the basis for calculating process indicators used for performance assessment. This database is also complemented by physician reports of patient outcomes for indicators related to diabetes control and management.

Reach of the programme

Which providers participate and how many people are covered?

The programme is implemented nationally by the three health insurance funds under the National Health Insurance Fund, which together cover the entire population. Prior to the new National Agreement, about 16,000 primary care physicians enrolled in the programme between May 2009 and November 2011, which represented nearly 40 per cent of eligible primary care physicians.

Since July 2011, ROSP in theory covers GPs, as physicians are automatically registered into ROSP, unless they opt out or do not provide the data requested. In 2012, of 115,000 private physicians, 110,000 were eligible to participate to the programme, and only 3300 have formally refused to participate,[3] or less than three per cent (Caisse Nationale d'Assurance Maladie, 2013).

Improvement process

How is the programme leveraged to achieve improvements in service delivery and outcomes?

The NHIF has facilitated the improvement process by feeding information back to providers on their performance, information that was not readily available to providers prior to the introduction of the programme. Physicians now can access information on their performance and activity directly on the NHIF website using their professional accounts[4] (Chevreul et al., 2010).

For each indicator, individual physician information is compiled and stored on a quarterly basis. Individual physicians can track their scores over time and also benchmark them against national targets, and regional and national averages. The extent to which physicians use the performance information to improve their practice of care is still unclear, however. Physicians have to actively log-in their professional space on the website to access the information, which is not accessible via their routine electronic patient management software. Nonetheless, local Insurance Fund offices send delegates to discuss performance scores with physicians and suggest possible improvements. There is no public disclosure of performance scores.

Results of the programme

Has the programme had an impact on performance, and have there been any unintended consequences?

Programme monitoring and evaluation

The NHIF has conducted two main evaluation studies of the ROSP and CAPI:

- The first study compares the evolution of the performance indicators between CAPI signers (including 12,000 physicians) and a comparison group of 23,700 physicians who did not sign the CAPI between March 2009 and 2012.
- The second study compares the performance indicators before and after the change to ROSP in January 2012 (December 2011 to December 2012).

The evaluation of CAPI (Caisse Nationale d'Assurance Maladie, 2010)

A first analysis of individual performance data was conducted by the NHIF comparing CAPI signers and non-signers between 2009 (prior to the introduction of the new National Agreement) and 2012. Propensity matching scores were used to match the two groups, using patient and physician demographic information (e.g. urban vs. rural, incidence of chronic conditions, socio-economic information on the areas of practice).

Before the introduction of CAPI, differences between the two groups were not significantly different at the one per cent threshold for almost all indicators.

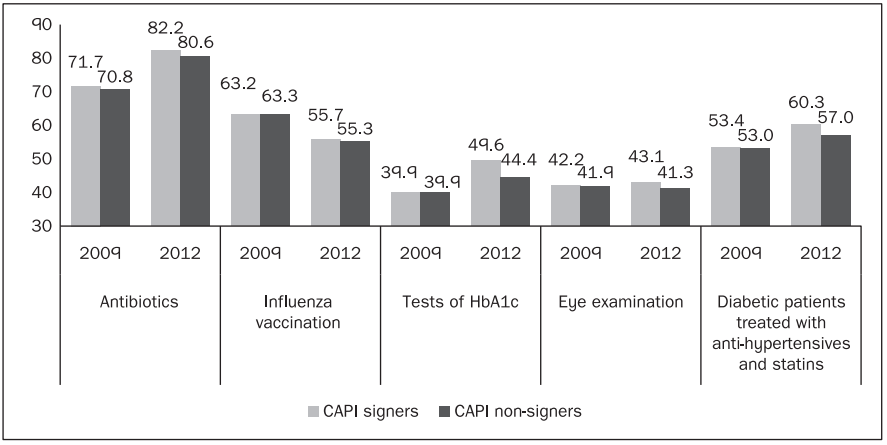


Figure 8.1 Achievement rates for quality indicators for CAPI signers and non-signers in France between March 2009 and 2012

Source: Caisse Nationale d'Assurance Maladie (2013).

Although greater improvements were observed for most of the indicators in the group of CAPI signers, differences between the two groups remained modest (see Figure 8.1). The exception was diabetes disease management indicators, for which improvement was greater in the group of CAPI signers. The indicator related to HbA1c tests showed a nine percentage point increase in compliance for CAPI physicians compared with only a four percentage point increase among non-CAPI physicians. Nonetheless, the differences between the two groups, albeit small for some indicators, were all significant at the one per cent threshold. A trend of improvement is observed in the two groups for nearly all of the indicators.

The evaluation of ROSP (Caisse Nationale d'Assurance Maladie, 2013)

The NHIF also compiled data from 2011 and 2012, which showed that all indicators recorded under the ROSP improved in comparison with the previous year. Nonetheless, both the clinical and prevention domains showed mixed results. While some indicators improved importantly (e.g. prescription of statins for high risk patients, proportion of patients older than 65 treated with vasodilators during the year), there is still room for improvement in a number of areas of care. For instance, despite the substantial increase of HbA1c tests, only half of diabetes patients receive the appropriate number of blood monitoring tests, which is below the 65 per cent target set by the NHIF (see Table 8.2). Moreover, although generic prescribing has increased in share, improvements remain insufficient: as of December 2012, the share of generic prescriptions in all statin prescriptions stood at 53 per cent against the 70 per cent target set within ROSP.

Table 8.2 Sample of results of the France ROSP in December 2011–12 against objectives set by the National Health Insurance Fund

	Indicator	*Target*	*Dec. 2011*	*Dec. 2012*	*Progress in points*
Flu vaccination	Proportion of patients 65 and older with flu immunization.	>= 75%	57.8%	56.4%	–1.4
Cervical cancer screening	Proportion of women aged between 25 and 65 with at least one pap smear during the last three years.	>= 80%	58.7%	57.5%	–1.2
Diabetes	Proportion of diabetic patients who had three or four tests of HbA1c during the last year.	>= 65%	45.9%	48.7%	2.7
Diabetes	Proportion of type II diabetic patients whose HbA1c test <8,5%.	>= 90%	Na	85%	Na
Diabetes	Proportion of type II diabetic patients whose LDL cholesterol test <1.3g/l.	>= 80%	Na	74%	Na
Antibiotics	% generic antibiotics dispensed / total antibiotics dispensed (number of items).	>= 90%	78.6%	80.9%	2.3
Statins	% generic statins dispensed / total statins dispensed (number of items).	>= 70%	38.2%	53.8%	15.6

Source: Caisse Nationale d'Assurance Maladie, 2013.

Information on practice organization was not available for the first year of ROSP. According to the NHIF, however, physicians have responded that ROSP gave them the opportunity to upgrade their computer equipment and software. As of December 2013, 73 per cent of French GPs had an electronic medical patient file system consistent with HAS recommendations, and 71 per cent were able to provide a yearly synthesis report of individual patient records. However, only 64 per cent had installed certified software to assist prescription.

Costs and savings

Since the new National Agreement, 75,444 physicians (GPs and specialists) have received bonuses for their first year of participation in the programme. The main reason for not receiving bonuses was that some physicians treated too few patients to calculate the indicator. For GPs who act as the gatekeeping doctors for more than 200 patients, the yearly bonuses received from ROSP amounted to € 5365, which represents about 5–7 per cent of their annual income (Caisse Nationale d'Assurance Maladie, 2013).

At the national level, the NHIF spent approximately €250 M for the ROSP in 2012 (Caisse Nationale d'Assurance Maladie, 2013). Initially, the NHIF intended to make the programme cost-neutral by offsetting the costs of the incentive

payments and programme administration with savings generated by the replacement of branded medicine by generic prescribing. Before the launch of CAPI, simulations had shown that a limited improvement in generic prescribing could contribute to financing the programme. Evaluation of this part of the P4P programme has proved to be challenging, however, as other programmes providing incentives for generic prescribing were also implemented in recent years, and the list of generic medication has changed significantly since implementation.

Provider response

Initially, physicians strongly opposed the idea of linking performance to payment, and the implementation of the CAPI was considered highly controversial and supported by none of the unions of general practitioners. More importantly, the Order of Doctors highly opposed the programme on the basis that it interfered with the principle of independence in prescribing and that it could damage the patient–physician relation (Or, 2010). Moreover, some unions were also concerned that the programme would penalize doctors working in poorer and more difficult neighbourhoods where targets could be harder to achieve (Or, 2010). Finally, the Federation of Medical Unions argued that the traditional use of collective bargaining is a fairer approach to improving clinical practice rather than the individualized nature of the contracts. The French union of pharmaceutical industry (Les Enterprises du Médicament) also opposed the implementation of the programme, asserting that it would 'reduce doctors' liberty to prescribe and will put a brake on innovation, all in the name of improving public health' (Senior, 2009).

Despite the initial strong opposition, implementation of the programme proceeded in the initial years was without major obstacles, and close to 40 per cent of French GPs voluntarily opted for the programme after one year. With the current relative popularity of the programme, the Union of Doctors revised their position and began negotiations to include a P4P pillar in the New National Agreement to be applied to all GPs.

Overall conclusions and lessons learned

Has the programme had enough of an impact on performance improvement to justify its cost?

In April 2013, the NHIF released a first assessment of ROSP that considered that the programme has led to some progress in the quality of care of patients. In the area of disease management, diabetes indicators have shown some improvement. Such results, however, cannot be generalized to all areas of care rewarded under ROSP.

Nonetheless, in only four years of existence, the CAPI and ROSP have achieved considerable progress: from a voluntary programme, the programme was expanded to 97 per cent of GPs treating almost the entire French population;

the range of indicators has been refined and broadened to include more aspects of care; and similar payments are already applied to some specialists. In line with other international experience, modest improvements in processes of care have been recorded for some indicators following the introduction of targeted financial incentives. Moreover, negotiations with unions of doctors and other professional bodies have worked towards successfully integrating a P4P pillar in the new National Agreement; complementing the physician payment model historically based on fee-for-service, and setting ground for future programmes.

Several factors, however, may limit the overall effectiveness and impact of the programme in the future. *The organization of primary care in France relies mainly on solo practice, which does not provide much scope for improving coordination of care*. The design of the incentive payments is adapted to the French organization of primary care practices and ambulatory settings for specialists. While small group practices have developed in the past decade, now reaching almost half of the primary care practices, the cooperation between physicians often remains limited to shared accommodations. Within this framework, the role of ROSP in supporting care coordination may appear limited. The first assessment of ROSP, however, showed a positive impact of financial incentives on the practice organization as far as new communication technologies and the computerized patient file are concerned.

Coordination of care is also being addressed under a separate reform initiative within the 2008 Finance Law for Social Security, and in the experiment with new payment systems Expérimentation de Nouveaux Modes de Rémunération (ENMR), which started in January 2010 in six regions in France. ENMR is designed for medical homes and medical centres contracted by regional health agencies. An evaluation is currently underway to assess the performance of these organizations against individual practices.

Data collection provides limited scope to assess patient health. Recent studies have shown that the use of different source of data (claims versus patient file) do not have significant impact on the way data is reported (Van Herck et al., 2010; Eijkenaar, 2011, 2012). Nevertheless, the move towards more outcome-oriented indicators will require access to clinical data that is not available in claims. At the moment, French physicians do not have the capacity to provide individual clinical data in a consistent and systematic manner. While equipment of practices with certified electronic patient management software is being encouraged within ROSP, it is not expected that such data would be used to compile performance indicators in the foreseeable future. In particular, data on patient health status are an important missing piece. Outcome indicators currently used in ROSP mainly rely on self-reporting by providers, with no data verification process.

Although the basis for ROSP was developed four years ago, the programme is still at a relatively experimental stage. The NHIF has worked towards the development of other interventions in the area of management of chronic conditions and promotion of quality standards in primary care. ROSP is now supported and complemented by other initiatives, such as a diabetes disease management programme in place since 2009, and financial incentives for patients and pharmacists to support the use of generic drugs since 2012.

Working towards better integration of such initiatives and mixing financial and non-financial incentives in a consistent way should be one of the priorities of the French national health insurance system.

Notes

1 Arrêté du 22 septembre 2011 portant approbation de la convention nationale des médecins généralistes et spécialistes (http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000024803740&dateTexte=&categorieLien=id).
2 Cardiologists and gastroenterologists have their own sets of indicators since 2013. The other specialists are marginally involved in the programme: they are assessed against only four performance indicators of the organizational domain.
3 Of which 53% of GPs and 47% of specialists.
4 Each physician can create an account on the website of the French National Insurance Fund: *ameli.fr*.

References

Aubert, J. M., and Polton, D. (2009) *Le contrat d'amélioration des pratiques individuelles: un élément d'une stratégie d'efficience*. Presentation prepared for the conference 'Une révolution dans les relations conventionnelles: le contrat individuel rémunéré médecin-caisse', Paris, 4 March.

Caisse Nationale d'Assurance Maladie (2010) *Contrat d'Amélioration des Pratiques Individuelles (CAPI): une dynamique au bénéfice des patients* (http://www.ameli.fr/fileadmin/user_upload/documents/Dp_capi_16_09_2010_vdef.pdf, accessed January 2011).

Caisse Nationale d'Assurance Maladie (2013) *Rémunération des objectifs de santé publique: une mobilisation des médecins et de l'assurance maladie en faveur de la qualité des soins* (http://www.ameli.fr/fileadmin/user_upload/documents/DP_Bilan_ROSP_1_an_11042013_VDEF3.pdf, accessed June 2013).

Chevreul, K. et al. (2010) *Faut-il rendre publics les résultats des hôpitaux? Données scientifiques*. Presentation prepared for the 20èmes journées européennes de la Société Française de Cardiologie, Paris, 13–16 January.

Christianson, J., Leatherman, S. and Sutherland, K. (2007) *Paying for quality: understanding and assessing physician pay-for-performance initiatives*. Princeton, NJ: Robert Wood Johnson Foundation.

Clavreul L. (2012) Les médecins se convertissent au paiement à la performance, *Le Monde*, 2 January.

Degos, L. et al. (2008) Can France keep its patients happy? *British Medical Journal*, 336: 254–7.

Durand-Zalesky, I. (2008) The French health care system, in T. C. Fund, *Descriptions of health care systems: Denmark, France, Germany, the Netherlands, Sweden and the United Kingdom* (pp. 4–6). Washington, DC: The Commonwealth Fund.

Eijkenaar, F. (2011) Key issues in the design of pay for performance programs, *European Journal of Health Economics*, 14(1): 117–31.

Eijkenaar, F. (2012) Pay for performance in health care: an international overview of initiatives, *Medical Care Research and Review*, 69(3): 251–76.

Galvin, R. (2006) Pay-for-performance: too much of a good thing? A conversation with Martin Roland, *Health Affairs*, 25(5): 412–19.

Haute Autorité de Santé (n.d.) *About HAS* (http://www.has-sante.fr/portail/jcms/c_5443/english, accessed January 2011).

Martin, Jenkins and Associates Limited (2008) *Evaluation of the PHO peformance programme: final report*. Auckland: Martin, Jenkins and Associates Limited.

Or, Z. (2010) P4P for generalists: first results, *Health Policy Monitor* (http://www.hpm.org/survey/fr/a16/3, accessed January 2011).

Rodwin, V. (2003) The health care system under French national health insurance: lessons for health reform in the United States, *American Journal of Public Health*, 93(1): 31–7.

Sandier, S., Paris, V. and Polton, D. (2004) *Health systems in transition: France*. Copenhagen: WHO Regional Office for Europe on behalf of the European Observatory on Health Systems and Policies.

Schoen, C. et al. (2009) In chronic conditions: experiences of patients with complex health care needs in eight countries, 2008, *Health Affairs*, 28(1): 1–16.

Senior, M. (2009) Pay-for-performance hits France as part of cost-cutting measures, *Europharmatoday*, 12 October.

Sutton, M. and McLean, G. (2006) Determinants of primary medical care quality measured under the new UK contract: cross sectional study, *British Medical Journal*, 332: 389–90.

Van Herck, P. et al. (2010) Systematic review: effects, design choices, and context of pay-for-performance in health care, *BMC Health Services Research*, 10: 247.

World Health Organization (2000) *World health report 2000: health systems: improving performance*. Geneva: World Health Organization.

chapter nine

Germany: Disease management programmes

Y-Ling Chi

Introduction

In 2000 the World Health Organization ranked Germany's health care system as the twenty-fifth best performer in the world. This was considered to be a disappointing result, given that Germany also was ranked the third largest health spender among OECD countries, with more than 10 per cent of GDP spent on health in the same year (see Figure 9.1).

The results of the World Health Report (2000) were echoed by studies carried out by the German Advisory Council for the Concerted Action in Health Care,[1] which raised concerns over the efficiency and quality of the care delivered, especially in the area of prevention, diagnosis and management of chronic conditions and breast cancer. Inefficient and inadequate quality of care for chronic conditions was partly attributed to the increasing fragmentation of health care, especially the strict separation between inpatient care and primary and ambulatory care. Moreover, despite the development of clinical guidelines and protocols to manage chronic conditions, there was in practice no incentive for doctors to systematically implement and follow such guidelines, resulting in large inefficiencies and variations in quality (Busse, 2004). Diabetes was a particular area of increasing concern, as low quality of care usually translates into expensive hospitalization, development of complications and co-morbid conditions such as cardiovascular disease, and in turn higher mortality rates.

In addition, the structure and funding of the German Statutory Health Insurance system (SHI) created incentives for the insurers, sickness funds, to avoid patients with chronic conditions. Individuals are free to choose their sickness fund, and a risk-adjustment mechanism altered payment rates to funds based on the risk profile of their enrolled population. The risk adjustment was based on average spending by age and sex, but the higher cost of individuals with chronic conditions was not taken into account. Sickness funds therefore

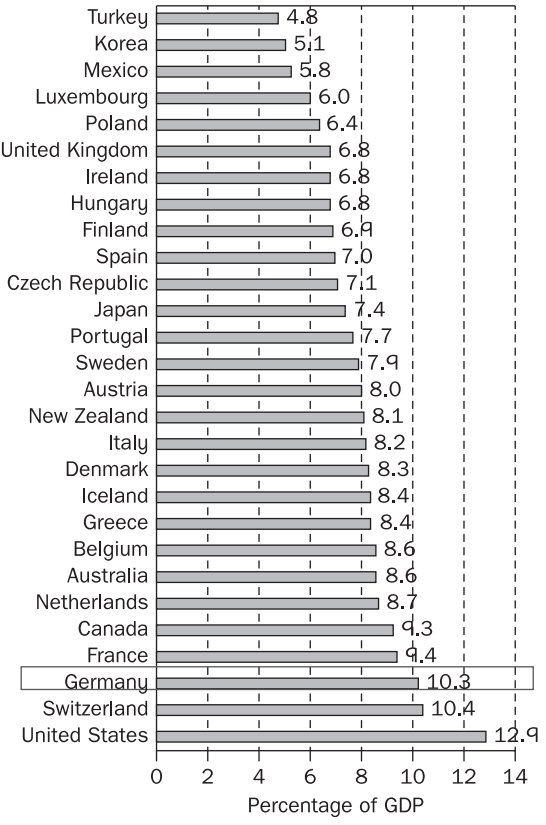


Figure 9.1 Health expenditure as a percentage of gross domestic product in Germany, 1998

Source: OECD, 2001.

had no incentive to try to develop specific initiatives targeted to the chronically ill (Busse, 2004).

To address these concerns, a number of changes were introduced between 2000 and 2002 to improve risk equalization in order to make it less costly for sickness funds to enrol individuals with chronic illness, and to give new opportunities and incentives for better care management (Busse, 2004; Szecsenyi et al., 2008). Disease Management Programmes (DMPs) were introduced in 2002 through legislation that mandates a national roll-out of DMPs to improve coordination and enhance quality of care for the chronically ill (Stock et al., 2011). Since that time, DMPs have been implemented to place primary care physicians as care coordinators for patients with chronic conditions, using financial incentives to reward better care quality.

Health policy context

What were the issues that the programme was designed to address?

Health system context

When the principle of free choice of sickness funds was introduced in 1996, the Risk Compensation Structure (RCS) was created to provide funding to sickness funds on a per capita basis adjusted mainly by age and sex. Thus, prior to the introduction of DMPs, chronic conditions as risk factors were not adequately accounted for in the payment mechanism, making such patients unattractive, as they were often associated with high-risk profiles and high consumption of medical services. Providing higher quality of care could put sickness funds at a disadvantage, because they would then possibly attract even more high-cost patients with chronic conditions. As a result, funds were concerned that more patients with chronic conditions would enrol in their pool of insurees.

DMPs were implemented after a series of reforms to the SHI and experiments to improve care coordination in the area of chronic conditions. Prior to DMPs, the 2000 SHI Reform prepared the ground for integrated care by establishing a Coordinating Committee to improve cooperation between ambulatory physicians and hospitals and allowing pilot projects for integrated care. Relatively few integrated care projects were implemented, however, as many legal, tax and organizational obstacles rendered contracting processes too lengthy and complicated. The 2000 SHI reform was considered too marginal to adequately address a broader health system financing problem (Busse, 2004).

Following discussion and debates within the government coalition, legislation for a more ambitious reform – including DMPs – was successfully passed and integrated into the Social Code in 2001. The goals of DMP were specified as follows:

- Enhance access to treatment and care for patients with chronic condition over the entire course of their lives.
- Successfully implement clinical guidelines to support physician practice of care in primary care setting.
- Create networks of supporting physicians across different levels in the health sectors, in particular enhance coordination of care between primary and ambulatory care.
- Provide financial support to sickness funds and encourage innovations to prevent, diagnose and care for chronic conditions.

Although cost savings was not stated explicitly as an initial goal, the rationale behind the definition of DMPs was to improve quality of care and in turn reduce costly complications, unplanned hospitalization and costs related to treatment and rehabilitation of complicated conditions. According to initial calculations from the IGES Institute for Health and Social Research,[2] the introduction of DMPs would avoid more than 5000 complications yearly, mainly in cardio and cerebro-vascular diseases (Figure 9.2).

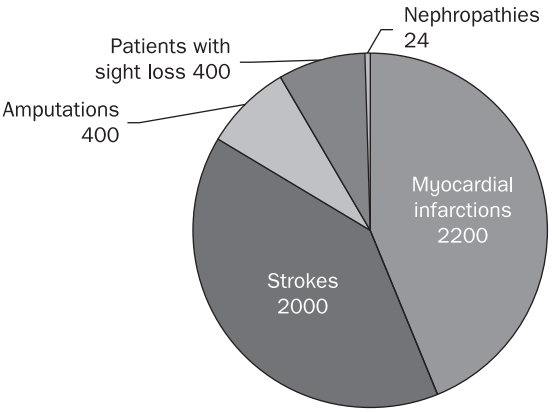


Figure 9.2 Estimated number of complications avoided yearly by the implementation of the Diabetes DMP in Germany

Source: Adapted from figures from Busse, 2004.

Stakeholder involvement

The Joint Federal Committee proposed the first four conditions for DMPs: Diabetes (type I and type II), breast cancer, obstructive pulmonary disease (asthma and COPD), and coronary heart disease. The Joint Federal Committee includes representatives from sickness funds, the Federal Association of SHI-Accredited Physicians, and the German Hospital Organization. The overarching institutional framework for DMPs was composed of national, federal and local actors. Figure 9.3 provides an overview of the overarching institutional arrangements for the DMPs.

The stewards of DMP include the Ministry of Health (MOH), Coordinating Committee, which was replaced by the Joint Federal Committee by 2004, and sickness funds. Disease-specific committees are created for each disease area composed of experts from universities and boards of medical associations. The disease-specific committees draft programme requirements based on evidence-based guidelines for each condition. Recommendations are then endorsed and adopted by the MOH, which issues local decrees that provide broad guidelines on the organization of DMPs, and serve as the basis for contracts between sickness funds and providers (Stock et al., 2011). Based on these recommendations, sickness funds define the care packages for each condition, which are then approved by the Federal Insurance Agency.

In some sickness funds and regions, individual physicians are highly engaged in the process of developing guidelines for DMPs. At the national level, the National Association of Physicians (Bundesärztekammer), the National Association of SHI Physicians, and the consortium of German Scientific Medical Associations were charged with drawing up 'national management guidelines' (Stock et al., 2011). The methodology for developing the guidelines is overseen by the Agency for Quality in Medicine, and it is a consensus process based on national and international literature on evidence-based medicine.

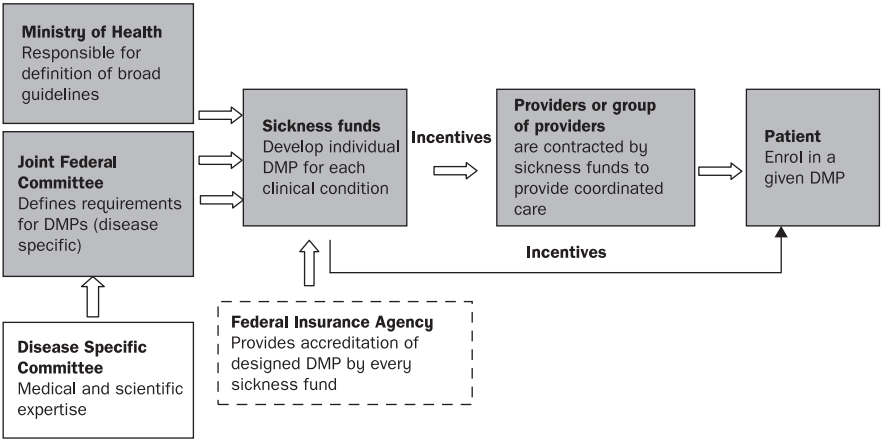


Figure 9.3 Summary of the DMP institutional framework in Germany

The Institute for Quality and Efficiency in Health Care also regularly checks the recommendations in the programmes against international norms (Stock et al., 2011).

Technical design

How does the programme work?

Sickness funds are free to design their own DMPs, but according to the law they must include the following elements: definition of enrolment criteria and enrolment process; treatment according to evidence-based care recommendations or best available evidence; quality assurance (e.g. feedback to physicians, patient reminders for preventive care, and peer review); physician and patient education; documentation in an electronic medical record; and evaluation (Stock et al., 2011). There are differences in the organizational features of DMPs across regions, as sickness funds individually define the organizational arrangements and implementation of DMPs (Stock et al., 2011). Example of the design of DMPs in two regions is presented in Box 9.1.

Sickness funds have a large degree of autonomy in the definition of contracting (including remuneration) of doctors. The Joint Federal Committee, assisted by the disease-specific committees, only defines broad clinical guidelines on the content of care packages. Rather than applying standardized measures and indicators across all sickness funds in all regions, the MOH has provided sickness funds with the necessary margin for manoeuvre to best organize care delivery to meet contract requirements and targets. This flexibility in implementation of DMPs is closely monitored by the Federal Insurance Agency, which has the mandate to validate all DMPs defined by sickness funds.

Sickness funds mostly contract directly with the regional Association of SHI Physicians and individual hospitals, which then in turn enrol voluntary

Box 9.1 Examples of DMPs: Rhineland and Hamburg regions

In the Rhineland and Hamburg regions, sickness funds collectively contract directly with voluntary physicians for each disease area to act as coordinating doctors.

According to guidelines issued by the Federal Joint Committee, the tasks of the coordinating doctor are as follows:

1. Referral to assigned specialist, coordination with the supporting contracted doctors (specialist diabetologist or dietician providing outpatient care).
2. Information, counselling and enrolment of the insured in accordance with clinical guidelines.
3. Collection of information about patient health status into a unique patient file shared between the assigned physicians.
4. Compliance with quality standards and clinical guidelines, especially with regards to cost-effective treatment choices.
5. Coordination of examinations and lab test performed for each patient.
6. Review of multiple treatment of each patient (especially in the case of co-morbidities) – in order to avoid treatment interaction and drug averse events.
7. Provision of documentation prepared for the purpose of patient education.

These guidelines are further translated into indicators used to pay coordinating doctors upon enrollment of patients in DMPs, and also on provision of specific services (example of such payments is provided below).

Documentation and coordination (per quarter)	
Information, advice, registration and preparation of initial documentation.	€25.00
Drafting of follow-up documentation.	€10.00
Follow-up of patients	
Care continuity and treatment of patients with diabetes type 2 (per quarter).	€22.50
Remuneration of additional services	
Comprehensive consultation for the diagnosis of diabetic neuropathy.	€38.35
Care of diabetic foot lesions per foot.	€16.70
Referral to nephrologists.	€2.05
Documentation of the ocular test.	€5.11
Remuneration of training	
Treatment and training programme of patients not on insulin therapy (four sessions with up to four patients within four weeks).	€25 per course per patient (max. €100 per patient)

Supporting materials for the training (without diabetes pass).	€9.00
Treatment and education programme for patients with hypertension.	€25.00 (max. per patient €100.00)

physicians in a network of supporting doctors for each disease area. In some instances, sickness funds contract directly with individual physicians, although this contracting arrangement is fairly rare. As part of this network, physicians are required to ensure continuity of care, provide patient education (treatment self-management and health lifestyles), and implement relevant clinical guidelines, both with regard to diagnosis and medical care of patients. These three guiding principles are issued by the MOH and the Joint Federal Committee; and further translated in a contractual arrangement between sickness funds, physicians, hospitals and sometimes rehabilitation clinics, which might vary between regions.

Sickness funds receive incentive payments for establishing DMPs and enrolling patients, and in turn provide incentive payments to physicians. Physicians participating in DMPs receive incentives in the form of reimbursement for additional services and materials such as documentation, patient education, and coordination of care. Some sickness funds offer incentives in the form of waived co-payments to patients.

Performance domains

DMPs aim to improve care coordination for chronic conditions and diseases that are highly prevalent and for which there are gaps in care (Stock et al., 2011). DMPs now encompass six large areas of chronic illness:

- diabetes – type I and II
- asthma
- chronic obstructive pulmonary disease
- coronary heart disease
- breast cancer.

Incentive payments to sickness funds

The incentives for sickness funds to offer DMPs to their enrolees have gone through significant changes with the evolving payment mechanism to sickness funds (Figure 9.4). Prior to the 2002 law initiating DMPs, the risk-adjustment formula of the RCS did not include health status of patients, but only age and sex of the patients. With the introduction of DMPs, sickness funds received a higher payment for patients diagnosed with a chronic condition and enrolled in a DMP. Special RCS groups were created for enrolled patients in the DMP clinical areas, for which standardized average costs were calculated and used to transfer higher payments to sickness funds.

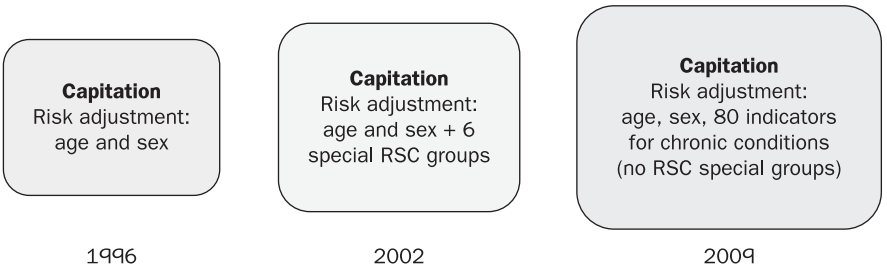


Figure 9.4 Changes in payment mechanism for sickness funds in Germany

In January 2009, a new reform of the RCS was introduced that added more morbidity-related risk factors beyond age and sex and DMP enrolment. The RCS is now composed of 80 indicators covering 80 morbidity groups and adjusted by age and sex, independent of patient participation to a DMP. Consequently, specific financial incentives for enrolling patients in DMPs were weakened. Sickness funds now only receive a flat-rate administrative fee of €152 for each patient enrolled in a DMP, which has been reduced yearly from €180 in 2010 (National Association of Statutory Health Insurance Funds, 2012).

Incentives for health care professionals

Physicians participate in DMPs on a voluntary basis, and receive financial incentives to encourage their participation in the form of additional payment for DMP-related services (see Box 9.1 for an example). For the care of each diabetes patient the physician receives a lump-sum payment of €15 per quarter in addition to the regular reimbursement for specific services. For referral of a patient to a diabetes specialist, he or she receives €5.11 per case (Stock et al., 2011). These incentives payments vary from regions to regions and can be quite high.

Patient incentives

Some sickness funds offer incentives to patients to enrol in DMPs in the form of waived practice fees and co-payments. Prior to the recent reform of practice fees in January 2013, patients were required to pay €10 per quarter (Nolte et al., 2008). A patient incentive used by sickness funds was to exempt DMP enrolees from this practice fee. Moreover, sickness funds also may reward participation in a DMP by a reduction of co-payments for some services and medicine. According to Stock et al. (2011), exemptions for those enrolled can be substantial, as payments for drugs, hospitalization, and physical therapy are very frequent for patients with chronic illnesses. There is no research on the impact of patient incentives in DMPs, however.

Reach of the programme

Which providers participate and how many people are covered?

As of January 2012, there were 10,618 DMPs implemented across all sickness funds and disease areas (Figure 9.5). The clinical area with the most DMPs in operation is coronary heart disease, followed by diabetes and asthma.

The Federal Insurance Agency compiles information on population coverage of DMPs. As of 2012, there were seven million participants in DMPs with six million people covered, as some individuals are enrolled in multiple DMPs

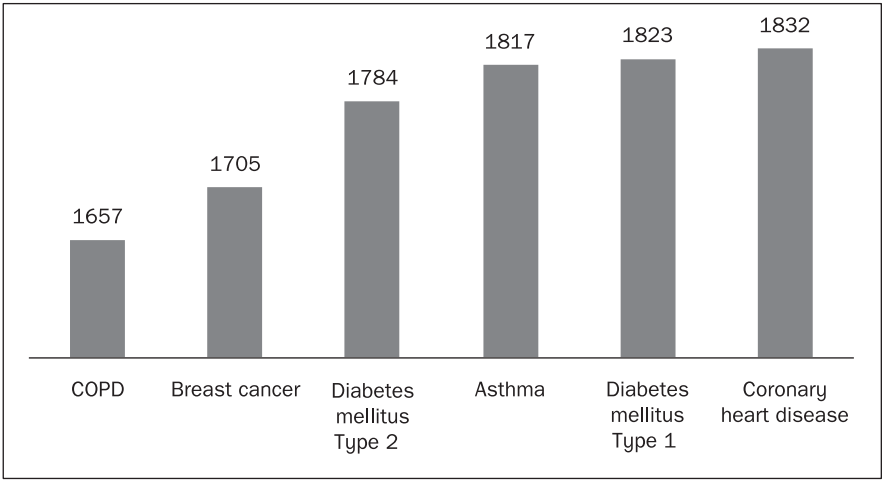


Figure 9.5 Number of DMPs developed by sickness funds by clinical condition in Germany

Source: Federal Insurance Agency, January 2012.

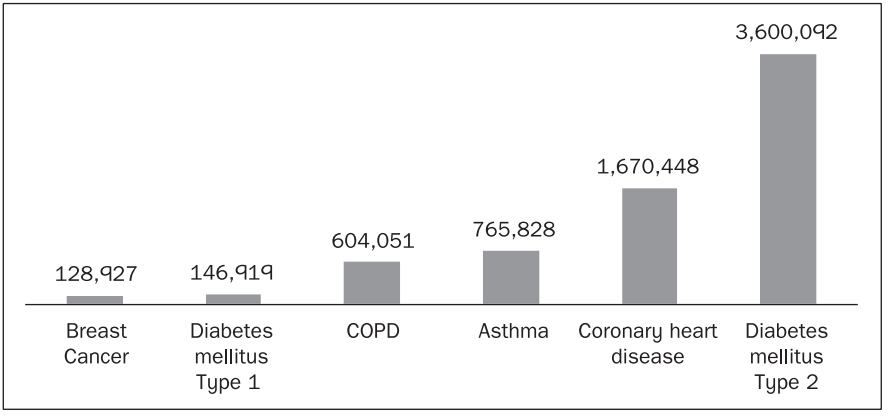


Figure 9.6 Number of DMP enrollees for each clinical condition in Germany

Source: Federal Insurance Agency, January 2012.

(see Figure 9.6 for distribution of number of enrolled by clinical condition). Programmes for type II diabetes have the most enrollees (more than 3.6 million), which is estimated to represent 70 per cent of all diabetic patients (Shaw et al., 2010; Federal Insurance Agency, 2012).

Participation in DMPs is voluntary for providers, and physicians are usually directly contracted by medical associations through individual contracts. There is only limited information on the share of providers participating in DMPs at the national level, but some regional information is available. According to Altenhofen et al. (2004), in the North Rhine region, for example, over 70 per cent of ambulatory physicians participate in a DMP.

Improvement process

How is the programme leveraged to drive service delivery improvement?

DMPs drive service delivery improvements for chronic disease management largely by providing sickness funds with tools to monitor care for chronic conditions and fostering fair competition between sickness funds to attract and serve individuals with chronic conditions. In 1996, reform of the SHI provided patients with the right to freely choose sickness funds, which created a competitive insurance market in Germany. In order to enrol as many patients as possible, sickness funds need to provide good quality care packages for all patients, sometimes including incentives to patients (such as reduction in fees, or exemptions to co-payment of certain services and medicine) and to contracted coordinating doctors.

In addition, dataflow generated by the programme was also a key feature of the programmes. The Federal Insurance Agency collects clinical and financial indicators of performance and sends it back to individual sickness funds. The implementation of electronic tools specifically designed to manage patients with chronic conditions alongside DMPs was reported to have benefited patients. For instance, in their review of the programme, Linder et al. (2011) found that doctors reported that reminders sent by insurance funds for patient monitoring purposes were very useful.

Results of the programme

Has the programme had an impact on performance, and have there been any unintended consequences?

Programme evaluation

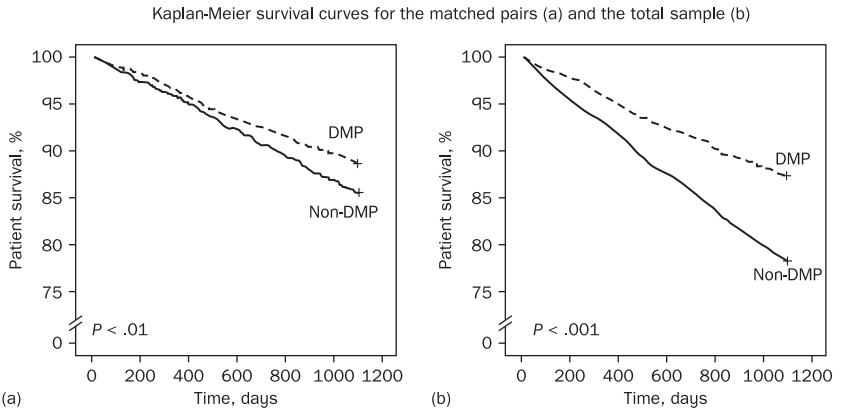
Evaluation of DMPs is mandatory and stipulated by law. Results for each disease area must be analysed for each region, insurance carrier, and patient cohort. Nevertheless, in the absence of a control group, it has been difficult to attribute changes in processes of care and outcomes to the programmes (Van Lente, 2012).

A large body of literature using surveys completed by sickness funds and independent research provide more robust methodologies and evidence. These studies point to some positive results related to processes of care and patient satisfaction. A survey carried out by one of the largest sickness funds (AOK) showed that patient satisfaction was high among individuals enrolled in DMPs. Ninety-nine per cent of patients reported that they spend more time in consultation with their coordinating doctor since they have enrolled in the DMP; 97 per cent understand their disease and treatment much better; and 87 per cent feel that they are in better control of their disease since they have enrolled in a DMP (Schoul & Gniostko, 2009). A review of claims data in North Rhine, North Wurttemberg and Hesse also show that drop-out rates of patients are very low overall across programmes (Fullerton et al., 2012).

Other external research has focused on patient outcomes and processes of care. A cohort study compared 444 patients in a diabetes DMP and 494 patients with diabetes not enrolled in a DMP serving as control group. The study used data collected in a baseline interview and a telephone follow-up after ten months. Results showed significantly better processes of care for DMP participants (Schäfer et al., 2010). Szecsenyi et al. (2008) also found that type II diabetic patients were more likely to receive patient-structured and coordinated care than similar patients not enrolled in DMPs, and that the implementation of DMPs has been followed by a change in services to be oriented more toward patient-centred care. A longitudinal population-based study of over 100,000 DMP participants between 2006 and 2010 in Bavaria also found improvements in quality of care for patients with asthma. The study showed an increase of both patient education (from 4.4 per cent to 23.4 per cent) and utilization of an individual self-management plan (from 40.3 to 69.3 per cent), as well as a reduction in the hospitalization rate (from 2.8 per cent to 0.7 per cent; Mehring et al., 2012).

Research also shows some modest improvement in health outcomes for patients enrolled in a DMP. A large evaluation of a type II diabetes DMP (with 195,225 enrolled patients) showed a reduction in the share of patients with blood sugar levels outside the target range from 8.5 to 7.9 per cent within a six-month period (Altenhofen et al., 2004). The study reported a positive effect on the treatment of hypertension among diabetic patients, which is in line with other external studies such as Berthold et al. (2011). These studies also concluded that the programme improved care provided to patients with diabetes, but came to mixed conclusions when looking at intermediate outcomes such as HbA1c or blood pressure level. Altogether, these results suggest that improvements in process indicators only partially translated into improvement in outcomes.

Finally, two recent evaluations use matched pairs of DMP participants and non-participants to control for possible underlying characteristics of DMP participants that may affect their outcomes independent of participation in the programmes. A nationwide evaluation that examined outcomes for 11,079 diabetic patients (including 1927 matched pairs) found that participation in a DMP was associated with a reduction in hospitalization rates and a lower three-year mortality rate as shown on Figure 9.7 (11.3 per cent vs. 14.4 per cent for non-participants; Miksch et al., 2010). A retrospective observational study of 19,888 matched pairs used routine administration data from Germany's largest



DMP indicates disease management programme.

Figure 9.7 Comparison of mortality rates between diabetes patients enrolled and not enrolled in a DMP in Germany, 2010

Source: Miksch et al., 2010.

sickness fund to compare outcomes and costs for diabetes DMP participants vs. non-participants (Drabik et al., 2012). The study found that participation in a DMP was associated with a modest increase in survival time over a three-year period (1045 days vs. 985 days) and lower costs per patient (€122 vs. €169 including DMP administration and service costs.

Costs and savings

In 2012 the German SHI system spent a total of €920 million on all DMP programmes, with an average expenditure of €153 per enrolled patient. About 52 per cent of the expenditure is allocated to fees paid to physicians for DMP-related services such as coordination and documentation, about 26 per cent is paid to physicians for patient education, and 22 per cent is paid to sickness funds for administration and data management (Van Lente, 2012).

Some studies have shown lower costs for patients enrolled in DMPs. Germany's largest insurer AOK reports net cost savings ranging from 8-15 per cent of total annual costs of care for enrolees with chronic conditions (Stock et al., 2011). More rigorous studies find even larger estimated cost savings when controlling for underlying characteristics of participants in DMPs vs. non-participants (Drabik et al., 2012). Linder et al. (2011) report higher costs of implementation of DMPs and question the extent to which the benefits of the programme fully justify the high operational costs.

Provider response

The implementation of DMPs was initially fiercely opposed by medical associations and physicians. As the first contract was about to be signed, a national assembly of all regional physicians associations passed a motion to

block the regional physician associations from entering into DMP contracts (Busse, 2004). The arguments against the DMPs included uncertainty about whether the law would be repealed after recent regulations, as well as concerns about the quality of clinical guidelines and possible misuse of patient data for financial advantage of the sickness funds. The opposition of the physician groups was overcome after the election and with some minor modifications to data requirements for DMPs (Busse, 2004). Since then, operation of DMPs has been without major obstacles. The response and satisfaction of providers after more than ten years of implementation, however, has not been studied.

Overall conclusions and lessons learned

Has the programme had enough impact on improvement to justify its cost?

Building new and efficient programmes to address the rise of chronic illness by improving care management is the challenge facing all health systems across the OECD and to a certain extent in the world. Initially, DMPs were introduced to compensate insurers and health care providers for care to higher need enrolees with chronic conditions and to create financial incentives targeted to physicians to follow evidence-based clinical guidelines. The direct financial incentive for sickness funds to enrol patients in DMPs was reduced with the 2009 reform, under which insurers are now compensated for morbidity-related risk factors independent of whether or not the patient is enrolled in a DMP. Sickness funds now only receive a flat-rate fee for each DMP enrolee to compensate for additional administrative processing of DMPs. This compensation fee is far from representing the real cost of DMPs.

A large body of external reviews has pointed to positive improvements in certain aspects of care and outcomes following the introduction of DMPs. These studies almost unanimously show improvement in care processes and high satisfaction rates of DMP enrolees, but they provide mixed results when looking at patient outcomes. This is consistent with the international research on disease management programmes, which show statistically significant but clinically modest effects of the programmes on outcomes (Mattke et al., 2007; Pimouguet et al., 2011). It is also important to note that DMPs were introduced within a broader range of initiatives targeted to improve quality of chronic disease management, such as the development of comprehensive clinical guidelines, standard referral processes, and ongoing quality assurance through definition of care standards and process indicators.

