

Monitoring and evaluating social protection systems



Lessons from the EU-SPS Programme

Monitoring and evaluating social protection systems

The European Union Social Protection Systems (EU-SPS) Programme

The European Union Social Protection Systems Programme (EU-SPS) is a European Union action co-financed by the Organisation for Economic Co-operation and Development (OECD) and the Government of Finland. The OECD Development Centre and the Government of Finland's National Institute for Health and Welfare (THL) manage its implementation. The EU-SPS supports low- and middle-income countries in building sustainable and inclusive social protection systems. The programme was implemented over the course of four and half years until April 2019 in partnership with national and regional social protection authorities, think-tanks and expert institutions in 10 countries.

The programme has three specific objectives: 1) To develop appropriate methodologies and tools with which to assess the social protection policies, programmes and capacities in ten selected partner countries; 2) To enhance administrative and technical capacity in the partner countries to support the development of affordable, sustainable and inclusive social protection systems; and 3) To generate evidence-based knowledge for future EU co-operation and for use by other development partners on the effectiveness of social protection in reducing poverty and vulnerability, addressing inequality and promoting social cohesion.

Disclaimer

The opinions expressed and arguments employed here are the sole responsibility of the authors and do not necessarily reflect those of the OECD, its Development Centre, or of their member countries.

This document, as well as any data and maps included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

Photo credit: Cover by the OECD Development Centre.

© OECD 2019

Abstract

This paper provides guidance on developing robust monitoring and evaluation (M&E) systems as tools to assess the effectiveness and potential areas of improvement of social protection systems. It highlights the importance of M&E, particularly under the current development agenda, and evaluates the pros and cons of existing methodological evidence-building approaches. This guidance note highlights the value added of a mixed-methods approach and the interaction between