The role that financial incentives play in the results achieved by Germany's DMPs is difficult to isolate. Financial incentives may have more to do with a better match of payment to providers with the services needed to effectively manage chronic conditions. External evaluations also only partly address the question of bias in enrolment in a DMP, which could affect study results. Moreover, most of these studies only look at one area of care (diabetes) and fail to interpret results in other areas of care for which DMPs might not be as

efficient. Another issue is the heavy bureaucratic and administrative system supporting the implementation and operation of DMPs (Linder et al., 2011). In spite of these limitations, some key lessons from the German DMP experience include the following:

1. Nationwide standards of care according to evidence-based guidelines, combined with a strong emphasis on quality assurance and primary care doctors as leaders in the process can form the backbone of better processes of care and outcomes for chronic conditions.
2. Aligning the incentives of the underlying payment mechanisms with the services and processes of care recommended in evidence-based guidelines makes it possible and more attractive for providers to follow clinical guidelines.
3. Engaging patients in the management of their conditions through financial incentives, patient education and self-management plans may be a particularly critical aspect of disease management programmes.

Notes

1 German name of the commission: *Sachverständigenrat für die Konzertierte Aktion im Gesundheitswesen*.
2 Institut für Gesundheits und Sozialforschung.

References

Advisory Council for the Concerted Action in Health Care (2001) *Appropriateness and efficiency: over-, under-, and misuse, report 2000/2001*, Vol. 3. Bonn: Federal Ministry of Health.

Altenhofen, L. et al. (2004) *Report on quality assurance – disease – management programs in North-Rhine* [in German]. Duesseldorf: KV Nordrhein (http://www.kvno.de/importiert/qualbe_dmp04.pdf, accessed July 2011).

Berthold, H.K. et al. (2011) Disease management programs in type 2 diabetes: quality of care, *American Journal of Managed Care*, 17(6): 393–403.

Busse, R. (2004) Disease management programs In Germany's statutory health insurance system, *Health Affairs*, 23(3): 56–67.

Drabik, A. et al. (2012) Life prolonging of disease management programs in patients with type 2 diabetes is cost-effective. Diabetes, *Diabetes Research and Clinical* Practice, 95(2): 194–200.

Eichenlaub, A. et al. (2004) Evaluation des Modellprogramms 'Diabetes Sachsen-Anhalt'. Köln: AOK Bundesverbandes und des IKK Bundesverbandes.

Federal Insurance Agency (2012) *Data on approval of disease management programs (DMP) by the Federal Insurance Office (BVA)* (http://www.bundesversicherungsamt.de/nn_1046154/DE/DMP/dmp__inhalt.html#doc1046158bodyText8, accessed April 2012).

Fullerton, B. et al. (2012) Predictors of dropout in the German disease management program for type 2 diabetes, *BMC Health Service Research*, 12(8).

KV Nordrhein – nordrheinische Krankenkassen-verbände (2010) *DMP-Vertrag Diabetes mellitus Typ 2 ab 01.01.2010* (http://www.kvno.de/downloads/dmp_diab2_vertrag.pdf, accessed June 2011).

Linder, R. et al. (2011) The benefit and efficiency of the disease management program for type 2 diabetes, *Deutsches Ärzteblatt international*, 108(10): 155–62.

Mattke, S., Seid, M. and Ma, S. (2007) Evidence for the effect of disease management: is $1 billion a year a good investment?, *American Journal of Managed Care*, 13: 670–6.

Mehring, M. et al. (2012) Disease management programs for patients with asthma in Germany: a longitudinal population-based study, *Respiratory Care*.

Miksch, A. et al. (2010) Is there a survival benefit within a German primary care-based disease management program?, *American Journal of Managed Care*, 16(1): 49–54.

National Association of Statutory Health Insurance Funds (2012), DMP-Programmkosten-pauschale (http://www.gkv-spitzenverband.de/DMP_Programmkosten.gkvnet, accessed December 2012).

Nolte, E., Knai, C. and McKee, M. (2008) *Managing chronic conditions. Experience in eight countries*. Copenhagen: European Observatory on Health Systems and Policies and WHO Regional Office for Europe.

OECD (2001) *Health at a glance*. Paris: Organization for Economic Co-operation and Development.

Pimouguet, C. et al. (2011) Effectiveness of disease-management programs for improving diabetes care: a meta-analysis, *Canadian Medical Association Journal*, 183(2): E115–27.

Schäfer, I. et al. (2010) The disease management program for type 2 diabetes in Germany enhances process quality of diabetes care – a follow-up survey of patient's experiences, *BMC Health Services Research*, 10(55).

Schoul, A.K. and Gniostko, C. (2009) *AOK Bundesverband: evaluation AOK Curaplan COPD* (http://www.aok-gesundheitspartner.de/imperia/md/gpp/bund/dmp/evaluation/befragung/auswahl_copd_ergebnisse_2009.pdf, accessed May 2013).

Shaw, J.E., Sicree, R.A. and Zimmet, P.Z. (2010) Global estimates of the prevalence of diabetes for 2010 and 2030, *Diabetes Research and Clinical Practice*, (87)1: 4–14.

Stock, S. et al. (2010) *German diabetes management programs improve quality of care and curb costs*, *Health Affairs*, 29(12): 2197–205.

Stock, S. et al. (2011) Disease-management programs can improve quality of care for the chronically ill, even in a weak primary care system: a case study from Germany, *The Commonwealth Fund, Issues in International Health Policy*, November.

Szecsenyi, J. et al. (2008) German diabetes disease management programs are appropriate for restructuring care according to the chronic care model: an evaluation with the patient assessment of chronic illness care instrument, *Diabetes Care*, 31(6): 1150–4.

Van Lente, E. (2012) 10 years experience with disease management programs (DMP) in German social health insurance: Presentation at ESIP Conference, Brussels, 23 May.

chapter ten

New Zealand: Primary health organization performance programme*

Cheryl Cashin

Introduction

Primary health care is the cornerstone of the health care system in New Zealand and has a long history of being at the centre of structural, and at times ideological, reforms. An unsuccessful attempt in the early 1990s to create a market-based system of competing purchasers and providers was followed, after the 1996 election, by the creation of a single national health care purchaser. In 1999, the political pendulum swung again with a new Labour-led coalition government that distanced itself from market-based approaches and initiated a new radical reform of primary health care (PHC) that moved toward greater control and financing by the government. General practitioners (GPs) have consistently maintained their independence to operate as private practices and the right to charge patients fees for their services despite these numerous fundamental reforms and swings of the political pendulum.

Throughout its evolution, PHC in New Zealand traditionally has been funded by a partial fee-for-service payment from the government for consultations and pharmaceuticals, supplemented by substantial co-payments from patients. The high levels of fees and co-payments have been an ongoing political issue in New Zealand, as the social inequalities in GP access are exacerbated by the fee-for-service payment and high co-payments. Despite some targeting of government subsidies to higher need populations, inequalities in access have persisted, with low-income groups and Māori populations often having higher health needs but using services at a lower rate than the rest of the population (Barnett & Barnett, 2004). Fee-for-service has not only created barriers for some high-need patients, but has also provided little incentive for collaborative approaches by GPs or linkages with other parts of the health sector (Barnett & Barnett, 2004).

A New Zealand Health Strategy was introduced in 2000 (Ministry of Health NZ, 2000), with a set of 13 population health priorities and three priorities for reducing specific health inequalities included. Under the umbrella of the New

Zealand Health Strategy, a separate Primary Health Care Strategy introduced population-based approaches to address growing inequalities, with a reduction in ethic health disparities an overarching goal of the strategy. A new organizational structure for service provision, primary health organizations (PHOs), was established to focus on the priority health areas identified in the New Zealand Health Strategy and to address problems of access to services and a lack of coordination between providers.

Under the umbrella of the 2000 New Zealand Health Strategy, a pay for performance programme was introduced in 2006. The PHO Performance Management Programme aimed to sharpen the focus of PHOs on the population health and inequality priorities. This programme is one element of an overall quality framework, and was designed by PHC representatives, District Health Boards (DHBs) and the Ministry of Health (MOH). The aim of the programme is to reinforce the combined health sector efforts to improve the health of enrolled populations and reduce inequalities in health outcomes by supporting clinical governance and rewarding quality improvement within PHOs.

Health policy context

What were the issues that the Programme was designed to address?

Until the 1990s, most GPs in New Zealand operated privately and independently, with little or no coordination between individual GPs. The major health system reforms of the early 1990s were aimed at introducing a market model into the health sector through new contracting arrangements between providers and newly formed government-funded purchasing agencies. In response, GPs organized themselves into Independent Practitioner Associations (IPAs), usually within geographically defined areas, to manage budgets for pharmaceuticals and diagnostic testing, enhance quality of care, and to pool savings to fund other local health initiatives. In 1999, over 80 per cent of GPs were members of IPAs, which ranged in size from six to eight physician members, to about 340 members in a large association in Auckland, Pro Care Health (French, Old & Healey, 2001). These organizations were formed mainly to protect the business interests of GPs, and taking a more population-based approach to primary care was a secondary objective for most IPAs.

When the next major structural reform of the health sector was introduced in 1996, a single purchaser was established (the Health Funding Authority), and IPAs began to consolidate to gather some bargaining power against the new single purchaser. This process of IPA consolidation in response to government reforms was then followed by further health system restructuring in 1999, when the new Labour-led coalition came to power. The Health Funding Authority was abolished and replaced by 21 (now 20) new DHBs to increase local involvement in the health system and improve the equity of financial allocations, and ultimately service utilization and outcomes (McAvoy & Coster, 2005). By that time, it was widely perceived that the current model of primary care was not

effectively addressing population health and equity and that the system was in need of investment and reform (Smith, 2009).

The 2000 New Zealand Health Strategy provided an overall framework for the heath sector, with the aim of strengthening health services in those areas that would provide the greatest benefit for the population, focusing in particular on reducing inequalities in health. The approach was to concentrate resources and efforts around these priorities, which were articulated in 13 population health objectives and three objectives to reduce specific inequalities (Table 10.1).

Within the subsequent PHC Strategy, the new PHO organizational structure was introduced to direct GPs towards the priorities identified in this New Zealand Health Strategy. PHOs are not-for-profit, non-governmental groups of individual GP practices that serve patients who enrol with them, usually within a geographic area. GPs can join PHOs on a voluntary basis, but they must be part of a PHO to receive some of the higher levels of government subsidies provided under the PHC Strategy.

The PHC Strategy also altered funding arrangements for primary health care to counteract the incentives of fee-for-service and encourage more population-based approaches. Under the PHC Strategy, the main mechanism for delivering

Table 10.1 Priorities in New Zealand's 2000 Health Strategy

Population health priorities	
1. Reduce smoking.	8. Reduce the incidence and impact of cardiovascular disease.
2. Improve nutrition.	9. Reduce the incidence and impact of diabetes.
3. Reduce obesity.	10. Improve oral health.
4. Increase level of physical activity.	11. Reduce violence in interpersonal relationships, families, schools and communities.
5. Reduce the rate of suicides and suicide attempts.	12. Improve the health status of people with severe mental illness.
6. Minimize harm caused by alcohol and illicit and other drug use to both individuals and the community.	13. Ensure access to appropriate child health care services including well child and family health care and immunization.
7. Reduce the incidence and impact of cancer.	
Priorities to reduce inequalities	
1. Ensure accessible and appropriate services for people from lower socio-economic groups.	3. Ensure accessible and appropriate services for Pacific peoples.
2. Ensure accessible and appropriate services for Maori.	

public funding to primary care changed from fee-for-service at the GP level, to capitation at the PHO level, with the intention of promoting a population-health approach and of promoting the role of non-GP health professionals. There is no requirement for PHOs to transmit the government subsidy to individual GPs by capitation, which makes it possible that some providers may still continue to receive public funding through the traditional fee-for-service payment. A study by Croxson, Smith and Cumming (2009), however, found that most PHOs were using the same capitation formula to pay GP practices that was used to calculate PHO payment. It was left up to individual practices to determine how they pay individual GPs and others working in the practice (Croxson, Smith & Cumming, 2009).

The community non-profit PHO model replaced the more profit-oriented IPA model as the vehicle for increased government subsidies to reduce patient co-payments (Gauld, 2008). The PHOs did not completely replace IPAs, however, and some of the larger PHOs rely on IPAs for management services (Gauld, 2008). The result has been a lack of clarity in the role of PHOs, and in particular how they relate to IPAs and DHBs, which has persisted since they were introduced in 2002 (Gauld, 2008; Smith, 2009). At one point there were more than 80 PHOs. A new government elected in 2008 encouraged consolidation of PHOs, and there are now 31 (Ministry of Health NZ, 2006, 2013).

The PHO Performance Management Programme, later renamed to PHO Performance Programme ('the Programme') was introduced in 2006 to sharpen the focus of PHOs towards the priorities of the 2000 Health Strategy and to manage unplanned expenditure growth (DHBNZ, 2005). The Programme, which includes a pay for performance component, aims to improve the health of populations and reduce inequalities through clinical governance and continuous quality improvement processes with PHOs and their contracted providers (PHO Performance Programme, 2010). The Programme is reinforced with financial incentives to record and pursue targets across the clinical, process and financial indicators, as well as creating an information feedback loop to give PHOs access to their own performance data to use in their improvement processes.

Stakeholder involvement

The Programme is administered by DHB Shared Services (DHBSS), formerly DHBNZ, which is the national organization representing the individual DHBs. The DHBSS unified the Programmes of the 21 DHBs into one national performance programme. The development and composition of performance indicators is overseen by a governance group, which includes mandated members representing providers (from the General Practice Leaders Forum), PHOs, DHBs, DHBSS, and the MOH (PHO Performance Programme, 2010). The governance group was established in 2008 in response to criticisms about the lack of clinical leadership in the Programme. There also is a Programme Advisory Group, which provides expert advice about the content of the programme, and ensures clinical relevance and business sustainability (PHO Performance Programme, 2009). The initial Programme indicators were

developed by DHBs/MOH as part of the Clinical Performance Indicator and Referred Services Management Projects.

Technical design

How does the Programme work?

The Programme pays PHOs a performance incentive per enrolled person based on the percentage of targets the PHO meets for ten performance indicators (see Table 10.2). To participate in the programme, PHOs must fulfil eligibility criteria demonstrating that they have a clinical governance structure in place to support the programme:

- Minimum of 85 per cent ethnicity recording.
- Minimum of 70 per cent valid identification number on patient register.
- Compliance with the fees agreement.
- Signed PHO Agreement.
- Complete practitioner information.
- Complete PHO reporting.
- Approved PHO performance plan.

DHBs provide start-up funding during the set-up phase of a PHO entering the programme. The 'establishment payment' includes a fixed amount of NZ$20,000 per PHO, and a variable amount of 60 cents per enrolled person in the PHO (Ministry of Health NZ, n.d.).

The PHO Performance Programme also has a strong focus on the priority of reducing health disparities, which is achieved through three different pathways:

1. Measuring performance separately for high needs populations where appropriate.
2. Weighting payments towards progress against targets for the high needs populations for those indicators relating to an area of health disparity.
3. A weighting for high needs population in the pharmaceutical and laboratory expenditure targets (Ministry of Health NZ, n.d.).

Performance domains and indicators

The Programme includes a set of ten performance indicators covering the domains of service coverage and quality (Table 10.2). The indicators have evolved since the beginning of the Programme, moving from prioritization given to process and financial indicators, to a greater emphasis on clinical indicators. Eleven indicators were used in Phase 1 (2006–2008), and these were updated and reduced to ten indicators in Phase 2 (2008–present). Each indicator is given an annual weight, which reflects the share of the total possible payment to a PHO in a year that is related to performance against that indicator.

The Programme also collects a set of indicators that are for information only and not tied to incentive payments (Table 10.3). Beginning in 2011 the efficiency indicators were no longer tied to incentives and collected for

Table 10.2 Current funded New Zealand PHO Performance Programme indicators, 2012

Indicator	*Weight*
Chronic conditions indicators	
Breast cancer screening coverage	
Total population	—
High needs	6.0 per cent
Cervical cancer screening coverage	
Total population	3.0 per cent
High needs	6.0 per cent
Ischemic cardiovascular disease detection	
Total population	2.5 per cent
High needs	5.0 per cent
Cardiovascular disease risk assessment	
Total population	8.0 per cent
High needs	12.0 per cent
Diabetes detection	
Total population	2.5 per cent
High needs	5.0 per cent
Diabetes follow-up after detection	
Total population	3.0 per cent
High needs	6.0 per cent
Smoking status recorded	
Other	2.0 per cent
High needs	5.0 per cent
Smoking cessation advice or support	
Other	4.0 per cent
High needs	9.0 per cent
Prevention of infectious diseases indicators	
Influenza vaccination in the elderly (>65 years)	
Total population	3.0 per cent
High needs	6.0 per cent
Per cent of children fully immunized	
Total population	2.0 per cent
High needs	4.0 per cent
Per cent of eligible children fully immunized by 8 months of age	
Other population	2.0 per cent
High needs	4.0 per cent
Total score	100 per cent

Table 10.3 New Zealand PHO Performance Programme indicators collected for information only, 2012

Indicator	*General population*	*High-need population*
Breast cancer screening coverage (per cent enrolled women age 50–69)	Yes	Yes
Smoking status and advice/support given	Yes	Yes
Inhaled corticosteroids		Yes
Investigation of thyroid function		Yes
Acute phase response measurements		Yes
Metformin: sulphonylureas ration		Yes
Utilization by high-need enrolees (doctor, nurse consultations)		Yes
GP referred laboratory expenditure		Yes
GP referred pharmaceutical expenditure		Yes
Per cent of diabetes patients with HbA1c test result of 8 per cent or less or 64 mmol/mol or less in the last year	Yes	Yes

information only, because of concerns that they are inconsistent with the main focus on screening, management and treatment (PHO Performance Programme, 2010).

Indicators were initially chosen based on data that were already available through claims data. For Phase 2, indicators were selected with a stronger focus on the agreed health priority areas for New Zealand, which meant that the Programme had to invest in the infrastructure required to generate new data directly from the GP practices, rather than drawing directly from claims data, which tended to underreport certain activities. For example, very little data were initially available on diabetes, hypertension, and smoking.

Through the evolution of the Programme, these indicators were considered to be increasingly important, so the Programme made investments to ensure that relevant data were available. The DHBs and MOH shared the cost of this infrastructure for the new data collection. The Programme also invested heavily in consultations with provider groups, and in automated data reporting that had previously been done manually. These steps ensured that new data collection requirements data for the Programme were not a burden to providers, helping to gain provider acceptance of the new indicators and reporting.

Some indicators are measured separately for 'high-need populations', and are rewarded at a higher rate. The PHO's high-needs population is defined by the sum of individual enrolled patients who are Māori (the indigenous population of New Zealand), Pacific Islanders or living in geographic areas with high relative socio-economic deprivation. To strengthen the incentives to reduce health inequalities, payments for performance are weighted more heavily when measuring progress and outcomes amongst the high-needs populations (Buetow, 2008).

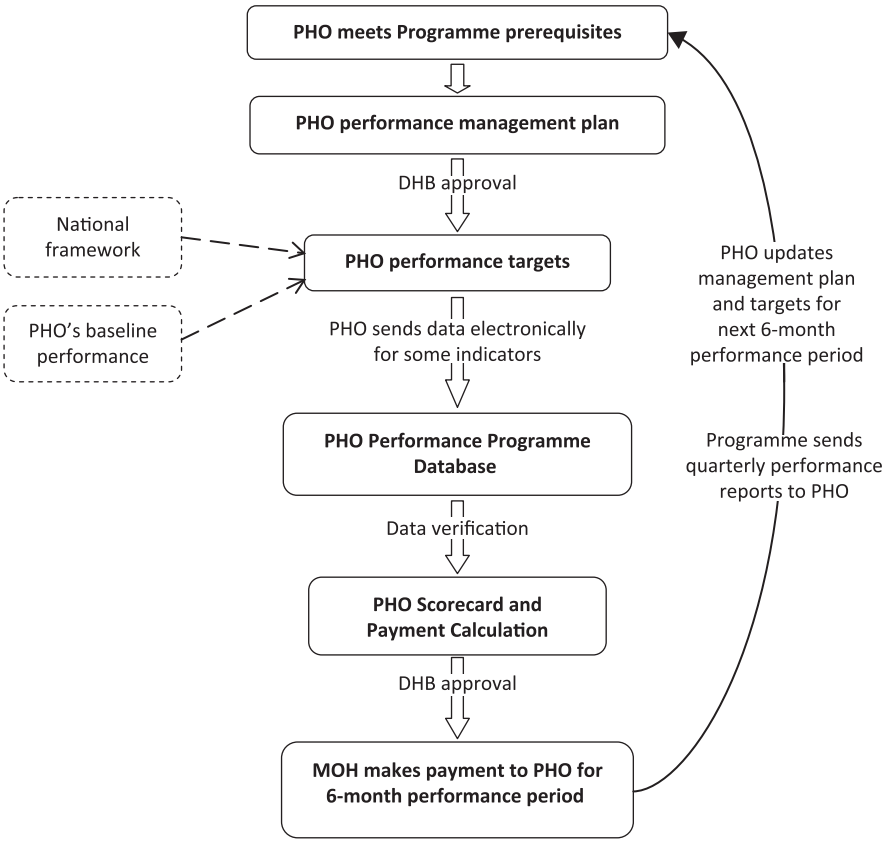


Figure 10.1 Design of the New Zealand PHO Performance Programme

Targets are set individually for each PHO using a national target setting framework and taking into consideration their baseline performance (Figure 10.1). Indicator targets are agreed on an annual basis for the two six-month performance periods (i.e. 1 July to 31 December and 1 January to 30 June). The second six-month targets may be renegotiated between PHOs and DHBs at the end of the first six-month period if PHOs had been unable to meet their targets (PHO Performance Management Programme, n.d.).

Incentive payments

Flat-rate payments for the majority of indicators are made to PHOs for each six-month performance period based on the percentage attainment of each target. Performance payment amounts are based on the following:

- population enrolled with PHO for the performance period;
- progress toward targets for each performance indicator;

- payment amount defined in the PHO Agreement per performance period per enrolled person.

The maximum available payment was initially set at NZ$6 per enrolled member, if all targets were achieved (Ministry of Health NZ, n.d.). This payment is not risk adjusted. The bonus was increased to NZ$9.27 in 2008, but it was reduced to NZ$6.13 in 2011. Each target is assessed independently for a predetermined fraction of the total flat-rate payment, so 'overachievement' against one target cannot be used to compensate for 'underachievement' in another (Buetow, 2008).

DHBs have the flexibility to support local needs through additional funding to support more indicators or reinforce national indicators by applying additional funds to either all or particular indicators (providing this does not exacerbate existing health inequalities) (Ministry of Health NZ, n.d.).

Data sources and flows

The Programme has a national database to enable the analysis and reporting of performance against targets. This database also calculates the performance payments for PHOs. Data for some indicators are sent electronically by PHOs using a standard form to the PHO Performance Programme team. Data for other indicators are retrieved by the Programme from existing databases, e.g. breast cancer screening register (PHO Performance Management Programme, n.d.).

A number of measures are taken to validate the data submitted by PHOs. Every quarter, information from PHOs is run through logic algorithms which include variation markers that highlight unusual changes in indicators. The Programme dedicates significant time and resources to make sure the data are accurate. If there are any discrepancies, PHOs have to justify unreasonable data, but no data are made available until they have been validated and agreement has been reached with the PHOs. If agreement is not reached, the Programme goes through a rigorous process to identify the reason for variation. Flu vaccination rates, for example, come from claims data, and even when a claim is rejected, the service is counted. Claims data in fact often underestimates actual coverage, as providers may not submit claims for every vaccination.

Quarterly progress reports are provided to PHOs, DHBs and the MOH. For each six-month performance period, DHBs review the PHO performance reports and the scorecards generated by the PHO Programme team, and they approve the payment amounts. Once DHB consent is received, the Programme generates a payment summary confirming the amount to be paid to the PHO and forwards to the MOH to make the payment (PHO Performance Management Programme, n.d.).

Reach of the Programme

Which providers participate and how many people are covered?

The Programme now covers all 31 PHOs, although participation is voluntary. Uptake was rapid, beginning with 29 PHOs participating in 2006 (36 per cent),

48 more PHOs entering by the following year (total of 95 per cent of PHOs), and 100 per cent coverage of the 81 PHOs by 2008, before they were consolidated into the current 31 (Ministry of Health NZ, n.d.). Nearly 100 per cent of GPs and primary care nurses participate in the Programme through the network of PHOs, covering about 98 per cent of the population. Total income from the Programme performance payments is small in relation to total PHO incomes (Buetow, 2008), and makes up less than one per cent of the government primary care budget.

Improvement process

How is the Programme leveraged to achieve improvements in service delivery and outcomes?

The key feature of this pay for performance programme is that it is a supporting component of the health sector's overall quality framework, aligned with other initiatives in order to enable and push the primary care system to reduce inequalities across the population and improve health outcomes for all New Zealanders (PHO Performance Programme, 2010). The financial incentives under the Programme are intended give better focus to the activities of PHOs and to provide some additional resources to enhance quality. The PHOs have discretion in how they use the bonus payments, but they require DHB approval, and there is an expectation that the PHOs use their bonus payments to deliver improved services in support of the objectives of the Primary Health Care Strategy, rather than to supplement the incomes of members of their practices, except perhaps to help to recruit or retain practice staff (Buetow, 2008).

There are no guidelines for distributing the bonus within the PHO, which has caused some ambiguity. Questions arose, for example, around whether it would be better to distribute the Programme bonus to high performing GP practices (reward) or low performers (investment) (Martin, Jenkins & Associates Limited, 2008). This ambiguity has led to tension in some cases, and meant that some PHOs did not allocate their performance payments. However, performance payments are typically spent, and practices often are involved in decisions about how best to use the incentive payment.

A case study of six PHOs found different approaches to allocating the bonus funds. One PHO did not distribute funds at all, one retained the funds at the PHO level to contract out for services such as education, and four PHOs shared the funds with GP practices. For these four PHOs, distribution ranged from 15–60 per cent of funds retained by PHO to compensate participation in a Clinical Advisory Group, to fund large initiatives, etc., and 40–85 per cent of funds distributed to GP practices. Three of the PHOs that shared the funding with GP practices distributed the funds on a capitation basis, with only one PHO distributing the funds based on achievement of targets (Martin, Jenkins & Associates Limited, 2008).

Some PHOs use the incentive payment for PHO-wide initiatives that benefit all practices, such as education or outreach initiatives. One PHO, for example, started a 'mammogram bus' for the enrolled members of its GP practices. In

some cases these global programmes may have more impact on improvement than spreading the bonus across all providers, which may not amount to significant payment for each individual practice.

The performance indicators and bonus payments are designed to be only a part of the PHO Performance Programme. Because of the low budget for the incentive, the Programme has had to find other ways to drive change and performance improvement. The Programme works directly with providers, looking to understand and address their individual performance issues. A large effort also has been made to feed information back to providers to use internally for performance improvement. Information is fed back to PHOs, using certain security measures, through the DHBSS website. PHOs also receive timely monthly reports for four of the indicators (flu, cervical and breast screening and immunization) and a flat file of information on a quarterly basis with the information used to calculate their indicators, which is information that was not previously available to them. The six-monthly DHB and PHO level reports also are made publicly available on the Programme website.

The Programme offers other services to support PHOs in their performance improvement processes. PHOs can receive individualized feedback reports on their own performance against the indicators as compared with benchmarks, nationally consistent education materials customized to their needs, and other services that may include the use of clinical facilitators. In spite of these efforts, however, it is not clear that the improvement process is moving beyond the PHO to the front-line GP practice level (Smith, 2009).

Results of the Programme

Has the Programme had an impact on performance, and have there been any unintended consequences?

Programme monitoring and evaluation

There has been no rigorous evaluation of the PHO Performance Programme. The efforts to monitor and evaluate the PHO Performance Programme have been largely ad hoc, relying on indicator analysis, small sample non-rigorous surveys, case study approaches, and anecdotal evidence. To help monitor early effects of the Programme, for example, the national DBSS produced a questionnaire for the managers of the first cohort of 29 participating PHOs, to which 16 responded. All respondents stated that as a result of the Programme, their PHO had developed an increased focus on quality improvement, including clinical facilitation, data collection, data quality and feedback to member practitioners, and clinical governance groups (Buetow, 2008). This survey did not, however, capture the perceptions of front-line providers.

An independent evaluation was completed in 2008 using a case study design. The key informant interviews were conducted with a purposive sample of six PHOs to include PHOs of different sizes, serving different types of populations. The evaluation found that the Programme is perceived as useful by PHOs but more as a reinforcement of existing objectives and initiatives than an

independent driver of improved quality (Martin, Jenkins & Associates Limited, 2008).

The PHO Performance Programme recently began issuing an annual report that assesses the contribution of the Programme based on its objectives, and provides a trend analysis of the performance indicators (PHO Performance Programme, 2009).

Performance related to specific indicators

All ten performance indicators have shown some improvement since the Programme was introduced in 2006, and increases in coverage are substantial in some cases. Breast cancer screening rates, for example, increased from 55 to 68 per cent for the total population between 2006 and 2012, and from 42 to 63 per cent for high-needs population. Cervical cancer screening increased from 66 to 74 per cent for the total population, but increased only from 63 to 66 per cent for the high-needs population. Cardiovascular screening increased from 30 to 50 per cent, and diabetes detection and follow-up rate increased from 46 to 72 per cent for the total population, and 50 to 70 per cent for high-needs population. Childhood vaccination rates increased from 60 to 90 per cent, but there was no change in flu vaccination rate (PHO Performance Programme, 2012). Some of the improvements in coverage of priority services is impressive, but these results do not control for underlying trends or the impacts of broader quality initiatives, so it is difficult to attribute the changes to the financial incentive. For example, some disease-specific initiatives were introduced during that time, which also could have contributed to these improvements, including the MOH 'Diabetes Care Improvement Package' to strengthen community-based diabetes care.

Equity

There appears to be little progress on the objective of reducing health disparities, as only one indicator clearly improved relatively more for high-need populations than for the population as a whole. The breast cancer screening rate for the high-needs population, for example, increased from 42 per cent to 58 per cent. This represents a 38 per cent improvement, as opposed to a 20 per cent improvement for the general population over the same period. Other indicators, however, do not suggest movement toward reducing heath disparities. The rates of diabetes detection and follow-up increased from 50 per cent to 70 per cent for the high-needs population, which is a smaller percentage improvement in coverage that was observed for the general population (PHO Performance Programme, 2010). The rate of cervical cancer screening increased by only three percentage points for the high-needs population, which is a much lower rate of improvement than for the general population.

Costs and savings

The total budget for the PHO Performance Programme was NZ$36.4 million in 2009, of which 93 per cent was intended for the incentive payments (PHO

Performance Programme, 2010). Of the total amount available for incentive payments, about 20 per cent was not allocated to PHOs as a result of the PHOs not fully achieving their set targets.

As a share of total government PHC expenditure, the cost of the Programme is relatively small, at less than one per cent. This does not take into account, however, the cost to providers of participating in the Programme. One large network of PHOs, for example, estimated that just under half of the funds it anticipated earning from the Programme would be needed to run the Programme (Buetow, 2008). For the most part, however, the PHOs are largely implementing the Programme with existing staff and structures, with senior PHO management overseeing the Programme (Martin, Jenkins & Associates Limited, 2008).

Provider response

Initially the PHO Performance Programme was perceived as being imposed from the top and bureaucratic. This perception was compounded by a more general problem surrounding the role of PHOs, which had never been fully clarified (or accepted) following the 2000 reforms (Smith, 2009). Some progress has been made, however, to garner the buy-in of GPs through a more participatory governance structure, investments by the Programme to support better data systems, and a process-oriented approach to interpreting and using performance information beyond simply calculating bonus payments. Other factors that are considered to be important for gradually increasing the buy-in of providers is that the indicators have evolved to have more of a clinical focus and based on clinical evidence, and that the Programme clearly is designed to be aligned with and supportive of the 2000 Health Strategy, which is widely accepted as definitive for setting the priorities and guiding principles for the development of the health sector (Gauld, 2008).

Overall conclusions and lessons learned

Has the Programme had enough of an impact on performance improvement to justify its cost?

In general, PHO Performance Programme is perceived as having made a positive contribution to furthering the objectives of the 2000 Health Strategy, even if the incentives themselves are too diluted to have been the real motivator of change. The Programme is perceived as aligning with and reinforcing overarching objectives of the strategy, which were agreed to by all of the stakeholders. The Programme is considered to have added value by focusing attention on priority areas and raising awareness. The Programme has also been regarded as successful at taking a comprehensive approach – providing resources, tools, and processes – in addition to incentives to change clinical practice (Martin, Jenkins & Associates Limited, 2008). The clinically credible indicators and collaborative governance have been key to this success.

As in the case of the UK QOF, an important positive spillover effect of the P4P programme is the improved collection and use of data for quality improvement purposes. There has also been an overall improvement in the clinical governance of the primary care sector. Establishing clinical governance structures and processes to engage professional members and achieve improvements is a condition for PHO participation in the Programme (Buetow, 2008). Furthermore, the Programme is overseen by a tripartite governance group consisting of representatives of providers, PHOs and DHBs. Overall governance of the PHC sector has become more participatory, as multiple stakeholders have remained actively involved in designing and shaping the Programme, and PHOs and providers have made ongoing investments in the Programme's governance structure (PHO Performance Programme, 2009). Although the PHO Performance Programme is playing an important role in reinforcing broader quality initiatives, the financial incentive itself is limited in its potential to drive changes in clinical practice, improvements in provider performance, and better outcomes. There are several main issues:

1. *The size of the incentive is small.* There are no good estimates of what percentage of PHO budgets or GP practice income are contributed by the PHO incentive, but total incentive payments make up less than one per cent of government primary care expenditures. This is a particularly small incentive in comparison to the UK Quality and Outcomes Framework (QOF) programme, which is often used as a comparison in discussions of the PHO Performance Programme. The QOF payments can represent up to 25 per cent of the annual income of GP practices in the UK (Campbell et al., 2007). In fact, the assessments of the Programme that have been done attribute any achievements to the compounding effect of the incentive rather than the incentive itself.
2. *There is a disconnect between programme management, payment of the incentive, and clinical providers.* The main criticisms of the Primary Health Care Strategy often centre on the need for more change at the practice level to bring about better care coordination (Smith, 2009). The structure of the Programme and the ambiguous relationship between PHOs and GP practices make it difficult for comprehensive performance improvement initiatives to reach day-to-day clinical practice. This is a more general problem related to the role of PHOs in the primary care system. Furthermore, the funds are not transparently distributed or reinvested, which can limit the motivational and recognition aspects of the Programme.

The investment in the PHO Performance Programme has been small, however, and there have been no reports of adverse consequences of the Programme or gaming by PHOs. Some large improvements in coverage of priority services have been achieved, which may be at least in part attributable to the Programme. It may be the case that improvements in data use, clinical governance, and population-based initiatives that have been motivated by the Programme are yielding sufficient system-wide benefits to make the investment in the PHO Performance Programme worthwhile. This conclusion could only be fully confirmed, however, following a rigorous evaluation of the Programme, or at the least more systematic monitoring and analytical assessments.

Note

* This case study is based on the 2011 report *RBF in OECD Countries: New Zealand: Primary Health Organization Performance Programme* prepared by Cheryl Cashin for the International Bank for Reconstruction and Development at the World Bank.

References

Barnett, R. and Barnett, P. (2004) Primary health care in New Zealand: problems and approaches, *Social Policy Journal of New Zealand*, 21: 49–66.

Buetow, S. (2008) Pay-for-performance in New Zealand primary health care, *Journal of Health Organization and Management*, 22(1): 36–47.

Campbell, S. et al. (2007) Quality of primary care in England with the introduction of pay for performance, *New England Journal of Medicine*, 357(2): 181–90.

Croxson, B., Smith, J. and Cumming, J. (2009) *Patient fees as a metaphor for so much more in New Zealand's primary health care system*. Wellington: Health Services Research Centre Victoria University of Wellington.

DHBNZ (2005) *PHO performance management. Data definitions for performance prerequisites*. Wellington: DHBNZ.

French, S., Old, A. and Healey, J. (2001) New Zealand health system in transition, *European Observatory on Health Care Systems Health in Transition Reports*, 3(19).

Gauld, R. (2008) The unintended consequences of New Zealand's primary health care reforms. *Journal of Health Politics, Policy and Law*, 33(1): 93–115.

McAvoy, B. and Coster, G. (2005) General practice and the New Zealand health reforms – lessons forAustralia?, *Australia and New Zealand Health Policy*, 2: 26–36.

Martin, Jenkins and Associates Limited (2008) *Evaluation of the PHO peformance programme: final report*. Auckland: Martin, Jenkins and Associates Limited.

Ministry of Health NZ (2000) *New Zealand health strategy*. Wellington: Ministry of Health.

Ministry of Health NZ (2006) *Implementing the New Zealand health strategy 2006*. Wellington: Ministry of Health.

Ministry of Health NZ (2013) *About primary health organisations* (http://www.health.govt.nz/our-work/primary-health-care/about-primary-health-organisations, accessed February 2014).

Ministry of Health NZ (n.d.) *Primary health care PHO performance programme* (http://www.moh.govt.nz/moh.nsf/indexmh/phcs-projects-phoperformance, accessed July 2010).

Ministry of Health NZ (n.d.) *Performance management programme: summary information for PHOs* (http://www.moh.govt.nz/moh.nsf/pagesmh/5817/$File/PHOperfmagmtsummaryinfo.pdf, accessed July 2010).

PHO Performance Management Programme (n.d.) *PHO performance management programme report user's guide* (http://www.dhbnz.org.nz/Site/SIG/pho/Old-Documents/Programme_Documents.aspx, accessed July 2010).

PHO Performance Programme (2009) *PHO performance programme annual report: 1 July 2008–30 June 2009*. Auckland: DHBNZ.

PHO Performance Programme (2010) *PHO performance programme annual report: 1 July 2009–30 June 2010*. Auckland: DHBNZ.

PHO Performance Programme (2012) *National summary of PHO performance 1 July 2012–31 December 2012*. Auckland: PHO Performance Programme.

Smith, J. (2009) *Critical analysis of the implementation of the primary health care strategy implementation and framing of issues for the next phase*. Auckland: Ministry of Health NZ.

chapter eleven

Turkey: Family medicine performance based contracting scheme*

Rekha Menon, Son Nam Nguyen, Aneesa Arur, Ahmet Levent Yener and Iryna Postolovska

Introduction

Prior to 2003, health outcomes in Turkey, including maternal and child health (MCH) outcomes, lagged behind those of OECD countries and of those of other middle-income countries. In 2002, the infant mortality rate was 28.5 deaths per 1000 live births compared to the OECD average of five. Life expectancy at 71.9 years was significantly lower than the OECD average of 78.6 years. The maternal mortality ratio in 2000 was more than five times the OECD average at 61 deaths per 100,000 live births in Turkey compared to the OECD average of 11.8.

Furthermore, within Turkey there were clear regional and rural–urban disparities. In 2003, the infant mortality rate (IMR) was 70 per cent higher in rural areas than in urban areas (39 and 23 deaths per 1000 live births, respectively). Infant mortality rates were higher than the national average of 29 deaths per 1000 live births in the North and East regions. Istanbul had the lowest rate (19 per 1000 live births), while Southeast Anatolia had the highest (38 per 1000 live births) (Hacettepe University, 2004).

National coverage rates for immunization masked significant variation across provinces. In 2003, the national coverage rate was around 70 per cent for BCG (Bacille Calmette-Guerin), DPT3 (Diptheria, Pertussis and Tetanus), measles, and HepB3 (Hepatitis B3) vaccines. In Şirnak province, coverage rates were as low as 29 per cent for BCG and 31 per cent for measles. In comparison, Tekirdağ and Gaziantep provinces had 100 per cent coverage rates for BCG, while Ankara and Tekirdağ had the highest coverage rates for measles (88 per cent).

Health service utilization was low. The average number of visits per capita to primary care facilities was 0.9 visits in 2002 (Ministry of Health, 2011). Over 18 per cent of women did not seek antenatal care during their pregnancy, and this indicator was significantly higher in rural areas, where 34.2 per cent of

women did not receive any antenatal care. More than 23 per cent of women first sought care after the first trimester (Hacettepe University, 2004).

These concerns set the stage for the Ministry of Health (MOH) of Turkey's wide-ranging reform agenda to improve access, efficiency and quality in the Turkish health sector through the Health Transformation Programme (HTP)[1]. A key element of these reforms was the introduction of family medicine within a performance-based contracting framework.

Health policy context

What were the issues the programme was designed to address?

A number of underlying health systems performance concerns contributed to the lagging MCH outcomes and regional disparities that motivated the Health Transformation Programme (HTP) (Ministry of Health, 2006). First, access to primary health services varied considerably across the country both between rural and urban areas, and also among provinces. These inequities were to a large extent driven by uneven distribution of health personnel. In 2002, population per general practitioner varied between 875:1 and 7571:1 among provinces (Vujicic, Sparkes & Mollahaliloglu, 2009). Governance concerns also existed at the service delivery level. A combination of low salaries and the absence of performance incentives led to staff absenteeism. This had ripple effects for higher level facilities, as patients responded to perceived poor quality at the primary care by bypassing primary care facilities and increasing patient loads at secondary and tertiary facilities. Only 38 per cent of the population in 2002 chose to utilize outpatient care at the primary care level (Ministry of Health, 2011).

Fragmentation in health service delivery, with several agencies providing care to different parts of the population, meant limited emphasis on preventive health. Centralized administration of service delivery from Ankara made it difficult to effectively manage for results, while distracting the MOH from paying full attention to its role as steward of the health sector. Public dissatisfaction with the health system was growing as a result of governance concerns and perceptions of poor quality.

Against this backdrop the 2002 elections provided the political impetus to drive health system reform, as the newly elected government perceived a clear mandate to improve social services. A key element of the MOH's response was the creation of a new primary care specialty and service delivery approach, bringing family physician salaries up to and exceeding those of specialists, promoting the use of clinical guidelines, implementing well-functioning health information and decision support systems and designing properly aligned financial incentives. Individual doctors and other clinical staff in the family medicine programme are contracted using performance based contracts. This model of primary care, the family medicine programme, was initially introduced as a pilot but now operates nationwide.

Technical design

How does the programme work?

The FM PBC is a performance-based contracting programme with a portion of contracted provider income contingent on performance against a set of targets, and the threat of contract cancellation if a threshold of performance violations is reached. The FM PBC is funded through general revenues within the budget of the MOH. However, for all practical purposes, purchasing and contract management is delegated to Provincial Health Directorates (PHDs) in each province.

Under the programme, each family medicine unit composed of family physicians, nurses and other ancillary staff is responsible for the health and well-being of an assigned group of patients and for coordinating patient care across the health system. Individuals are assigned to a specific family physician who is expected to act as the custodian of the health and well-being of his or her patients. People have the option of voting with their feet and choosing their family physician if dissatisfied with the one assigned to them. Family medicine clinical personnel are individually contracted by the PHD in each province to deliver an integrated package of preventive, promotive and curative services to patients assigned to their practice. Contracted family doctors are also responsible for managing health facilities and ensuring that their facilities meet service standards. PHDs have the day-to-day responsibility for managing and monitoring contracts, including managing payments. Community Health Centres (CHC) provide logistical and technical assistance to family medicine units and supervise and monitor FM PBC on behalf of the PHD. The MOH is the funder of the programme and plays an oversight role. The technical design of the FM PBC programme and the relationships among the different actors is shown in Figure 11.1.

The base payment for contracted providers is defined on a capitation basis, with a higher coefficient for certain categories of the population such as registered pregnant women (adjustment factor of 3), prisoners (adjustment factor of 2.25), children under four years and elderly over 65 years (adjustment factor of 1.6). In addition, if they work in an underserved area, contracted personnel can receive a 'service credit' or monthly bonus payment for location. The service credit is calculated on a sliding scale and could be as high as 40 per cent of the base capitation payment in the most underserved areas. As managers of their health facilities, doctors also receive an additional monthly lump sum payment to cover operating expenses such as rent and utilities, cleaning, office supplies, small repairs and medical consumables. The range of services and quality standards to be satisfied vary by the category of the family medicine unit. Depending on the category of the family medicine unit,[2] family physicians are paid an additional lump sum payment that ranges from 10 per cent of the maximum monthly base capitation payment for category D family medicine units to 50 per cent of the maximum monthly base capitation payment for category A family medicine units. FMP also receive an additional lump sum payment (1.6 per cent of the maximum base capitation payment) to defray the costs of providing mobile services,

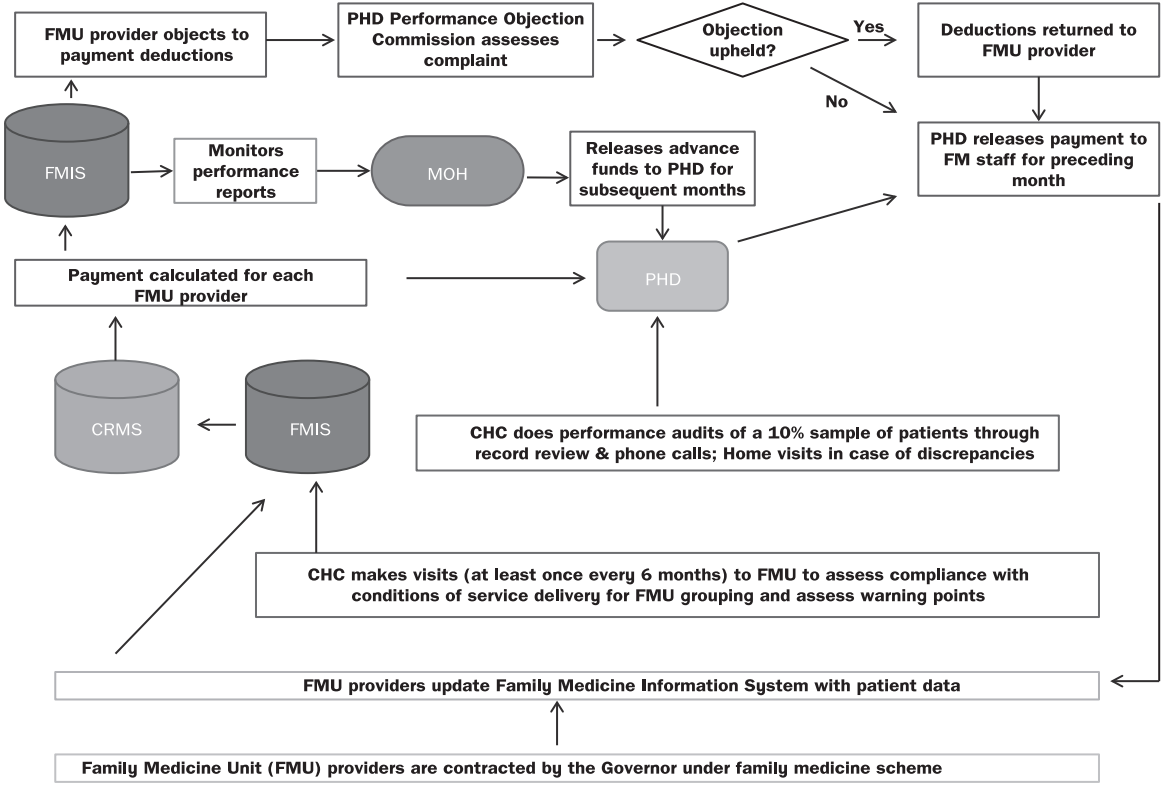


Figure 11.1 Design of the Turkey FM PBC

and the doctors are reimbursed for the expenditures they incur on laboratory tests.

The contractual framework also includes two performance levers. A *salary deduction system* wherein contracted providers risk up to 20 per cent of their base payment if their family medicine unit fails to meet coverage targets for key MCH indicators. The second performance lever relates to an *administrative system* of written admonitions or 'warning points' for failure to meet governance, service delivery or quality standards specified in a set of 35 indicators. If a provider accumulates 100 or more warning points over a contract period his or her contract can be terminated.

Performance domains and indicators

The *salary deduction system* includes eight indicators in one performance domain – coverage of priority MCH services. The performance indicators are:

- immunization coverage rate of registered children for each target vaccination (BCG, DPT3, Pol3, measles, HepB3, Hib3, each assessed separately);
- registered pregnant women with a minimum of four antenatal care visits according to schedule;
- follow-up visits of registered babies and children carried out according to the schedule.

Incentive payments

The salary deduction system

Under the FM PBC programme, performance penalties are applied to the salaries of family physicians and to family health unit staff, including managers, based on their team's performance. Providers risk up to 20 per cent of their individual base payments each month if their family medicine unit fails to meet at least 98 per cent of coverage targets for the performance indicators.

Deductions are made from the total monthly base payment of each FM provider on a sliding scale for each indicator that drops below the minimum target coverage rate of 98 per cent. The targets are applied uniformly across all indicators and family medicine units with a maximum total deduction of 20 per cent:

- A deduction of 2 per cent if the monthly coverage rate is 97 per cent to 98 per cent.
- A deduction of 4 per cent if the monthly coverage rate is 95 per cent to 96 per cent.
- A deduction of 6 per cent if the monthly coverage rate is 90 per cent to 94 per cent.
- A deduction of 8 per cent if the monthly coverage rate is 85 per cent to 89 per cent.
- A deduction of 10 per cent if the monthly coverage rate is lower than 85 per cent.

The administrative system

FM staff are evaluated against the 35 performance indicators and warning points are given for any violation. Each violation is linked to a pre-specified number of points based on the severity of the violation (Table 11.1). If a family medicine staff member accumulates 100 or more warning points over a single contract period, his or her contract is terminated and he or she is debarred from applying for a new contract for a year. Given the substantial increase in take-home pay in the FM PBC programme, this is a powerful incentive to maintain governance, quality and service coverage standards.

Repeated failure to meet performance targets for key MCH indicators could result in contract termination for all the family medicine staff in the unit, in addition to the payment deductions levied for each indicator falling below the performance target. Furthermore, it is also clear that the family medicine contracting programme is very concerned about the quality of reporting and patient privacy. Incorrect ('non-factual') reporting or failure to keep patient records secure can result in 50 warning points, implying that two violations over two years would result in contract termination. Another governance-related concern highlighted by the warning points is drunkenness on duty.

Table 11.1 Performance indicators and warning points in the Turkey FM PBC

Performance indicators	*Warning points*
Failing to comply with plan of work hours.	3
Absence without excuse (for every day not worked).	5
Not posting the posters and announcements duly.	5
Guidance signboards inside FHC and guidance signboards outside FHC not being in suitable form.	5
Using material containing drug advertisement during duty.	5
Not keeping regular records relating to duty or not informing the directorate or the Ministry.	10
Not transferring personal health records of registered persons.	10
Not replacing missing medical equipment of the Family Health Centre within ten days (for each missing material).	10
Exceeding the designated duration of absence for the trainings given.	10
Keeping drugs with expired dates.	10
Not protecting the drugs subject to green and red prescriptions.[3]	10
Admitting drug company representatives inside family health centre within working hours.	10
Illumination being not sufficient in waiting and treatment areas.	5
Not doing directly observed treatment of patients with tuberculosis or not having it done.	5

Not doing the portion of duty for home care services.	10
Retarding or not keeping with plan in ambulatory health services.	10
Not doing other duties given by regulations.	5
Not wearing uniform.	5
Failing to provide adequate security of personal health records.	20
Not providing security of personal health records intentionally.	50
Not making the minimum physical conditions of Family Health Centre suitable within ten days.	10
Not conforming with Regulation on control of medical wastes.	20
Not cooperating in audits, not presenting the desired data, making nonfactual statements.	20
Not declaring property as per regulation.	20
Not doing the imposed duty in preventive medicine implementations, making nonfactual statements.	20
Inoculation rates of each vaccination subject to performance below 90 per cent except cases of force majeure or in cases of denouncement.	10
Follow-up of pregnant women rates subject to performance below 90 per cent except cases of force majeure or in cases of denouncement.	20
Follow-up of baby–child rates, one of the preventive medicine implementations, below 90 per cent except cases of force majeure or in cases of denouncement.	20
Not abiding by cold chain rules.	20
Not abiding by patient rights and patient confidentiality as per provisions of respective legislation.	20
Not abiding by the Medical Deontology Code of practice or patient confidentiality.	20
Insulting colleagues or those receiving service or threaten them.	20
Coming drunk to work or taking alcoholic beverages at place of duty.	50
Preparing nonfactual report or document.	50

Data sources and flows

The family medicine programme is supported by two main information systems that are used to track performance on technical and managerial/budgetary parameters: (1) the Core Health Resource Management System (CRMS), a MOH-wide information system used to track budgets and expenditures; (2) the Family Medicine Information System (FMIS), which tracks health-related indicators relevant to family medicine services and is a decision-support system for health providers.

The CRMS includes data on parameters that determine payments to family medicine staff, including socio-economic development coefficients for each

district, expenditures on lab tests, staffing, expenditures on mobile services, etc. Not all districts were initially covered by the CRMS, however, and some did not input data correctly in the past. The PHD also manually tracks these data in the provinces to ensure data validity.

The FMIS was developed and introduced in conjunction with the family medicine model. The provider interface of the FMIS includes an electronic health record for each person registered with a family physician. This electronic health record can be updated directly by family medicine personnel and is a comprehensive record of patient characteristics and services received, including but not limited to MCH services that are specifically targeted by performance incentives in the FM PBC. The FMIS also provides decision support to family medicine staff by generating reminders or follow-up lists, and allows family medicine providers to track their progress for indicators that are linked to payment deductions.

FMIS data are updated on a central server and also can be accessed by authorized staff in the PHD and by the MOH. The PHD and MOH assess each individual family medicine unit's performance on targeted performance indicators linked to the deduction system by calculating service coverage rates among eligible population registered to each family medicine unit.

The information generated through these information streams is used by the PHD to assess the level of payments to be made to individual family medicine staff, compliance with standards and to identify whether contracts should be terminated. The exact payment due to each provider is calculated in the CRMS. The PHD uses data from the FMIS, CRMS and performance audit findings to release payments to providers by the fifteenth of each month. Providers are also informed of their calculated payments, and of possible deductions, each month through the FMIS by the thirteenth of that month. The FMIS also provides a source of data for the MOH to oversee the performance of the programme and release advance funds to each PHD related to expected performance.

Data verification

Verification of performance data is of utmost importance in Turkey's FM PBC programme, as performance indicators in the FMIS are entered into the information system by family medicine staff themselves and are therefore self-reported data. As the entity responsible for managing contracts, the PHD is responsible for verifying that these self-reported data on service coverage are accurate. Every month, approximately 10 per cent of family doctors are selected for data verification by the CHC. Staff from the CHC conduct a performance audit of the selected doctor through a combination of patient records review, phone calls or home visits. Approximately 10 per cent of the patients for an audited doctor are selected for participation in this audit. Findings from regular audits can trigger a more in-depth audit or investigation. Except under exceptional circumstances, no doctor is audited in two consecutive months.

In addition, each family medicine practice is visited by CHC staff at least once every six months to assess compliance with service delivery and

governance standards and identify any violations linked to warning points. These data are also used to verify that the family medicine unit delivers the range and meets quality standards associated with the family medicine unit's service classification (A–E). Any discrepancies identified during routine visits can trigger a more in-depth audit of individual family medicine staff.

Reach of the programme

Which providers participate and how many people are covered?

Established as part of the family medicine practice pilot programme under the Law on Piloting of Family Medicine (Number 5258), the FM PBC programme was initially implemented in Düzce province in September 2005. The programme was rolled out nationwide starting in 2006. By December 2010 all 81 provinces in Turkey had been included in the FM PBC programme, and in November 2011 the programme was designated a permanent programme of the government.[4]

As of the end of 2011, the family medicine practice programme covered the entire 74.7 million population of Turkey. At the time of preparation of the study a total of 20,243 family medicine practice doctors and 20,243 family health personnel (mainly nurses and midwives) worked in 6463 family health centres.[5] In addition, there were 13,476 staff members working in 960 community health centres, 2349 of whom are physicians. On average 3500 patients were registered for each FMP doctor but the number of registered patients per physician can be as high as 4500. The MOH goal is to reduce this number to 2000 by 2023.

Improvement process

How is the programme leveraged to improvements in service delivery and outcomes?

As noted above, the FM PBC programme was part of a comprehensive reform process to improve MCH outcomes. The programme aimed to improve maternal and child health directly through the performance components of the programme, and also through mechanisms to improve governance and accountability at the service delivery level. The design of the FM PBC programme includes a number of incentives for providers to focus their efforts on reaching pregnant women and children through the following performance levers:

1. *Payments held 'at risk' conditional on performance.* The contracting framework for family medicine staff specifies that providers risk up to 20 per cent of their base payment if their unit fails to meet minimum coverage targets of 98 per cent for vaccinations, antenatal care and follow-up of mothers and babies. A portion of their individual salaries may be deducted if critical MCH performance targets are not met by the family medicine unit. This creates strong incentives to focus on immunizations, ensuring that antenatal care services are delivered and mothers and children are followed up.

2. *Performance conditions linked to contract termination.* Contracts can be terminated if a family medicine provider accumulates 100 or more warning points over a single contract period (i.e. a maximum of two years). Failure to maintain vaccination rates, follow-ups of pregnant women and infant and child follow-ups at 90 per cent or higher results in 10, 20 and 20 warning points per violation, respectively, so that five failures to meet performance targets could, in principle, result in contract termination.

   Furthermore, a survey of 38 provinces that had implemented FM PBC for three or more years found that failure to meet performance targets was the most frequent reason for assigning warning points in the first year of family medicine in 47.8 per cent of family medicine provinces. This risk has created strong incentivizes for providers to focus their efforts on improving MCH services.

   By 2011, significant improvements had been achieved in MCH services and the share of provinces reporting failure to meet performance targets as the most frequent reason for assigning warning points decreased to 29.2 per cent. The second most common reason for the issuance of warning points in the first year of implementation was the failure to comply with working hours (13 per cent). By 2011, it has become the most common reason for issuance of warning points.
3. *Incentives created by the capitation-based formula used to calculate Family Medicine provider salaries.* The base capitation payment assigns higher weights to enrolling pregnant women and children to motivate providers to improve access to care among these categories of the population. In effect this gives incentives to family medicine personnel to proactively seek out pregnant women and register children under the age of five.
4. *Uniform absolute performance targets – rather than targets that are relative to baseline – reflect the MOH's policy objective of closing geographic gaps in performance.* Uniform targets give Family Medicine providers in areas with lagging performance the incentive to work harder to prevent salary deductions for failure to meet these performance thresholds.
5. *Team performance is assessed as a unit rather than individual performance to reduce fragmentation and increase accountability.* Although family medicine staff members are contracted individually and performance sanctions are applied to each individual, the performance of the team is assessed as a unit to incentivize cooperation and coordination within the team. Under the family medicine model, the family physician is expected to coordinate the care of his or her patients across levels of the health system, therefore creating a single point of responsibility for primary care services and reducing the fragmentation in service delivery at the primary care level.

Further, under the programme, payment rates for family medicine general providers were made attractive enough to induce them to leave government positions and join as a contracted family medicine physician. In fact, family medicine doctors are now paid on average almost the same as specialists working in hospitals, and about 1.6 times what general practitioner doctors in hospitals are paid.

Mechanisms to improve governance and accountability combined with autonomy

Governance concerns at the service delivery level prior to the launch of the HTP meant that improving accountability was a key health system objective. The warning points, performance points, complaint mechanism and institutional arrangements aimed to achieve this objective in a number of ways. Warning points help provincial health authorities to hold family medicine providers accountable for maintaining basic service standards related to structural aspects of quality. The system also helps maintain expected standards of behaviour for health professionals. Supervisors visit family medicine units to assess whether warning points must be awarded. This direct link between independently assessed performance along predetermined parameters and contract termination is an important mechanism in the programme for improving accountability, for ensuring that services meet basic quality standards, and for improving service delivery governance. There is also peer-to-peer learning and an open platform to share experiences. The MOH conducts annual meetings with the Family Practitioner Association to understand and resolve issues and grievances.

High levels of public dissatisfaction with health services due to perceptions of poor quality and staff absenteeism before family medicine was introduced meant that improving service delivery to meet user expectations was an important reform objective for the Turkish MOH. Under the FM PBC, the population has the option of choosing another family physician if dissatisfied with the one assigned to them. Since family medicine providers are paid based on the number of people registered with them, this creates incentives for providers to be more responsive to their registered population.

Complaint mechanisms are also an important feature of improving responsiveness. The MOH has a national toll-free hotline that people can call to lodge their complaint. Hotline complaints are investigated by the PHD and the CHC in the province, independently of the family medicine providers, and can trigger an audit. The separation of purchaser and provider created by the contracting framework helps to maintain the independence of provincial-level authorities who are effectively responsible for holding family medicine providers accountable. Findings from key informant interviews with provincial regulators and contracted providers suggest that investigations based on complaints are taken seriously.

A focus on results with management flexibility to attain them gives providers and PHDs the space to achieve results. The FM PBC programme holds family medicine providers in a unit jointly accountable for achieving contractually specified results while giving providers management autonomy. Contracts specify service standards that must be met, but providers are given flexibility in organizing their work hours, recruiting non-clinical support staff, and maintaining physical premises of their facilities (for which they receive a lump sum payment). Similarly, PHDs have the autonomy to exercise their contract management role within the guidelines specified by the MOH.

Results of the programme

Has the programme had an impact on performance, and have there been any unintended consequences?

Health outcomes, service utilization, and patient satisfaction

Turkey has seen significant improvements in key health outcomes (mainly MCH and malaria) in the period surrounding the introduction of the family medicine programme. The infant mortality rate fell from 28.5 to 10.1 deaths per 1000 live births between 2003 and 2010. The maternal mortality ratio fell from 61 to 16.4 deaths per 100,000 live births over that period. The average national vaccination coverage rate for DPT3 rose to 97 per cent in 2010 from 68 per cent in 2003, while regional disparities narrowed. In addition, more and more pregnant women have at least four antenatal care visits in line with WHO standards. A trend analysis of FMIS data on the number of antenatal care visits indicates that the national average increased from 3.8 visits in 2003 to 4.6 visits in 2010. Further by 2010, 20 provinces had an average of less than four antenatal care visits and only two had an average of less than three visits compared with 50 provinces with less than four antenatal care visits and 20 with less than three visits in 2003 respectively.

The number of primary health care consultations increased with the implementation of family medicine from 1.9 outpatient visits per capita in 2005 to 2.8 in 2009. The number of visits per capita to PHC facilities was significantly higher in provinces that had implemented family medicine – 2.9 visits per capita in FM provinces compared to 2.1 in non-FM provinces. In fact, a fixed-effects regression controlling for both province and year shows that the introduction of family medicine is associated with an increase in per capita consultations of 0.28, a 14 per cent increase in visits over this short time span. Further, the share of population that chose to utilize outpatient services at the primary care level rose from 38 per cent in 2002 to 51 per cent in 2010 (Ministry of Health, 2011).

Patients are more satisfied with the health system since the family medicine reforms in Turkey. Surveys conducted in using the EUROPEP scale to investigate patient satisfaction along a number of dimensions in 2008 and 2011 allow for comparisons of patient satisfaction in provinces that had implemented the FM programme and in those that were yet to do so. Satisfaction rates were statistically significantly higher in provinces that had implemented the FM programme. Between 2008 and 2011, satisfaction rates among new reformers, i.e. provinces that adopted the FM PBC programme after 2008, rose from 80.8 per cent to 90.2 per cent.

In addition to the Family Medicine programme roll-out in 2005, many other measures have been initiated since 2003 to reduce infant and maternal mortality, improve immunization coverage, and increase the number of antenatal and postnatal visits. In order to inform family planning decisions and detect pregnancies at an early stage, women between 15 and 49 years old are now followed up twice a year by primary health care and family medicine providers. Prenatal and postnatal care management guidelines have been developed, and standards have been set for the minimum number and timing of

antenatal and postnatal care visits. Beginning in 2005, free iron supplements are distributed to infants and pregnant women as part of the Iron-Like Turkey and Iron Supplement for Pregnant Women programmes. Spending on vaccination increased more than 19-fold between 2002 and 2010.

The HTP was also accompanied by an increase in public resources for primary care in absolute and relative terms. Spending on primary care doubled between 2002 and 2010. The primary care reforms associated with the family medicine model accounted for nearly 50 per cent of primary care spending and 5.6 per cent of public spending on health in general.

While it is difficult to disentangle the impact of the family medicine performance based payment system given the significant investments in the sector that were undertaken just prior to its implementation, as shown below it is evident that a comprehensive reform of how MCH services are delivered in Turkey has resulted in significant improvements in key performance indicators, which the programme reinforces.

Provider response

Health providers are important stakeholders in any health reform effort. Managing provider expectations and supporting provider performance by responding to their legitimate needs is essential to ensure that health reform yields good results. In 2008, a health care employee satisfaction survey was conducted in public health facilities and university hospitals to evaluate providers' views on job satisfaction, motivation and commitment. In this survey, providers were asked to rate their responses on a scale ranging from one being the most favourable option to six being the least favourable option. Job satisfaction was highest among family physicians (average score of 2.32 compared to an average score of 2.64 among specialists). Motivation and commitment were also highest among family physicians – 2.86 and 2.60 respectively, compared to 3.25 and 2.90 among specialists.

Costs and savings

As the FM PBC is a negative incentive programme, there is no cost related to payment of incentives. Data on the administrative cost of the programme are not available. Although this has not been measured, there may also be net savings to the health system as a result of higher utilization of primary care services of better quality, which may reduce more costly inpatient services.

Overall conclusions and lessons learned

Has the programme had enough of an impact on performance improvements to justify its cost?

While it is difficult to disentangle the impact of the FM PBC programme given the significant investments in the sector that were undertaken just prior

to its implementation, it is evident that Turkey's experience of successfully strengthening primary care over a period of less than ten years has yielded significant results. Provider performance has improved, as have health outcomes, and performance gaps between regions have narrowed. As a result, both user and provider satisfaction improved significantly. This has been achieved through a carefully designed combination of measures including increased human and financial resources and properly aligned financial incentives. Higher remuneration for family physicians has attracted much needed personnel to join family medicine practices. This higher remuneration was accompanied by accountability measures and performance incentives. Using incentives and performance targets to focus provider efforts and hold them accountable through the FM PBC programme reinforced this strategy.

The FM PBC programme also makes an important and direct contribution to the health sector through the FMIS, which provides a robust and comprehensive source of data on service coverage and health outcomes, while the incentives for accurate and timely reporting embedded in the programme increase the likelihood that data are of good quality. From an institutional perspective, the purchaser–provider split introduced by the contracting mechanism facilitates greater use of these data for stewardship of the sector. Moving forward, there are still a few areas where the system could be strengthened further, which include the following:

- *Reorienting the performance agenda to address outstanding and new challenges.* The administration portion of the FM PBC programme currently includes a number of indicators that mainly capture structural aspects of quality of care focused on the basic minimum prerequisites for service delivery. The system does not directly incentivize the clinical process dimension of quality. Therefore it can be said that the FM PBC programme in Turkey started out with a mostly 'pay for quantity' approach for MCH. Quality checks are an integral part of Turkey's quality improvement programme in primary health care. As such, it may be timely to include quality indicators for MCH services in the performance contracts, with a focus on clinical processes to support the ongoing quality improvement efforts.

  A second emerging health sector challenge is non-communicable diseases (NCDs). Cognizant of these concerns, Turkey plans to incorporate performance incentives to prevent and manage NCDs into the FM PBC programme. While the programme has not been designed as yet, the current intent is to design positive incentives for family medicine providers to address NCDs rather than negative incentives, as is the case for MCH. Positive incentives are considered by the MOH to be important to motivate case finding, which is a key issue with chronic conditions, but positive incentives will have to be designed so that they do not increase the total family medicine budget.
- *Standardizing monitoring of FM providers.* At the moment, there are substantial variations among provinces in the way FM providers are monitored and how performance is verified (for instance, how doctors are selected for audits or how warning points are assessed). This MOH has

begun to address this issue with the introduction of standard monitoring tools and guidelines.

- *Strengthening performance feedback to FM providers.* While annual meetings are held between MOH and the Family Practitioners Association to resolve issues and complaints, currently, no standard guidelines on feedback between providers and provincial health departments exist. Standardizing these strategies across provinces and strengthening the dialogue between providers and PHDs could make an important contribution to further improving the performance of family medicine providers.
- *Improving use of peer-to-peer learning networks for quality improvement.* Peer-to-peer learning networks for quality improvement are used as a provider-driven tool to improve quality in many health care settings. Turkey has taken advantage of the availability of good internet connectivity in most provinces, which presents a cheap and potentially effective option and has created an internet-based 'open platform' for peer-to-peer learning. This can be a good mechanism for training forums and other modes of peer-to-peer learning. Within a relatively short time, Turkey successfully introduced and rolled out nationwide a family medicine model, of which performance-based contracting is an integral component. MCH indicators have significantly improved as a result of this concerted strategy.

Performance-based contracting was appropriate to meet the priority needs of the sector at the time of implementation. The institutional arrangements, accountability structures, as well as an elaborate and functioning monitoring and evaluation system are in place to form the basis for performance-based payments. This combination of supporting structures has shown to be effective in preventing doctors from gaming the system, as well as improving accountability. As progress is made towards the original challenges that framed the FM PBC programme, incentives should evolve to be aligned with the most important current challenges. There is also scope for fine-tuning the institutional arrangements for implementation.

Notes

* This case study is based on the 2013 report *Turkey: performance based contracting scheme in family medicine – design and achievements* prepared by the World Bank with the support of the Public Health Institution of the Ministry of Health, Turkey.

1 Prior to 2003, various governments had made considerable efforts to restructure health service delivery and financing, and these are well documented in MOH (2011). In fact, the National Health Policy prepared by the MOH in 1993 included among its reform policies the development of the primary care services within the framework of family medicine. However, in 2003, there was a unique opportunity to pursue policies to strengthen primary care when the government outlined its reform objectives under a Health Transformation Program (HTP), which highlighted the need for a broad 'transformation' in the way health care was financed, delivered, organized, and managed, particularly in extending health coverage to the entire population and reducing the inequalities in access to and utilization of services across the country.

2 Family medicine units are divided into four categories (A, B, C and D). These categories specify the range of services and quality standards, including equipment

and personnel, which must be satisfied by a family medicine unit classified in each category.

3 Turkey has a colour coded prescription system. Red prescriptions are for opioids, green for sedatives and opioid derivatives and white prescriptions for all others.

4 Decree in Force of Law no. 663 on the Organization and Duties of the Ministry of Health and Its Affiliates.

5 A family health centre is defined as a health care organization which provides family health care services through one or more doctor (family physician) and at least an equal number family health personnel (midwives/nurses).

References

Hacettepe University (2004) *Turkey demographic and health survey 2003*. Ankara: Hacettepe University.

Hacettepe University (2009) *Turkey demographic and health survey 2008*. Ankara: Hacettepe University.

Ministry of Health (2006) *Family medicine: the Turkish model*. Ankara: Ministry of Health.

Ministry of Health (2011) *Turkey health transformation program evaluation report 2003–2010*. Ankara: Ministry of Health.

Ministry of Health (2012) *Turkish health sector: human resources for health vision 2023*. Ankara: Ministry of Health.

Vujicic, M., Sparkes, S. and Mollahaliloglu, S. (2009) *Health workforce policy in Turkey: recent reforms and issues for the future*. Health, Nutrition and Population (HNP) discussion paper. Washington, DC: World Bank.

World Bank (2013) *Turkey: performance based contracting scheme in family medicine – design and achievements*. Report No: 77029-TR, Human Development Sector Unit Europe and Central Asia Region. Washington, DC: World Bank.

chapter twelve

United Kingdom: Quality and outcomes framework*

Cheryl Cashin

Introduction

Since it was established in 1948 the United Kingdom's single-payer National Health Service (NHS) has effectively provided universal coverage with high-quality care and cost containment. The NHS model is widely considered to be an international best practice in primary care-centred health services delivery, and the focus on primary care has contributed to the cost containment and efficiency of the system.

By 1997 when the Labour government came to power, however, the cost containment efforts of the NHS appeared to be overly successful. The UK's total expenditure on health was only 6.6 per cent of its gross domestic product (GDP), as compared with 10.3 per cent in France at that time (World Bank, 2013). Per capita total health spending was only $1813 in the UK, compared with $2387 in France, $2580 in Canada, $2780 in Germany, and $4540 in the United States. As a consequence of the relative underspending, UK health care infrastructure was becoming outdated, there were not enough health professionals, and waiting times for routine surgeries were unacceptably long (Stevens, 2004). Primary health care was particularly underresourced.

In its 2000 NHS Plan for Reform and Investment, the UK government made a historic commitment to investing in the NHS (Government of the UK, 2000). Over the next ten years, spending on the NHS was increased by 43 per cent in real terms, and total health spending increased to 8.7 per cent of GDP by 2008, close to the OECD average of 9.0 per cent (OECD, 2010). This infusion of resources into the NHS was accompanied by measures to increase accountability and set standards for providers. 'National service frameworks' were developed to specify standards for key conditions such as heart disease and diabetes. A health technology assessment agency, the National Institute for Clinical Excellence (NICE), was established in 1999 to issue binding recommendations on services to be funded by local NHS authorities.

Noting that the NHS 'currently lacks the incentives many private sector

organizations have to improve performance', the 2000 NHS Plan also called for a significant extension of quality-based contracts for GPs (Government of the UK, 2000). The Plan called for changes throughout the NHS that would move from the existing incentives for improved performance that were too narrowly focused on efficiency and 'squeezing more treatment from the same resources' to incentives that support quality, patient responsiveness and partnership with local authorities.

Performance targets, some of which were tied to financial incentives, became a key feature of the approach to reforming the NHS. The 2000 NHS plan called for a National Health Performance Fund, which would be held and distributed regionally, to allow for each health authority to reward progress against annually agreed objectives. The publication in 2001 of the first NHS Performance Ratings for NHS Trusts providing acute hospital services and the *NHS Performance Indicators 2001/02 for Primary Care Organisations* represented further steps in performance measurement and accountability (UK Department of Health, n.d.). GPs already had some experience with financial incentives from the limited use of incentive programmes that were initiated in 1990 (Middleton & Baker, 2003).

Against this backdrop, in 2004 a new General Medical Services (GMS) contract between Primary Care Organizations (PCOs)[1] and General Practitioner (GP) practices (Figure 12.1) was negotiated with the GP Committee of the British Medical Association. The new contract made a number of changes, including ending responsibility to provide services outside of operating hours, as well as a voluntary P4P programme based on the Quality and Outcomes Framework (QOF). The initial programme included 146 targets in four domains (clinical, organizational, patient experience, and additional services), which are revised periodically. The cost of the QOF, around £600 million in the first year, and around £1 billion thereafter, formed part of the planned increased investment in primary health care services over the first three years of the new contract.

Health policy context

What were the issues that the programme was designed to address?

Nearly all GP practices in the UK are private entities contracted by PCOs under the NHS. GP practices are paid by capitation (a flat payment rate per enrolled individual) for basic services. Prior to the 2004 revision of the contract, GPs were facing an increasing workload, as they were required to manage chronic conditions from secondary care and make their services available 24 hours a day, seven days a week. There was growing concern about the low status and pay of GPs, which was leading to low morale and problems with recruitment and retention (McDonald, 2009).

Because of concerns about morale and retention of GPs, a number of concessions were made in the 2004 contract revision. The capitation payment was supplemented by a Minimum Practice Income Guarantee for any practice

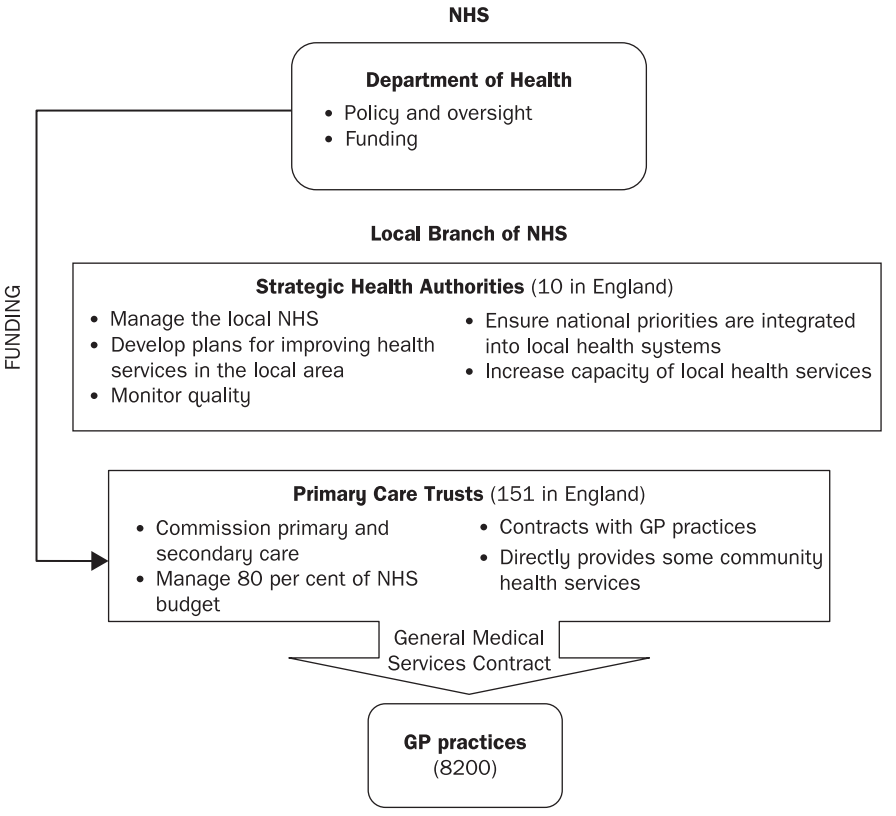


Figure 12.1 Structure of the primary care system in the NHS in England in 2004

that would have lost income under the new payment formula that was introduced with the new contract. GP practices could now opt out of providing additional services and out-of-hours care in exchange for a reduction in their capitation payments. On the other hand, the way GP practices were paid previously was considered to be partially responsible for the problems in primary care observed prior to the 2000 reforms, particularly low morale. The 2000 NHS Plan stated:

> 'The way family doctors are rewarded today remains largely unchanged from 1948. GP fees and allowances are related to the number of patients registered with them and insufficiently to the services provided and the quality. The GP remuneration system has failed to reward those who take on additional work to make services more responsive and accessible to patients and to relieve pressures on hospitals. The system has not succeeded in getting the right level of primary care services into the poorest areas which need them most.'
>
> (Government of the UK, 2000)

The QOF pay for performance programme was implemented to correct these failings of the current capitated payment system and reward more

activity and better quality of care. The programme was consistent with the approach outlined in the 2000 NHS strategy of infusing the NHS with additional resources, but also tying those resources to greater accountability and more rigorous performance standards. Given the deeper problems in the NHS and the primary care sector as a whole, the QOF had objectives that extended beyond improving performance and quality of care. The overall aims of the QOF P4P programme were to:

- increase productivity;
- redesign services around patients;
- improve the skill mix in primary care;
- create the culture and governance structure to improve quality of care;
- extend the range of services available;
- improve recruitment, retention and morale (UK National Audit Office, 2008).

Stakeholder involvement

The QOF is implemented solely by the NHS. PCOs manage the contracts under the supervision of the Strategic Health Authority (SHA), the local representation of the NHS. PCOs assess performance and calculate scores for the bonus payments. In 2009, NICE took over a new role in advising on future indicators for the QOF. A crucial part of the new process is the creation, by NICE, of an independent Primary Care QOF Indicator Advisory Committee, which is reviewing existing indicators and will recommend new ones in a participatory way (Rawlins & Moore, 2009). Negotiations between the NHS Employers and the General Practitioners Committee decided which indicators were eventually adopted into the 2011/12 QOF (NICE, 2010).

Technical design

How does the programme work?

Performance domains and indicators

The 2011/12 QOF includes 142 indicators in four domains, with targets that are uniform across GP practices. Each indicator has a maximum point value. Practices accumulate quality points according to their performance on the indicators, up to a maximum of 1000 points. Achievement of points for many of the indicators is triggered at lower and upper target thresholds of attainment (per cent of eligible patients reached) for each performance indicator. Upper thresholds are set below 100 per cent of patients to allow for practical difficulties attaining 100 per cent of patients listed on the disease register (Mason et al., 2008). For other indicators, payment is received when an action is confirmed, for example, production of a relevant disease register. The contract is renegotiated annually, and QOF indicators and targets are updated as agreed between the negotiating parties. The domains covered by QOF indicators include the following (NHS Employers, 2011):

- *Clinical care*: the domain consists of 87 indicators across 20 mostly chronic disease clinical areas (e.g. coronary heart disease, heart failure, hypertension) for a maximum of 661 points. Several indicators are related to whether chronic diseases are well controlled (e.g. per cent of patients with coronary heart disease with their blood pressure under control).
- *Organizational*: the domain consists of 45 indicators across five organizational areas – records and information; information for patients; education and training; practice management and medicines management. The organizational domain has a maximum total of 262 points.
- *Patient experience*: the domain consists of one indicator worth up to 33 points that is related to the length of GP consultations.
- *Additional services*: the domain consists of nine indicators across four service areas, which include cervical screening, child health surveillance, maternity services, and contraceptive services. The additional services domain has a maximum of 44 points.

Examples of indicators in each domain with their point values are presented in Table 12.1. The points are distributed to weight indicators more heavily that have a higher estimated workload, many of which are closer to outcomes.

Table 12.1 Examples of indicators in the four performance domains of the UK QOF, 2011–12

Clinical care (example – secondary prevention of coronary heart disease)

- The practice can produce a register of patients with coronary heart disease (4 points).
- For patients with newly diagnosed angina, the per cent who are referred for specialist assessment (7 points).
- The per cent of patients with coronary heart disease whose last measured total cholesterol (measured in the previous 15 months) is 5 mmol/l or less (17 points).
- The per cent of patients with coronary heart disease with a record in the preceding months that aspirin, an alternative anti-platelet therapy, or an anti-coagulant is being taken (7 points).
- The per cent of patients with coronary heart disease who are currently treated with a beta blocker (7 points).
- The per cent of patients with a history of myocardial infarction currently treated with an ACE inhibitor (or ARB if ACE intolerant), aspirin or an alternative anti-platelet therapy, beta-blocker and statin (10 points).
- The per cent of patients with coronary heart disease who have had influenza immunization in the preceding 1 September to 31 March (7 points).

Organizational (example – practice management)

- Individual health care professionals have access to information on local procedures relating to child protection (1 point).
- There are clearly defined arrangements for backing up computer data, back-up verification, safe storage of back-up tapes and authorization for loading programmes where a computer is used (1 point).
- The hepatitis B status of all doctors and relevant practice-employed staff is recorded and immunization recommended if required in accordance with national guidance (0.5 points).

(*continued*)

Table 12.1 Examples of indicators in the four performance domains of the UK QOF, 2011–12 (*continued*)

- The practice offers a range of appointment times to patients, which as a minimum should include five mornings and five afternoons per week, except where agreed with the PCO (3 points).
- The practice has systems in place to ensure regular and appropriate inspection, calibration, maintenance and replacement of equipment (3 points).
- The practice has a protocol for the identification of carers and a mechanism for the referral of carers for social services assessment (3 points).
- There is a written procedures manual that includes staff employment policies (2 points).

Patient experience

- The length of routine booked appointments with the doctors in the practice is not less than ten minutes (If the practice routinely sees extras during booked surgeries, then the average booked consultation length should allow for the average number of extras seen in a surgery session. If the extras are seen at the end, then it is not necessary to make this adjustment). For practices with only an open surgery system, the average face-to-face time spent by the GP with the patient is at least eight minutes. Practices that routinely operate a mixed economy of booked and open surgeries should report on both criteria (33 points).

Additional services (example – cervical screening)

- The per cent of women aged from 25 to 64 in England and Northern Ireland, 20 to 60 in Wales, and from 20 to 64 in Wales whose notes record that a cervical screening test has been performed in the last five years (11 points).
- The practice has a system for informing all women of the results of cervical smears (2 points).
- The practice has a policy for auditing its cervical screening service, and performs an audit of inadequate cervical smears in relation to individual smear-takers at least every two years (2 points).
- The practice has a protocol that is in line with national guidance and practice for the management of cervical screening, which includes staff training, management of patient call/recall, exception reporting and the regular monitoring of inadequate smear rates (7 points).

Source: NHS Employers, 2012.

For example, identifying patients with coronary heart disease is worth four points, while the percentage of patients with specific diagnostic information recorded is worth seven points, and the percentage of patients with measured blood pressure below an acceptable threshold is worth 17 points. The patient experience indicator has a high point value (33 points), while organizational indicators tend to have point values below 10.

The overall distribution of points across domains (and organizational sub-domains) is shown in Figure 12.2. The points, which all carry equal monetary value, are heavily distributed toward clinical indicators, with 67 per cent of all possible points in this domain (up from 52 per cent when the QOF began in 2004). The indicators and achievement thresholds were revised substantially by NICE for the 2012/2013 QOF, with a number of indicators retired and updated.

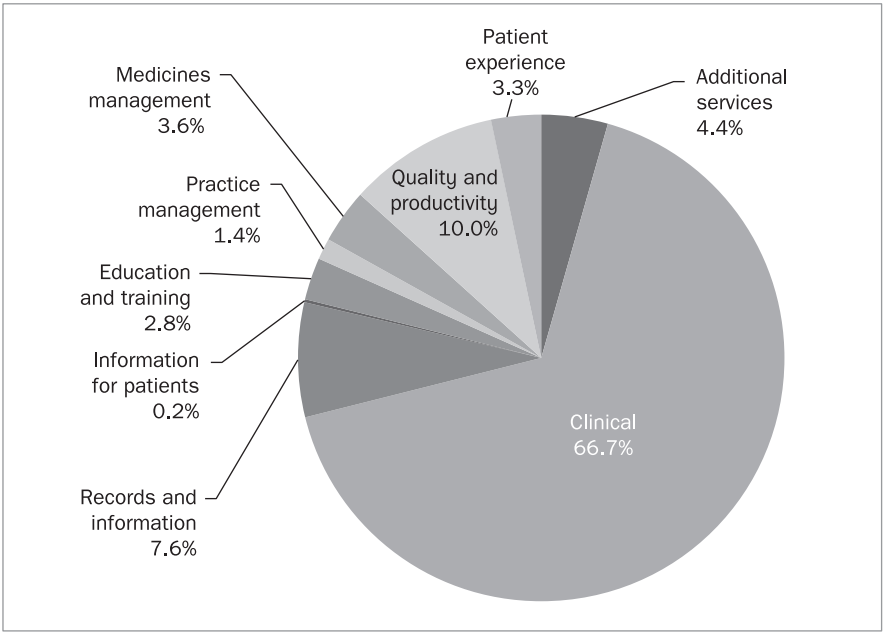


Figure 12.2 Distribution of points across performance domains in the UK QOF, 2011–12

Incentive payments

Incentive payments are made to GP practices on an annual basis. Practices are paid a flat rate for each point they achieve (£127 per point in 2010/11 increased to £133.76 in 2012/13). The reward is capped at a maximum of 1000 points and the corresponding total bonus amount. Payments are adjusted for practice size and disease prevalence relative to national average values (Mason et al., 2008), but the programme has been criticized for not adequately compensating the extra work required to achieve quality targets in deprived areas (Hutchinson, 2008).

The QOF does allow GP practices to 'exception-report', or exclude certain patients from the calculation of achievement scores. Exceptions are intended to avoid penalizing practices for reaching out to more complicated patients who could potentially reduce their indicator scores, and to exclude patients who are not suitable for the standard course of treatment rewarded by the QOF. Patient exception reporting applies to those indicators in the clinical domain where the level of achievement is determined by the percentage of patients receiving the designated level of care. Exception reporting also applies to one cervical screening indicator in the additional services domain. Patients can be exception-reported from individual indicators if, for example, they do not attend appointments or where the recommended treatment is judged as inappropriate by the GP (such as medication cannot be prescribed due to side

effects). Some exception-reporting is done automatically by the electronic data systems that are used for the QOF, specifically for patients who are recently registered with a practice or who are recently diagnosed with a condition (Health and Social Care Information Centre, 2011). The average exception rate overall is approximately five per cent of patients (NHS Information Centre, Prescribing Support Unit, 2009b; NHS Information Centre, Prescribing & Primary Care Services, 2011).

Data sources and flows

Data to calculate achievement scores mainly are extracted from electronic medical records into the Quality Management Analysis System (QMAS), a national system developed by NHS Connecting for Health specifically to support the QOF. Providers enter patient-level data directly into the electronic medical records during the consultation, which is fed into the information sent to QMAS (McDonald, 2009). Reports are run by the QMAS to calculate individual practices' QOF achievement and reward payments. Other supporting information is submitted by the GP practices to the PCOs as needed.

Data relating to most of the organizational indicators cannot be automatically extracted, and the practices must enter much of the information manually on the QMAS website. The QOF guidance documents outline the types of evidence required for non-clinical indicators, which includes, for example, a 'report on the results of a survey of a minimum of 50 medical records of patients who have commenced a repeat medication', and a report of 'the results of a survey of the records of newly registered patients'. There are at least 15 such reports specified in the guidance documents, with about half that need to be generated each QOF period and half that are one-off reports of policies and procedures which would not change every QOF period (NHS, 2010).

There is no patient-specific data in QMAS, because this is not required to support the QOF. For example, QMAS captures aggregate information for each practice on patients with coronary heart disease and on patients with diabetes, but it is not possible to identify or analyse information about individual patients (NHS Information Centre, Prescribing Support Unit, 2009a). The achievement scores are calculated automatically by specialized software (Checkland, Marshall & Harrison, 2004). PCOs are currently required to carry out pre-payment verification checks on all practices and formally audit a five per cent sample of practices (UK National Audit Office, 2008).

Reach of the programme

Which providers participate and how many people are covered?

The QOF is a national programme and although it is a voluntary programme, nearly all GP practices in the UK participate. In 2011/12 the programme covered 8123 GP practices and almost 100 per cent of registered patients (NHS Information Centre, Prescribing Support Unit, 2009a; NHS Information Centre, Prescribing & Primary Care Services, 2011).

The reach of the QOF is also significant as a source of financing for GP practices. The average additional income from the QOF per GP practice was £74,300 in 2004–05 and £126,000 in 2005–06. The QOF continues to make up on average 20 per cent of annual GP practice income. The size of the reward is considered to be large by international standards. In fact, no other country experimenting with quality incentives is tying as large a proportion of provider income to quality of care (Campbell et al., 2007). GP partners benefited most from the new income, with individual incomes rising by 58 per cent in the first three years. Incomes of salaried GPs and nurses have not increased significantly (UK National Audit Office, 2008).

Improvement process

How is the programme leveraged to achieve improvements in service delivery and outcomes?

Unlike most other P4P programmes, the QOF attempts to establish a traceable pathway between the incentives in the QOF, provider performance for specific processes of care, and better outcomes. For example, for 2011/12 indicators related to coronary heart disease covered primary prevention (two indicators), recording of patients who have been diagnosed (one indicator), diagnosis and initial management (one indicator), ongoing management (four process indicators), and clinical outcomes (two indicators). Although it is assumed that better clinical outcomes (such as controlled blood pressure) translate into better health outcomes (reduced emergency services, and hospital admissions, and mortality), this has not been supported empirically (Downing et al., 2007). It also has been argued that the interventions which receive higher point values are not those interventions that bring the greatest health gain (Fleetcroft & Cookson, 2006).

GP practices have made internal changes to orient their services more clearly around the targets set in the QOF. New staff structures and the more prominent role of IT seem to be the main vehicles for this change. The NHS does not provide any guidance on how bonus payments are used or distributed among the staff of GP practices (UK National Audit Office, 2008). Some of the additional funding is being reinvested by GP practices to improve patient care, although it is not possible to quantify how much of overall reinvestment by practices in patient services is attributable to their increased QOF income. A portion of the additional funding is also being used by the GP practices to employ more staff to specifically focus on some of the QOF targets, such as increased employment of nurses for chronic disease management, data entry clerks to manage additional data collection processes, and 'health care assistants' to carry out health promotion (Roland, 2006). Most practices set up 'QOF teams' to ensure the systems are in place to collect the necessary data, conduct internal audits to ensure targets are being met, and setting up call and recall systems for patients.

The upgrading of computer systems and increased role of IT in GP practices has been supported by the QOF, which has been used to a large extent in the

quality improvement process within the practices. In 2004 alone 30 million GBP additional capital funding was made available to PCOs to support the upgrading of clinical data systems and to provide systems for non-computerized practices (NHS, 2004). The process of recording and using data to manage patient care has had benefits beyond the clinical areas rewarded by the QOF. One study found that rates of recording increased for all risk factors (i.e. including those not incentivized by QOF), with a 'spillover' effect of 11 per cent increased recording rate for other, unincentivized factors in targeted patients (Sutton et al., 2010). There also has been an increase in the use of computerized templates to guide clinicians and to assist in collecting data during consultations (Campbell et al., 2007).

The GP practices get some direct external support for their improvement processes through the annual QOF verification visit by the PCO team. In addition to verifying the practice's records, the visit is used to discuss the practice's future plans within the QOF, including the following year's goals. This part of the visit can also include discussion of the learning, support and development needs of the practice to achieve higher quality (NHS, 2004; Cashin & Vergeer, 2013).

Finally, the public reporting of GP practice performance within the QOF is used as an additional lever to drive performance improvement. The NHS Information Centre for health and social care (NHS IC) maintains an online database to allow public access to the performance of GP practices against QOF indicators (UK National Health Service, n.d.).

Results of the programme

Has the programme had an impact on performance, and have there been any unintended consequences?

Performance related to specific indicators

Since the QOF began in 2004, the GP practices have consistently achieved high scores relative to performance targets. The achievement rate in England was 91 per cent in 2004/05 and increased to 96.2 per cent in 2005/06, and it has remained at 94–97 per cent ever since. The achievement rate across performance domains for England in 2008–2012 is presented in Figure 12.3. All of the domains show achievement rates above 95 per cent, with the exception of patient experience. The patient survey-based indicator was retired at the end of 2010, leaving only one indicator for patient experience related to average consultation length. The achievement rate increased immediately to nearly 99 per cent with this change.

It is not clear whether the high rates of performance achievement for the QOF translate into improved overall patient care and health outcomes. Some data suggest the introduction of the QOF has shown moderate improvements in processes and outcomes for patient care in some long-term conditions such as asthma and diabetes (Campbell et al., 2007; Vamos et al., 2011). A more recent study found that the introduction of financial incentives was associated

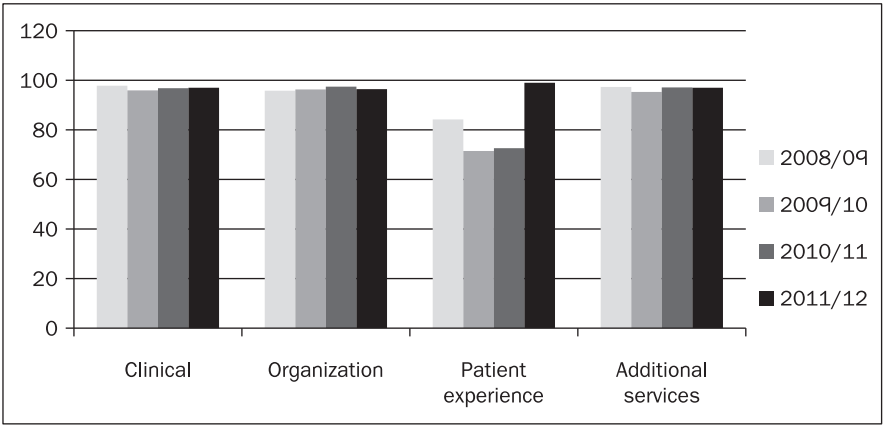


Figure 12.3 Share of total points achieved by GP practices across UK QOF performance domains, 2008–12

Source: NHS Information Centre, 2009–2012.

with improvements in the quality of diabetes care in the first year, but these improvements mostly related to documentation of recommended aspects of clinical assessment, not patient management or outcomes of care. Improvements in subsequent years were more modest (Kontopantelis et al., 2013). There is no evidence of an effect on health outcomes. One study assessed the impact of the incentives and targets on quality of care and health outcomes for 470,000 British patients with hypertension and found that they had no impact on rates of heart attacks, kidney failure, stroke or death (Serumaga et al., 2011).

For coverage of preventive services, there is evidence only that Influenza immunization rates increased significantly since the QOF began. Influenza immunization increased from 67.9 to 71.4 per cent between 2003/04 and 2006/07. Rates of increase were higher for populations with previously low immunization rates (e.g. up to 16 percentage point increase for individuals under 65 years of age with previous stroke (Norbury, Fawkes & Guthrie, 2011).

Programme monitoring and evaluation

A 2008 study by the National Audit Office (NAO) assessed the performance of the QOF against the expected benefits listed in the business case for the new GP contract, including the QOF. The NAO study found that progress so far had been modest overall. 'Good progress' was found only for participation of GP practices in the QOF programme and the effect on recruitment and retention of GPs. 'No progress' was found for the objectives of increasing NHS productivity and redesigning services around patients. 'Some progress' was found for the remaining areas, including rewarding high quality care (UK National Audit Office, 2008).

Aside from the few published studies that analyse the effect of a subset of indicators, there is no comprehensive time series (pre- and post- measures)

or control group evaluation available for the QOF, so it has been difficult to determine the extent to which QOF has rewarded GPs for what they were already doing, what they would have done anyway, what they would have done on the basis of transparent feedback alone, and what they did in response to financial incentives (Hutchinson, 2008). The changes that have been observed since the QOF began in 2004 are further confounded by the overall increase in funding for primary care and other quality improvement measures (such as new standards of care) that accompanied the incentive programme.

Equity

Although not an explicit objective of the QOF P4P programme, there may be some positive impacts on equity in health care. QOF performance is slightly lower in deprived areas (UK National Audit Office, 2008), but there is evidence of some 'catch-up' (Doran et al., 2008). The difference in mean QOF score in the least and most deprived quintiles fell from 64.5 points (2004/05) to 30.4 (2005/06) (Ashworth et al., 2007). A systematic review of the equity effects of the QOF found small but significant differences that favoured less deprived groups, but these differences were no longer observed after correcting for practice characteristics (Boeckxstaens et al., 2011).

Costs and savings

The QOF is expensive, about £1 billion per year, and has in the past contributed to higher than expected increases in GPs' personal take-home pay. Budget overruns were a particular problem in the initial years when achievement rates were significantly higher than expected. The QOF was not piloted before it was introduced and there were no baseline estimates for the indicators, so the performance levels and potential budget requirements were underestimated. Expenditures have remained at around £1 billion per year, and with better planning budget overruns have steadily declined.

Even when the QOF appears to drive better processes of care, there is no evidence of related cost savings. For example, although providers are rewarded under the QOF for prescribing medicines that are cost effective, higher quality scores related to prescribing are not associated with lower spending on medicines (Fleetcroft et al., 2011). In fact, higher quality scores were associated with slightly higher costs in five prescribing areas: influenza vaccination, beta blockers, angiotensin converting enzyme inhibitors, lipid lowering, and antiplatelet treatment. Higher quality scores were associated with slightly lower prescribing costs only for hypertension and smoking cessation.

Provider response

There are mixed conclusions about how GPs have perceived the QOF based on several small surveys and qualitative studies. One small qualitative study found that most physicians had a generally positive view of the QOF.

The GPs regarded the incentive payment as a financial reward in return for extra work. They also recognized the value of the incentive and believed that the quality targets had improved patient care by focusing attention on necessary clinical activities that might have been neglected (Campbell, MacDonald & Lester, 2008). On the other hand, the physicians interviewed for that study also noted the emergence of potentially competing 'agendas' during office visits if patient concerns do not relate to activities that are tied to the incentive.

Some candid responses in the qualitative study and data reported by the NAO show that in fact GPs may be compensated disproportionately more than the extra work required by the QOF, and much of that extra work is being passed on to nurses and other staff. The NAO study found that GPs are working, on average, almost seven hours *less* per week and their pay has significantly increased. On the other hand, the total number of consultations in GP practices has increased, and the average length of a GP consultation has increased. The main reason for this change is that the total number, and overall proportion, of consultations carried out by practice nurses has increased (UK National Audit Office, 2008). There is some evidence that GP practices may be diverting resources away from activities that are not rewarded under the QOF. The NAO study found that 75 per cent of GPs believed that they spend more time on areas which attract QOF points and significantly less time on areas which were less likely to be rewarded under QOF (UK National Audit Office, 2008). Furthermore, although there is no evidence of widespread gaming of the QOF, there have been cases of classifying patients with borderline clinical measures or laboratory values as having a condition covered by the QOF (Mangin & Troop, 2007), and inappropriate exclusion of patients for whom GPs have missed (or are likely to miss) the QOF targets (Doran et al., 2006; UK National Audit Office, 2008).

Overall conclusions and lessons learned

Has the programme had enough of an impact on performance improvement to justify its cost?

Overall, the aims of the UK QOF are being met in terms of some improvements in disease-specific processes of patient care and physician income, as well as improved data availability and use. Furthermore, the QOF is not implemented in isolation, but rather as part of a comprehensive strategy to improve provider performance and quality throughout the NHS. The costs are high, but a large investment in primary care was planned in the 2000 NHS Plan, and the QOF serves to link this investment to more rigorous performance standards and greater accountability.

The investment in infrastructure to generate and use better data has been an important underpinning and outgrowth of the programme. In fact, one of the most widely acknowledged positive spillover effects of the QOF P4P programme is a general improvement in available data, which can be used to improve quality overall (Galvin, 2006; Cashin & Vergeer, 2013). The increased

use of computerized templates to guide clinicians and to assist in collecting data during consultations also could have more general positive impacts on overall quality of care (Campbell et al., 2007).

The QOF has taken root, and if there is widespread opposition or discontent on the part of providers, it has not been voiced in an organized way. The perceived validity of most of the indicators, which are based on accepted clinical guidelines, and general professional commitment to evidence-based practice have contributed to the acceptance of the programme (Wilson, Roland & Ham, 2006; McDonald, 2009). The involvement of NICE in indicator refinement may further strengthen the clinical validity of the indicators and acceptance by providers. In addition, the ground had already been prepared for a significant pay for performance component to be added to the GP contract. The QOF was layered on a series of quality initiatives beginning in the 1990s that were associated with substantial improvements in quality of care during the period leading up to the QOF (Campbell et al., 2005), and GPs already had some experience with financial incentives from the limited use of incentive programmes that were initiated in 1990 (Campbell et al., 2007). The major concerns about the QOF, however, include the following:

1. *The high cost of the programme and large share of physician income tied to the incentives.* The absence of a pilot programme and adequate forecasting led to budget overruns in the initial phase of the QOF. A large budget was set aside for the QOF, and even so the lack of a pilot or financial risk forecasting led to overruns. The QOF overexpenditure may be crowding out expenditure on other quality initiatives (UK National Audit Office, 2008), and the cost of this trade-off has not been assessed. Furthermore, the programme represents a large share of physician income, so the incentives that are created have the potential not only to drive performance improvement, but also to distort provider behaviour and practice management.
2. *The enormous scale of the programme, both in absolute expenditure and relative share of GP income, is not linked to improved health outcomes.* There is still no evidence that the high expenditure on QOF can be linked to improvements in health outcomes. The high expenditure on the programme makes it critical to be sure that the performance improvement is not achieved at the expense of other more valuable initiatives, services, or non-measurable aspects of patient care.

A rigorous evaluation of the QOF that can provide a satisfactory assessment of whether the QOF overall provides value for money has not been done so far. The studies that have been completed have failed to show more than modest effects on quality and patient outcomes. In general, there is the opinion that the NHS has paid more than necessary to achieve high performance against the targets. One of the benefits of the QOF, however, has been the transparent processes that have been put in place to constantly improve the programme, and specifically the indicators. There is an entire infrastructure in place to provide tools for PCOs and providers to make better use of the QOF. These processes and tools may allow the QOF to continue to evolve in order to better exploit the potential of the resources, information and incentives in the programme to improve patient care beyond.

Note

* This case study is based on the 2011 report *RBF in OECD Countries: United Kingdom: Quality and Outcomes Framework* prepared by Cheryl Cashin for the International Bank for Reconstruction and Development and The World Bank.

1 The general term Primary Care Organization (PCO) is used throughout QOF guidance documents, because the organization responsible for contracting primary care services is different in the three different countries. In England the organization is Primary Care Trusts (PCTs), Local Health Boards in Scotland and Wales, and Health and Social Care Board in Northern Ireland.

References

Ashworth, E. et al. (2007) The relationship between social deprivation and the quality of primary care: a national survey using indicators from the UK Quality and Outcomes Framework, *British Journal of General Practice*, 57: 441–8.

Boeckxstaens, P. et al. (2011) The equity dimension in evaluations of the quality and outcomes framework: a systematic review, *BMC Health Services Research*, 11: 209–24.

Campbell, S. et al. (2005) Improvements in the quality of clinical care in English general practice: longitudinal observational study, *British Medical Journal*, 331: 1121–3.

Campbell, S. et al. (2007) Quality of primary care in England with the introduction of pay for performance, *New England Journal of Medicine*, 357(2): 181–90.

Campbell, S., MacDonald, R. and Lester, H. (2008) The experience of pay for performance in English family practice: a qualitative study, *Annals of Family Medicine*, 6(3): 228–34.

Cashin, C. and Vergeer, P. (2013) *Verification of performance in results-based financing: the case of the U.K. QOF*. Washington, DC: World Bank.

Checkland, K., Marshall, M. and Harrison, S. (2004) Re-thinking accountability: trust versus confidence in medical practice, *Quality and Safety in Health Care*, 13: 130–5.

Doran, T. et al. (2006) Pay-for-performance programs in family practices in the United Kingdom, *New England Journal of Medicine*, 355: 375–84.

Doran, T. et al. (2008) Exclusion of patients from pay-for-performance targets by English physicians, *New England Journal of Medicine*, 359: 274–84.

Downing, A. et al. (2007) Do the UK government's new quality and outcomes framework (QOF) scores adequately measure primary care performance? A cross-sectional survey of routine healthcare data, *BMC Health Services Research*, 7: 166–72.

Fleetcroft, R. and Cookson, R. (2006) Do the incentive payments in the new NHS contract for primary care reflect likely population health gains?, *Journal of Health Services Research and Policy*, 11(1): 27–31.

Fleetcroft, R. et al. (2011) Correlation between prescribing quality and pharmaceutical costs in English primary care: national cross-sectional analysis, *British Journal of General Practice*, 61(590): e556–64.

Galvin, R. (2006) Pay-for-performance: too much of a good thing? A conversation with Martin Roland, *Health Affairs*, 25(5): 412–19.

Government of the UK (2000) *The NHS plan*. London: Stationery Office.

Health and Social Care Information Centre (2011) *Quality outcomes framework (QOF): frequently asked questions (FAQs)*. Leeds: Health and Social Care Information Centre.

Hutchinson, B. (2008) Pay for performance in primary care: proceed with caution, pitfalls ahead, *Health Care Policy*, 4(1): 10–15.

Kontopantelis, E. et al. (2013) Recorded quality of primary care for patients with diabetes in England before and after the introduction of a financial incentive scheme: a longitudinal observational study, *BMJ Quality and Safety*, 22(1): 53–64.

McDonald, R. (2009) Pay for performance in primary care in England and California: comparison of unintended consequences, *Annals of Family Medicine*, 7(2): 121–7.

Mangin, D. and Troop, L. (2007) The quality and outcomes framework: what have you done to yourselves?, *British Journal of General Practice*, 57: 435–7.

Mason, A. et al. (2008) *The GMS quality and outcomes framework: are the quality and outcomes framework (QOF) indicators a cost-effective use of NHS resources*? York: Centre for Health Economics, University of York.

Middleton, E. and Baker, D. (2003) Comparison of social distribution of immunisation with measles, mumps, and rubella vaccine, England, *BMJ*, 326: 854.

NHS (2004) *General guidance on annual QOF review process*. London: UK National Health Service.

NHS (2010) *QOF management guide*, Vol. 4. London: UK National Health Service.

NHS Employers (2011) *Changes to QOF 2011/12*. London: UK National Health Service.

NHS Information Centre, Prescribing Support Unit (2009a) *Quality and outcomes framework achievement data 2008/09*. Leeds: Health and Social Care Information Centre.

NHS Information Centre, Prescribing Support Unit (2009b) *Quality and outcomes framework exception data 2008/09*. Leeds: Health and Social Care Information Centre.

NHS Information Centre, Prescribing and Primary Care Services (2011) *Quality and outcomes framework achievement data 2010/11*. Leeds: Health and Social Care Information Centre.

NICE (2010) *National Institute for Health and Clinical Excellence* (http://www.nice.org.uk/aboutnice/qof/indicatorsindevelopment.jsp, accessed May 2010).

Norbury, M., Fawkes, N. and Guthrie, B. (2011) Impact of the GP contract on inequalities associated with influenza immunisation: a retrospective population-database analysis, *British Journal of General Practice*, 61(588): e379–85.

OECD (2010) *OECD health data 2010* (http://www.oecd.org/document/11/0,3343,en_21571361_44315115_45549771_1_1_1_1,00.html, accessed November 2010).

Rawlins, M. and Moore, V. (2009) Helping to provide high quality care in primary care, *Ulster Medical Journal*, 78(2): 82–3.

Roland, M. (2006) Financial incentives to improve the quality of primary care in the UK: predicting the consequences of change, *Primary Care Research and Development*, 294: 18–26.

Serumaga, B. et al. (2011) Effect of pay for performance on the management and outcomes of hypertension in the United Kingdom: interrupted time series study, *British Medical Journal*, 342: d108.

Stevens, S. (2004) Reform strategies for the English NHS, *Health Affairs*, 23(3): 37–44.

Sutton, M. et al. (2010) Record rewards: the effects of targeted quality incentives on the recording of risk factors by primary care providers, *Health Economics*, 19: 1–13.

UK Department of Health (n.d.) *Performance ratings* (http://www.performance.doh.gov.uk/performanceratings/index.htm, accessed November 2010).

UK National Audit Office (2008) *NHS pay modernisation: new contracts for general practice services in England*. London: The Comptroller and Auditor.

UK National Health Service (2004) National programme for information technology, *GMS IT-QMAS Bulletin*.

UK National Health Service (n.d.) *Quality and outcomes framework: online GP practice results database* (http://www.qof.ic.nhs.uk/, accessed May 2010).

Vamos, E. et al. (2011) Association of practice size and pay-for-performance incentives with the quality of diabetes management in primary care, *CMAJ*, 183(2): e809–16.

Wilson, T., Roland, M. and Ham, C. (2006) The contribution of general practice and the general practitioner to NHS patients, *Journal of the Royal Society of Medicine*, 99: 24–8.

World Bank (2013) *World development indicators* (http://data.worldbank.org/indicator/SH.XPD.TOTL.ZS/countries?page=3, accessed November 2013).

chapter thirteen

United States: California integrated healthcare association physician incentive programme

Meredith Rosenthal

Introduction

Approximately two-thirds of Americans obtain their health insurance coverage through private companies. Across the nation, there are hundreds of private insurance carriers that market and sell thousands of different insurance products. The regulation of private insurance, which is generally focused on acceptable benefit packages and underwriting practices, is left largely to the states, and there is little standardization among private insurers in terms of the method or amount of payment to health care providers. Such a fragmented financing environment poses a substantial challenge to any payer seeking to employ financial or other incentives to encourage providers to improve quality, reduce waste, or achieve other objectives.

Like the public sector Medicare programme, which is the largest payer in the US, most private insurers reimburse physicians based on fee schedules, the levels of which vary both among payers and within a payer across providers. Most insurers pay hospitals and other facilities separately, mainly using case-based payment systems (e.g. the Diagnosis Related Group system) or per diem (bed-day) payment. It is widely acknowledged that these volume-based payment approaches fail to encourage the delivery of high quality care. In some geographic and product markets, US insurers use capitation to pay providers for all or most services. Capitation alone, however, is unlikely to encourage high quality, because incentives to control costs are more likely to produce short-run efforts to eliminate costly services rather than investments in prevention, which might pay off more slowly. This is particularly true because patient populations in the US change insurance carriers and providers frequently, which limits the ability of insurers to take a longer view of health care investments and costs.

For decades, research in the US has documented a shortfall in health care quality along a number of dimensions, including primary and secondary prevention, patient safety, patient experience, and equity. In 2001, the Institute

of Medicine, an influential quasi-governmental body, issued the *Crossing the Quality Chasm* report summarizing evidence of pervasive quality problems in the US delivery system and offering a series of recommendations. One of these recommendations was to address the failure of current provider payment systems to reward quality and value. The *Crossing the Quality Chasm* report was extremely influential not only with public sector payers, but also in the private health care purchasing sector. While government programmes have moved methodically towards pay for performance (also known as value-based purchasing in the US), some of the first major initiatives to experiment with these new incentives were organized by private insurers.

One of the first, and perhaps the largest, private pay for performance (P4P) initiatives of this era was launched by the Integrated Healthcare Association (IHA) in 2001 with eight health plans representing ten million members in California. IHA, a multi-stakeholder organization, is responsible for convening participants to establish measurement and reporting rules, collecting data, applying a common set of performance measures, and reporting results for several hundred physician groups.[1] The IHA programme is of particular interest not only because of its size, but also because it has been sustained for more than a decade and has been independently evaluated.

Health policy context

What were the issues that the programme was designed to address?

Health policy context

While the Institute of Medicine report awakened health purchasers in the US to widespread quality problems, there was at the same time a so-called 'backlash' against the concept of managed care and, in particular, the use of financial incentives for providers to limit high-cost care. Thus, the IHA P4P programme can be viewed as an attempt to address both the specific quality deficits that had been identified by experts, and also the perception that payers and providers were excessively focused on cost control. Consistent with these goals, the initial focus of the IHA programme was on addressing underuse of evidence-based care (e.g. childhood immunization and screening), patient experience, and the adoption of health information technology.

Concurrently with these changes in the targeting of financial incentives, there was increasing emphasis both in California and nationally on quality measurement and public reporting for physicians and other health care providers. In California, a large and influential employer purchasing coalition (the Pacific Business Group on Health) had begun collecting and reporting comparative data on physician groups. Individual health insurers also had developed their own public report cards to encourage quality improvement and

spur informed consumer choice of provider. These payer-specific report cards typically relied on claims data and varied widely in terms of measure selection and method of presentation.

The proliferation of competing quality measurement and performance reporting systems caused concern among physician groups about the potential for confusion on the part of consumers and dilution of focus for provider quality improvement efforts. Moreover, single-payer measurement systems were more likely to encounter small sample problems, since most insurers in California captured relatively small market shares across the providers in their networks. By early 2000, there was growing support for aligning the various health plan and purchaser performance measurement and incentive efforts in California. Under the auspices of the IHA, payers, providers, and a variety of other stakeholders began to build a coordinated statewide initiative to measure and reward quality.

Overall the IHA programme aims to achieve quality improvement using three tactics: (1) a common set of measures; (2) a public report card; (3) health plan incentive payments that vary across payers but are aligned to a high degree. The adoption of a common set of performance measures for use by all health plans as the basis for reward and recognition reduces confusion and increases the impact of each payer's incentives. Moreover, the aggregation of data across all participating health plans improves not only the statistical properties of measurement due to sample size enhancements, but also the confidence of physician groups in the results.

Stakeholder involvement

The planning phase and design for a statewide P4P initiative were completed in late 2001, with funding and leadership by the California HealthCare Foundation, a charitable organization whose mission is to support ideas and innovations to improve the health of Californians (Integrated Healthcare Association, 2006). Six health plans initially endorsed the IHA initiative, agreeing to a common set of measures and uniform reporting: Aetna, Blue Cross of California, Blue Shield of California, CIGNA HealthCare of California, Health Net, and PacifiCare (now UnitedHealthcare). The group was later joined by Western Health Advantage and the Permanente Medical Group. Permanente is a physician organization that exclusively contracts with the Kaiser Foundation Health Plan, and it participates in public reporting only (Integrated Healthcare Association, 2006).

The IHA P4P programme is a private, voluntary initiative with government involvement limited to the public reporting of results for consumer use. Programme governance is provided by the IHA Board of Directors. The programme is managed by IHA P4P Steering, Executive, and Technical Committees, with assistance from the Pacific Business Group on Health (PBGH), the National Committee for Quality Assurance (NCQA), and other technical experts.[2] Within IHA, there is prominent representation of physicians, including the leadership of the major participating groups, insurers, other large purchasers (e.g. PBGH), and consumer groups.

Technical design

How does the programme work?

The IHA programme is a framework for pay for performance, which includes measure selection, technical specification, a data aggregation process, public reporting, and high-level guidelines about payment methodology.[3] The extent to which individual insurers use the IHA measures and data is optional and varies across plans, with the programme's intent to encourage harmonization where possible.

Performance domains and indicators

The initial measurement set included three domains with 25 individual measures in the areas of clinical quality, patient experience, and health information technology use. Over time, the number of measures has increased and broadened in scope. For measurement year 2012, there are four domains that are recommended for use in P4P, including clinical quality, health information technology, patient experience and resource use (Table 13.1).

While the initial measure set focused largely on process measures of quality associated with underuse of evidence-based care, the current version includes intermediate health outcomes (such as blood pressure control) and overuse measures such as appropriate antibiotic treatment. IHA also recommends weighting for each domain as part of the effort to harmonize P4P across payers. Domain weights have changed over time as well, with increasing emphasis on clinical quality. Rewards associated with resource use measures are framed in terms of 'shared savings' with payers rather than as a component of bonuses.

Shared savings approaches typically calculate rewards as a percentage of the amount by which actual spending is lower than expected, using an actuarial formula that takes patient characteristics in the assigned population and secular trends in health care spending into account.

Incentive payments

The consolidated performance results are used by health plans to calculate bonuses distributed each year. Each plan determines its own P4P budget and methodology for calculating bonus payments to the physician groups. The IHA suggests a Standard Payment Methodology, in which physician groups are scored on both attainment and improvement for each measure. The higher of the two is summed across all measures in the domain to calculate a domain total, which is then weighted as described in Table 13.1.

Each year, the IHA releases a 'transparency report' detailing the measures and methodology used by each insurer to calculate incentive payments. The report documents the percentage of each plan's aggregate P4P payments accounted for by IHA measures, total dollars paid, and the specific formula used to compute payments, among other details. These reports suggest that

Table 13.1 Approved measurement set in the California IHA, 2012

Domain	*Weighting*	*Measures approved for payment*[4]
Clinical Quality	50 per cent	Cardiovascular 1. Annual Monitoring for Patients on Persistent Medications – ACEI/ARB, Digoxin and Diuretics 2. Cholesterol Management – LDL Screening 3. Cholesterol Management – LDL Control < 100 4. Proportion of Days Covered by Medications – ACEI/ARB 5. Proportion of Days Covered by Medications – Statins Diabetes Care 1. HbA1c Testing 2. HbA1c Poor Control > 9.0 per cent 3. HbA1c Control < 8.0 per cent 4. HbA1c Control < 7.0 per cent for a Selected Population 5. LDL Screening 6. LDL Control < 100 7. Nephropathy Monitoring 8. Blood Pressure Control < 140/90 9. Optimal Diabetes Care Combination 1 – LDL < 100, HbA1c < 8.0 per cent, Nephropathy Monitoring 10. Proportion of Days Covered by Medications – Oral Diabetes Medications Musculoskeletal 1. Use of Imaging Studies for Low Back Pain Prevention 1. Childhood Immunization Status – 24-mo Continuous Enrollment: Combination of all Antigens 2. Immunizations for Adolescents – Tdap 3. HPV Vaccination for Female Adolescents 4. Chlamydia Screening in Women – Ages 16–24 5. Evidence-Based Cervical Cancer Screening – Appropriately Screened 6. Breast Cancer Screening – Ages 50–69 7. Colorectal Cancer Screening Respiratory 1. Asthma Medication Ratio – Ages 5–64 2. Appropriate Testing for Children with Pharyngitis 3. Appropriate Treatment for Children with Upper Respiratory Infection 4. Avoidance of Antibiotic Treatment of Adults with Acute Bronchitis
Meaningful Use of HIT	30 per cent	1. Use CPOE for medication orders directly entered by any licensed healthcare professional who can enter orders into the medical record per state, local and professional guidelines

(*continued*)

Table 13.1 Approved measurement set in the California IHA, 2012 (*continued*)

Domain	*Weighting*	*Measures approved for payment*[4]
		2. Implement drug-drug and drug-allergy interaction checks 3. Maintain up-to-date problem list of current and active diagnoses 4. Generate and transmit permissible prescriptions electronically (eRx) 5. Maintain active medication list 6. Maintain active medication allergy list 7. Record demographics 8. Record and chart changes in vital signs 9. Record smoking status 10. Report ambulatory clinical quality measures 11. Implement one clinical decision support rule relevant to specialty or high clinical priority, along with the ability to track compliance with that rule 12. Provide patients with an electronic copy of their health information 13. Provide clinical summaries for patients at each office visit 14. Capability to exchange key clinical information 15. Protect electronic health information created or maintained by the certified EHR technology 16–20. Any (5) CMS/ONC Menu set measures 21. Chronic Care Management for Diabetes, Depression and one other Clinically Important Condition 22. Within-PO Performance Variation
Patient Experience	20 per cent	1. Doctor–Patient Interaction Composite for PCPs 2. Doctor–Patient Interaction Composite for Specialists 3. Coordination of Care Composite 4. Timely Care and Service Composite for PCPs 5. Timely Care and Service Composite for Specialists 6. Overall Ratings of Care Composite 7. Office Staff Composite 8. Health Promotion Composite
Appropriate Resource Use	No weight-shared savings recommended	1. Inpatient Utilization: Acute Care Discharges PTMY 2. Inpatient Utilization: Bed Days PTMY 3. Inpatient Readmission Within 30 days 4. Emergency Department Visits PTMY 5. Outpatient Procedures Utilization: per cent Done in Preferred Facility 6. Generic Prescribing: SSRIs/SNRIs 7. General Prescribing: Statins 8. Generic Prescribing: Anti-Ulcer agents

9. General Prescribing: Cardiac-Hypertension and Cardiovascular
10. Generic Prescribing: Nasal Steroids
11. General Prescribing: Diabetes – Oral
12. Generic Prescribing: Anxiety/Sedation – Sleep Aids
13. Total Cost of Care
14. Frequency of Selected Procedures – Back Surgery
15. Frequency of Selected Procedures – Total Hip Replacement
16. Frequency of Selected Procedures – Total Knee Replacement
17. Frequency of Selected Procedures – Bariatric Weight Loss Surgery
18. Frequency of Selected Procedures – PCI
19. Frequency of Selected Procedures – Carotid Catheterization
20. Frequency of Selected Procedures – CABG
21. Frequency of Selected Procedures – Cardiac Endarterectomy

in practice alignment of P4P through the IHA has been only partially accomplished. Payments for IHA measures as a per cent of a payer's total P4P payments varied from 13.7 to 87 per cent in 2010. In the same year, three of seven insurers reported that payments were calculated using the IHA Standard Payment Methodology. Other insurers used variations on the Standard Payment Methodology, which varied the way in which attainment and improvement scores determined the bonus, although all considered attainment and improvement in some way.

Total quality incentive payouts from health plans to California physician groups started at US$38 million in 2004, peaked at US$65 million in 2007, and have levelled off at about US$50 million for the last several years (Table 13.2). While these total figures appear substantial, the average P4P payouts amounted to two per cent or less of the total capitation payments made to participating groups (Integrated Healthcare Association, 2010). Per member per month payments across insurers ranged from only US$ 0.28 to US$ 1.32.

Data sources and flows

IHA produces a measurement manual including technical measure specifications, along with data collection and reporting guidelines. Through a vendor, the IHA generates quality measure performance scores on an annual basis using its uniform measure set and data submitted by both health insurers and physician groups. Physician groups may choose to self-report across all payers and patients, or they may rely on the health insurers to report data

Table 13.2 Annual payouts in the California IHA, 2004–10

Payout year	*Measurement year*	*Total payout**	*Number of physician organizations***	*Number of health plan members***
2004	2003	USD 38M	215	6.4M
2005	2004	USD 54M	230	8.8M
2006	2005	USD 55M	228	11.2M
2007	2006	USD 65M	235	11.2M
2008	2007	USD 52M	233	10.9M
2009	2008	USD 52M	229	10.5M
2010	2009	USD 49M	221	9.9M

* Total payouts are for seven health plans using P4P results for incentive payments.
** Includes Permanente Medical Group Northern California starting in MY 2004. Includes Permanente Medical Group Southern California starting in MY 2005. Permanente Medical Group participates in public reporting only.

Source: Integrated Healthcare Association, 2010.

on their behalf. All data must be derived from standardized electronic sources that are subject to audit. The majority of data are derived from encounter records (also known as shadow claims, because they mimic billing data but are not used for payment) and laboratory billing data. If data for a particular measure are reported both by the insurers and the physician group, scoring is based on the more favourable of the two. For resource use, all measures are evaluated based on insurer billing data only. Patient experience surveys are conducted with samples of patients for each physician group by a survey vendor using a validated instrument. Finally, information on use of health information technology is collected by survey and validated by an accrediting organization.

In June of each year, the IHA issues preliminary reports to both physician groups and insurers. Either party may appeal these preliminary reports within a narrow time frame (approximately three weeks). Final clinical quality, patient experience, and health information technology performance data are released to the public through the office of the Patient Advocate, a state government agency.

Reach of the programme

Which providers participate and how many people are covered?

Over 200 California physician groups participate in the IHA programme, representing approximately 35,000 physicians. These groups provide care for

about ten million health maintenance organizations or point of service plan members. Seven California health plans contribute data and provide incentive payments based on the aggregated P4P results.

Improvement process

How is the programme leveraged to improvements in service delivery and outcomes?

The IHA programme relies primarily on financial incentives, which explicitly incorporate measures of improvement in scoring, as well as both private and public reporting of all-payer data to spur improvement in service delivery and outcomes. There is no ongoing technical assistance or separate investments in physician group capabilities for quality improvement.

Most physician groups that participate in the programme, however, are large, sophisticated entities with the capability to engage physicians in quality improvement and implement systems to manage population health (Rosenthal et al., 2001) Surveys with the leadership of participating physician organizations suggest that the IHA initiative has spurred a variety of investments and policy changes, including increased patient outreach and use of data for internal quality improvement (Figure 13.1). A gradient in performance that favours large groups, however, suggests that the modest financial incentives provided by the programme may not be sufficient for some entities that lack infrastructure to close the performance gap (Damberg et al., 2009).

Results of the programme

Has the programme had an impact on performance, and have there been any unintended consequences?

Performance related to specific indicators

More generally, IHA's own monitoring reports give a mixed picture of performance improvement over time (Table 13.3). Performance measures included in the IHA P4P programme have improved modestly and unevenly across measures, with no evidence of 'breakthroughs' in quality improvement (Damberg et al., 2009). Moreover, because these analyses do not attempt to control for secular trends in quality improvement, it is unclear the extent to which any gains can be attributed to P4P rather than other trends.

Programme monitoring and evaluation

An independent evaluation of the IHA programme was funded alongside its implementation, and a number of impact and implementation studies also have been published (Rosenthal et al., 2005; Damberg et al., 2009; Mullen, Frank & Rosenthal, 2009). In addition, each year, the IHA publishes its own report

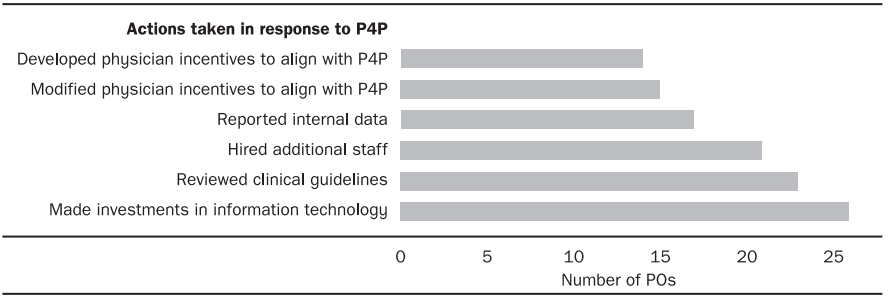


Figure 13.1 Survey of physician group responses to the California IHA, 2007

Source: Damberg et al., 2009.

on programme results, which includes trends in performance measures and payments.

Two controlled studies provide the strongest evidence of impact of the IHA initiative. These analyses are limited to measures for which pre-intervention data were available and one payer with contemporaneous data for a set of comparison practices from neighbouring states. These studies find that not all targeted clinical process measures of quality improved. Among the measures that could be analysed, only cervical cancer screening improved differentially among the IHA participants, and improvement was modest at best, approximately 3.5–6 percentage points depending upon the statistical model used (Mullen, Frank & Rosenthal, 2009). One of these studies also examined the impact of the IHA initiative on performance indicators that were not

Table 13.3 Average clinical quality achievement rates in the California IHA, 2006–09

Measure	*2006*	*2007*	*2008*	*2009*
Breast cancer screening	66.8	68.0	69.4	72.0
Childhood immunization	88.4	88.9	90.6	89.8
Chlamydia screening	42.5	46.7	51.1	51.8
Colorectal cancer screening	–	43.3	47.5	51.0
Appropriate treatment for upper respiratory infection	82.4	87.5	87.7	89.5
Cholesterol screening for CVD	83.9	86.1	86.2	87.2
Cholesterol control for CVD	50.4	52.3	54.9	59.8
HbA1c screening	77.1	79.8	81.0	83.4
HbA1c poor control	46.2	46.7	47.1	42.0
Cholesterol screening for diabetes	74.3	77.4	79.0	81.0
Cholesterol control for diabetes	32.9	33.5	37.0	40.5
Nephropathy monitoring for diabetes	73.7	75.9	78.5	79.1

Note: Lower rates indicate better performance for HbA1c control.

Source: Integrated Healthcare Association, 2009.

included in the programme in an attempt to detect both positive and negative spillover effects. In these analyses, no clear pattern emerged to suggest that non-targeted measures either benefited or suffered from the presumed focus on targeted measures.

Equity

While there has been no systematic analysis of the impact of the IHA programme on equity, several empirical clues suggest that P4P may not have distributed its benefits equally. First, while there has been some compression in the distribution of performance scores, physician groups that performed poorly on quality measures at the launch of the programme have not caught up with high performers and overall have received only a small share of payments (Damberg et al., 2009). Second, there is substantial geographic variation in performance, which may be associated with factors such as socio-economic status and local health care delivery system capacity (Integrated Healthcare Association, 2010). Finally, interviews with physician group leaders revealed some concerns that the P4P programme has caused groups to avoid patients whose health or health behaviour would negatively affect the group's performance.

Provider response

Physician leaders have expressed favourable opinions of the IHA programme and belief that it plays an important role in quality improvement efforts in California (Damberg et al., 2009). A survey of the general population of primary care physicians also found generally positive attitudes about P4P in theory, but in practice some expressed concerns about their ability to understand the IHA programme details, the size of the bonuses, and the impact on health care quality (Figure 13.2).

Costs and savings

Evaluations of the IHA P4P programme have concentrated on the early years of the programme when resource use and costs were not directly targeted by the programme. While no formal analyses have been reported, it is unlikely that improvements in clinical quality, health information technology, and patient experience (to the extent they have occurred) would generate savings for payers. This, in part, may have motivated the recent evolution of the programme towards inclusion of resource use measures and a shared savings component. Measures related to resource use will explicitly be included in the IHA payouts for 2013.

	Rating scale (1–5), mean by performance level				
	Overall	High	Medium	Low	p value[a]
Views on P4P design elements					
Importance of the IHA P4P program to the PO (1 = not important, 5 = very important)	4.0	4.2	3.8	4.0	0.636
Importance of public reporting (1 = litttle/no importance, 5 = very important)	3.0	3.3	2.9	2.5	0.427
Increasing the incentive as percent of total capitation (1 = lower, 3 = about right, 5 = higher)	4.4	4.7	4.4	3.8	0.030
Measurement domains (1 = little/no importance, 5 = very important)					
Clinical measures	4.6	4.7	4.5	4.7	0.583
Patient experience	3.9	4.1	3.7	3.8	0.560
Information technology capability	4.1	4.4	4.4	3.3	0.018
Views on quality environment and support for quality					
Organizational culture of quality (1 = weak, 5 = strong)	4.7	4.8	4.6	4.7	0.500
Support organization dedicates to addressing quality issues (1 = very little/no support, 5 = strong support)	4.1	4.4	4.2	3.6	0.081
Success in monitoring POs' quality performance (1 = not successful, 5 = very successul)	3.9	4.1	3.7	3.7	0.246

[a] Based on the Kruskal-Wallis nonparametric test adjusted for ties.

Figure 13.2 Survey of physician group perceptions of the California IHA, 2007

Source: Damberg, et al., 2009.

Overall conclusions and lessons learned

Has the programme had enough of an impact on performance improvement to justify its cost?

The continued commitment to the IHA P4P programme by payers and physician groups alike, despite acknowledgement of weak performance improvement, suggests that there is a perception that on the whole, the programme is worth supporting. While no formal cost-effectiveness analysis has been undertaken, the estimates of impact on specific performance measures described above are unlikely to be sufficiently valuable to offset the economic costs of data collection, auditing, and reporting. In a broader sense, however, the IHA programme may be worth its cost. Observers have commented in particular on the importance of the initiative for establishing the basis for collaboration and trust among the participants.

The underwhelming performance improvements that have been seen under the IHA programme have raised questions about obstacles to broader and deeper quality improvement. One is whether the magnitude of incentives needs to be increased or greater emphasis placed on performance improvement. Some observers have suggested that incentives need to be closer to 10 per cent of total revenues to stimulate improvement (as compared to the <2 per cent offered currently). There is some concern, however, that increased rewards might bring increased adverse effects, including patient dumping.

While alignment of measurement and P4P programme design was a central goal of the IHA initiative, the annual transparency reports suggest that such alignment has been imperfect. Variation in the extent to which participating insurers have relied on IHA measures and guidelines may have diluted the effect of the programme, although some degree of flexibility may be desirable (and necessary from an antitrust perspective).

Another possible explanation for weak results may be the continued expansion of the measure set and the difficulty physician organizations face in making investments in quality improvement when the targets are continuously moving. There is an obvious tension here with the desire to include a comprehensive set of measures to avoid 'teaching to the test', a narrow focus that causes providers to concentrate on a small subset of tasks at the expense of unrewarded domains, and to incorporate the best available measurement science over time.

Finally, P4P alone almost surely will be insufficient to mobilize improvement for all physician groups in California. It appears that some groups may lack the capacity or knowledge to improve their performance in the absence of technical assistance or investments in infrastructure and human resources.

While questions remain about how to increase the effectiveness of the IHA programme, a number of important lessons about the implementation of P4P programmes in a context like California's were distilled by evaluators (Damberg et al., 2009). First, the involvement of a neutral convener seems to have been important to bring payers and providers to the table around measure selection and programme design. Likewise, use of a third party data aggregator was essential to ensure uniformity in measurement and confidence in the results. The IHA has also modelled an effective measurement evolution process that introduces 'testing measures' prior to adoption of new measures so that measurement and validity issues may be identified prior to inclusion in P4P. Finally, effective communication with all stakeholders about modifications to the measure set and recommended payment algorithms, as well as about the process by which decisions have been made, has been critical to maintain engagement and commitment to the programme.

Notes

1 The delivery system in California is largely organized around medical groups and independent practice associations (IPAs). These entities may be more or less formally integrated but typically contract together, include 100 or more primary care physicians as well as major specialties, and share accountability for costs and quality.
2 http://www.iha.org/program_governance.html.
3 For antitrust reasons, payers cannot openly align the details of provider payment. The details of the IHA Standard Payment Methodology are limited to weighting of measures and domains as well as treatment of attainment (absolute performance level) vs. improvement (the change over time in an individual provider's performance relative to history).
4 The IHA measurement set also includes other measures to be collected and reported only to providers, to be collected and publicly reported (but not for payment), and testing measures.

References

Damberg, C. et al. (2009) Taking stock of pay for performance: a candid assessment from the front lines, *Health Affairs*, 28(2): 517–25.

Institute of Medicine (2001) *Crossing the quality chasm: a new health system for the 21st century*. Washington, DC: National Academies Press.

Integrated Healthcare Association (2006) *Advancing quality through collaboration: the California pay-for-performance program*. Oakland, CA: Integrated Healthcare Association.

Integrated Healthcare Association (2009) *Measurement year 2009 results report*. Oakland, CA: Integrated Healthcare Association.

Integrated Healthcare Association (2010) *California pay for performance program – incentive payouts*. Oakland, CA: Integrated Healthcare Association.

Integrated Healthcare Association (2010) *Pay for performance report of results for measurement year 2010*. Oakland, CA: Integrated Healthcare Association.

Integrated Healthcare Association (2010) *Transparency report*. Oakland, CA: Integrated Healthcare Association.

Mullen, K., Frank, R. and Rosenthal, M. (2009) Can you get what you pay for? Pay for performance and the quality of healthcare providers, *RAND Journal of Economics*, 41(1): 64–91.

Rosenthal, M. et al. (2001) Scale and scope of capitated physician organizations in California, *Health Affairs*, 20: 109–19.

Rosenthal, M. et al. (2005) Early experience with pay for performance: from concept to practice, *Journal of American Medical Association*, 294(14): 1788–93.

Inpatient care

chapter fourteen

Brazil: São Paulo: Social organizations in health

Y-Ling Chi and Emily Hewlett

Introduction

Brazil has made significant strides in improving the organization and financing of its health system since the constitutional change establishing the right to health care in 1988. Government health financing was consolidated, and the public delivery system was decentralized to states and municipalities and organized into a country-wide system (Unified Health System, or SUS). Programmes such as the Basic Health Package and Family Health Programme have helped to shift the focus from a hospital-heavy system to basic primary care. Other major improvements have been facilitated by improvements in health human resources and infrastructure and other advances in the public sphere (Paim et al., 2011).

In spite of these advances, however, many challenges remain in Brazil's health care system. Health spending in Brazil has been increasing faster than in neighbouring countries, especially after 2004, reaching close to nine per cent of GDP in 2009. Although SUS is expected to provide coverage for nearly 80 per cent of the population, less than half of total health spending is contributed by the government (43.6 per cent) (WHO, 2012). Given that only 22 per cent of the population is covered by private health insurance plans (Economist Intelligence Unit, 2010), the government total health expenditure seems low relative to the size of the population it covers. The massive scale of SUS (serving 87 million people) is also supported by a complex governance, management and financing structure, combining multiple (and sometimes competing) service providers and purchasers both from the public and the private sector. Historically, the development of the health care system in Brazil has been geared towards the provision of services in private hospitals and clinics, operating alongside the public sector through contracting out arrangements (Paim et al., 2011). Furthermore, Brazil's federal structure and the decentralized nature of the SUS make the financial flows difficult to track and monitor, thus limiting accountability. For instance, La Forgia and Couttolenc (2008) point out that it

is only until recently that estimates of hospital spending are available at all at the aggregate level.

Concerns about inefficiency and poor performance, particularly in public facilities, have motivated new innovative management and organizational approaches, including a range pay for performance programmes. One such innovative approach is a performance-based contracting arrangement in São Paulo between the government health system and a private non-profit management group called Social Organizations in Health (OSS). Under the OSS model, the State Secretariat of Health (SES) negotiates a performance contract with the OSS that provides a global budget to manage the hospitals, and the OSS commits to specific volume and performance targets. The OSS managers are granted greater flexibility than their counterparts in traditional state hospitals to run the hospital in the best way to meet their performance targets.

Health policy context

What were the issues that the programme was designed to address?

Policy objectives

The Brazilian hospital sector could at best be described as diverse, innovative, inventive and at the cutting edge of developing and providing excellent treatment in some areas of care, and at worst could be described as disorganized, overly bureaucratic, rigid, underfunded and inefficient. Great disparities are observed not just between regions, between local areas and between hospitals, but are also often apparent within a single hospital for different conditions. Describing the Brazilian public hospital administration model is in itself a challenging task given how widely hospitals tend to differ in administration, funding, and autonomy (La Forgia and Couttolenc, 2008). In almost all cases, though, hospital administration lacks quality and performance monitoring systems, both of which are frequently lost in the many layers of administration. A large majority of public hospitals are directly administered by either the federal, state or municipal government. Directly administered hospitals have been criticized, however, for being highly inefficient. As a result, autonomous organizational models for hospitals started to gain importance in the 1990s as an alternative to direct administration.

Social organizations were created by the Law 9637 in 1998 during the Reform of Public Administration as not-for-profit civil entities, which manage public organizations in a large range of areas such as health care, education, culture or research. These organizations were created to increase efficiency and civil participation, thereby reducing deficits and limiting waste. These goals translate into very practical arrangements making social organizations accountable for results, closely monitored and transparent. In line with the federal model of Social Organizations, the Law 846 also, passed in 1998, established the

Social Organizations in Health (OSS) as new entities to manage hospitals in the state of São Paulo. Initially, OSS were created to operate in newly built general hospitals, serving more disadvantaged and vulnerable populations on the periphery of São Paulo. These hospitals typically offer services in four priority services: surgery, gynaecology and obstetrics, internal medicine, and paediatrics. Both inpatient and outpatient services are available in most of the general hospitals managed by OSS, as well as ambulatory surgery services and psychiatric inpatient care. Since January 2011, all public hospitals have the opportunity to switch to OSS management and become self-managed units. At the time of writing, however, only a handful of hospitals under direct public management have changed their status to adopt OSS management. Changing the hospital administration model would involve a complex process of converting all hospital employee contracts. Moreover, the OSS performance contracting model is also being applied in private hospitals (not-for-profit and for-profit hospitals) that have a service agreement with SUS.

OSS can be thought of as a public–private partnership arrangement, in which OSS are completely autonomous organizations acting as operators to manage public facilities. OSS manage hospitals autonomously and operate under a high degree of flexibility, and OSS are not regulated by public sector laws. OSS are contracted by the SES through a five-year renewable contract, depending on performance. The SES of São Paulo negotiates a hospital management contract directly with OSS which specifies the volume of different services to be performed annually, as well as other performance targets used for payments.

Since 1998, the number of OSS has been steadily increasing, with all newly opened hospitals after that time automatically placed under OSS management. Initially, 15 hospitals in poor areas were selected to be managed by OSS, but these rules were later reformed. At the time of writing, OSS cover 37 hospitals, 38 clinics, a referral centre for outpatient specialist care, two pharmacies and three clinical laboratories in the State of São Paulo (Governo De Estado do São Paulo, n.d.).[1]

Stakeholder involvement

Following the creation of OSS, two core monitoring institutions were established. A contract management unit was created within the São Paulo SES that is responsible for negotiating with the OSS on the annual performance and volume targets. An Independent Assessment Commission (AIC) was created in 2001 that reviews the performance indicators and calculates the level of penalties, if needed, quarterly. The AIC is composed of representatives of the SES, the legislative branches, and other members of the civil society. Payments to OSS are based on the assessment by the AIC of the performance and volume targets of each hospital. A state audit agency is also in charge of a financial and technical audit of the OSS.

Technical design

How does the programme work?

Performance domains and indicators

The payment of the global budget to OSS hospitals is contingent upon achievement of both volume and performance targets. Volume targets are based on the preceding year's level of service, and apply across departments within the hospital. Volume is measured by either bed days, consultations, admissions, or number of procedures (see Box 14.1 for an example of OSS contracting terms with Pirajussara Hospital).

Performance targets are usually classified in four domains: (1) quality of care; (2) patient satisfaction; (3) information quality; (4) efficiency. In one example provided in La Forgia and Couttolenc (2008), there are nine performance indicators across the four domains, and indicators in the quality domain are weighted more heavily, accounting for 70 per cent of the performance target (Table 14.1).

Performance of the hospitals is assessed against numeric targets and on general assessment by the SES. For instance, in 2010, targets for most hospitals were related to sending the information to the SES (performance reporting compliance), analysis of the trends for quality indicators, and description of the measures developed by facilities to drive quality improvements, if applicable. In addition, the AIC verifies the quality of data and conducts technical audit on a yearly basis.

Table 14.1 Performance indicators of the Brazil OSS, 2002–04

Category	*Weight*	*Indicator*
Quality of care	0.7	Mortality, ethics and infection control commissions fully operational.
		Percentage of deaths analysed by mortality commission.
		Percentage reduction in hospital infection rate.
Patient satisfaction	0.1	Percentage of patient complaints addressed.
		Completion of patient satisfaction surveys.
Information quality	0.1	Medical records contain secondary diagnoses.
		Place of residence codes completed in patient records.
		Reason for Caesarean sections provided.
Efficiency	0.1	Average length of stay for specific services (without secondary diagnosis).

Source: La Forgia & Couttolenc, 2008.

Incentive payments

Every OSS is required to sign a performance contract with the São Paulo SES that aims to increase service delivery and care standards. This contract is linked to target objectives for output and quality of care within a global budget in the following manner:

1. *Volume component*: 90 per cent of the OSS budget is allocated monthly based on achievement of volume targets as follows:
    - if hospital achieves between 85–100 per cent of the volume target, the budget is fully disbursed;
    - if the hospital achieves between 75–85 per cent of the volume target, the monthly allocated budget can be reduced by up to 10 per cent;
    - if the hospital achieves less than 75 per cent of the volume target the monthly allocated budget can be reduced by up to 30 per cent (World Bank, 2006).
2. *Performance component*: 10 per cent of the maximum possible budget is held in a 'retention fund', which is disbursed quarterly, depending on achievement related to agreed performance indicators.

Volume targets and performance indicators are agreed between hospitals and the São Paulo SES on a case-by-case basis. OSS can then organize service delivery and input use in the best way to achieve their targets. OSS have the autonomy to decide on the level of all inputs (procurement of all types of medical staff, purchase of medical equipment and drugs, outsourcing of medical services to outpatient specialized services, etc.), with the exception of capital investments, for which the OSS has to refer to the SES.

Incentive payments do not take into account costs incurred for additional investments in medical equipment (capital costs decided with the SES yearly) or data systems. In addition, OSS-managed hospitals are only authorized to charge privately insured patients for out-of-pocket fees, as stated in every contracting arrangement.

Box 14.1 Example of OSS contracting terms with Pirajussara Hospital, 2011

The Pirajussara hospital was one of the very first OSS-managed hospitals in São Paulo. Inaugurated in 1999, the hospital initially covered a population area of about 500,000 patients, mainly through outpatient specialist visits. Since 1999, Pirajussara has grown to be one of the largest hospitals in the area, providing a wide range of services in 46 specialties from obstetrics and gynaecology to neurosurgery and cardiac surgery. The hospital also now provides services to patients in rehabilitation. Pirajussara is managed by the OSS São Paulo State Association for Development of Medicine, the Associação Paulista para o Desenvolvimento da Medicina, which operates 22 hospitals in the state.

The latest contract between the São Paulo SES and the Pirajussara hospital was signed on the 20 December 2011 for payments for the following year (2012). The contract stated that the hospital budget for the year 2012 would be a maximum of R$92,700,000, composed of the production target payments and the retention fund.

Payments for production targets are made in 12 monthly instalments of R$6,952,000 each (amounting to R$83,430,000). Volume targets apply to 32 specialities (out of a total of 46 specialities in the hospital) and are divided in five assessment areas: (i) hospitalization; (ii) day and ambulatory surgery; (iii) ambulatory specialist care (consultations); (iv) emergency services; (v) diagnostic and therapeutic activities (CT scans, radiology, endoscopy, etc.). Payments are made monthly, but following assessment of volume in February, May, August and November, adjustments for penalties can be made according the payment mechanism detailed above.

A separate quality assessment is performed in April, July and October. Disbursement of the retention fund is conditional on the quality assessment of the hospital activities, mainly focusing on recording of economic and financial data on hospital service costs, publication of such information on the website of São Paulo State SUS, preparation and publication of monthly reports on hospital activities for each specialty, notably on issues such as patient safety and hospital infection. Quarterly analysis of these reports is performed by the AIC and serves as basis to payment of the retention fund. As for all OSS-managed hospitals, retention funds amount to 10 per cent of hospital total payments, i.e. R$9,270,000. Quality indicators are reviewed and subject to revision every year.

In addition, under the OSS contracting arrangement, São Paulo SUS can issue a warning to the hospital, or even temporarily suspend the hospital (or its units) from running for at maximum two years. At any time, OSS are also allowed to withdraw from the contracting arrangement, and return the hospital management functions to SUS.

Pijurassa Hospital was included in external reviews of the OSS model in São Paulo, and has been shown to have significantly better hospital efficiency reports than its counterpart. In 2003, the hospital was accredited by the National Accreditation Organization (Universade Federal de São Paulo, 2004).

Source: Estado de São Paulo, 2011.

Data sources and flows

Prior to the start of the contract, a three-year start-up phase is launched to put the data systems in place for performance management. During this start-up phase, the SES and the OSS set up a standardized cost accounting and data collection system, collect the information on volume and performance indicators, and test the contractual arrangements. The collected information serves as baseline data. No penalties are imposed during this time. Data to

implement the performance contracts are then collected mainly though this standardized cost accounting system. For some indicators (e.g. hospital-acquired infections), reports are prepared by OSS and sent to the SES for separate assessment (Radesca, 2010).

The assessment of the performance data is carried out every three months by the AIC, which discusses the results with the hospitals. A yearly report is also published in the State's Official Diary and analysed by the Court of Accounts of the State of São Paulo (Barata & Mendes, 2007; Radesca, 2010).

One of the other key elements of the São Paulo performance contracting is the recognition that services delivered by hospitals should be tailored to the health needs of the population covered. In this sense, contracting on the basis of yearly consultation and negotiation on volume and quality targets between hospitals and the SES was more suitable than fixed pre-established common targets applied to all hospitals enrolled. In addition, systematic review of reports on provision of services and regular consultations with the SES creates an ongoing dialogue to support performance improvement.

The OSS can retain any surpluses generated by incentive payments and efficiency gains, which can be used only within the hospital. There is little information, however, about whether and how the incentive payments and surpluses are used by the hospitals to improve quality of care. According to the contracting arrangements, the incentive payments can only be used to upgrade facilities (renovations, purchasing of additional equipment), or to pay for additional human resources. Managers in the OSS receive a fixed salary and cannot personally benefit from the incentive payments, nor do they personally incur losses linked to performance.

Results of the programme

Has the programme had an impact on performance, and have there been any unintended consequences?

Programme monitoring and evaluation

External reviews of the São Paulo experience have been carried out, the most extensive of which are La Forgia and Couttolenc (2008) and the World Bank (2006). These reviews show that OSS-managed hospitals appear to be more efficient and also more productive than their counterparts. The World Bank evaluation focused on the managerial tools provided to OSS and concluded that greater autonomy was the key element to the success of performance contracting. In particular, decision making related to human resources was critical, as it not only enabled hospitals to hire the necessary staff, but also to retain the staff that performed and adapted best to the model (World Bank, 2006).

La Forgia and Couttolenc (2008) compared the performance of the two types of hospitals using data reported by 12 hospitals operated by OSS and 12 hospitals under direct administration serving as a comparison group. Hospitals were matched on the basis of hospital characteristics (size, number of physicians

per bed, discharges, spending) and case mix. The authors compared efficiency scores generated by Data Envelope Analysis (DEA). The results showed that autonomous hospitals are more efficient than directly administered hospitals, and even more efficient than private hospitals (Figure 14.1). According to the authors, publicly managed hospitals require approximately 60 per cent more resources to produce an equivalent output (La Forgia & Couttolenc, 2008).

Hospitals operated by OSS performed better along other measures of efficiency, including bed turnover rate, average length of stay, bed occupancy rate, and expenditure per discharge. The bed turnover rate was about 60 per cent higher in hospitals operated by OSS, and average lengths of stay were about 20 per cent shorter in OSS-managed hospitals (Figure 14.2). The bed occupancy rate was 81 per cent in OSS hospitals, in comparison to only 63 per cent in directly administered hospitals. Overall, expenditure per discharge was about 50 per cent lower in OSS-managed hospitals.

The comparison between the two types of hospitals also suggests that these gains in efficiency were not made at the expense of quality of care. Mortality rates in general, surgical and paediatric units was much lower in OSS-managed than in directly administered hospitals. Barata et al. (2009) also compared Caesarean-section rates between OSS-managed hospitals and other South-eastern directly administered public hospitals and showed that only hospitals managed by OSS do perform caesarean-section rate below the WHO recommended level of 25 per cent of deliveries.

The World Bank also investigated the impact of performance contracting and showed that in combination to greater autonomy, OSS-managed hospitals had greater incentives to improve managerial techniques and reduce red tape. According to the report, these changes in internal organizational characteristics have been successfully implemented, mostly without paying higher salaries than public hospitals, or using performance incentives for medical staff (Matzuda

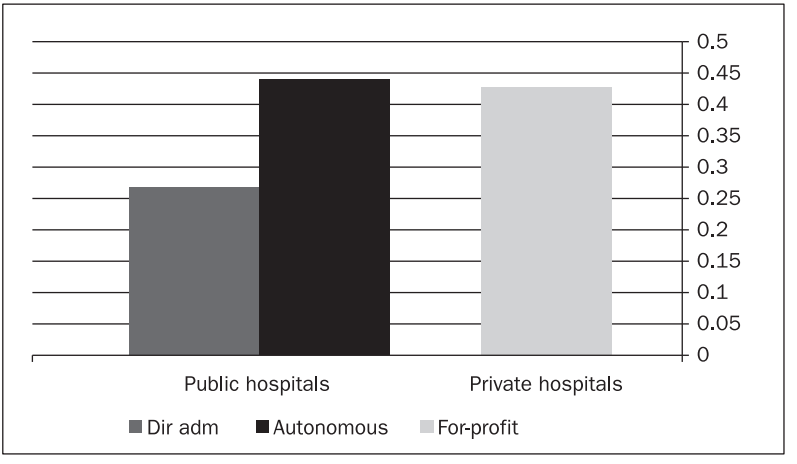


Figure 14.1 Data Envelopment Analysis (DEA) efficiency scores for hospitals following the implementation of the Brazil OSS, 2002

Source: La Forgia and Couttolenc, 2008.

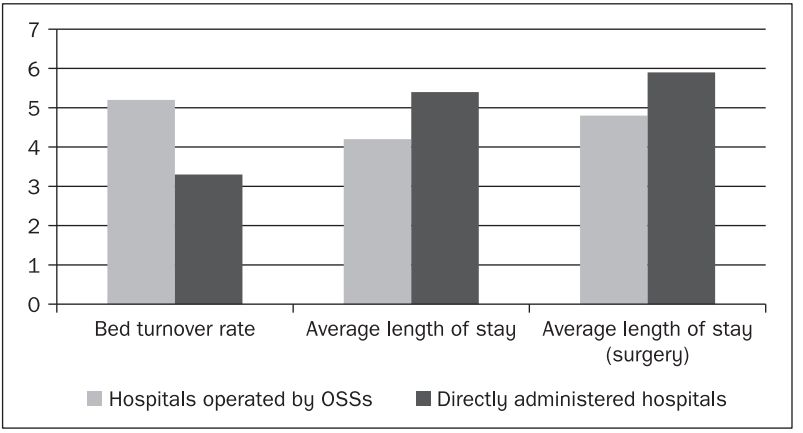


Figure 14.2 Bed turnover rate and average length of stay for OSS and non-OSS hospitals in Brazil, 2002

Source: La Forgia and Couttolenc, 2008.

et al., 2008). Managerial autonomy mainly enabled hospitals to recruit medical staff with a more balanced skill mix, resulting in a smaller but more efficient workforce composition. These findings have also been confirmed by Barata et al. (2009). The report emphasizes some caveats in the comparison between publicly administered hospitals and OSS-managed hospitals, particularly that publicly managed hospitals seem to treat more expensive and difficult cases than OSS-managed facilities (even though the OSS-managed hospitals serve more disadvantaged parts of the suburban São Paulo).

One of the main weaknesses of these external reviews, however, is the robustness of the techniques used to compare OSS-managed and directly administered hospitals. Beyond comparing population characteristics and case mix, such evaluations do not control for hospital infrastructure and characteristics. Since OSS-managed hospitals were initially only newly built hospitals, directly administered hospitals might differ in other characteristics compared to OSS-managed hospitals, which are not accounted for in these studies and could drive lower outcomes and efficiency scores.

Implementation of OSS and performance contracting also encouraged some hospitals to get accreditation by the National Accreditation Association. In 2009, 11 OSS hospitals out of 32 received full accreditation, while none of the directly administered hospitals have sought accreditation (Mendes & Bittar, 2010). Out of these 11 accredited hospitals, three hospitals have achieved level III, which corresponds to accreditation with distinction. Despite the implementation of a comprehensive national accreditation programme, only 103 hospitals are accredited by the National Accreditation Association, and only 32 have received accreditation with distinction. Hospitals have little financial incentive to complete the accreditation requirements (La Forgia & Couttolenc, 2008).

Equity

The OSS hospital contracting initiative had an inherent equity objective, as it aimed to provide high quality inpatient services at reasonable cost in more vulnerable communities located outside of the São Paulo metropolitan area. The programme was accompanied by construction of 37 new hospitals, which increased access to inpatient care by expanding services to 5.2 million inhabitants. OSS-managed hospitals are not permitted to charge patient fees to supplement public revenues, which may further improve financial protection and equity.

Costs

Estimates of performance payments are currently not available, as the payment of bonuses is blended into the broader hospital payment mechanism. It is known that in 2007–2008 a part of the retention fund was withheld for non-compliance to performance targets for a number of entities.

Some information is available on the costs of OSS hospitals from an external evaluation comparing data from 2005 on matched pairs of OSS and non-OSS hospitals. The evaluation concluded that while hospitals managed by OSS received on average 8 per cent more revenues than directly administered hospitals, they also produced a higher volume of services than their counterparts, resulting in a 24 per cent lower cost per bed day than publicly managed hospitals (Barata & Mendes, 2007). The study was repeated using 2006 data and showed that the average cost per bed was 9.8 per cent lower in OSS-managed hospitals, indicating that publicly managed hospitals have reduced the productivity gap overtime.

Overall conclusions and lessons learned

Has the programme had enough impact on improvement to justify its cost?

External reviews and evaluations consistently showed that hospitals managed by OSS performed better along efficiency and quality measures than directly administered hospitals, controlling for case mix. The OSS performance-based contracting model has led to improved governance, planning and monitoring capacity of hospital managers by providing not only managerial autonomy to OSS, but also by implementing a standardized cost-accounting and volume-managing computer-based system. Nonetheless, it is important to note that the existing evidence is relatively outdated and does not necessarily account for recent developments of the model. Moreover, even the results of the largest evaluation have technical limitations, e.g. failure to control differences in OSS-managed and directly administered hospitals in terms of facilities quality and overall infrastructure.

Full managerial autonomy under the performance-based contracting has been accompanied by greater accountability of both individual providers to

hospitals and of hospitals to patients through publication of quality indicators and reports on the website of the São Paulo SUS. Hospital activity measures (e.g. utilization, length of stays, mortality rates) are also now systematically measured and monitored. These improvements are likely to have an impact on governance and productivity of hospitals, patient satisfaction, and ultimately health outcomes. Therefore, considered against the initial objectives of improving hospital management, responding to population needs in vulnerable peripheral areas of São Paulo, and improving health outcomes (through increased medical care utilization rates), the OSS initiative seems to have been relatively successful. Such positive results are in line with those of other experiences of OSS in health and in other sectors in Brazil.

Although these external studies have put forward evidence on the link between increased managerial autonomy and improvements in efficiency and quality scores of OSS-managed hospitals, it is hard to assess the specific role played by the performance-based payment mechanism. Financial transparency, a key element of P4P programmes, has not been investigated by external reviews. Simple measures to track the use of penalties (delays in disbursement of the retention fund, for instance) and performance payment are not publicly available. The lack of information on financial flows and on performance measures is surprising, given the limited number of participating hospitals and the single payer model. Moreover, there is little information on how the global budgets are initially calculated for each hospital, suggesting that incentives are not clearly defined in relation to the base payment.

While external studies have presented tangible evidence of improvements in quality and efficiency under this management model, further analysis should examine the differences in management practices between directly and OSS-administered hospitals, and within the group of OSS-administered hospitals. Given the lack of common indicators applied across all intervention hospitals, it is important to gain a better understanding of how hospitals respond to financial incentives and adapt to different targets.

Note

1 http://www.saude.sp.gov.br/ses/acoes/organizacoes-sociais-de-saude-oss.

References

Barata, L. and Mendes, J. (2007) *Organizações sociais de Saúde: a experiência exitosa de gestão pública de Saúde do Estado de São Paulo*, São Paulo: Secretaria de Estado da Saúde, Gabinete do Secretário.

Barata, L. R. et al. (2009) Comparação de grupos hospitalares no Estado de São Paulo, *Revista de Administração em Saúde*, 11(42): 8–15.

Barreto, J.O. (2011) *Gestão por resultados em Piripiri*. Piripiri: s.n.

Economist Intelligence Unit (2010) *Broadening healthcare access in Brazil through innovation*. London: Economist Intelligence Unit.

Ibañez, N. et al. (2001) Organizações sociais de saúde: o modelo do Estado de São Paulo, *ABRASCO*, 6(2): 1–28.

Instituto Brasileiro de Geografia e Estatística (n.d.) (http://www.ibge.gov.br/estadosat/perfil.php?sigla—g#, accessed June 2011).

La Forgia, G.M. and Couttolenc, B.F. (2008) *Hospital performance in Brazil – the search for excellence*. Washington, DC: World Bank.

La Forgia, G.M. and Harding, A. (2009) Public–private partnerships and public hospital performance in São Paulo, Brazil, *Health Affairs*, 4: 1114–26.

Matzuda, Y. et al. (2008) Brazil: enhancing performance in Brazil's health sector: lessons from innovations in the state of São Paulo and the city of Curitiba, *World Bank En Breve*, 116.

Mendes, J.D. and Bittar, O.J. (2010) Saúde pública no estado de São Paulo – informações com implicações no planejamento de programas e serviços, *Revista de Administração em Saúde – Edição Especial*, January, 5–71.

Paim, J. et al. (2011) The Brazilian health system: history, advances, and challenges, *The Lancet*, 377(9779): 1778–97.

Paulo Borem, E.V. et al. (2010) *Pay-for-performance in Brazil: UNIMED-Belo Horizonte physician cooperation*, USAID Health Systems 20/20.

Radesca, E. (2010) *ENC: respostas solicitadas* (Interview 23 July 2010).

Secretaria de Atenção à Saúde (2006) *Law Nº 197 – DOU de 13/10/06*. Brasilia (Distrito Federal, Brasil): Diário Oficial – Impresa Nacional.

Secretaria Municipal de Saúde, n.d.. *Notícias* (http://extranet.saude.prefeitura.sp.gov.br/noticias/sao-paulo-exporta-modelo-de-gestao-da-saude, accessed June 2011).

Shepherd, G. (2010) *A civil service which performs: primary healthcare in Curitiba, Brazil*. Phnom Penh: Cambodia Development Resource Institute and Oxford Policy Institute, 25.

Universidade Federal de São Paulo (2004) *Pirajussira Hospital indice*. São Paulo: Universidade Federal de São Paulo.

World Bank (2006) *Brazil: enhancing performance in Brazil's health sector, lessons from innovations in the state of Sao Paulo and the city of Curitiba* (Report No. 35691-BR). Washington, DC: World Bank.

World Health Organization (2012) *Global health expenditure database* (who.int/nha/expenditure_database, accessed December 2013).

chapter fifteen

Republic of Korea: Value incentive programme

Raphaëlle Bisiaux and Y-Ling Chi

Introduction

The Republic of Korea has undergone a remarkable transformation of its health care system in the past decades and has consequently realized impressive gains in health outcomes. Korea now has one of the highest life expectancies in the world, at an average of 80.3 years in 2009 compared to 52.8 years in 1960 (OECD, 2011). These gains in life expectancy have been achieved through a combination of rapid expansion of health care services and expansion of coverage through the national health insurance system. Within two decades, health coverage was made universal. Korea also has benefited from relatively favourable demographic conditions and population lifestyle behaviour.

The improvement in health coverage and outcomes in Korea has been accompanied by a significant growth in health spending. The increase in health care professionals and health care infrastructure, particularly in the hospital sector, and the introduction of numerous new technologies and treatment modalities has resulted in one of the highest health spending growth rates across OECD countries at 8.6 per cent per year between 2000 and 2009 (Figure 15.1).

With the growing complexity of health care and the health care system, the need to assure high standards of quality of care and ensure sustainability of health spending has been put forward as a major priority for policymakers and stakeholders. Concerns about quality and financial sustainability are particularly serious for the hospital sector, which accounted for 34 per cent of total health expenditure in 2009 in Korea (OECD, 2011). Between 2000 and 2009, expenditures on inpatient care rose by 6.4 per cent compared to an OECD average of about 3.2 per cent, the third highest increase amongst OECD countries (OECD, 2011). It is likely that rising chronic disease rates driven by a rapidly ageing population and changing lifestyle habits will challenge the well-performing Korean health system in the future. In particular, the rise in prevalence and the high mortality rates associated with cardiovascular

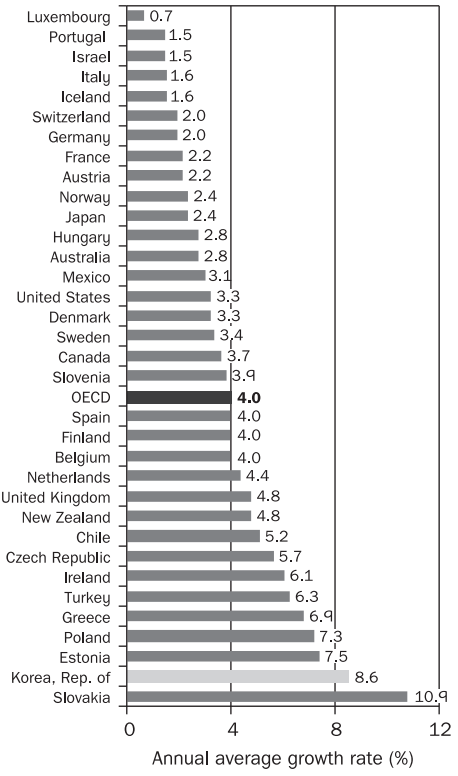


Figure 15.1 Annual average growth rate in health expenditure per capita in real terms in the Republic of Korea, 2000–09 (or nearest year)

Source: OECD, 2011.

diseases embody some of the current concerns related to value for money in the Korean health care system.

Against this backdrop, in 2007 the Ministry of Health and Welfare (MOHW) launched the Value Incentive Programme (VIP), a pay for performance (P4P) programme covering 44 tertiary teaching hospitals and aimed at improving care in two strategic areas: acute myocardial infarction (AMI) and Caesarean sections. The programme was designed based on the Hospital Quality Incentive Demonstration (HQID) implemented by the Center for Medicare and Medicaid Services (CMS) in the United States. However, the VIP and HQID differ in size and scope. In HQID, performance indicators cover more areas of care (e.g. heart attack, heart failure, pneumonia, coronary artery bypass graft, and hip and knee replacements) and bonus payments (and penalties) are usually higher.

The VIP programme in Korea has been implemented as part of a broader effort to contain health spending and ensure quality of care in the hospital sector. Similar pay for performance schemes have also recently been considered for long-term care hospitals and primary care.

Health policy context

What were the issues that the programme was designed to address?

The rapid health expenditure growth has been a major concern for health policymakers in Korea, and a number of initiatives are aimed at managing cost escalation and ensuring value for money. The VIP P4P programme was implemented within a broader reform effort, which started with the Reformed National Health Insurance Act of 2000. The health insurance law mandated the integration of numerous health insurance funds into a single payer system, the National Health Insurance Corporation (NHIC). The NHIC established a solid legal basis for strategic health purchasing, including quality assessment and monitoring of providers, and adjusting provider payment based on performance (Kim et al., 2012). The Health Insurance Review Agency (HIRA) was established in 2000 to review provider payment systems and fee schedules, conduct health technology assessments for the benefit package, manage information submitted by health care provider institutions, and conduct research. HIRA also carries out quality assessments of health care providers, which include a number of measures to help health care institutions improve the quality of care and reduce lower costs. All of these reforms and new institutional roles laid the groundwork for experimentation with payment models and for pay for performance initiatives.

Policy objectives

The goal of the VIP is to improve the overall quality of care and decrease the quality gaps among health care institutions (Kim et al., 2012). HIRA decided to focus initially on two conditions – acute myocardial infarction (AMI) and Caesarean sections (C-sections). Performance data suggest that quality of care for both of these conditions may be lagging behind other OECD countries. The prevalence and death rates by ischaemic heart disease in Korea are still relatively low compared to other OECD countries (OECD, 2011). However, while in most OECD countries mortality from ischemic heart disease has declined in the past decades, Korea's mortality rates for the condition have been steadily increasing, peaking at 29.5 per 100,000 population in 2007 (Statistics Korea, 2007). The 30-day case-fatality rates for AMI are also among the highest in OECD countries (Figure 15.2). These figures suggest that low quality of acute care for AMI might result in premature deaths, while ischaemic heart disease is a disease area where research has provided physicians and hospitals with evidence-based clinical and practice guidelines that lead to good quality of care (Figure 15.2).

The rate of C-sections is also higher than the average for OECD countries and well above WHO recommendations. HIRA has reported institutional C-section rates annually since 2001, and the rate was more than 35 per cent of deliveries in 2009 (OECD, 2011). The WHO recommendations suggest that C-section deliveries should account for about 15 per cent of all deliveries (Figure 15.2).

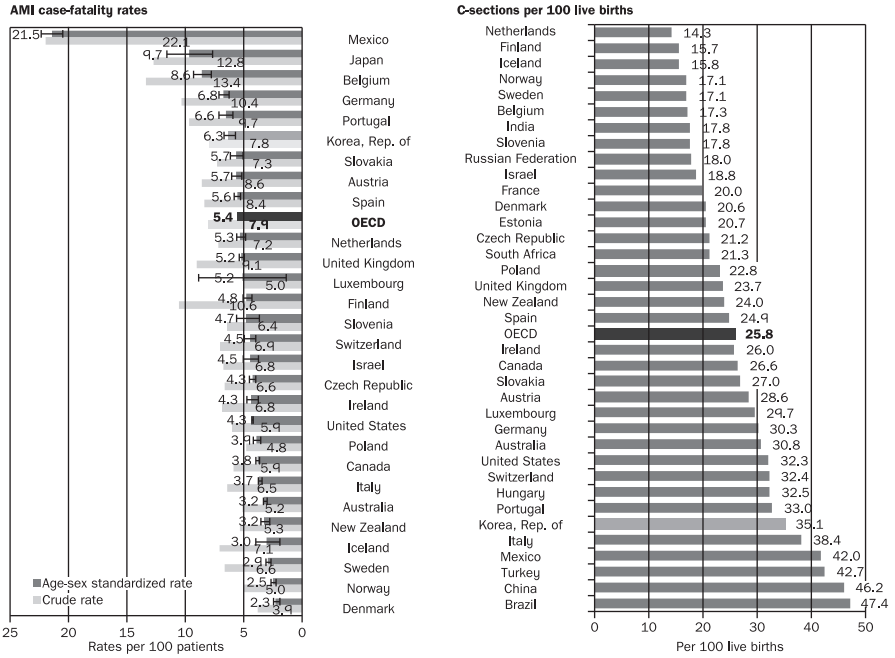


Figure 15.2 In-hospital case-fatality rates within 30 days of admission for AMI and C-section in OECD countries, 2009

Source: OECD, 2011.

Stakeholder involvement

Korea's VIP was designed by HIRA without involving hospitals and other key stakeholders. Both the Korean Medical Association and the Korean Hospital Association were opposed to any P4P programme, which they viewed as government interference or control over health care organizations and an infringement on autonomy (Lee et al., 2012). The VIP is a mandatory programme, which may have further increased resistance by medical professionals to the programme.

Technical design

How does the programme work?

The VIP programme was designed after the Premier Hospital Quality Incentive Demonstration Project of the United States Centers for Medicare and Medicaid Services (Kim et al., 2012). The top tier of performing hospitals receives a bonus payment, and the bottom tier is penalized. The programme has one domain – clinical quality – with seven indicators across two clinical areas (AMI and

C-sections). Bonus payments and penalties are based on the relative ranking of tertiary hospitals in different groups according to composite quality scores for each clinical area.

Performance domains and indicators

The C-section indicator is the number of C-sections per live deliveries in the hospital. For the AMI clinical area, five process indicators and one outcome indicator are used to measure quality:

1. Fibrinolytic therapy received within 60 minutes of hospital arrival (30 minutes as of 2010).
2. Primary percutaneous coronary intervention (PCI) received within 120 minutes of hospital arrival (60 minutes as of 2010).
3. Administration of aspirin at arrival.
4. Aspirin prescribed at discharge.
5. Beta-blocker prescribed at discharge.
6. Risk-adjusted 30-day mortality rate.

The AMI and C-section performance indicators are each translated into composite quality scores through a fairly complex formula (Table 15.1). The AMI composite quality score is calculated using a formula that weights the indicators: timely interventions upon arrival at the hospital (1 and 2) are weighted by a factor of 4.5, appropriate prescription of drugs is weighted by a factor of 2.5, and the case-fatality rate is weighted by a factor of three. The C-section indicator is translated into a composite quality score by calculating the difference between the observed C-section rate and the expected rate estimated from a regression analysis controlling for the 15 risk factors.

HIRA is planning to expand the VIP by including two additional clinical domains: acute stroke and use of prophylactic antibiotics for surgical care. Currently, measurements of the baseline performance in these two areas are being tested in several hospitals. Performance will be assessed using the routine data collection for the indicators presented in Table 15.2.

Incentive payments

Prior to 2011, incentive payments were calculated by ranking hospitals according to five grades and applying single thresholds for the incentive payment and the penalty. With the recent expansion of the VIP to general hospitals, the incentive payment mechanism was changed slightly. Since 2011, bonus payments are distributed on the basis of quality improvement relative to a baseline survey at the beginning of each year. Hospitals are ranked into nine grades based on their composite quality scores, which are disclosed both to hospitals and to the public.

High-performing hospitals (Grade 1) receive a payment amounting to 2 per cent of the payment by the National Health Insurance Corporation (NHIC) for the disease area, the second highest performing group (Grade 2) receives a 1 per cent payment. Penalties are applied when hospitals fail to reach either

Table 15.1 Calculation of composite quality scores in the Korea VIP

AMI composite quality score				*C-section rate quality score*
Process measures	*Numerator*	*Denominator*	*Weight*	C-section rate quality score = (O-E)/SE
Fibrinolytic therapy received within 60min of arrival	a	a'		O: observed C-section rate
P.PCI received within 120min of hospital arrival	b	b'	4.5	E: expected C-section rate (based on 16 clinical risk factors)
Reperfusion group (A) = a+b/a'+b'	a + b	a' + b'		SE: Standard Error
Aspirin at arrival	c	c'		**Risk factors**
Aspirin prescribed at discharge	d	d'		*Maternal factors*: age, bleeding, cord prolapse, diabetes mellitus, dystocia, pre-eclampsia & eclampsia, gynaecologic malignancy, placenta abruptio, placenta previa, venereal disease
Beta-blocker prescribed at discharge	e	e'		
Medication group (B) = c + d + e/c' + d' + e'	c + d + e	c' + d' + e'	2.5	
Outcome measure				
Adjusted 30-day mortality rate: Survival index (C)			3.0	*Neonatal factors*: malpresentation, macrosomia, multiple gestation, fetal abnormality
CQS = [[(A × 4.5)+(B × 2.5)+(C × 3.0)]/10] × 100				*Other factors*: previous C-section, premature birth

Source: Cho et al., 2010.

of the two thresholds for the composite quality score. Those hospitals with scores below the lower threshold receive a penalty of 2 per cent, and those hospitals below the upper threshold but above the lower threshold receive a penalty of 1 per cent (Figure 15.3). Grade Five (lowest performing grade before 2011), which were performing below the penalty threshold were to be subject to penalties starting from 2009.

The incentives awarded to the hospitals amounted to KRW 857 million in total between 2008 and 2010, or approximately US$740,000. In the second year, KRW 453 million (about US$360,000) was paid to 21 hospitals, and in the third year KRW 404 million (about US$380,000) was awarded to 26 hospitals. The majority of bonus payments are made to tertiary hospitals for the AMI domain and to general hospitals for the C-section domain (Figure 15.4).

Tertiary hospitals received bonus payments every year from 2008 onwards, and penalties in 2009 and 2010. No hospitals received penalties in 2009 or 2010,

Table 15.2 Indicators for acute stroke care and use of prophylactic antibiotics collected by HIRA

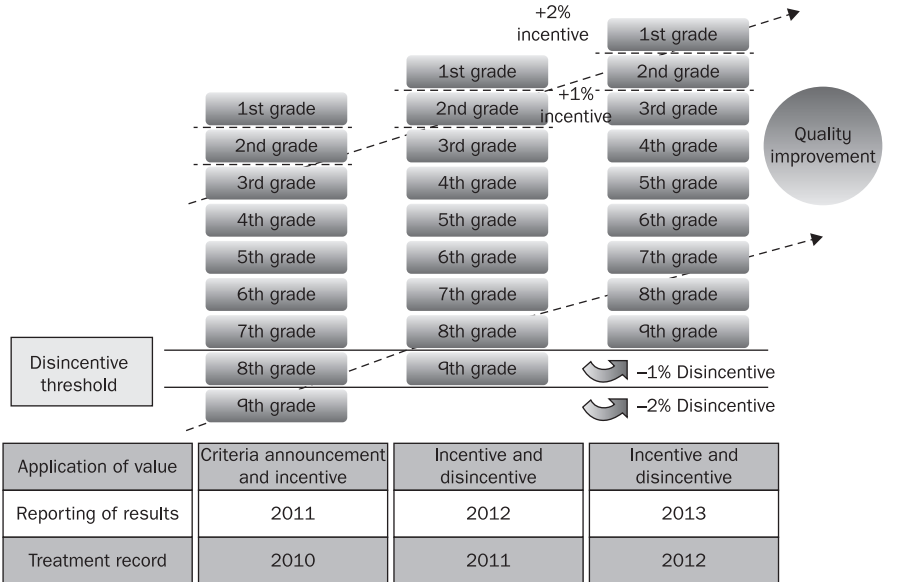
Acute stroke	*Use of prophylactic antibiotics*
• Organization of specialist personnel (specialists in the neurology, neurosurgery, and rehabilitation departments)	• Initial prophylactic antibiotics within one hour before skin incision
• Documentation rate of smoking history (doctor's records)	• Prophylactic antibiotics administration rate before proximal tourniquet inflation (applied to total hip replacement arthroplasty)
• Neurological examination rate (Category: consciousness, motor and sensory functions, cranial nerve exam, reflex function)	• Administration rate of aminoglycosides
• Dysphagia examination rate (within two days)	• Administration rate of third or later generation cephalosporin antibiotics
• Initial diagnosis	• Prophylactic antibiotics combination rate
• Diagnostic brain imaging rate (within 24 hours)	• Antibiotics prescription rate at discharge
• Blood lipids test rate	• Total average prophylactic antibiotics administration days (Administered at hospital + prescription at discharge)
• Initial treatment	• Documentation rate of information related to surgery
• Consideration rate of t-PA intravenous injection	• Documentation rate of information related to antibiotics administration
• Antithrombotics administration rate (within 48 hours)	• Documentation rate of history of antibiotics allergy
• Antithrombotics prescription rate at discharge	• Documentation rate of ASA class
• Anticoagulants prescription rate (atrial fibrillation patient)	

even given that for both indicators a large share of hospitals performed below the penalty threshold at the time of the baseline study. This points to some improvements following the implementation of the VIP.

As part of the VIP incentives, quality scores for each hospital are made public on the HIRA website. The US experience has shown that public disclosure of hospital scores can also be a good lever for quality improvement. There is little evidence on the impact of such non-financial incentives on provider behaviour in Korea. Given the highly competitive nature of the hospital market in Korea, however, public disclosure of performance scores could possibly be an important non-financial incentive to drive improvements in quality of hospital services.

Data source and flows

The performance data for the VIP come from HIRA's highly integrated claims database. The integration of the numerous health insurance funds under a



Application of value	Criteria announcement and incentive	Incentive and disincentive	Incentive and disincentive
Reporting of results	2011	2012	2013
Treatment record	2010	2011	2012

Figure 15.3 Ranking of hospitals by performance in the Korea VIP

Source: Kim et al., 2012.

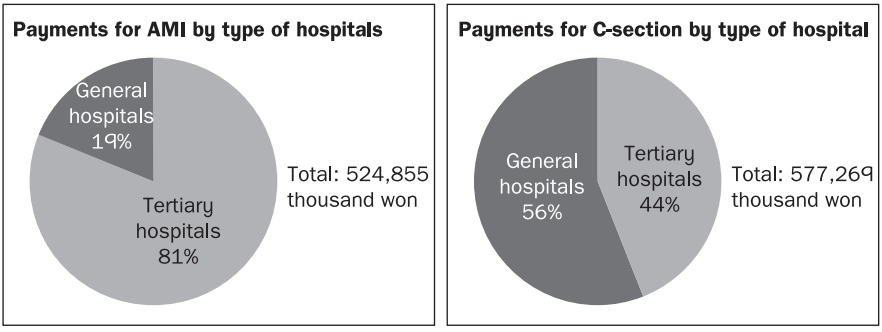


Figure 15.4 Bonus payments disbursed under the Korea VIP by type of hospital, 2011

Source: Kim et al., 2012.

single-payer system has led to a more integrated health information system in which every patient is identified through different levels of care using a unique patient identifier. This unique patient identifier now allows comprehensive data on patient health status and service use to be linked through reimbursement claims data. Data collected by HIRA include a broad range of indicators covering process and outcomes. Every year a quality assessment report is prepared by HIRA, which reviews patient claims in a wide range of areas of care in addition to the VIP performance domains (e.g. acute diseases, chronic diseases, health care utilization, long-term care).

The information needed for the AMI quality assessment is gathered from the claims data warehouse and supplemented by medical records data through a web-based hospital quality data acquisition system. The date of death for the case-fatality indicator is supplied by the Ministry of Public Administration and Security. The C-section rate is calculated only with the claims data warehouse.

Data are validated by direct inspection once a year to confirm the quality of the claims data. Claims data are cross-checked by survey data on a random 5 per cent sample of cases, with a maximum 20 cases per year per hospital). In 2011, 97.4 per cent of the performance data were found to be valid (HIRA, 2011).

Reach of the programme

Which providers participate and how many people are covered?

At the beginning of the programme, 44 tertiary teaching hospitals were mandated to participate in the VIP. Only one hospital is a public hospital owned by the Ministry of Health and Welfare, nine are national university hospitals belonging to the Ministry of Education, Sciences and Technology, and 34 are private university hospitals owned by university foundations (Chun et al., 2009).

In 2011 the programme was expanded to include general hospitals that treat AMI cases and that have at least 200 C-sections. For the AMI domain, 71 general hospitals (49 per cent of the total) were mandated to participate, and 50 of those hospitals also were mandated to participate in the C-section domain (Kim et al., 2012).

Results of the programme

Has the programme had an impact on performance, and have there been any unintended consequences?

Programme monitoring and evaluation

An evaluation was conducted by HIRA using the claims data for over 12,665 cases of AMI between July 2007 and December 2010 for all five performance domains and one outcome indicator. For the purpose of data cross-checking, survey data were also collected on a random sample of seven patient cases per hospital. Results were compared between years one and three to estimate improvement trends for hospital performance under the VIP programme. The results show improvement in all process indicators, although the baseline achievement levels were already high (Figure 15.5). The most notable improvement is shown on the indicators related to fibrinolytic therapy and timely PCI. For the drug administration indicators (administration of aspirin on arrival, prescription of aspirin and beta-blocker at discharge) achievement rates were high at the time of baseline data collection, and they have slightly increased over the period. The overall composite score for AMI increased 5.3 percentage points, from 92.1 per cent to 97.4 per cent (HIRA, 2010b).

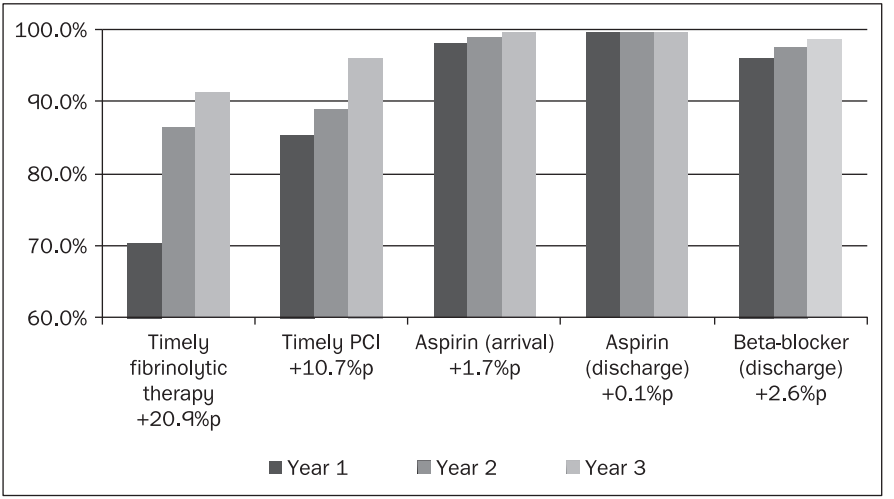


Figure 15.5 Improvement in process indicators in the Korea VIP, 2007–09

Source: HIRA 2010b.

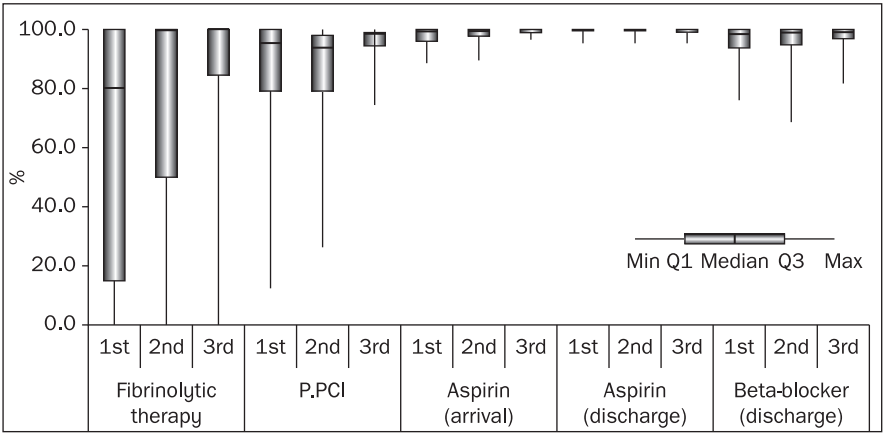


Figure 15.6 Decrease in variance in process indicators among hospitals in the Korea VIP, 2009

Source: HIRA, 2010b.

HIRA's evaluation also found that the gap in performance across hospitals has narrowed since the VIP was initiated. There has been a decrease in the variance among hospitals with respect to the two indicators on timely fibrinolytic therapy and timely PCI (Figure 15.6), with the lowest performing hospitals raising their standards of care and improving quality.

The impact of the VIP on C-section rates was also evaluated by HIRA and found to be only modest. Claims data for 64,887 deliveries between 2007

and 2010 were examined by HIRA using claims data. An analysis carried out between 2007 and 2009 showed that the overall composite score for this area of care decreased by only 1.6 points, although improvement did occur in the lowest performing group (HIRA, 2010b). Moreover, in practice, none of the hospitals scored below the penalty threshold, meaning that only bonus payments were distributed in the first two years of the programme (Figure 15.7).

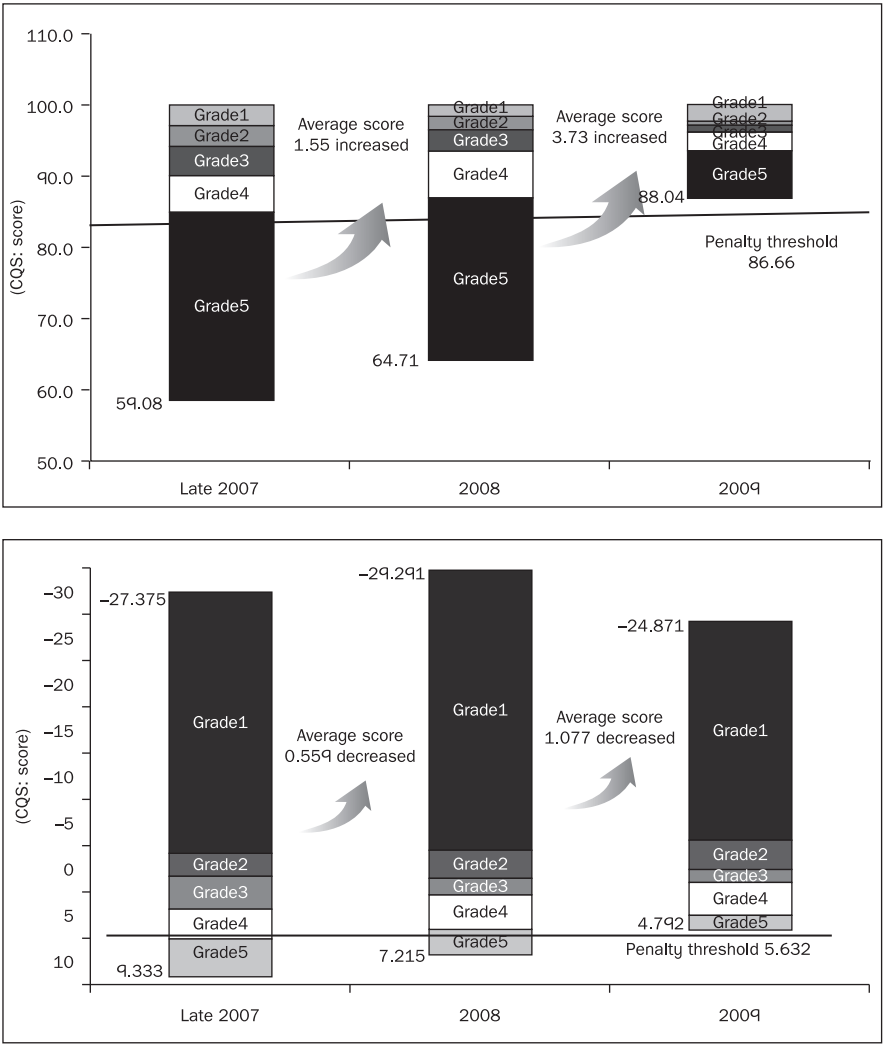


Figure 15.7 AMI composite quality score (high) and C-section (low) in the Korea VIP, 2007–09

Source: HIRA, 2010b.

Costs and savings

Estimating the total cost of the implementation and administration of the VIP is difficult, as HIRA routinely collects and monitors performance of providers on a regular basis, as part of the general assessment of hospitals. No additional data collection system or administrative layer was introduced following the implementation of the VIP, but there may be some administrative costs.

The cost of the bonus payments may be offset by lower costs of care in some cases. According to an economic evaluation carried out by HIRA, for example, the reduction in C-section rates amounted to a cost reduction of up to KW 1.14 billion in 2011, while the payment incentives amounted to KW 296 million for this area of care (HIRA, 2011). This estimate takes into account costs reductions associated with increased vaginal delivery and indirect economic impact (mainly complications) (HIRA, 2011).

There is no existing information on the use of bonuses by hospitals, although anecdotal reports suggest that additional payments are distributed to resident doctors. No study has been conducted to determine how bonus payments were used in tertiary hospitals to further drive quality improvements.

Provider response

Although there was initial opposition to the VIP by provider groups, after more than five years of implementing the programme, the hospitals that have been included have grown more supportive. A recent study found that more than 70 per cent of hospitals surveyed that currently participate in VIP are supportive of the programme (Lee et al., 2012). Nearly half of surveyed tertiary hospitals reported that the VIP has no significant financial effect on their institution, but 78 per cent reported that the programme has led to behaviour change among the providers.

Among those health care provider institutions without experience with the programme, however, both awareness of the programme and support are much lower. Although 96 per cent of general hospitals are aware of the VIP, only 38 per cent responded that they are supportive of the programme. Among clinics, only 35 per cent of respondents were even aware of the programme. These results suggest some potential challenges with stakeholder acceptance as the VIP expands beyond tertiary hospitals (Lee et al., 2012).

Overall conclusions and lessons learned

Has the programme had enough of an impact on performance improvement to justify its cost?

The evaluation of the VIP yielded mixed results. While some of the indicators related to care of AMI have improved, other indicators have only shown a small change. C-section rates appear to have decreased only marginally since the introduction of the VIP, and rates remain high compared to other OECD countries and far from the WHO recommendation. On the other hand, it does

appear that the VIP has contributed to reducing the variation in quality across hospitals for the clinical domains covered by the programme, and overall composite quality scores have improved, especially for the lowest performing grades.

The expansion of the VIP to general hospital aims to address shortcomings in quality of care beyond tertiary hospitals. So far, the uptake of the VIP among general hospitals has been relatively high, with 71 and 50 hospitals enrolled for the AMI and C-section clinical areas, respectively, within a year of expansion. However, the decision to collectively assess general and tertiary hospitals together might be viewed as unfair to general hospitals, as they tend to have less capacity to drive improvements and receive less funding from the NHIC. The NHIC pays an additional 30 per cent to tertiary hospitals to support costs related to investment in high-technology medical equipment and infrastructure. The current design of the programme might risk redirecting an even greater share of funding toward larger and better equipped facilities and exacerbate inequalities in funding and disparities in quality between tertiary and general hospitals.

Overall, the implementation of the VIP has shown some positive results, with no evidence on unintended consequences. Improvements in quality of some aspects of care have been achieved with relatively small bonuses. The implementation was also largely facilitated by the transition to electronic data interchange technology for submitting claims and the introduction of unique patient identifier (Kelly, Gray & Minges, 2003). The VIP is a good example of the use of routine data collection to assess performance and link it to financial incentives. The administrative and additional costs linked to the VIP cannot be properly calculated, as assessment of hospitals' performance through patient files and claims data has become a routine procedure in HIRA.

In light of the modest impact of the VIP on hospital quality indicators, HIRA and NHIC should consider a broader approach to improving quality of hospital care. Beyond process indicators, HIRA could consider a more comprehensive quality assessment tool. Moreover, HIRA or another agency involved in monitoring the VIP should attempt to understand how improvement is driven in individual hospitals, and seek to play a greater role in disseminating good practices. This new function would be important for general hospitals, in particular, which tend to start the VIP from a lower baseline performance level and could learn from the experience of tertiary hospitals. In addition to financial incentives, raising quality standards and applying clinical guidelines for AMI and C-sections in clinics and general hospitals should also be a priority to policymakers.

References

Centers for Medicare and Medicaid Services (2009) Medicare demonstrations show paying for quality health care pays off, *Medicare News*, 17 August.

Cho, Y.H. et al. (2010) The impact of pay for performance demonstration project for AMI care in Korea, *HIRA*, October.

Chun, C.B. et al. (2009) Republic of Korea health system review, *Health Systems in Transition*, 11(7): 1–184.

HIRA (2010a) *Comprehensive quality report of national health insurance: 2009*. Seoul: HIRA.

HIRA (2010b) *Report on value incentive program 2010*. Seoul: HIRA.

HIRA (2011) *Report of pay for performance (P4P) demonstration project 2011*.

Kelly, T., Gray, V. and Minges, M. (2003) *Broadband Korea: internet case study*. Geneva: International Telecommunication Union.

Kim, B.Y. (2010) *Quality assessment and pay for performance in Korea*. HIRA, May.

Kim, S.M. et al. (2012) Korean national health insurance value incentive program: achievements and future directions, *Journal of Preventive Medicine and Public Health*, 45(3):148–55.

Korean Statistical Information Service (2007) *Statistics Korea* (http: //kosis.kr/eng, accessed December 2013).

Lee, J. et al. (2012) Healthcare organizations' attitudes toward pay-for-performance in Korea, *Health Policy*, 108: 277–85.

National Institute for Clinical Excellence (2002) *Guidance on the use of drugs for early thrombolysis in the treatment of acute myocardial infarction*. London: National Institute for Clinical Excellence.

OECD (2009) *Achieving better value for money in health care*. Paris: OECD Publishing.

OECD (2010) *OECD health data*. Paris: OECD Publishing.

OECD (2011) *Health at a glance: OECD indicators*. Paris: OECD.

Park, W. and Yu, G. (2002) HRM in Korea: transformation and new patterns, in Z. Rhee and E. Chang, *Korean business and management: the reality and the vision*. Seoul: Hollym Corp.

Weil, A. (2007) Patient lives saved as performance continues to improve in groundbreaking CMS/Premier pay-for-performance project, *Premier*, 26 January.

chapter sixteen

United States: Maryland hospital acquired conditions programme

Robert Murray

Introduction

Since 1977, the State of Maryland in the United States (US) has operated a unified prospective hospital payment system in which all payers – public and private – pay the same rates for the same service at a given hospital. This all-payer system has been used as the foundation for pay for performance (P4P) programmes since 2009. One of the P4P programmes, the Maryland Hospital Acquired Conditions Programme (MHAC), links payments to hospital performance on a set of 49 potentially avoidable hospital acquired complications across all payers and patients in the state.

Maryland is the only state in the US operating a unified hospital payment system across all payers (Reinhardt, 2011; Murray, 2012). Because of its unique legal authority and relative political independence, the Maryland all-payer system operates quite differently from the general US model of health care financing and delivery. The Health Services Cost Review Commission (HSCRC) is the state government agency charged with the responsibility for establishing uniform payment rates ('all-payer rates') for all inpatient and outpatient services provided by Maryland's acute care hospitals. The HSCRC is governed by seven volunteer commissioners serving four-year staggered terms and appointed by Maryland's governor. The HSCRC's broad rate-setting authority has enabled it to establish consistent payment incentives for hospitals, which is in contrast to the more common situation in the US where prices for similar services in the same hospital vary considerably across payers (New Jersey Commission on Rationalizing Health Care Resources, 2009; Coakley, 2011). The participation of the government health insurance programmes Medicare and Medicaid in Maryland's all-payer system is made possible by a federal waiver, which exempts Maryland hospitals from national Medicare and state Medicaid fee schedules.

In 2008, the Center for Medicare and Medicaid Services (CMS) made plans to implement national hospital pay for performance (P4P) programmes to

promote the use of evidence-based process measures and reduce hospital acquired complications. The Medicare HAC programme was designed to penalize hospitals financially and thus encourage them to eliminate avoidable complications. The policy eliminates payment under Medicare's Inpatient Prospective Payment System for eight complications acquired by a patient during hospitalization that Medicare thought should be 100 per cent preventable.[1] Before the implementation of the Medicare HAC policy the presence of these complications would have (in most cases) resulted in a higher weighted Diagnosis Related Group (DRG) assignment for that patient and thus a higher payment for the hospital. CMS hoped that not paying extra for potentially expensive avoidable complications would provide hospitals with a disincentive to provide poor quality care.

The State of Maryland used its all-payer system as the basis for developing its own versions of the programmes to be applied to all 46 acute care hospitals in the state. These programmes were the Quality Based Reimbursement (QBR) programme and the MHAC programme. The QBR programme allocated rewards and penalties for hospitals based on their performance on evidence-based clinical process of care measures for heart attack, heart failure, pneumonia, and surgical infection prevention. The MHAC programme adjusts hospital payment based on performance related to potentially preventable complication rates.

Both Maryland programmes were facilitated by the national mandate for P4P programmes to improve hospital quality, the well-established unified hospital payment system, and the extensive data infrastructure created by the HSCRC for the development and implementation of its all-payer system. Through the QBR and MHAC programmes, Maryland has built on its all-payer system and data sources to use financial incentives to change the behaviour of hospitals to be in line with the primary policy goals of the HSCRC, namely, cost control, equity in payment, improved access to care, accountability and financial predictability and stability.

Health policy context

What were the issues that the programme was designed to address?

Policy objectives

While Maryland's all-payer hospital rate system was performing well against its stated objectives (to control the growth in cost per hospital admission, ensure access to life-saving hospital care and improve equity in payment), the impact on health care quality was not well documented (Murray, 2009). As a result there were concerns about the general incentive created by case-based payment systems to discharge patients too early and cut back on quality in other ways. Most states implementing case-based hospital payment in the 1980s based their systems on diagnosis-related groups (DRGs) as the unit of payment. DRG-based payment systems establish an average payment rate for all cases in

a DRG, and thus they provide strong incentives for hospitals to reduce length of stay, ancillary service use, and the intensity of service per inpatient stay. While these payment systems helped reduce unnecessary hospital services per case, there was concern that these financial incentives may have a negative effect on hospital outcomes. At the same time that these concerns about the incentives of DRG payment systems were growing, the US Institute of Medicine's landmark reports (*To Err is Human* in 1999 and *Crossing the Quality Chasm* in 2001) brought quality of care to the forefront of discussions on provider payment.

The literature on all-payer systems in the US was mixed, including the experience of a number of states in addition to Maryland that experimented with partial-payer or all-payer systems during the 1970s and 1980s. While some studies did find a correlation between the presence of rate-setting systems and higher mortality rates, other studies found no substantive difference between rate-setting states and non-rate-setting states in terms of overall hospital quality (Shortell & Hughes, 1988; Smith et al., 1993). Given the absence of accepted metrics on the quality of hospital care, however, the HSCRC had not been able to actively promote quality through a restructuring of the underlying incentives of the HSCRC's payment system.

As hospital quality concerns related to DRG-based payment and all-payer systems reached the forefront and national P4P programmes were being announced by CMS, the state of Maryland faced a policy imperative to actively promote better quality of care in hospitals. The HSCRC used its platform of unified hospital payment to develop P4P initiatives to promote hospital quality.

Stakeholder involvement

The HSCRC assembled work groups for the design of both the QBR and MHAC P4P programmes. The work groups included clinical and financial representatives of the full-time professional staff of the HSCRC and representatives of hospitals and private and public insurers. The HSCRC staff carried out the foundational analytical work and prepared draft recommendations for each P4P programme. The work groups then met over nine to twelve months to discuss and amend the original HSCRC recommendations on the evidence-based process measures for the QBR programme and the hospital acquired conditions in the MHAC programme. This process led to a near consensus of all those involved on the final recommendations for both P4P programmes presented to the HSCRC for their approval.

Prior to the announcement of CMS that P4P programmes would be implemented on a national scale, the HSCRC encountered considerable resistance from various industry stakeholders. When the national programmes were announced, both HSCRC staff and the stakeholders realized that Maryland would need to craft its own incentive-based approach to improve hospital quality, or it would face having the broader national programme imposed on the state at some later date. There was agreement that it would be better to develop a system that was more responsive to Maryland's circumstances, and therefore potentially more effective, than to have a system imposed by the federal government.

Reach of the programme

Which providers participate and how many people are covered?

Both the QBR and MHAC are mandatory programmes and applied to inpatient care provided by all of the state's 46 acute care hospitals. The hospital system in Maryland has a diverse array of acute care facilities, ranging from small 20–30 bed facilities in rural parts of the state to large 1000-bed and premier academic medical centres, such as Johns Hopkins Hospital and University of Maryland Medical Center in Baltimore. This system accounts for hospital revenues in excess of US$13 billion per year under the HSCRC's regulatory authority. Because the two P4P programmes apply only to inpatient care provided by hospitals, they directly impact about 700,000 inpatient cases per year, accounting for approximately US$9 billion in annual expenditures.

Technical design

How does the programme work?

In 2003, as the federal government was implementing its initial quality-based Pay-for-Reporting programme for Medicare, the HSCRC began to develop its QBR programme, which it ultimately launched in 2008. The Maryland programme provides financial incentives – both rewards and penalties – in Maryland hospital payment rates to encourage improvements in process-of-care measures, such as giving heart attack patients aspirin upon arrival at the hospital, or administering blood-thinning agents to surgery patients following certain surgical procedures. In creating the programme, the HSCRC worked with hospital and private payer representatives to develop a programme that mirrored the proposed federal Value-Based Purchasing (VBP) initiative but also could be implemented in the context of Maryland's all-payer rate system. Maryland's programme initially included 19 core CMS and Joint Commission process measures in the following four care domains: heart attack, heart failure, pneumonia, and surgical infection prevention.

For patients admitted for a heart attack, for example, hospitals are evaluated on the basis of the frequency with which they administered the following evidence-based processes of care. For heart attack patients these measures included: (1) aspirin at arrival; (2) aspirin prescribed at discharge; (3) angiotensin converting enzyme inhibitors or angiotensin receptor blockers for left ventricular systolic dysfunction; (4) adult smoking cession counselling; (5) beta blocker prescribed at discharge; (6) beta blocker at arrival. For each heart attack patient, hospitals receive credit for every time one of these six processes of care was administered.

Under the QBR programme, rewards and penalties are distributed to hospitals through their regulated payment levels, in a revenue-neutral manner with a linear distribution function. In other words, the net increases in rates for better performing hospitals are funded entirely by net decreases in rates for poorer performing hospitals. The worst performing hospital loses 0.5 per cent of its total inpatient revenue. In fiscal year 2012, Maryland reallocated

US$7.5 million among its 46 hospitals. A more detailed description of the QBR methodology can be found at the HSCRC's website.[2]

Performance domains and indicators

The HSCRC's MHAC programme has one performance domain – clinical quality – with 49 indicators for the rate of actual versus expected hospital acquired conditions. The HACs are derived from a list of 64 potentially preventable conditions (PPCs) developed by 3M Health Information Systems based on their clinical appropriateness and significant cost implications when they occur. PPCs are defined as harmful events (e.g. accidental laceration during a procedure) or negative outcomes (e.g. hospital acquired pneumonia) that develop after hospital admission and may result from processes of care and treatment rather than from natural progression of the underlying illness and are therefore potentially preventable (Hughes et al., 2006).

The HSCRC chose its performance domain and indicators to be consistent with the national P4P programmes for Medicare implemented by CMS. The HSCRC began developing the Maryland HAC programme as CMS was developing and preparing to implement its HAC programme for hospitals paid through Medicare in the rest of the country. The CMS HAC programme would deny hospitals extra payment for complications acquired during the hospital stay and not present on admission. Section 5001(c) of the Deficit Reduction Act of 2006 required the Secretary of Health and Human Services to identify at least two target conditions for the programme. The criteria for selecting the target conditions were as follows:

1. High cost, high volume, or both.
2. Result in the assignment of a case to a DRG payment group that has a higher payment when present as a secondary diagnosis.
3. Could have been prevented through the application of evidence-based guidelines.

As noted, under the federal CMS programme, hospitals would not receive additional payment for cases (i.e. a higher DRG payment) when any of CMS's selected conditions was coded as a secondary diagnosis if it was not present on admission.

The original intent of CMS was not to pay for complications that were expensive and thought to be 100 per cent preventable, as CMS believed this would place a hospital at some financial risk for poor quality. The degree to which payments were reduced, however, depended on: (1) whether the deletion of a hospital acquired condition code changes the DRG assignment for a particular patient; (2) the magnitude of the change in payment; (3) the number of conditions and patients to which the policy applies. In practice, the presence of one of the eight HACs identified by CMS did not always result in a payment reduction. In addition, the very limited scope of the CMS HAC programme (only eight conditions that occur relatively infrequently) meant that the financial impact on hospitals was very limited. This result caused many to question whether the federal HAC programme would have the intended impact on hospital behaviour (McNair et al., 2009).

The existence of the Medicare waiver made it possible for Maryland to experiment with variations on the general themes outlined by the federal government. The HSCRC convened the MHAC Payment Policy Group, comprising hospital industry and payer stakeholders, to review the Medicare HAC list and the 3M Health Information Systems list of 64 PPCs. The group chose a subset of 11 PPCs from the 3M list that were thought to be the most preventable because of the wide variation in hospital performance rates for those conditions. The initial set was eventually expanded to 49 PPCs in response to concerns about possible unintended consequences if hospitals had the incentive to focus disproportionately on too narrow a set of conditions. The set of HACs used by the Maryland HAC programme is shown in Table 16.1.

Table 16.1 Maryland HAC categories

Hospital Acquired Conditions	
Stroke & Intracranial Haemorrhage	Urinary Tract Infection
Extreme CNS Complications	GU Complications Except UTI
Acute Pulmonary Edema and Respiratory Failure without Ventilation	Renal Failure without Dialysis
Acute Pulmonary Edema and Respiratory Failure with Ventilation	Renal Failure with Dialysis
Pneumonia & Other Lung Infections	Diabetic Ketoacidosis & Coma
Aspiration Pneumonia	Post-Haemorrhagic & Other Acute Anaemia with Transfusion
Pulmonary Embolism	In-Hospital Trauma and Fractures
Other Pulmonary Complications	Post-Operative Infection & Deep Wound Disruption Without Procedure
Shock	Post-Operative Wound Infection & Deep Wound Disruption with Procedure
Congestive Heart Failure	Moderate Infections
Acute Myocardial Infarction	Septicaemia & Severe Infections
Major Gastrointestinal Complications with Transfusion or Significant Bleeding	Acute Mental Health Changes
Other Cardiac Complications	Decubitus Ulcer
Ventricular Fibrillation/Cardiac Arrest	Cellulitis
Peripheral Vascular Complications Except Venous Thrombosis	Reopening Surgical Site
Venous Thrombosis	Other Surgical Complication – Mod
Major Gastrointestinal Complications without Transfusion or Significant Bleeding	Post-Operative Haemorrhage & Hematoma with Haemorrhage Control Procedure or I&D Proc
Cardiac Arrhythmias & Conduction Disturbances	Accidental Puncture/Laceration During Invasive Procedure
Major Liver Complications	Accidental Cut or Haemorrhage During Other Medical Care
Other Gastrointestinal Complications without Transfusion or Significant Bleeding	Post-Operative Haemorrhage & Hematoma without Haemorrhage Control Procedure or I&D Proc

Encephalopathy	Other Complications of Medical Care
Inflammation & Other Complications of Devices, Implants or Grafts Except Vascular Infection	Gastrointestinal Ostomy Complications
Iatrogenic Pneumothorax	Infections due to Central Venous Catheter
Mechanical Complication of Device, Implant & Graft	Obstetrical Haemorrhage with Transfusion
Infection, Inflammation & Clotting Complications of Peripheral Vascular Catheters & Infusions	

Incentive payments

After extensive deliberations with stakeholders, the HSCRC staff initially recommended a payment methodology that mirrored Medicare's proposed 'payment denial' approach, where the presence of a post-admission complication would result in a lower DRG payment for that case. Hospital representatives and clinicians raised concerns, however, about possible unintended consequences of the punitive approach adopted by CMS. Also, because this methodology focused only on complication categories that were thought to be 100 per cent or nearly 100 per cent preventable, it limited the number of complication categories that could be included in the programme. In response to these concerns, the HSCRC revised the MHAC proposed policy to shift the focus of the programme away from a case-specific approach to a hospital's rate of actual versus expected hospital acquired condition, with the expected rates defined by the case mix of patients the hospital treats. This change, which emphasized rates of complications, allowed the HSCRC to include complication categories that were not always 100 per cent preventable and to expand the list of HACs from 11 to 49. The initial year of the programme used 2009 as the base year, 2010 as the performance year, and adjusted hospital payment rates for 2011.

To calculate bonuses/penalties for each hospital, all hospitals are ranked based on the total impact of their HACs, which is determined by both the incidence of HACs and the amount of excess charges they created, as a percentage of their total inpatient charges. The incidence of complications is the count of each HAC adjusted for the patient case mix, which is calculated using All Patient Refined Diagnosis Groups (APR-DRG) and Severity of Illness (SOI) categories. This method calculates the hospital's expected incidence of complications given the severity of its patient case mix based on the defined performance criteria (state average in the previous year), and compares expected values to the observed rates. The amounts of additional charges for each HAC are estimated using a state-wide regression analysis of standardized charges for all of the 3M PPCs in the previous year, which controls for the admission diagnosis and severity. The total amount of additional charges for HACs is aggregated across poorer performing hospitals and redistributed to better performing hospitals. In this way, the programme is budget neutral, or does not add any additional costs to the system for the incentive payments.

The final ranking of hospitals is based on the overall additional resource use due to complication rates for each hospital as a percentage of their total inpatient charges. Table 16.2 provides the results of the HSCRC's estimate of the relative cost of each of the 64 PPCs identified by the 3M Health Information System methodology.

All hospitals that perform better than the state-wide average (or overall expected rate of complications weighted by the charge weights of each HAC) receive bonus payments. Although a state-wide normative standard was used as the basis for assessing hospital performance (i.e. hospitals below the state-wide average were penalized and those above it were provided rewards), the hospitals were generally accepting of this approach because of their long-standing confidence in the risk-adjustment mechanism used by the HSCRC.[3] Adjusting an individual hospital's performance to compare its rate of complications to that of the state as a whole by DRG and SOI subcategory allows for an analysis that matches each hospital's performance (given their mix of patients) to state-wide performance for the same mix of patients.

Once the final ranking of hospitals is established, the HSCRC allocates a predetermined amount of revenue (or percentage of net patient revenue) to be 'at risk.' This predetermined at-risk percentage is then applied to the revenue of hospitals performing less favourably than the state-wide average. The resulting dollar amount is then reallocated to hospitals performing better than the state-wide average. The HSCRC has gradually increased the amount of hospital revenue at risk for penalties and rewards for the MHAC programme, reflecting more emphasis on outcome-based P4P. In the first year, HSCRC reallocated only the revenue from the annual payment increase to account for inflation, resulting in a very modest US$2.1 million total amount reallocated from poorer performing hospitals to better performing hospitals. The total amount reallocated increased to US$13.3 million in the second year and an estimated US$20.1 million in the third year.

Table 16.3 illustrates how revenue is reallocated across hospitals to create budget-neutral incentive payments. Individual hospital rewards and penalties were either added to (in the case of rewards) or subtracted from (in the case of penalties) each hospital's annual inflation adjustment. The annual inflation adjustment is an amount approved by the HSCRC as an increase to the base rates of all hospitals to cover the expected inflation of inputs to the hospital production process (e.g. salaries and benefits, utilities, capital, contractual services, supplies, etc.). For example, if the HSCRC established an annual system-wide inflation adjustment of 2.5 per cent in a given year, a hospital (such as Peninsula Regional in Table 16.3) that performed well on the MHAC P4P programme would have its annual inflation update increased by the magnitude of its MHAC reward. In this example, Peninsula Regional would receive an increase to its hospital payment rates of 3.343 per cent (the 2.5 per cent update applied to all hospitals plus Peninsula Regional's own 0.843 per cent MHAC reward). Likewise, a poorer performing hospital (such as the University of Maryland) would have its MHAC penalty of –0.76 per cent applied to its 2.5 per cent inflation update, resulting in a net rate increase of only 1.74 per cent (the 2.5 per cent base update less 0.76 per cent MHAC penalty specific to University of Maryland).

The precise amount of the total revenue available to be reallocated as incentive payments is determined by the distribution of hospitals with positive and negative performance, and the relative size of hospitals. The presence of larger hospitals in the poorer performing group would effectively 'free up' a larger amount of revenue to be allocated to the better performing hospitals. Table 16.3 is a simulation of the FY 2012 reallocation of revenue based on a previous year's actual performance.

Data sources and flows

The data for the MHAC come from the information system created by the HSCRC to operate the all-payer system. The HSCRC created an extensive data infrastructure, first collecting uniform cost data from the hospitals and then assembling a robust patient-level case-mix data set, containing detailed demographic, financial and clinical data on every inpatient and outpatient hospital encounter. HACs are identified based on the information for secondary diagnoses in the hospital discharge abstract data set submitted to HSCRC. The data set is also the basis for the HSCRC inpatient DRG-based prospective payment system and enables hospitals to report up to 30 secondary diagnosis and 15 procedure codes for each patient. Thus, it is in the financial interest of the hospitals to submit complete and accurate discharge data to the HSCRC to ensure appropriate payment. The HSCRC has established administrative and chart review processes to audit the diagnosis coding on an ongoing basis using screening algorithms to assess accuracy. In addition, the HSCRC implemented selected audits of medical records to determine the accuracy of hospital coding of the Present on Admission indicator (which identifies whether a particular condition/complication developed prior to the admission or whether it was a result of substandard hospital care during the hospitalization.[4]

Improvement process

How is the programme leveraged to achieve improvements in service delivery and outcomes?

An important feature of the MHAC programme is that it created a specific tool for discussing, assessing and evaluating overall and relative quality of care. The use of a uniform method for categorizing complication rates provides a useful communication tool to all professionals (financial, clinical and coding personnel), which has been essential to achieving behaviour changes to reduced complications over time. This is similar in a way to how the adoption of DRGs for payment purposes created a mechanism for financial and clinical personnel within a hospital to communicate hospital performance that had both a financial and clinical dimension.

The HSCRC provides state-wide performance data to each hospital at the beginning of the year, which shows each hospital's position relative to state-wide performance by complication category. The HSCRC also provides quarterly

Table 16.2 Example of Maryland HAC categories and estimated resource use (charges) by category, fiscal year 2009

	MHACs	*Rate per 1000*	*Charge/cost per episode*	*Total charge/cost*
Infection-related hospital acquired conditions				
Infections due to Central Venous Catheters	569	0.67	$27,198	$15,475,662
Post-Operative Wound Infection & Deep Wound Disruption w/Procedure	118	0.51	$19,361	$2,284,598
Septicemia & Severe Infections	3,659	4.87	$16,731	$61,218,729
Post-Operative Infection & Deep Wound Disruption w/o Procedure	1,380	0.16	$14,422	$19,902,360
Pneumonia & Other Lung Infections	5,072	6.25	$14,895	$75,547,440
Moderate Infections	1,317	1.04	$13,849	$18,239,133
Infection, Inflammation & Clotting Complications of Peripheral Catheters	525	0.73	$13,959	$7,328,475
Aspiration Pneumonia	1,953	2.91	$11,181	$21,836,493
Inflammation & Other Complications of Devices, Implants or Grafts	1,380	1.88	$9,024	$12,453,120
Urinary Tract Infection	7,416	10.85	$8,038	$59,609,808
Cellulitus	1,464	1.83	$4,474	$6,549,936
Subtotal	24,853		$12,089	$300,445,754

Top 10 other hospital acquired conditions				
Acute Pulmonary Edema & Respiratory Failure w/Ventilation	940	3.08	$23,062	$21,678,280
Other Complications of Medical Care	822	1.12	$18,945	$15,572,790
Shock	2,038	3.54	$17,634	$35,938,092
Decubitus Ulcer	1,312	1.81	$17,951	$23,551,712
Ventricular Fibrillation/Cardiac Arrest	908	3.22	$16,433	$14,921,164
Venous Thrombosis	1,908	2.52	$12,347	$23,558,076
Other Pulmonary Complications	5,177	4.58	$8,306	$43,000,162
Renal Failure w/o Dialysis	8,576	11.49	$7,451	$63,899,776
Post-Operative Hemorrhage & Hematoma	3,921	14.74	$5,952	$23,337,792
Acute Pulmonary Edema & Respiratory Failure w/Ventilation	6,274	3.08	$5,318	$33,365,132
Subtotal	31,876		$9,375	$298,822,976
All Other MHACs (PPCs) included in the MHAC methodology (28)	20,247		$7,950	$160,970,480
Other PPCs not yet included in the MHAC methodology (13)	9,892		$2,900	$28,685,986
State-wide Totals FY2010	86,868	2.38	$9,082	$788,925,196

Table 16.3 Example of the revenue-neutral reallocation (or 'scaling') of rewards/penalties in the Maryland HAC programme

Scaling $20,109,000 Statewide on the Basis of Rates of Maryland Hospital Acquired Conditions (risk adjusted)							
	Hospital name	*% Revenue at risk*	*Rate of complications*	*Rank*	*Percentile*	*Adjustment to AI* rate update*	*$ Impact to hospital*
Worst performing hospitals	Prince Georges Hospital	$178,190,982	4.35%	46	0.0%	–1.50%	–$2,672,865
	Montgomery General Hospital	$100,257,704	2.60%	45	4.0%	–0.90%	–$899,312
	Shady Grove Adventist Hospital	$213,531,084	2.38%	44	7.0%	–0.82%	–$1,753,090
	Doctors' Community Hospital	$114,901,908	2.30%	43	9.0%	–0.79%	–$911,172
	University of Maryland Medical Center	$560,470,190	2.21%	42	11.0%	–0.76%	–$4,270,783
	Sinai Hospital	$365,194,980	1.97%	41	13.0%	–0.68%	–$2,479,674
	:		:				
	:		:				
	:		:				
State-wide average			0.00%				
	:		:				
	:		:				
	:		:				
Best performing hospitals	Calvert Memorial Hospital	$61,553,830	–1.26%	7	87.0%	0.82%	$506,588
	Peninsula Regional Medical Center	$262,717,273	–1.29%	6	89.0%	0.84%	$2,214,707
	Carroll Hospital Center	$143,028,893	–1.31%	5	91.0%	0.86%	$1,224,327
	Mercy Medical Center	$205,914,768	–1.43%	4	93.0%	0.94%	$1,925,303
	Garrett County Municiple	$20,456,088	–1.47%	3	96.0%	0.96%	$196,583
	Dorchester General Hospital	$30,163,278	–1.57%	2	98.0%	1.03%	$309,475
	Bon Secours Hospital	$74,581,886	–2.26%	1	100.0%	1.48%	$1,101,574

Note: All 46 hospitals were included in the analysis, however, not all 46 hospitals are shown here (table truncated).
* AI = Annual Inflation.

updates to hospitals so they can track their performance during the course of the year. These data are usually available 60 days after the end of each quarter, so although not immediate feedback, this information does give hospitals some ability to adjust their efforts during the course of the year. Providing hospitals with data showing their relative performance by category provides clinical and financial staff with the actionable intelligence they need to first identify areas of concern and then systematically target these areas, with the overall goal of reducing the frequency of hospital acquired complications. Because hospital performance on HACs overall is weighted by the relative costliness of each HAC, positive or negative relative performance on more expensive conditions would have a proportionately larger impact on a hospital's overall score. This gives the incentive to hospital personnel to first focus improvement efforts on HACs with both higher frequency and higher cost.

The relative breadth of the MHAC programme, 49 complications across nearly all product lines of a full service hospital, while daunting to some hospital personnel, has also provided an incentive to implement systematic approaches to reducing complications across the hospital in general. This counteracted the potential unintended consequence of many P4P programmes of providers targeting or reallocating resources to certain quality metrics now 'under the spotlight,' also known as 'teaching to the test'.

In addition to providing financial incentives for hospitals to improve their rates of hospital acquired complications, the HSCRC presents the results of annual hospital performance on its website, with relative rankings and some indication of whether hospitals are performing better than, worse than, or at an average level of relative performance. The Maryland Hospital Association, which is largely an advocacy organization on behalf of the 46 acute care hospitals, was not involved in the development of the HSCRC's web-based reports and indicated their opposition to public reporting, despite their previous endorsement of the overall MHAC P4P incentive programme design.

Results of the programme

Has the programme had an impact on performance, and have there been any unintended consequences?

The HSCRC has noted improvements in patient outcomes and costs based on data from the initial two years of MHAC, as shown in Table 16.4. Based on regression estimates for 2009, preventable hospital acquired conditions as defined by the HSCRC accounted for US$789 million, about 8 per cent of inpatient revenue (Calikoglu et al., 2012). This figure is consistent with national studies that estimate that complications of hospital care account for 15 per cent of inpatient costs nationally of which about half are thought to be preventable (McNair et al., 2009). Over the first two years that the MHAC incentive programme was in place, complication rates declined by 15 per cent in Maryland, resulting in US$110.9 million savings in the system. The improvements were consistent across the Maryland HACs, with 75 per cent of HACs included in the programme declining in both years (Calikoglu et al.,

Table 16.4 State-wide trends in hospital acquired conditions in Maryland, fiscal years 2009–11

Potentially preventable complication (PPC) number and name	*Risk adjusted complication rates/1000*			*Annual change*		
	FY 2009	*FY 2010*	*FY 2011*	*FY 2009–10*	*FY 2010–11*	*Charge/cost*
State-wide	*2.38*	*2.17*	*2.02*	*–8.85%*	*–7.03%*	*–$110,957,872*
PPC 1 Stroke & Intracranial Hemorrhage	1.56	1.54	1.51	–1.28%	–1.95%	–$250,565
PPC 2 Extreme CNS Complications	0.63	0.56	0.51	–11.11%	–8.93%	–$968,065
PPC 3 Acute Pulmonary Edema & Resp. failure w/o Ventilation	7.9	7.41	6.64	–6.20%	–10.39%	–$4,739,899
PPC 4 Acute Pulmonary Edema & Resp. failure w/Ventilation	3.08	3.07	2.83	–0.32%	–7.82%	–$2,231,164
PPC 5 Pneumonia & Other Lung Infections	6.25	5.54	4.98	–11.36%	–10.11%	–$10,286,330
PPC 6 Aspiration Pneumonia	2.91	2.73	2.66	–6.19%	–2.56%	–$2,052,555
PPC 7 Pulmonary Embolism	1.1	0.96	0.98	–12.73%	2.08%	–$357,218
PPC 8 Other Pulmonary Complications	4.58	4.17	3.95	–8.95%	–5.28%	–$1,466,468
PPC 9 Shock	3.54	3.61	3.09	1.98%	–14.40%	–$3,654,322
PPC 10 Congestive Heart Failure	3.82	3.25	2.59	–14.92%	–20.31%	–$2,636,381
PPC 11 Acute Myocardial Infarction	3.02	2.58	2.3	–14.57%	–10.85%	–$2,332,140
PPC 12 Cardiac Arrythmias & Conduction Disturbances	295.78	284.11	286.01	–3.95%	0.67%	–$44,424
PPC 13 Other Cardiac Complications	0.65	0.48	0.4	–26.15%	–16.67%	–$364,816
PPC 14 Ventricular Fibrillation/Cardiac Arrest	3.22	2.83	2.76	–12.11%	–2.47%	–$5,566,386
PPC 15 Peripheral Vascular Complications except Venous Thrombosis	0.51	0.4	0.31	–21.57%	–22.50%	–$1,402,443
PPC 16 Venous Thrombosis	2.52	2.04	2.06	–19.05%	0.98%	–$2,414,855
PPC 17 Major Gastro. Complications w/Transfusion or Significant Bleeding	1.5	1.16	1.16	–22.67%	0.98%	–$2,641,855

PPC 18 Major Gastro. Complications w/Transfusion or Significant Bleeding	0.46	0.48	0.43	4.35%	–10.42%	–$156,733
PPC 19 Major Liver Complications	0.58	0.55	0.53	–5.17%	–3.64%	–$338,033
PPC 20 Other Major Gastro. Complications w/o Transfusion or Significant Bleeding	0.53	0.51	0.49	–3.77%	–3.92%	$107,935
PPC 22 Urinary Tract Infection	10.85	7.91	6.96	–27.10%	–12.01%	–$17,254,363
PPC 23 GU Complications Except UTI	0.72	0.63	0.5	–12.50%	–20.63%	–$468,867
PPC 24 Renal Failure w/o Dialysis	11.49	11.1	10.84	–3.39%	–2.34%	–$1,905,891
PPC 25 Renal Failure w/Dialysis	0.23	0.23	0.19	0.00%	–17.39%	–$461,888
PPC 26 Diabetic Ketoacidosis & Coma	0.08	0.09	0.09	12.50%	0.00%	$35,470
PPC 27 Post-Hemorrhagic & Other Acute Anemia w/Transfusion	2.57	2.15	1.94	–16.34%	–9.77%	–$608,184
PPC 28 In-Hospital Trauma and fractures	0.25	0.24	0.19	–4.00%	–20.83%	–$266,330
PPC 31 Decubitus Ulcer	1.81	1.37	1.38	–24.31%	0.73%	–$5,554,086
PPC 33 Cellulitis	1.83	1.49	1.45	–18.58%	–2.68%	–$798,443
PPC 34 Moderate Infections	1.04	0.91	0.74	–12.50%	–18.68%	–$1,626,652
PPC 35 Septicemia & Severe Infections	4.87	3.87	3.11	–20.53%	–19.64%	–$16,564,123
PPC 36 Acute Mental Health Changes	0.16	0.93	0.79	481.25%	–15.05%	–$258,851
PPC 37 Post-Operative Infection & Deep Wound w/o Procedure	4.62	4.4	3.89	–4.76%	–11.59%	–$992,140
PPC 38 Post-Operative Wound Infection & Deep Wound Disruption w/Procedure	0.51	0.49	0.34	–3.92%	–30.61%	–$448,209
PPC 39 Reopening Surgical Site	0.8	1.2	1.26	50.00%	5.00%	$1,850,052
PPC 40 Post-Operative Hemorrhage & Hematoma w/Hemmorrhage Control proc. Or I&D proc.	14.74	13.12	11.18	–10.99%	–14.79%	–$4,154,100

(*continued*)

Table 16.4 State-wide trends in hospital acquired conditions in Maryland, fiscal years 2009–11 (*continued*)

Potentially preventable complication (PPC) number and name	*Risk adjusted complication rates/1000*			*Annual change*		
	FY 2009	*FY 2010*	*FY 2011*	*FY 2009–10*	*FY 2010–11*	*Charge/cost*
State-wide	*2.38*	*2.17*	*2.02*	*–8.85%*	*–7.03%*	*–$110,957,872*
PPC 41 Post-Operative Hemorrhage & Hematoma w/Hemorrhage Control Proc or I&D proc	1.24	1.24	1.27	0.00%	2.42%	$134,742
PPC 42 Accidental Puncture/Laceration during Invasive procedure	7.49	6.28	5.96	–16.15%	–5.10%	–$1,254,462
PPC 43 Accidental Cut or Hemorrhage during other Medical Care	0.16	0.17	0.15	6.25%	–11.76%	$29,824
PPC 44 Other Surgical Complication – Moderate	2.31	1.9	1.6	–17.75%	–15.79%	–$1,600,776
PPC 47 Encephalopathy	1.38	1.24	1.23	–10.14%	–0.81%	–$1,543,462
PPC 48 Other Complications of Medical Care	1.12	1.02	1.15	–8.93%	12.75%	–$216,874

2012). Infection-related HACs declined much faster than the rest of the HACs in the MHAC programme, which may indicate the impact of other clinical quality improvement projects implemented in the state. Maryland hospitals participate in national programmes to eliminate central line-associated bloodstream infections, catheter-associated urinary tract infections, and a state-wide hand hygiene collaborative.

Extrapolated to the Medicare Fee for Service payments nationally, similar results could have resulted in cost reductions of approximately US$1.3 billion. For all hospital care (across all payers, public and private in the broader US hospital system), such a programme could have resulted in an estimated US$5.3 billion reduction in costs associated with the reduction of preventable complications over two years (assuming that 58 per cent of hospitals spending, or US$814 billion, was for inpatient care) (Centers for Medicare and Medicaid Services, 2011).

To test whether the observed changes in MHACs are due to the P4P programme or related to other changes occurring in the health care market, the changes in the PPCs used for MHAC were compared to changes in the excluded PPCs as a control group. Of the 64 control PPCs, 15 PPCs were excluded due to lack of significant added costs or clinical concerns.

While PPCs used in MHAC declined by 18.6 per cent in two years, the excluded PPCs increased by 2.8 per cent (Calikoglu et al., 2012). A further analysis is needed to explain why excluded PPCs increased while the PPCs used in MHAC declined in the first two years of the programme. The increase in the excluded PPC rates may reflect real changes in these complications, and it may also partially be the result of improvements in documentation and coding, which might have differential impact on the excluded PPCs. Finally, the increase in the hospital acquired conditions excluded from MHAC may be the result of hospitals shifting the focus of their quality efforts.

Overall conclusions and lessons learned

Has the programme had enough of an impact on performance improvement to justify its cost?

Maryland's QBR and MHAC P4P initiatives show that the application of consistent and clear financial incentives can help promote hospital care quality in addition to improved efficiency. The core payment system was used as a powerful 'change agent' to create moderate to strong financial incentives to drive hospital care improvement efforts. The declaration by CMS that it planned to implement its own P4P programmes for Medicare nationally helped overcome the reluctance to tackle this issue by both regulators and providers at the state level.

The MHAC focus on hospital acquired infections has shown to be a more acceptable way to link quality of care to financial incentives than the more process-oriented measures of the QBR programme. While there were some operational and political advantages to start the HSCRC's P4P efforts with a focus on evidence-based process measures, the HSCRC staff has since

generated a number of substantive concerns over the effectiveness of the QBR. First, the focus on promoting the use of a set of processes of care measures appears to be a highly proscriptive approach to improving quality and requires extensive preparatory work to define the appropriate measures. For instance, in the case of an Acute Myocardial Infarction patient, hospitals have strong incentives to provide all seven evidence based processes of care – aspirin upon admission, beta-blockers upon discharge – whether that patient truly needs these prescribed interventions or not. More importantly, the literature analysing the link between selected process measures and patient outcomes provides very limited evidence that these measures are related to improvements in patient outcomes (Bradley et al., 2006; Ryan, 2009; Morse et al., 2011; Shahian et al., 2012).

Moreover, the process-based QBR P4P programme continues to be implemented without any concerted attempt to assess the unintended consequences of the programme. The four areas of clinical focus (Acute Myocardial Infarction, Pneumonia, Heart Failure and Surgical Infection Prevention) cover relatively narrow domains of hospital services. There remains a distinct possibility that hospitals reallocate limited resources away from other quality assurance efforts to focus on these care domains and selected process measures. Although average composite quality scores of hospitals increased across all clinical areas during the programme, the lack of comprehensive time-series (pre- and post-measures) or control group evaluation creates a challenge to determine the extent to which improvements were directly attributable to participation in the programme. The time lag between the year performance is measured and the financial results are known, and by the time rewards and penalties are determined, hospitals have four months left in the next measurement year, which may weaken the incentive. By contrast, because hospitals generate their own MHAC-related data, they are able to monitor their year-to-year performance on a more or less real-time basis. Also, as noted, the HSCRC provides running quarterly analyses of state-wide performance to show how hospitals are performing relative to all other facilities in the state.

These data and analyses are available 60 days after the end of any given quarter. Because of these and other limitations to the QBR, the HSCRC moved expeditiously toward the use of risk-adjusted outcomes measures and increased financial incentives in outcome-based P4P. The use of hospital acquired complications as an outcome measure remedied many of the concerns outlined above. Because of the broad scope of the programme and the need only to do well 'on average' within a given complication category, there seems to be a lower likelihood that hospitals would adopt a 'teach to the test' strategy and reallocate resources inappropriately, because all areas of inpatient care are included in the analysis. The consistency and the strength of the incentives applied, which were progressively increased, provided more compelling reasons for hospital personnel to engage in these communications and coordinate activities to reduce their rates of complications over time. On the other hand, using a rate-based approach introduced considerable methodological challenges, such as whether the existing risk-adjustment method used for hospital DRG-based payment was sufficient, and whether and

how to incorporate complications that were thought to be less than 100 per cent preventable.

The inclusive and deliberative process established by the HSCRC that included HSCRC staff, payer, hospital financial and clinical staff proved to be a key element in the final acceptance of the MHAC methodology by virtually all stakeholders. Broad acceptance of the risk-adjustment methodology employed by the HSCRC in the context of the larger payment system, also helped reduce opposition to the implementation of MHACs.

Another concern raised by the hospitals was that the MHAC programme constituted an 'unfunded' mandate by the HSCRC. However, hospitals were generating savings by eliminating avoidable complications, while their DRG payments remained unchanged the majority of the time. According to the HSCRC's analysis, the assignment of cases to different DRGs when complications were eliminated only resulted in reduced payment about 40 per cent of the time. Thus 60 per cent of the time hospital DRG revenue remained the same, even while their resource costs to the hospital were reduced with the removal of a preventable complication. This in combination with the potential to generate P4P rewards, made it possible for a hospital to earn significant positive returns on their investment in improving quality.

P4P programmes such as the MHAC do rely on the availability of timely and accurate administrative claims data. Many clinicians have been critical of the use of administrative data in P4P schemes because of concerns about the accuracy of coding (Pronovost & Liford, 2011). Yet, these concerns are reminiscent of complaints lodged against rate-setting agencies in New Jersey, Maryland and the Medicare programme when DRGs were first introduced. Over time the use of DRGs in an incentive payment system leads to substantial improvements in both the accuracy and the depth of coding. As a result, hospital discharge data sets that have been used in payment systems are arguably far more accurate and complete than claims data sets that are routinely collected for just monitoring purposes. Hospitals produce more accurate and complete data, because payment levels are negatively impacted if the information on secondary diagnoses and procedures are inaccurate or incomplete.

Although there was evidence of improvement in process of care measures under the QBR, there was not clear evidence that Maryland improved faster than the nation (which had a pay-for-reporting and later P4P process measure initiative in place over this period).

However, the reductions in hospital acquired conditions experienced by Maryland hospitals provides some evidence that the employment of consistent and powerful financial incentives can motivate focused efforts on the part of hospital personnel to improve outcomes. The MHAC programme also appeared to offer some distinct advantages over CMS's HAC initiative. First, by virtue of its risk-adjusted rate of complication approach, the Maryland programme includes 49 complication categories (including complications that are not considered 100 per cent preventable) vs. only eight for the federal programme. Second, the adoption of a programme that allowed for the application of significant rewards and penalties (arguably sufficient to change behaviour) based on performance may have added to the success and overall acceptance of the programme by

hospitals. Third, the incorporation of increasingly stronger financial incentives with each subsequent year, applied uniformly through each hospital's rate base (which covers all-payers, but public and private), was thought to be a source of considerable motivation (by the third year of the programme, as shown in Table 16.3, some hospitals stood to gain or lose as much as 1.5 per cent of their total revenue base). Finally, the use of an interactive vetting process with hospital clinical representatives (which resulted in a refinement of the methodology), along with the use of the HSCRC's extensive clinical coding and case-mix data infrastructure added credibility to the effort and broad acceptance by the hospital industry.

Although state-based all-payer rate-setting programmes have been effective in controlling the costs of hospital admissions, the literature is mixed on whether rate setting has a negative impact on quality of care (Atkinson, 2009). Yet, just as rate-setting systems have been effective in structuring incentives to improve operational efficiency per case, they also can be effective in the same way to structure incentives to improve quality performance when broad-based outcome measures such as hospital acquired avoidable complications are linked to payment. In this way, rate-setting systems can perhaps provide a more powerful mechanism to promote systematic improvements in quality, because the incentives applied under an all-payer rate system are consistent and can be applied in a progressively stronger fashion over time. Hospitals outside of Maryland are faced with myriad performance measures being applied in P4P programmes of different payers, which may result in unclear or even conflicting incentives.

Notes

1 The original CMS Hospital Acquired Conditions included: foreign object retained after surgery; air embolism; blood incompatibility; stage III and IV pressure ulcers; in-hospital falls; catheter-associated urinary tract infection; vascular catheter-associated infection; and surgical-site infection (mediastinits) after Coronary Artery Bypass Graft.
2 http://www.hscrc.state.md.us/init_qi_qbr.cfm.
3 The HSCRC has long used All-Patient Refined (APR)-DRGs developed by 3M Health Information Systems. APR-DRGs are a severity adjusted grouping system where each of the 314 DRGs has four severity of illness subcategories.
4 A recent analysis by an outside evaluator indicated that 98 per cent of Maryland hospitals are correctly coding for present on admission, compared to 53 per cent for hospitals nationally (Michael Pine and Associates was contracted by the HSCRC to review present on admission coding in all hospitals in the state).

References

Atkinson, J. (2009) *Hospital rate setting revisited*, *Commonwealth Fund* (http://www.commonwealthfund.org/Publications/Issue-Briefs/2009/Oct/State-Hospital-Rate-Setting-Revisited.aspx, accessed February 2011).

Bradley, E.H. et al. (2006) Hospital quality for acute myocardial infarction: correlation among process measures and relationship with short-term mortality, *Journal of the American Medical Association*, 296(1): 72–8.

Calikoglu, S., Murray, R. and Feeney, D. (2012) Hospital pay-for-performance programs in Maryland produced strong results, including reduced hospital-acquired conditions, *Health Affairs*, 31(12): 2649–58.

Centers for Medicare and Medicaid Services (2011) *National health expenditures 2010 highlights*. Baltimore, MD: Centers for Medicare and Medicaid Services.

Coakley, M. (2011) *Examination of health care cost trends and cost drivers: report for annual public hearing*. Boston, MA: Office of Massachusetts Attorney General Martha Coakley (http://www.mass.gov/Cago/docs/healthcare/2011_HCCTD.pdf, accessed June 2013).

Hughes, J.S. et al. (2006) Identifying potentially preventable complications using a present on admission indicator, *Health Care Financing Review*, 27(3): 63–82.

Institute of Medicine (1999) *To err is human: building a safer health system*. Washington, DC: National Academies Press.

Institute of Medicine (2001) *Crossing the quality chasm: a new health system for the 21st century*. Washington, DC: National Academies Press.

McNair, P.D., Luft, H.A. and Bindman, A.B. (2009) Medicare's policy not to pay for treating hospital-acquired conditions: the impact, *Health Affairs*, 28: 1485–93.

Morse, R.B. et al. (2011) Hospital-level compliance with asthma care quality measures at children's hospitals and subsequent asthma-related outcomes, *Journal of the American Medical Association*, 306(13): 1454–60.

Murray, R.B. (2009) Setting hospital rates to control costs and boost quality: the Maryland experience, *Health Affairs*, 28: 1395–1405.

Murray, R.B. (2012) The case for a coordinated system of provider payments in the United States, *Journal of Health Politics, Policy and Law*, 37: 679–95.

New Jersey Commission on Rationalizing Health Care Resources (2009) *Final report [Internet]*. Trenton, NJ: The Commission (http://www.nj.gov/health/rhc/finalreport/documents/entire_finalreport.pdf, accessed May 2013).

Pronovost, P.J. and Liford, L. (2011) Analysis and commentary: a road map for improving the performance of performance measures, *Health Affairs*, 30(4): 569–73.

Reinhardt, U.E. (2011) The many different prices paid to providers and the flawed theory of cost shifting: is it time for a more rational all-payer system?, *Health Affairs*, 30: 2125–33.

Ryan, A.M. (2009) Effects of the premier hospital quality incentive demonstration on Medicare patient mortality and cost, *Health Services Research*, 44: 821–42.

Shahian, D.M. et al. (2012) Association of National Hospital Quality Measure adherence with long-term mortality and readmissions, *British Medical Journal*, 21(4): 325–36.

Shortell, S. and Hughes, E. (1988) The effects of regulation, competition and ownership on mortality rates among inpatients, *New England Journal of Medicine*, 318(17): 1100–7.

Smith, D.W., McFall, S.L. and Pine, M.B. (1993) State rate regulation and inpatient mortality rates, *Inquiry*, 30: 23–33.

chapter seventeen

United States: Hospital quality incentive demonstration

Cheryl Cashin

Introduction

Medicare is the primary source of health insurance for the elderly and disabled in the United States (US), and so it is by far the largest payer in the US health care system accounting for 21 per cent of total national health expenditure in 2011 (CMS, 2013). The growth in Medicare spending has been one of the most significant burdens on the federal budget, reaching 13 per cent of total federal spending in 2010 (US Office of Management and Budget, 2009). Medicare spending growth is not only driven by rapid improvements in medical technology and the ageing of the 'baby boom' population, but also by the traditional provider payment systems that have failed to contain costs even after the move from traditional fee-for-service to diagnosis-related group (DRG) payment for hospitals.

In the midst of projections of sharply rising costs and even possible insolvency for the Medicare system, by 1997 Medicare reform became a focus of national policy debate. The National Bipartisan Commission on the Future of Medicare was established by the Balanced Budget Act (BBA) of 1997. Following the commission process, President Clinton put forward his own proposal for Medicare reform in June 1999. Although all of the various proposals attempted to address Medicare expenditure growth, proposed reforms of provider payment under the Medicare programme were modest and focused largely on restricting growth in payment rates, without fundamentally addressing the role of Medicare as a prudent health purchaser. At the same time alternative approaches to securing the future financial sustainability of Medicare were being debated, a number of hospital quality improvement initiatives were started under Medicare in response to the alarming Institute of Medicine (1999) report on preventable medical errors. The link between provider performance and payment under Medicare began with a 'pay-for-reporting' programme to begin assembling reliable data on quality indicators. As part of the Medicare Prescription Drug, Improvement, and Modernization Act (MMA)

of 2003, Congress put in place financial incentives for hospitals to participate in the public reporting of quality information. This initiative laid the foundation for the first attempts to link the quality of hospital services with Medicare payment.

In July 2003, Premier Inc. (a network of private non-profit hospitals) and the Centers for Medicare and Medicaid Services (CMS) launched the Hospital Quality Incentive Demonstration Project (HQID), a three-year programme designed to determine if financial incentives are effective at improving the quality of inpatient care for beneficiaries covered by the traditional Medicare programme. The traditional Medicare programme that is run by the federal government covers about three-quarters of Medicare beneficiaries, while the other 25 per cent opt for Medicare Advantage, which is the privately administered alternative.

In the HQID project, CMS measured performance and paid incentives to participating hospitals that achieved high levels of quality in five clinical areas of acute care: acute myocardial infarction (AMI), heart failure, pneumonia, coronary artery bypass graft, and hip and knee replacements. The incentive system was competitive, with hospitals in the two highest deciles of performance for a condition receiving a bonus, while those with the poorest performance risked financial penalty (Centers for Medicare and Medicaid Services, n.d.). HQID was the first and largest federally sponsored pay for performance (P4P) programme in the US (Glickman et al., 2007). The initial three-year demonstration was extended for another three years through 2009, and the programme has formed the cornerstone of a recent widespread proposal to move toward P4P models through 'value-based purchasing' in the US Medicare and Medicaid programmes (US Department of Health and Human Services, 2007).

Health policy context

What were the issues that the programme was designed to address?

The acceleration of the quality movement in the US health care system can be traced back to the Institute of Medicine's (IOM) publication *To Err is Human: Building a Safer Health System* (Institute of Medicine, 1999). The watershed report made public the widespread preventable medical errors in hospitals that led to between 44,000 and 98,000 deaths each year. That report was followed by *Crossing the Quality Chasm: A New Health System for the 21st Century*, which exposed that health care in the US routinely deviated from clinical guidelines best practices (Institute of Medicine, 2001). A key recommendation of that report was that payment incentives for providers needed to be realigned to support quality improvement. These alarming reports prompted a rush of congressional hearings and proposals for creating a culture of quality in hospitals that coincided with the focus on reforming Medicare.

Although linking payment to performance was not at the forefront of the proposals of the late 1990s to reform Medicare and contain costs, support for

experimenting with such an approach emerged following the release of the IOM studies. There were a number of P4P schemes operational in the private sector by 2002, but these initiatives typically remained small and experimental. The first large-scale private sector P4P was initiated by the Integrated Healthcare Association in California in 2003 and is still ongoing (see Chapter 13). The programme covers 35,000 physicians participating in integrated medical groups or independent practice associations who care for more than ten million patients. The IHA programme continues to be unique in the US because of the attempt to align performance measures across seven major payers. Before 2003, there was no precedent for such programmes in the government Medicare or Medicaid health insurance programmes. The first steps toward measuring quality under the federally funded health insurance programmes began with 'pay for reporting' programmes, which were a significant step at that time.

The Hospital Quality Alliance (HQA) collaborative, which made information about hospital quality performance available to the public, was the first step toward quality measurement and reporting in the Medicare system. In December 2002, the Department of Health and Human Services announced a partnership with several collaborators to promote hospital quality improvement and public reporting of hospital quality information. In July 2003, CMS began the National Voluntary Hospital Reporting Initiative, which later became the HQA. The HQA provides data to a CMS database that initially included ten measures of clinical quality among three conditions: AMI, heart failure, and pneumonia. Initially 4200 hospitals voluntarily submitted data on the quality measures. All acute care hospitals were invited to participate. The initiative was strengthened with financial incentives under the MMA of 2003, which legislated that hospitals that did not report on ten measures of quality receive a 0.4 per cent reduction in their annual Medicare payment update for inpatient hospital services. By linking participation in the programme to Medicare payment, CMS was able to achieve participation rates of more than 98 per cent (Lindenauer, 2007).

Taking the step to link *achievement* on reported quality measures with payment under Medicare, however, was considered a more drastic move, and a pilot approach was required. Demonstration projects undertaken under the CMS demonstration authority are a well-institutionalized approach in CMS to test and measure the effects of potential programme changes before they are launched nationwide. CMS's demonstration authority allows the agency to waive certain Medicare payment rules that determine what services are covered and how they are paid in order to test potential improvements. Most major payment reforms under Medicare, including DRG payment for hospital services, were initiated after demonstration projects (CMS, 2010).

The P4P model tested in the HQID demonstration project included financial incentives and public recognition for top-performing hospitals, as well as financial penalties for hospitals that did not improve above a predefined quality measure threshold by the third year of the project. The objective was to test the hypothesis that quality-based incentives would raise the entire distribution of hospitals' performance on selected quality metrics, and to evaluate the impact of incentives on quality process and outcomes, as well as costs.

Stakeholder involvement

The HQID demonstration project was designed jointly between CMS and Premier Inc., a nationwide organization of not-for-profit hospitals and other providers. Premier Inc. is a provider-owned health care alliance that represents more than 2300 US hospitals and over 70,000 health care sites in 38 states. Premier submitted an unsolicited proposal to CMS for the demonstration programme and was selected as the sole programme partner with CMS because of its database of hospital performance benchmarks (CMS, 2009). Premier's Perspective database, the largest clinical comparative database in the nation, was already in place and able to track hospital performance in several clinical areas. Although other agencies were not involved in the design of HQID, the specific measures included were largely based on those already developed and in use by government and private organizations, such as the National Quality Forum (NQF), the American Hospital Association (AHA) and the Leapfrog Group.

Technical design

How does the programme work?

Performance domains and indicators

HQID linked incentive payments to 34 nationally defined, standardized, risk-adjusted measures covering both processes of care to reflect compliance with clinical guidelines (e.g. administration of prophylactic antibiotic prior to surgery), and patient outcomes (e.g. mortality). Performance was measured for the five acute clinical conditions available in Premier's database: AMI, coronary artery bypass graft, heart failure, community-acquired pneumonia, and hip and knee replacement. By the final year of the demonstration, two additional clinical areas were added: surgical care improvement and ischaemic stroke (Premier Inc., 2009).

The quality measures were based on indicators widely accepted and in use by nationally recognized health institutions. For example, the indicators included all ten indicators from the starter set of the National Voluntary Hospital Reporting Initiative, 27 indicators were National NQF indicators, 15 indicators were Joint Commission Core Measures indicators, and four indicators were the patient safety indicators of the Agency for Healthcare Research and Quality (AHRQ). The set of performance measures is present in Table 17.1.

The eligible patient populations for each clinical area were identified by the ICD-9-CM diagnosis codes and/or procedures codes associated with their admissions. Hospitals were required to participate in all five of the clinical areas, but if there were fewer than 30 cases in a clinical area the hospital was excluded from that area. CMS calculated a Composite Quality Score annually for each clinical area for each demonstration hospital with the minimum sample of 30 cases. Separate scores were calculated for each clinical condition by 'rolling-up' individual process and outcome measures into an overall quality score. Performance measures that represented patient outcomes were risk adjusted using well-established methods (Premier Inc., 2006). CMS then ranked

Table 17.1 Performance indicators in the US HQID

Clinical area and indicators
Acute Myocardial Infarction (AMI)
Aspirin at arrival
Aspirin prescribed at discharge
Angiotensin converting enzyme inhibitor (ACEI) for left ventricular systolic dysfunction (LVSD)
Smoking cessation advice/counselling
Beta blocker prescribed at discharge
Thrombolytic received within 30 minutes of hospital arrival
PCI received within 120 minutes of hospital arrival
Inpatient mortality rate
Coronary Artery Bypass Graft (CABG)
Aspirin prescribed at discharge
CABG using internal mammary artery
Prophylactic antibiotic received within 1 hour prior to surgical incision
Prophylactic antibiotic selection for surgical patients
Prophylactic antibiotic discontinued within 24 hours after surgery end time
Inpatient mortality rate
Post operative haemorrhage or hematoma
Post operative physiologic and metabolic derangement
Heart Failure (HF)
Left ventricular function assessment
Detailed discharge instructions
ACEI for LVSD
Smoking cessation advice/counselling
Community Acquired Pneumonia (PN)
Percentage of patients who received an oxygenation assessment within 24 hours prior to or after hospital arrival
Initial antibiotic consistent with current recommendations
Blood culture collected prior to first antibiotic administration
Influenza screening vaccination
Pneumococcal screening/vaccination
Antibiotic timing, percentage of pneumonia patients who received first dose of antibiotics within four hours after hospital arrival
Smoking cessation advice/counselling
Hip and Knee Replacement (HKR)
Prophylactic antibiotic received within 1 hour prior to surgical incision
Prophylactic antibiotic selection for surgical patients
Prophylactic antibiotics discontinued within 24 hours after surgery end time
Post operative haemorrhage or hematoma
Post operative physiologic and metabolic derangement
Readmission 30 days post discharge

the quality scores of individual hospitals into deciles to identify top performers for each condition, which were published on the HQID website. All of the hospitals in the top 50 per cent of hospitals were reported as top performers on the website. Those hospitals in the top first or second deciles received the financial bonus. Quality incentive payments were limited to only Medicare

patients, but the quality scores were based on measures of care for all adults within the clinical areas (Premier Inc., 2006).

Incentive payments

In the first phase of the demonstration project, the top-performing 20 per cent of all hospitals within each clinical area were eligible for a quality incentive payment. If the hospital was in the top decile of performers, the incentive payment was two per cent of their Medicare payment for all Medicare patients treated for that specific clinical condition. For hospitals in the second decile, the incentive payment was one per cent of their Medicare payment.

The incentive payment system also included a penalty for poor performers. Hospitals that did not score above the ninth decile threshold in any of the clinical areas received a one per cent reduction of their Medicare payment. Hospitals that did not score above the tenth decile threshold in any of the clinical areas received a two per cent reduction. Because a hospital would have to be in the lowest decile in all of the clinical areas to be penalized, few hospitals were penalized, with only three receiving a penalty in 2007.

In the second phase of the demonstration project (2006–2009), the incentive payment structure was revised to reward performance improvement. Hospitals could receive a financial reward in each clinical area in three ways: (1) attaining the median level of performance; (2) achieving a performance level in the top 20 per cent of hospitals; (3) achieving significant improvement (in the top 20 per cent of improvers). CMS allocated 40 per cent of its budget to attainment awards, and 60 per cent to top performer and top improvement awards. The penalty system remained the same in the second phase of the demonstration project. This change in the payment model significantly increased the number of hospitals eligible for an incentive payment each year.

The per-patient payment amounts were uniform across hospitals and across clinical conditions. The payment rates were calculated by dividing the available 40 per cent of the incentive award budget by the total number of discharged patients of hospitals attaining the median for the attainment award, and by dividing the available 60 per cent of the incentive award budget by the total number of discharged patients of hospitals eligible for the top performance and improvement awards.

Incentive payment amounts for individual hospitals were based on the number of cases identified by CMS as being traditional Medicare beneficiaries who received care within the demonstration year within the clinical area, as determined by the principal diagnosis or principal procedure code. All awards in year five were based on the change in the hospital composite quality score in the performance year compared to two years prior (year three to year five). Participants were eligible to receive a maximum of 12 awards.

Data sources and flows

To participate in the demonstration, hospitals were required to allow Premier to submit patient-level data and hospital-level quality data to CMS for all discharges from the five clinical areas. The first step in the data submission

process was to send the monthly discharge summary file to Premier. This file includes the patient account number, patient demographic information, physician information, payer information, and all applicable ICD-9-CM diagnosis and procedure codes, which were used to group the patients into clinical conditions. Next, Premier grouped the hospital data into the HQID clinical conditions and populated the Premier Quality Measures Web Tool. Once the patients were grouped to clinical conditions, hospitals were required to submit specific additional data from the patient records on the care provided. Premier's web-based tool applied over 200 business rules to audit the quality of the data, with any errors identified sent to the hospital for correction (Premier Inc., 2006). Once the error correction process was complete, Premier sent reports to the hospitals for review and validation before sending the data to CMS's QualityNet warehouse. QualityNet is the CMS-approved privacy-protected website for secure communications and health care quality data exchange between quality improvement organizations, hospitals, physician offices, other health care providers and networks, and data vendors (QualityNet, 2010).

Hospitals were required to pass a data validation process to be eligible for quality incentive payments. CMS validated the data by pulling a sample of seven patients from each hospital and requesting copies of the patient records from the hospitals for review by the Clinical Data Abstract Center (CDAC). CDAC re-abstracted the medical record data into a CMS tool and compared it to the hospital abstracted data results submitted to the warehouse. The demonstration project has a second validation process for rate calculations. After the patient-level data was submitted to the QualityNet warehouse, CMS and Premier calculated the hospital-level payment rates and together verified the accuracy.

Reach of the programme

Which providers participate and how many people are covered?

Between 222 and 273 acute care hospitals across 38 states participated in HQID during each of its six years, covering about 400,000 patients annually. Participating hospitals tended to be large and urban, with more than 80 per cent of them located in urban areas and 40 per cent having more than 300 beds (Premier Inc., 2006). In 2006 only 14 per cent of participating hospitals were teaching hospitals.

The financial incentive is considered to be modest, at only one to two per cent of Medicare payment for only five clinical conditions. The incentive is further diluted by the fact the hospitals in the US receive revenue from a multitude of different payers, most of them private insurers that often pay higher rates than Medicare. Nonetheless, hospital margins are typically slim at under ten per cent in the US, with a large share of hospitals operating with negative margins (AHA, 2013), so one–two per cent of Medicare payment is not trivial for hospitals. Also, the absolute level of the incentive payment to individual hospitals was quite large in some cases, often reaching over $100,000 per hospital. The top performers could earn a total of up to nearly $1 million across all clinical areas (Butcher, 2007).

Improvement process

How is the programme leveraged to achieve improvements in service delivery and outcomes?

Achieving the quality measures in HQID required a concerted effort on the part of hospitals to increase the consistency in processes of care in the five clinical areas. Hospitals adopted strategies such as forming collaborative work groups across several hospitals or hiring additional staff to collect data (Grossbart, 2006). Some hospitals viewed their participation in HQID as an opportunity to implement a tracking system, identify areas for improvement and see how they stacked up against other hospitals (Butcher, 2007). Top performing HQID hospitals cited commitment from leadership and administrative staff, comprehensive physician engagement, and the involvement of interdisciplinary teams in designing and implementing care delivery standards as critical success factors (Finarelli, 2009). The organization and communication among hospitals in the Premier alliance also provided an important channel to disseminate best practices. Through site visits and meetings with top-performing hospitals, Premier documented the ways in which those hospitals reported creating a culture of quality (Premier Inc., 2006).

Results of the programme

Has the programme had an impact on performance, and have there been any unintended consequences?

Performance related to specific indicators

Over the first five years of HQID for which data are available, the average composite quality score increased in all five clinical areas, and the variation in performance across hospitals was reduced, although the starting point was relatively high in most cases. The change in the average composite quality score between October 2003 and September 2008 is shown in Figure 17.1. Additional research by Premier showed that by September 2008 HQID participants scored on average 6.4 percentage points higher (95.05 per cent to 88.64 per cent) than non-participants across 19 measures. The details of the research are not available, however, and it is not possible to assess the extent to which participating hospitals differed from non-participating hospitals.

Programme monitoring and evaluation

As is typical for CMS demonstration projects, an external evaluation was commissioned after the first three years of HQID. The evaluation examined whether the programme had an independent effect on quality measures in three of the five clinical areas (AMI, heart failure, and pneumonia), as well as the effect of the demonstration on Medicare outlays and beneficiary average length of hospital stay. The evaluation found that although average composite scores

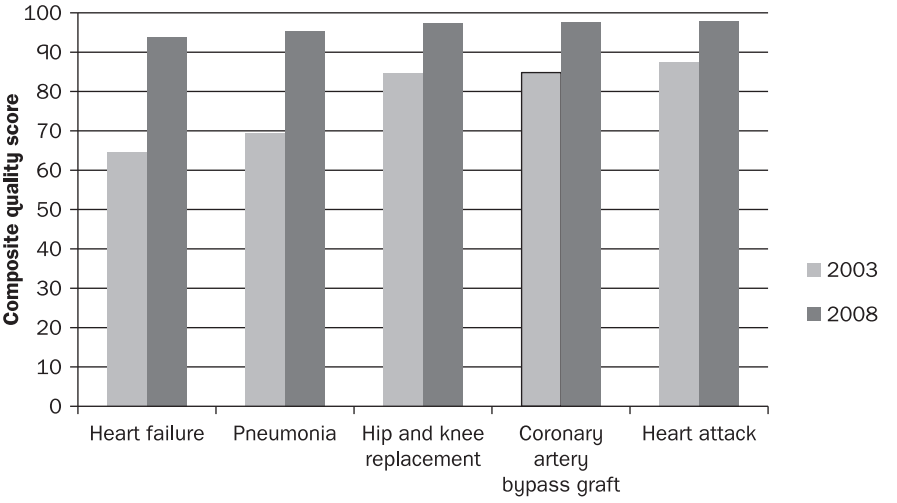


Figure 17.1 Change in the average composite quality score in the US HQID between 2003 and 2008

Source: Premier Inc., 2009 (http://www.premierinc.com/about/news/09-aug/hqid081709.jsp).

increased in all three clinical areas studied, the increases that could be attributed to the programme were only 0.7, 3.8 and 2.4 percentage points for AMI, heart failure and pneumonia, respectively (Kennedy et al., 2008). The evaluation also found that the programme was not budget neutral, as the outlays for the bonus payments were not offset by lower outlays or penalties (see below).

Aside from the early evaluation, no continued monitoring of the programme or final evaluation of the demonstration were completed. Premier Inc. published the results of the performance measures on its website for the first five years of the demonstration, but after the second year of the programme, no monitoring reports were produced to accompany the indicator tabulations. Furthermore, no attempt was made to assess any unintended consequences of the programme. For example, the indicator for whether a pneumonia patient receives the first dose of antibiotics within six hours after arrival at the hospital has been criticized for possibly leading to overuse of antibiotics and contributing to drug resistance (Wachter, 2006). No attempt was made to assess the impact of this or other indicators on overprovision of some services that are related to bonus payments and underprovision of others.

Several independent studies provide almost no evidence of an effect of HQID participation on quality of care. A peer-reviewed study comparing HQID with comparable hospitals that only publicly reported their quality results found that HQID hospitals had slightly greater improvements in quality over a two-year period than did those receiving no financial incentives (Lindenauer, 2007). Another independent study of four acute care hospitals that participated in HQID and five that did not, however, found that the performance of the participating hospitals accelerated in year one of the programme, but that the

scores of the two groups converged over three years (Grossbart, 2008). A recent study examining the impact of the change in the incentive structure to reward performance improvement found that in practice the new incentive design resulted in the strongest incentives for hospitals that were already performing above the median (Ryan, Blustein & Casalino, 2012).

Participation in HQID has not been found to lead to improvements in outcomes for any of the covered clinical areas. A study of the effect of the programme on AMI outcomes found that participation in HQID over the three-year period 2003-2006 was not associated with a significant improvement in the quality of processes of care or outcomes (Glickman et al., 2007). Another study found no evidence that HQID improved 30-day mortality rates for AMI, heart failure, pneumonia, or CABG (Ryan, 2009). Finally a study of the long-term impact of HQID on 30-day mortality for the five clinical areas in the programme found no evidence that the programme led to a decrease in 30-day mortality over the period 2003 to 2009 (Jha et al., 2012).

There is also no evidence that the process measures used by HQID themselves are associated with outcomes. A study of hospital quality process measures reported on the CMS website 'Compare', a subset of which was used in HQID, found that hospital performance along those measures predict only small differences in hospital risk-adjusted mortality rates (Werner & Bradlow, 2006). Another independent study found that a higher composite quality score for the hip and knee replacement surgery measures was not associated with lower rates of complications or mortality (Bhattacharyya et al., 2009).

Equity

There is some evidence that the HQID did not worsen equity by penalizing hospitals that serve a larger share of indigent patients, and in fact the gap in performance may have narrowed. An independent peer-reviewed study showed that HQID hospitals caring for a higher proportion of poor patients improved at a more rapid rate than those not participating in the project. Among both P4P hospitals and those in a national sample, hospitals with a higher share of indigent patients had lower baseline performance than did those with fewer indigent patients. A higher share of indigent patients was associated with greater improvements in performance for AMI and pneumonia, but not for congestive heart failure, and the gains were greater among hospitals that participated in HQID than among the national sample. After three years, hospitals that had a higher share of indigent patients and received financial incentives caught up for all three conditions, whereas those with more indigent patients among the national sample continued to lag (Jha, Orav & Epstein, 2010).

Costs and savings

CMS budgeted about $12 million per year for the incentive payments. Overall, CMS awarded more than $48 million over five years to top providers. With an average participation of 250 hospitals per year and the top 20 per cent receiving bonus payments, the availability of funds for incentive payments averaged

$240,000 per rewarded hospital. The 2008 evaluation commissioned by CMS found that any improvements in quality that could be attributed to HQID did not lead to any reductions in total payment per Medicare episode. The programme was found to create a net increase in costs of $41 per episode over all clinical areas, or a three-tenths of one per cent increase (Kennedy et al., 2008).

Some leaders of hospitals participating in HQID claimed that the bonus money did not cover the administrative costs that the project imposes on their institutions (Hospitals and Health Networks, 2007). Premier Inc. on the other hand claimed that their analyses showed cost savings to hospitals related to the quality improvements driven by the programme. Premier's analysis after the first three years of the project estimated that the median hospital costs per patient for HQID participating hospitals declined more than $1000 across the first three years of the project while the median mortality rate decreased by 1.87 per cent during the same time frame (Remus, 2006). The Premier study, however, was descriptive and did not control for patient factors or hospital characteristics that may be associated with both costs and quality measures. An independent study failed to find any impact of HQID, either positive or negative, on Medicare's costs (Ryan, 2009).

Provider response

There are no surveys available to shed light on the hospital experience with HQID, but some interviews in the press with hospital administrators indicate a perception that participation in HQID sharpened the hospitals' focus on specific quality improvements, but that participation came at a relatively high cost. In fact, the number of hospitals participating declined over the period of the demonstration, from a high of 273 in the first year to closer to 220 by the end of HQID. The reasons for the declining participation are not clear, but because Premier, Inc. required that hospitals renew their subscription to the relatively expensive database tool as a condition for participation, the cost to hospitals of participation was seen as a limiting factor for expanding the reach of HQID (Grossbart, 2008).

Overall conclusions and lessons learned

Has the programme had enough of an impact on performance improvement to justify its cost?

Premier Inc. and CMS have claimed that the HQID P4P programme was a striking success (Manos, 2009). HQID was deemed a success in spite of the lack of monitoring of the various aspects of the programme design and implementation and an early evaluation that showed minimal impacts on improvements in processes of care. Improvement in the performance indicators over the life of the project were taken as de facto evidence of the effectiveness of HQID in improving hospital quality, and as such value for Medicare money spent. The programme has been used as the blueprint for a large-scale CMS

proposal for P4P under its 'value-based purchasing' initiative. In the Deficit Reduction Act of 2005, Congress mandated that CMS develop a plan for hospital value-based purchasing, which was released in November 2007. It is not clear how important the biased reports of HQID success were in influencing the decision made by Congress to expand the programme nationwide.

The conclusions about the success of HQID trumpeted by Premier Inc., which were embraced uncritically throughout both the mainstream media and some of the professional health care media in the US, have not been supported by the results of a number of rigorous peer-reviewed studies. There are other reasons to question the validity of a national roll-out of the HQID model. First, although there is no doubt that improvements in processes of care for the five clinical areas were observed over the life of HQID, this has not been associated with better patient outcomes. Furthermore, the hospitals that participated in HQID were mostly large urban hospitals, which may not be representative of most hospitals serving Medicare beneficiaries. The participation rate of hospitals declined significantly over the life of the demonstration, the reasons for which have not been explored publicly.

Finally, the possibilities for a conflict of interest in the evaluation of HQID should not be ignored. Premier Inc. submitted an unsolicited proposal to CMS and was chosen as the sole participant in the demonstration. The provider alliance received nearly $50 million in bonus payments over the six years of the demonstration. Premier Inc. and CMS drew conclusions about the results of the programme largely based on descriptive analyses by Premier Inc. itself. The analytical methods and the validity of the conclusions have not been critically assessed, and they have not been supported by peer-reviewed published studies. Given that the programme relied on an expensive web-based quality reporting tool, which hospitals must subscribe to for a fee paid to Premier Inc., it is possible that Premier could benefit greatly if the model is replicated nationwide.

One area that deserves further attention is the evidence that participation in HQID may have helped close the performance gap between hospitals serving a larger share of indigent patients and those serving higher income communities. Understanding how the incentive programme may have disproportionately benefitted hospitals serving low-income patients may have implications for equity within the Medicare system. Overall, however, the six-year demonstration has shed remarkably little light on whether and how the P4P programme drove improvement along the process of care measures, whether the incentive payments were too high or too low relative to the results achieved, and the extent to which the programme had positive or negative spillover effects that should be harnessed or mitigated in a national roll-out. More rigorous analysis of such questions would be beneficial before large-scale expansion of the model is undertaken.

References

AHA (2013) *Chartbook: trends affecting hospitals and health systems* (http://www.aha.org/research/reports/tw/chartbook/ch4.shtml, accessed June 2013).

Bhattacharyya T. et al. (2009) Measuring the report card: the validity of pay-for-performance metrics in orthopedic surgery, *Health Affairs*, 28(2): 2526–32.
Butcher, L. (2007) CMS pays dividend for hospitals, but not always financially, *American College of Physicians Hospitalist*, September (accessed November 2010).
Centers for Medicare and Medicaid Services. (n.d.) *Hospital quality initiatives* (http://www.cms.hhs.gov/HospitalQualityInits, accessed November 2010).
CMS (2009) *Premier hospital quality incentive demonstration fact sheet*. Washington, DC: Centers for Medicare and Medicaid Services.
CMS (2010) *Demonstration projects & evaluation reports* (http://www.cms.gov/DemoProjectsEvalRpts/01_Overview.asp#TopOfPage, accessed November 2010).
CMS (2013) *NHE fact sheet* (http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NHE-Fact-Sheet.html, accessed June 2013).
Damberg, C. et al. (2009) Taking stock of pay-for-performance: a candid assessment from the front lines, *Health Affairs*, 28(2): 517–25.
Finarelli, H. (2009) Community need analyses: what they tell us about physician shortage, *Strategies & Solutions*, November.
Glickman, S. et al. (2007) Pay for performance, quality of care, and outcomes in acute myocardial infarction, *Journal of the American Medical Association*, 2373–80.
Grossbart, S. (2006) What's the return? Assessing the effect of 'pay-for-performance' initiatives on the quality of care delivery, *Medical Care Research and Review*, 63(1): 29S–48S.
Grossbart, S. (2008) Effectiveness of pay for performance as a quality improvement strategy, *Prescriptions for Excellence in Health Care* (3): 1–3.
Hartman, M.A. (2009) National health spending in 2007: slower drug spending contributes to lowest rate of overall growth since 1998, *Health Affairs*, 28(1): 246–61.
Hospitals and Health Networks (2007) Inside the Premier/CMS pay-for-performance project: an H&HN exclusive: insights from the 5 top hospitals, *Hospitals and Health Networks*, March.
Institute of Medicine (1999) *To err is human: building a safer health system*. Washington, DC: National Academies Press.
Institute of Medicine (2001) *Crossing the quality chasm: a new health system for the 21st century*. Washington, DC: National Academies Press.
Jha, A., Orav, J. and Epstein, A. (2010) The effect of financial incentives on hospitals that serve poor patients, *Annals of Internal Medicine*, 153(5): 299–306.
Jha, A. et al. (2012) The long-term effect of Premier pay for performance on patient outcomes, *New England Journal of Medicine*, 366(17): 1606–15.
Kennedy, S. et al. (2008) *Evaluation of the Premier hospital quality incentive demonstration: impacts on quality, Medicare reimbursements, and Medicare lengths of stay*. Cambridge: Abt Associates.
Lindenauer, P. R. D. (2007) Public reporting and pay for performance in hospital quality improvement, *New England Journal of Medicine*, 356: 386–496.
Manos, D. (2009) *Positive outcomes for federal P4P demos build case for change* (http://www.healthcarefinancenews.com/news/positive-outcomes-federal-p4p-demos-build-case-change, accessed June 2013).
Premier Inc. (2006) *Premier Inc. centers for Medicare and Medicaid services (CMS)/ Premier hospital quality incentive demonstration project: project overview and findings from year one* (http://www.allhealth.org/briefingmaterials/CenterformedicareandMedicaidservices(CMS)PremierHospitalQualityandincentiveDemonstrationProject-Premier-5.15.pdf, accessed June 2013).
Premier Inc. (2009) *Year 6 specifications – clinical conditions and measures for reporting and incentives* (http://www.premierinc.com/quality-safety/tools-services/

p4p/hqi/HQID-yr-6/Y6_HQID_Measure_Specifications_April_September_%202009_update%20050709.pdf, accessed March 2011).

QualityNet (2010) *About QualityNet* (http://www.qualitynet.org/dcs/ContentServer?c=Page&pagename=QnetPublic%2FPage%2FQnetHomepage&cid=1120143435363, accessed November 2010).

Remus, D. (2006) *Exploring the nexus of quality and cost*. Charlotte, NC: Premier Inc.

Ryan, A. (2009) Effects of the Premier hospital quality incentive demonstration on Medicare patient mortality and cost, *Health Services Research*, 44(3): 821–41.

Ryan, A., Blustein, J. and Casalino, L. (2012) Medicare's flagship test of pay-for-performance did not spur more rapid quality improvement among low-performing hospitals, *Health Affairs*, 31(4): 797–805.

US Department of Health and Human Services (2007) *Report to Congress: plan to implement a Medicare hospital value-based purchasing program*. Washington, DC: Centers for Medicare and Medicaid Services.

US Office of Management and Budget (2009) *Fiscal year 2010 budget: budget summary by category*. Washington, DC: OMB.

Wachter, R. (2006) Expected and unanticipated consequences of the quality and information technology revolutions, *Journal of the American Medical Association*, 295(23): 2780–83.

Werner, R. and Bradlow, E. (2006) Relationship between Medicare's hospital Compare performance measures and mortality rates, *Journal of the American Medical Association*, 296(22): 2694–2702.

Index

Locators shown in *italics* refer to figures, tables, boxes and appendices.

Aboriginal Medical Services (AMS), 120
absolute measures
  significance for P4P performance reward, 32
accountability, health care
  characteristics and importance, 41–3
  mechanisms to improve in Turkey's FM PBC scheme, 199
  role of P4P in enhancing, 43–51, *44–50*, 60
Advisory Council for Concerted Action in Health Care (Germany), 157
Aetma (USA), 225
After Hours Incentive (Australia), *114*, 120
Aged Care Access Incentive (Australia), *114*
Agency for Healthcare Research and Quality (AHRQ), 6, *7*
Agency for Quality and Medicine (Germany), 160–61
Altenhofen, L., 166, 167
AMS (Aboriginal Medical Services), 120
ANAO (Australia National Audit Office), 85, 86–7, 117, 118–20
AOK, 167
Association of Physicians (Germany), 160
Association of SHI Physicians (Germany), 160, 161–2
Asthma Incentive (Australia), *113*
Australia
  absolute measures of P4P performance in, 32
  basis and nature of reward for P4P performance, 32, 36, 37
  clinical domains and indicators of P4P in, *24*, 29, 31
  evaluation of practice incentives programs, *69*, 74
  features, processes and results of PIP, 110–23, *112–15*, *118*
  impact of P4P on health strategies and objectives, 51–4
  need for change to enhance health governance and performance, 58–60
  overview of health services, 109–10
  P4P objectives in primary care, 12, *12*
  quality, equity and effectiveness of P4P, 84–9, *90*, 96–101
  role of data flows in monitoring health performance, 54–8, *57*
  role of P4P in enhancing governance and performance, 43–51, *44*, 60
  significance of clinical quality, 24
  summary of P4P in, *5*, *25*
Australian Medical Association, 111
Australian National Audit Office (ANAO), 85, 86–7, 117, 118–20

autonomy, service delivery
  mechanisms to improve in Turkey's FM PBC scheme, 199

Barata, L., 246, 247
beneficiaries, of P4P
  significance for P4P evaluation, 75–6
Berthold, H.K., 167
bias
  significance for P4P evaluation, 73–4
Blue Cross (California), 225
Blue Shield (California), 225
Brazil
  adoption of notion of P4P, 4
  basis and nature of P4P performance reward or penalty, 35, 36, 37
  evaluation of social organizations in health, *70*
  features, processes and results of OSS programme, 240–49, *242*, *243–4*, *246–7*
  impact of P4P on health strategies and objectives, 51–4
  overview of health services, 239–40
  quality, equity and effectiveness of P4P, 84–9, *90*, 96–101
  role of data flows in monitoring health performance, 54–8, *57*
  role of P4P in enhancing health governance and performance, 43–51, *44*, 58–60
  summary of P4P in, *5*, *25*

California
  basis and nature of P4P performance reward or penalty, 31, 32, 37
  clinical domains and indicators of P4P in, 30
  features, processes and results of IHA Physician Incentive programme, 224–35, *227–9*, *230*, *232*, *234*
  overview of health services, 223–4
  summary of P4P in, *5*, *27*
California HealthCare Foundation, 225
capacity, health care
  as element of health strategies and objectives, 53–4
  impact on performance variation, 87, 96
  significance for IHA physician incentive programme, 233
  significance for PIP, *25*, 111–12, *113–14*
  significance for VIP, 263
Capacity Stream e-health Incentive (Australia), *113*
CAPI (Contracts for Improved Individual Practice) (France), 142, 143, 149–50, 152–3
care, health *see* health care
case studies
  Australia's PIP, 110–23, *112–15*, *118*
  Brazil's OSS programme, 240–49, *242*, *243–4*, *246–7*
  California's IHA, 224–35, *227–9*, *230*, *232*, *234*
  Estonia's QBS, 129–39, *130*, *131*, *134*, *135*, *137–9*
  evaluation of P4P programmes, 66–72, *67–71*
  France's ROSP, 142–54, *145–7*, *150*, *151*
  Germany's DMP, 159–70, *160*, *161*, *162–3*, *164*, *165*, *168*
  impact of P4P on health strategies and objectives, 51–4
  Korea's VIP, 253–63, *254*, *256*, *257*, *258*, *260*, *261*
  Maryland's HAC programme, 266–84, *270–71*, *274–6*, *278–80*
  New Zealand's PHO performance programme, 173–86, *175*, *178–9*, *180*
  parameter matrices of research on P4P, 14, *16–19*
  quality, equity and effectiveness of P4P, 84–101, *90–95*
  role of data flows in monitoring health performance, 54–8, *57*
  role of P4P in enhancing health governance and performance, 43–51, *44–50*, 58–60
  summary of P4P programme, 5–6, *5*
  Turkey's FM PBC scheme, 190–203, 192, 194–5
  United Kingdom's QOF, 206–18, *207*, *209–10*, *211*, *215*
  United States of America's HQID programme, 287–98, *291*, *295*
Center for Global Development, 6, *7*
Center for Medicare and Medicaid Services (CMS), 6, 7, 265–6, 267, 268, 269, 271, 281, 283
Cervical Screening Incentive (Australia), *112–13*
change
  need for to enhance health performance, 58–60
CHCs (Community Health Centres) (Turkey), 191, 196–7, 199
CHMS (Core Health Resource Management System) (Turkey), 195–6
China
  adoption of notion of P4P, 4
CIGNA HealthCare (California), 225
claims data
  role in monitoring health performance, 55
  significance of reporting and verification of, 38
clinical performance domains, 51–3
Clinical Practice Improvement Payment System (Australia), 109
CMS (Center for Medicare and Medicaid Services), 6, 7, 265–6, 267, 268, 269, 271, 281, 283

communication
improved P4P programme purchaser and provider, 97
Community Health Centres (CHCs, Turkey), 191, 196–7, 199
complexity, health structures
significance for weakening P4P, 100
consequences, unintended
significance for P4P evaluation, 74–5
see also impacts, short- and long-term
context, health policy
significance for Australia's PIP, 110–11
significance for Brazil's OSS programme, 240–41
significance for California's IHA, 224–5
significance for Estonia's QBS, 129
significance for France's ROSP, 142–4
significance for Germany's DMP, 159–61, *160*, *161*
significance for Korea's VIP, 253–4, *254*
significance for Maryland's HAC programme, 266–7
significance for New Zealand's PHO performance programme, 174–7, *175*
significance for Turkey's FM PBC scheme, 190
significance for UK's QOF, 206–8, *207*
significance for USA's HQID programme, 288–90
significance for P4P programmes, 98
Continuous Data Quality Improvement Programme (Australia), 58, 116
Contracts for Improved Individual Practice (CAPI, France), 142, 143, 149–50, 152–3
Core Health Resource Management System (CHMS, Turkey), 195–6
costs and savings
case study findings on P4P, 88–9
extent in Australia's PIP, 121
extent in Brazil's OSS programme, 248
extent in California's IHA, 233–4
extent in France's ROSP, 151–2
extent in Germany's DMP, 168
extent in Korea's VIP, 262
extent in New Zealand's PHO performance programme, 184–5
extent in Turkey's FM PBC scheme, 201
extent in UK's QOF, 216
extent in USA's HQID programme, 296–7
justification for in Australia's PIP, 122–3
justification for in Brazil's OSS programme, 248–9
justification for in California's IHA, 234–5
justification for in Estonia's QBS, 136
justification for in France's ROSP, 152–4
justification for in Germany's DMP, 169–70
justification for in Korea's VIP, 262–3
justification for in Maryland's HAC programme, 281–4
justification for in New Zealand's PHO performance programme, 185–6
justification for in Turkey's FM PBC scheme, 201–3
justification for in UK's QOF, 217–18
justification for in USA's HQID programme, 297–8
net costs of P4P programmes, 15
significance for P4P evaluation, 77
Couttolenc, B., 239–40, 242, 245–6
Crossing the Quality Chasm (Institute of Medicine, USA), 4, 223–4
Croxson, B., 176
Cumming, J., 176
cycles, health performance, 42–3, *43*

data, health care
improved generation and use of P4P, 96
role in enhancing health governance and performance, 43–51, *44–50*, 58–60
role in monitoring health performance, 54–8, *57*
sources and flows of FM PBC scheme, 195–6
sources and flows of HQID programme, 292–3
sources and flows of IHA, 229–30
sources and flows of MHAC, 273
sources and flows of OSS programme, 244–5
sources and flows of PHO performance programme, 181
sources and flows of PIP, 116
sources and flows of QBS, 133
sources and flows of QOF, 212
sources and flows of ROSP, 148
sources and flows of VIP, 257–9
see also information; reporting, data; verification, data
see also type e.g. claims data
delivery, health
features, processes and results in Australia, 110–23, *112–15*, *118*
features, processes and results in Brazil, 240–49, *242*, *243–4*, *246–7*
features, processes and results in California, 224–35, *227–9*, *230*, *232*, *234*
features, processes and results in Estonia, 129–39, *130*, *131*, *134*, *135*, *137–9*
features, processes and results in France, 142–54, *145–7*, *150*, *151*
features, processes and results in Germany, 159–70, *160*, *161*, *162–3*, *164*, *165*, *168*
features, processes and results in Korea, 253–63, *254*, *256*, *257*, *258*, *260*, *261*
features, processes and results in Maryland, 266–84, *270–71*, *274–6*, *278–80*
features, processes and results in New Zealand, 174–86, *175*, *178–9*, *180*
features, processes and results in Turkey, 190–203, *192*, *194–5*

features, processes and results in UK, 206–18, *207*, *209–10*, *211*, *215*
features, processes and results in USA, 288–98, *291*, *295*
see also incentives, service delivery
see also players e.g. patients; providers, health care
design and implementation
and context of P4P programmes, 98
features that weaken P4P, 100–101
P4P case study, *25–7*
Diabetes Incentive (Australia), *112*, 120
Disease Management Programme (DMP, Germany), 32, 36, *46*, 52, 60, 74, 85, 86, 88, *92–3*, 99, 159–70, *160*, *161*, *162–3*, *164*, *165*, *168*
disparities, health
significance as element of P4P, 29
domains, performance
case study of Australia's PIP, 111–15, *112–15*
case study of Brazil's OSS programme, 243, *243*
case study of California's IHA, 226
case study of Estonia's QBS, 130–32, *130*, *131*, *137–9*
case study of France's ROSP, 144, *145–7*
case study of Germany's DMP, 163
case study of Korea's VIP, 255, *256*, *257*
case study of Maryland's HAC programme, 269–71, *270–71*
case study of New Zealand's PHO performance programme, 177–80, *178–9*, *180*
case study of Turkey's FM PBC scheme, 193
case study of UK's QOF, 208–11, *209–10*, *211*
case study of USA's HQID programme, 290–92, *291*
impact of P4P on health strategies and objectives, 51–4
significance as element of P4P, 23–31, *24*, *30*
significance for P4P programme incentives, 100–101
Domestic Violence Incentive (Australia), 111, *115*, 117, 120
Donadegian, A., 72
Doran, T., 75

effectiveness
factors contributing to P4P programme, 98–100
effects, spillover and substitution
P4P programmes, 14–15
positive for provider activity and health systems, 96–7
significance for P4P evaluation, 74–5
efficiency, health
case study findings on P4P, 88–9
significance as element of P4P, 28
EHIF (Estonia Health Insurance Fund), 52, 127–8, 129, 133, 136
ENMR (Expérimentation de Nouveaux Modes de Rémunération), 153
EPP (Evaluation of Professional Practice) (France), 143
equity
case study findings on P4P, 86–7
extent in Australia's PIP, 120
extent in Brazil's OSS programme, 248
extent in California's IHA, 233
extent in New Zealand's PHO performance programme, 184
extent in UK's QOF, 216
extent in USA's HQID programme, 296
significance as element of P4P, 29
significance for P4P evaluation, 75–6
Estonia
basis and nature of P4P performance reward or penalty, 36, 37
domains and indicators of P4P in, *24*, 28, 30, 31
evaluation of disease management programmes, *68*
features, processes and results of QBS, 129–39, *130*, *131*, *134*, *135*, *137–9*
impact of P4P on health strategies and objectives, 51–4
outcome measures of P4P performance, *11*
overview of health services, 127–9, *128*
quality, equity and effectiveness of P4P, 84–9, 91, 96–101
role of data flows in monitoring health performance, 54–8, *57*
role of P4P in enhancing health governance and performance, 43–51, *45*, 58–60
significance of clinical quality, 24
summary of P4P in, *5*, *11*, *25*
Estonia Health Insurance Fund (EHIF), 52, 127–8, 129, 133, 136
European Observatory on Health Systems and Policies, xvii
EUROPEP scale (Turkey), 200
evaluation, payment for performance
extent in Australia's PIP, 118–19
extent in Brazil's OSS programme, 245–7, *246–7*
extent in California's IHA, 231–3, *232*
extent in Estonia's QBS, 134–6, *135*
extent in France's ROSP, 149–52, *150*, *151*
extent in Germany's DMP, 166–8
extent in Korea's VIP, 259–61, *260–61*
extent in New Zealand's PHO performance programme, 183–4
extent in UK's QOF, 215–16

extent in USA's HQID programme, 294–6
importance of process, 63
issues to consider in process of, 72–7
overview of impact of social interventions, 64–6
practice in OECD countries, 66–72, *67–71*
Evaluation of Professional Practice (EPP, France), 143
ex-ante evaluation, 72
ex-post evaluation, 72
expenditure, health care
German health care, 157–8, *158*
see also costs and savings
see also specific programmes e.g. Quality Bonus System; Payment for Public Health Objectives programme; Practice Incentives Programme
expenditure, sickness
role in developing DMP guidelines, 160–61
significance for Germany's DMP, 161–4, *162–3*, *164*
see also factors affected e.g. providers, health care
experiences, patient
case study findings on P4P, 87–8
Expérimentation de Nouveaux Modes de Rémunération (ENMR), 153
experiments
role in P4P evaluation, 65

Family Medicine Information System (FMIS, Turkey), 56, 195–6, 200, 202
Family Medicine Performance Based Contracting (FM PBC) scheme (Turkey), 28, 31, 35, 36, 37, *47*, 56, 60, *94*, 190–203, *192*, *194–5*
Federal Insurance Agency (Germany), 161, 165, 166
Federation of Medical Unions (France), 152
feedback
role in enhancing health sector governance, 43–51, *44–50*
Finance Law for Social Security (2008, France), 143, 153
Flodgren, G., 66
flows, data
and sources of FM PBC scheme, 195–6
and sources of HQID programme, 292–3
and sources of IHA Physician Incentive programme, 229–30
and sources of MHAC, 273
and sources of OSS programme, 244–5
and sources of PHO performance programme, 181
and sources of PIP, 116
and sources of QBS, 133
and sources of QOF, 212
and sources of ROSP, 148
and sources of VIP, 257–9
role in monitoring health performance, 54–8, *57*
FM PBC (Family Medicine Performance Based Contracting) scheme (Turkey), 28, 31, 35, 36, 37, *47*, 56, 60, *94*, 190–203, *192*, *194–5*
FMIS (Family Medicine Information System) (Turkey), 56, 195–6, 200, 202
France
basis and nature of P4P performance reward or penalty, 31, 32, 34, 35, 36, 37
clinical domains and indicators of P4P in, *24*, 28, 29, 31
features, processes and results of ROSP, 142–54, *145–7*, *150*, *151*
impact of P4P on health strategies and objectives, 51–4
outcome measures of P4P performance, *11*, 24, 32
overview of health services, 141–2
P4P objectives in, 12, *12*, *68*
quality, equity and effectiveness of P4P, 84–9, *91*, 96–101
role of data flows in monitoring health performance, 54–8, *57*
role of P4P in strengthening health governance and performance, 43–51, *45*, 58–60
significance of clinical quality, 24
summary of P4P in, *5*, *12*, *25*
Fullerton, B., 167
funds and funding, health care
German health care, 157–8, *158*
see also costs and savings
see also specific programmes e.g. Quality Bonus System; Payment for Public Health Objectives programme; Practice Incentives Programme
funds and funding, sickness
role in developing DMP guidelines, 160–61
significance for Germany's DMP, 161–4, *162–3*, *164*
see also factors affected e.g. providers, health care

General Practice Financing Group (GPFG, Australia), 111
General Practice Immunization Incentive (GPII) programme (Australia), 109, 116
Germany
adoption of notion of P4P, 4
basis and nature of P4P performance reward or penalty, 32, 36
clinical domains and indicators of P4P in, 28
definition of P4P in, 8
evaluation of disease management programmes, *67*
evaluation of quality bonus systems, *68*
features, processes and results of DMP, 159–70, *160*, *161*, *162–3*, *164*, *165*, *168*
impact of P4P on health strategies and objectives, 51–4

outcome measures of P4P performance, *11*
overview of health services, 157–8, *158*
quality, equity and effectiveness of P4P, 84–9, *92–3*, 96–101
role of data flows in monitoring health performance, 54–8, *57*
role of P4P in strengthening health governance and performance, 43–51, *46*, 58–60
significance of clinical quality, 24
summary of P4P in, *5*, *26*
governance, health care
case study of change to enhance, 58–60
characteristics and importance, 41–3
identification of need for change to enhance, 58–60
mechanisms to improve in Turkey's FM PBC scheme, 199
role of P4P in enhancing accountability of, 60
role of P4P in strengthening, 43–51, *44–50*
GPFG (General Practice Financing Group) (Australia), 111
GPII (General Practice Immunization Incentive) programme (Australia), 109, 111
groups, comparison
significance for P4P evaluation, 73–4
guidelines, health care
need for P4P adherence to, 98

HAS (National Authority for Health) (France), 142–3
health care
characteristics and importance of governance of, 41–3
identification of need for change to enhance, 58–60
impact of P4P on strategies and objectives of, 51–4
role of data and information flows in monitoring of, 54–8, *57*
role of P4P in enhancing accountability of, 60
role of P4P in strengthening governance of, 43–51, *44–50*
significance of priority, 28
see also data, health care; delivery, health; improvement, health care; objectives, health care; outcomes, health; performance, health care; utilization, health care
see also aspects e.g. funding, health care; Practice Incentives Programme; quality, health care
see also factors affecting e.g. capacity, health care; guidelines, health care; incentives, service delivery; structures, service delivery
see also players e.g. patients; practitioners, health care; providers, health care; purchases, health care
Health Funding Authority (New Zealand) 174–5
Health Insurance Reform Act (2004, France), 142, 143
Health Insurance Review Agency (HIRA, Korea), 253, 255, 257–8, 259, 260–61, 262
Health Net (USA), 225
Health Service Cost Review Commission (HSCRC, USA), 53, 265, 266, 267–9, 270, 271, 272, 273–4, 277, 281–2, 283, 284
Health Services Organization Act (Estonia), 51, 128
Health Strategy (New Zealand, 2000), 173–6, *175*, 182, 185
Health System Priorities When Money is Tight (OECD), xv
Health Transformation Programme (HTP, Turkey), 190, 199, 201
Health Transformation Programme (Turkey), 98
HIRA (Health Insurance Review Agency) (Korea), 253, 255, 257–8, 259, 260–61, 262
Hospital, Patients, Health and Territories Law (2009, France), 143
Hospital Acquired Conditions programme (Maryland)
features, processes, extent and outcomes, 266–84, *270–71*, *274–6*, *278–80*
Hospital Association (Korea), 254
Hospital Quality Incentive Demonstration (HQID programme (USA), 28, 31, 33, 34, 35, 36, *48*, *50*, 55, 56, 66, 74, 86, 87, 89, *95*
Hospital Quality Incentive Demonstration programme (USA)
features, processes, extent and outcomes, 287–98, *291*, *295*
HSCRC (Health Service Cost Review Commission) (USA), 53, 265, 266, 267–9, 270, 271, 272, 273–4, 277, 281–2, 283, 284
HTP (Health Transformation Programme) (Turkey), 190, 199, 201

IHA Physician Incentive Programme (California), 4, 30, 31, 32, *49*, 52, 54, 55, 60, 85, 87, 88, 89, *95*, 96, 97, 100, 101, 224–35, *227–9*, *230*, *232*, *234*
impacts, short- and long-term
significance for P4P evaluation, 76–7
see also consequences, unintended
implementation and design
and context of P4P programmes, 98
features that weaken P4P, 100–101
P4P case study, *25–7*
improvement, health care
as outcome of DMP process, 166
as outcome of FM PBC process, 197–9
as outcome of HQID process, 294

as outcome of IHA process, 231
as outcome of MHAC process, 273, 277
as outcome of OSS programme, 248–9
as outcome of PHO performance programme process, 182–3
as outcome of PIP process, 117–18, *118*
as outcome of QBS process, 133–4, *134*
as outcome of QOF process, 213–14
as outcome of ROSP process, 149
as outcome of VIP, 262–3
design and implementation features weakening, 100–101
importance for P4P, 99–100
significance as basis for P4P performance reward, 32–3
see also performance, health care; ranking, relative
see also factors influencing e.g. costs and savings; payments, incentive
incentives, performance
nature and size of P4P, 34–8, *35*
significance and characteristics of P4P, *24*, 25–*7*, 31–4
significance for health strategies and objectives, 51–4
Independent Practitioner Associations (IPAs, New Zealand), 174, 176
India
adoption of notion of P4P, 4
indicators, performance
case study of Australia's PIP, 111–15, *112–15*
case study of Brazil's OSS programme, 243, *243*
case study of California's IHA, 226
case study of Estonia's QBS, 130–32, *130*, *131*, *137–9*
case study of France's ROSP, 144, *145–7*
case study of Germany's DMP, 161–4, *162–3*, *164*
case study of Korea's VIP, 255
case study of Maryland's HAC, 269–71, *270–71*
case study of New Zealand's PHO performance programme, 177–81, *178–9*, *180*
case study of Turkey's FM PBC, 191–7, *192*, *194–5*
case study of UK's QOF, 208–11, *209–10*, *211*
case study of USA's HQID programme, 290–92, *291*
role in health care improvement, 294, *295*
significance for P4P evaluation, 72–3
Indigenous Health Incentive (Australia), *113*
information
role of flows of in monitoring health performance, 54–8, *57*
see also data, health care
Institute for Health and Social Research (Germany), 159, *160*
Institute for Quality and Efficiency in Health (Germany), 159, *160*
Institute of Medicine (USA), 4, 223–4
Integrated Healthcare Association Physician Incentive programme (IHA, California), 4, 30, 31, 32, *49*, 52, 54, 55, 60, 85, 87, 88, 89, *95*, 96, 97, 100, 101, 224–35, *227–9*, *230*, *232*, *234*
interventions, social
impact on P4P evaluation, 64–6
IPAs (Independent Practitioner Associations) (New Zealand), 174, 176

Korea, Republic of
basis and nature of P4P performance reward or penalty, 33, 34, 36, 37
clinical domains and indicators of P4P in, 28
evaluation of incentives programmes, *71*
features, processes and results of VIP, 253–63, *254*, *256*, *257*, *258*, *260*, *261*
impact of P4P on health strategies and objectives, 51–4
outcome measures of P4P performance, *11*
overview of health services, 251–2, *252*
P4P objectives in primary care, 12, *12*
role of data flows in monitoring health performance, 54–8, *57*
role of P4P in enhancing health governance and performance, 43–51, *46*, 58–60
summary of P4P in, *5*, *26*

La Forgia, G., 239–40, 242, 245–6
Law on Piloting of Family Medicine (Turkey) 197
Linder, R., 166, 168

Martin, Jenkins and Associates Ltd, 183–4
Maryland
basis and nature of P4P performance reward or penalty, 33, 34, 36, 37
clinical domains and indicators of P4P in, 28, 29, 30
overview of health services, 265–6
summary of P4P in, *5*, *27*
Maryland Hospital Acquired Conditions Program (MHAC), 28, 29, 30, 33, 35, 36, 37, *50*, 53, 55, 59, 88, *95*, 97, 98, 99, 266–84, *270–71*, *274–6*, *278–80*
Maryland Hospital Association, 37–8
Mattke, S., 169
Maynard, K., 77
MBS (Medicare Benefits Schedule) (Australia), 110, 110, 119
measures, performance
calculation of achievement of, 33–4
significance as element of P4P, 23–31, *24*, *30*

see also type e.g. absolute measures; improvement, health care; ranking, relative
Medical Association (Korea), 254
Medicare and Medicade (USA), 6, 28, 35, *44*, *50*, 55, 56, *95*, 223, 252, 254, 265–6, 268, 269–70, 271, 281, 283
Medicare Australia, 44, 58, 110, 111, 114–15, 116, 121
Medicare Benefits Schedule (MBS, Australia), 109, 110, 119
Mehring, M., 167
Mental Health Nurse Incentive Programme (Australia) 116
MHAC (Maryland Hospital Acquired Conditions Program), 28, 29, 30, 33, 35, 36, 37, *50*, 53, 55, 59, 88, *95*, 97, 98, 99, 266–84, *270–71*, *274–6*, *278–80*
monitoring, health
role of data and information flows in process, 54–8, *57*
monitoring, P4P
extent in Australia's PIP, 118–19
extent in Brazil's OSS programme, 245–7, *246–7*
extent in Estonia's QBS, 134–6, *135*
extent in France's ROSP, 149–52, *150*, *151*
extent in Germany's DMP, 166–8
extent in IHA Physician Incentive programme, 231–3, *232*
extent in Korea's VIP, 259–61, *260–61*
extent in New Zealand's PHO performance programme, 183–4
extent in UK's QOF, 215–16
extent in USA's HQID programme, 294–6

National Accreditation Association (Brazil), 247
National Authority for Health (HAS, France), 142–3
National Committee for Quality Assurance (NCQA, USA), 225
National Health Insurance Corporation (NHIC, Korea), 253, 254, 255, 263
National Health Insurance Fund (NHIF, France), 88, 141–2, 143–4, 148, 149–54
National Health Performance Fund (UK), 206
National Institute for Clinical Excellence (NICE), 53, 205–6, 208, 210, 218
NCQA (National Committee for Quality Assurance) (USA), 225
New Zealand
basis and nature of P4P performance reward or penalty, 32, 36, 37
clinical domains and indicators of P4P in, 29, 31
evaluation of primary health organizations, *71*
features, processes and results of PHO performance programme, 174–86, *175*, *178–9*, *180*
impact of P4P on health strategies and objectives, 51–4
outcome measures of P4P performance, *11*
overview of health services, 173–4
P4P objectives in primary care, 12, *12*
quality, equity and effectiveness of P4P, 84–9, *93–4*, 96–101
role of data flows in monitoring health performance, 54–8, *57*
role of P4P in enhancing health governance and performance, 42–51, *47*, 58–60
summary of P4P in, *5*, *26*
NHIC (National Health Insurance Corporation) (Korea), 253, 254, 255, 263
NHIF (National Health Insurance Fund) (France), 88, 141–2, 143–4, 148, 149–54
NHS Pay Modernization (NAO, 2008), 215–16, 216–17
NHS Plan for Reform and Assessment (2000), 205–6, 207, 217
NICE (National Institute for Clinical Excellence), 53, 205–6, 208, 210, 218
non-clinical performance domains, 53–4
non-transparency, health structures, 100

objectives, health care
as basis for health performance, 51–4
need for alignment of in P4P programmes, 98
observation
role in P4P evaluation, 65–6
Opportunidades programme (Mexico), 63
Order of Doctors (France), 152
Organization for Economic Cooperation and Development
summary of P4P usage in, 10–12, *11*, *12*
see also countries e.g. Australia; Brazil; Estonia; France; Germany; Korea, Republic of; New Zealand; Turkey; United Kingdom; United States of America
OSS (Social Organizations in Health) programme (Brazil), 35, 43, *44*, 56, 59, 88, *90*, 96, 99, 100, 240–49, *242*, *243–4*, *246–7*
outcomes, health
case study findings on P4P, 86
DMP leverage to achieve, 166
FM PBC scheme leverage to achieve, 197–9
HQID programme leverage to achieve, 294
IHA Physician Incentive programme leverage to achieve, 231
impact of DMP, 166–9, *168*
impact of FM PBC scheme, 200–201
impact of IHA Physician Incentive programme, 231–4, *232*, *234*
impact of MHAC, 277–81, *278–80*

impact of OSS programme, 245–8, *246–7*
impact of PHO performance programme on, 183–5
impact of PIP, 118–21
impact of QBS, 134–6, *135*
impact of QOF, 214–17, *215*
impact of ROSP, 149–52, *150*, *151*
impact of VIP, 259–62, *260*, *261*
MHAC leverage to achieve, 273, 277
PHO performance programme leverage to achieve, 182–3
PIP leverage to achieve, 117–18, *118*
QBS leverage to achieve, 133–4, *134*
QOF leverage to achieve, 213–14
ROSP leverage to achieve, 149

P4P *see* 'Payment for Performance'
Pacific Business Group on Health (PBGH), 225
PacifiCare (USA), 225
participation, P4P programme
significance for P4P incentives, 100–101
patients
case study findings on P4P experiences of, 87–8
incentives within Germany's DMP system, 164
see also satisfaction, patient
payment, health care
impact of methods of on quality, 3–4
need for integration of in P4P programmes, 99
'Payment for Performance' (P4P)
basis for reward or penalty, 31–8, *35*
characteristics and elements, 23–38, *24*, *25–7*, *30*, *35*
definitions, 6–9, *7*
design, implementation and contextual factors, 98
effect on health system performance, 14–15
factors contributing to effectiveness of, 98–100
features that weaken incentives of, 100–101
findings on quality, equity and effectiveness, 84–95, *90–95*
impact on health strategies and objectives, 51–4
literature review of, 13–14
origin and purpose, 4
positivity of spillover effects of, 96–7
role in enhancing health governance and accountability, 43–51, *44–50*, 60
summary of case studies of, 5–6, *5*
summary of use in OECD, 10–12, *11*, *12*
theory underlying, 9–10
unintended consequences of, 97–8
see also evaluation, payment for performance
Payment for Public Health Objectives (ROSP) programme (France), 31, 32, 34, 35, 36, *45*, 52, 55, 60, 85, 88, *91*, 97, 99, 100, 142–54, *145–7*, *150*, *151*
payments, incentive
characteristics of P4P performance, 36–7
organization of DMP, 163–4, *164*
organization of FM PBC scheme, 193–5, *194–5*
organization of HQID programme, 292
organization of IHA, 226–9, *227–9*, *230*
organization of MHAC, 271–3, *274–7*
organization of OSS programme, 243–4, *243–4*
organization of PHO performance programme, 180–81
organization of PIP, 115–16
organization of QBS, 132
organization of QOF, 211–12
organization of ROSP, 144, 148
organization of VIP, 255–7, *258*
penalties, performance
nature of P4P programme, 34–8, *35*
significance and characteristics of P4P, *24*, *25–7*, 31–4
performance, health care
effect of P4P on, 14–15
identification of need for change to enhance, 58–60
impact and effect of DMP on, 166–9, 168
impact and effect of FM PBC scheme on, 197–203
impact and effect of HQID programme on, 294–7, *295*
impact and effect of IHA Physician Incentive programme on, 231–5, *232*, *234*
impact and effect of MHAC on, 277–81, *278–80*
impact and effect of OSS programme on, 245–9, *246–7*
impact and effect of PHO performance programme on, 182–6
impact and effect of PIP on, 118–21
impact and effect of QBS on, 134–6, *135*
impact and effect of QOF on, 213–17, *215*
impact and effect of ROSP on, 149–52, *150*, *151*
impact and effect of VIP on, 259–63, *260*, *261*
impact of P4P on health strategies and objectives, 51–4
need for concentration upon problems of, 99
role of data and information flows in monitoring of, 54–8, *57*
role of P4P in strengthening, 43–51, *44–50*, 60
see also domains, performance, evaluation, P4P; improvement, health care; indicators, performance; measures,

performance; penalties, performance; rewards, performance
see also elements contributing e.g. capacity, health care
see also factors impacting e.g. costs and savings
Permanente Medical Group (USA), 225
Pharmaceutical Benefits Programme (Australia), 109–10
PHDs (Provincial Health Directorates) (Turkey), 191, 196, 199, 203
PHO (Primary Health Organization) performance programme (New Zealand), 31, 32, 36, 37, *47*, 51, 52, 55–6, 58, 59, 84, 86, 87, 89, *93–4*, 174–86, *175*, *178–9*, *180*
Physician Group Practice Demonstration (Medicare), 28
Pimouguet, C., 169
Pirajussara Hospital (Sao Paulo)
social organization contracting terms, 243–4
Practice Incentives Program (PIP, Australia), 31, 32, 36, 37, *44*, 52, 55, 60, 85, 86, *90*, 99, 100–101, 109–23, *112–15*, *118*
Practice Nurse Incentive (Australia), 85, *90*, 111–12, *113–14*, 120
practitioners, health care
incentives within Germany's DMP system, 164
see also delivery, health
Primary Health Care Strategy (New Zealand), 173–4, 182, 186
Primary Health Organization (PHO) performance programme (New Zealand), 31, 32, 36, 37, *47*, 51, 52, 55–6, 58, 59, 84, 86, 87, 89, *93–4*, 174–86, *175*, *178–9*, *180*
Productivity Commission (Australia), 121
programmes, 'Payment for Performance' *see* 'Payment for Performance'
Progresa programme (Mexico), 63
providers, health care
extent and scope of Australia's PIP, 116–17
extent and scope of California's, 230–31
extent and scope of Estonia's QBS, 133
extent and scope of France's ROSP, 148
extent and scope of Germany's DMP system, 165–6, *165*
extent and scope of Korea's VIP, 259
extent and scope of Maryland's HAC, 268
extent and scope of New Zealand's PHO performance programme, 181–2
extent and scope of Turkey's FM PBC scheme, 197
extent and scope of UK's QOF, 212–13
extent and scope of USA's HQID programme, 293
improved communication of P4P, 97
response to Australia's PIP, 121
response to California's IHA, 233
response to France's ROSP, 152
response to Germany's DMP, 168–9
response to Korea's VIP, 262
response to New Zealand's PHO performance programme, 185
response to Turkey's FM PBC scheme, 201
response to UK's QOF, 216–17
response to USA's HQID programme, 297
see also responsibilities e.g. performance, health care
Provincial Health Directorates (PHDs, Turkey), 191, 196, 199, 203
purchasers, health care
improved communication of P4P, 97

QBR (Quality Based Reimbursement) programme (USA), 266, 267, 268–9, 281–2, 283
QBS *see* Quality Bonus System
QMAS (Quality Management and Analysis System), 56, *57*, 60
QOF *see* Quality and Outcomes Framework
quality, clinical
significance as element of P4P, 24
quality, health care
case study findings on P4P, 84–6
impact of payment methods, 3–4
significance as element of PIP, 111–12, *112–13*, 115–16
Quality and Outcomes Framework (QOF, UK), 30–31, 32, 34, 35, 36, 37, 39, 43, *48*, 52, 53, 54, 55, 56, *57*, 58, 60, 73, 84–5, 86, 87–9, *94*, 96, 97–8, 929–100, 101, 206–18, *207*, *209–10*, *211*, *215*
Quality Based Reimbursement (QBR) programme (USA), 266, 267, 268–9, 281–2, 283
Quality Bonus System (QBS) programme (Estonia), 30, 31, 36, *45*, 52, 55, 84, 85, 88, *91*, 96, 98, 129–39, *130*, *131*, *134*, *135*, *137–9*
Quality Management and Analysis System (QMAS), 56, *57*, 60
Quality Measures Web Tool (Premier), 56
quasi experiments
role in P4P evaluation, 65, 73–4

RACGPs (Royal Australian College of General Practitioners), 110, 111, 116
RAND Corporation, 6, *7*
randomized control trials (RCTs)
role in P4P evaluation, 64–5, 73
ranking, relative
significance as basis for P4P performance reward, 33
records, electronic
role in monitoring health performance, 55–6
Reformed National Health Insurance Act (2000, Korea), 253

reporting, data
  significance as element of P4P, 38
reporting, exception, 34–5, *35*
reporting, public
  role in enhancing health sector governance, 43–51, *44–50*
research
  on P4P, 13–14
  case study of design of, *25–7*
rewards, performance
  nature and size of P4P, 34–8, *35*
  significance and characteristics of P4P, *24*, 25–*7*, 31–4
  significance for health strategies and objectives, 51–4
ROSP (Payment for Public Health Objectives) programme (France), 31, 32, 34, 35, 36, *45*, 52, 55, 60, 85, 88, *91*, 97, 99, 100, 142–54, *145–7*, *150*, *151*
Royal Australian College of General Practitioners (RACGP's), 110, 111, 116
rural areas
  significance as element of PIP incentives, 112, *114–15*
Rural Doctors Association of Australia, 111
Rwanda
  adoption of notion of P4P, 4

São Paulo
  features, processes and results of OSS programme, 240–49, *242*, *243–4*, *246–7*
satisfaction, patient
  case study of effect of FM PBC scheme, 200–201
  significance as element of P4P, 29
savings *see* costs and savings
Schäfer, I., 167
Scheffler, R.M., *24*
Scott, A., 66, 74, 75
self-reported data
  role in monitoring health performance, 56–8
services, health care *see* health care
SHI (Statutory Health Insurance) System (Germany), 157–8, 159
sickness, funding of
  role in developing DMP guidelines, 160–61
  significance for Germany's DMP, 161–4, *162–3*, *164*
  see also factors affected e.g. providers, health care
Smith, J., 176
Society of Family Physicians (Estonia), 52, 129, 136
Social Organizations in Health (OSS) programme (Brazil), 35, 43, *44*, 56, 59, 88, *90*, 96, 99, 100, 240–49, *242*, *243–4*, *246–7*
specificity
  need for to enhance P4P, 99
stakeholders
  involvement in Australia's PIP, 111
  involvement in Brazil's OSS programme, 241
  involvement in California's IHA programme, 225
  involvement in Estonia's QBS, 129
  involvement in France's ROSP, 143–4
  involvement in Germany's DMP, 160–61
  involvement in Korea's VIP, 254
  involvement in Maryland's HAC, 267
  involvement in New Zealand's PHO performance programme, 176–7
  involvement in UK's QOF, 208
  involvement in USA's HQID programme, 290
  role in enhancing health sector governance, 43–51, *44–50*
  see also name e.g. patients; practitioners, health care
Standards for General Practices (RACGP's), 110, 116
Statutory Health Insurance (SHI) System (Germany), 157–8, 159
Stock, S., 164, 168
strategies, health care
  as basis for health performance, 51–4
  need for alignment to in P4P programmes, 98
structures, health governance
  characteristics and importance, 41–3
  role of P4P in strengthening, 43–51, *44–50*
structures, service delivery
  importance for P4P, 99–100
support, rural
  significance as element of PIP incentives, 112, *114–15*
Survey on Health System Characteristics (OECD), 10–11
Sutton, M., 75
Szecsenyl, J., 167

Teaching Incentive (Australia), *114*
To Err is Human (Institute of Medicine), 4
trials, randomised control
  role in P4P evaluation, 64–5, 74
Turkey
  basis and nature of P4P performance reward or penalty, 31, 35, 36, 37
  clinical domains and indicators of P4P in, 28
  features, processes and results of FM PBC scheme, 190–203, *192*, *194–5*
  impact of P4P on health strategies and objectives, 51–4
  outcome measures of P4P performance, *11*
  overview of health services, 189–90
  quality, equity and effectiveness of P4P programme, 84–9, *94*, 96–101

role of data flows in monitoring health performance, 54–8, *57*
role of P4P in enhancing health governance and performance, 43–51, *47*, 58–60
summary of P4P in, *5*, *26*

United Kingdom
adoption of notion of P4P, 4
basis and nature of P4P performance reward or penalty, 32, 34, 35, *35*, 36, 37
clinical domains and indicators of P4P in, 28, 29–30, *30*
definition of P4P in, 8
evaluation of quality and outcomes framework, *68–9*
features, processes and results of QOF, 206–18, *207*, *209–10*, *211*, *215*
impact of P4P on health strategies and objectives, 51–4
outcome measures of P4P performance, *11*
overview of health services, 205–6
P4P data reporting and verification, 38
P4P objectives in primary care, 12, *12*
quality, equity and effectiveness of P4P, 84–9, *94*, 96–101
role of data flows in monitoring health performance, 54–8, *57*
role of P4P in enhancing health governance and performance, 43–51, *48*, 58–60
summary of P4P in, *5*, *27*
United States of America
adoption of notion of P4P, 4
basis and nature of P4P performance reward or penalty, 31, 33, 34, 35, 36, 37
clinical domains and indicators of P4P in, 28
evaluation of California integrated health care association incentives, *70*
evaluation of hospital quality incentives, *67*
evaluation of Maryland acquired hospital infections, *69*
features, processes and results of HQID programme, 288–98, *291*, *295*
impact of P4P on health strategies and objectives, 51–4
outcome measures of P4P performance, *11*
overview of health services, 287–8
P4P objectives in primary care, 12, *12*
quality, equity and effectiveness of P4P programme, 84–9, *95*, 96–101
role of data flows in monitoring health performance, 54–8, *57*
role of P4P in enhancing health governance and performance, 43–51, 49–50, 58–60
summary of P4P in, *5*, *27*
US Agency for International Development (USAID), 6, *7*

Value Incentive Programme (VIP, Korea), 28, 33, 34, 36, *46*, 55, 86, 99, 253–63, *254*, *256*, *257*, *258*, *260*, *261*
Van Herck, P., 66
Västra, K., 136
verification, data
FM PBC scheme performance, 196–7
role in monitoring health performance, 55–6
significance as element of P4P, 38

Western Health Advantage (USA), 225
World Bank, 6, *7*, 245, 246
World Health Organization (WHO), 141, 157, 253–4
World Health Report (2000) 157

Open University Press

McGraw - Hill Education

Paying for **Performance** in **Health Care**

Implications for health system performance and accountability

Health spending continues to grow faster than the economy in most OECD countries. Pay for performance (P4P) has been proposed in many OECD countries as an innovative solution to the value-for-money challenge in health care.

However, to date, evidence that P4P in fact increased value for money, boosted quality in health care, or improved health outcomes, has been limited.

This book explores the many questions surrounding P4P such as whether the potential power of P4P has been over-sold, or whether the largely disappointing results to date are more likely rooted in problems of design and implementation or inadequate monitoring and evaluation. The book also examines the supporting systems and process, in addition to incentives, that are necessary for P4P to improve provider performance and to drive and sustain performance improvement.

The book utilises a substantial set of case studies from 12 OECD countries to shed light on P4P programs in practice. Featuring both high and middle income countries, cases from primary and acute care settings, and a range of national, regional and pilot programmes, each case study features:

- **Analysis of the design and implementation of decisions, including the role of stakeholders**
- **Critical assessment of objectives versus results**
- **Examination of the of 'net' impacts, including positive spillover effects and unintended consequences**

The detailed analysis of these 10 case studies together with the rest of the analytical text highlight the realities of P4P programs and their potential impact on the performance of health systems in a diversity of settings. This book provides critical insights into the experience to date with P4P and how this tool may be better leveraged to improve health system performance and accountability.

Dr. Cheryl Cashin Results for Development Institute

Ms. Y-Ling Chi University of Oxford

Dr. Peter Smith Imperial College London and the Centre for Health Policy in the Institute of Global Health Innovation

Dr. Michael Borowitz Global Fund to Fight AIDS, Tuberculosis and Malaria

Dr. Sarah Thomson WHO Regional Office for Europe, London School of Economics and Political Science and European Observatory on Health Systems and Policies.

European Observatory on Health Systems and Policies
a partnership hosted by WHO

OECD
BETTER POLICIES FOR BETTER LIVES

L'Assurance Maladie
Caisse Nationale

ISBN-13: 978-0-3352-6438-4
ISBN-10: 0-3352-6438-7

9 780335 264384

www.openup.co.uk
http://www.healthobservatory.eu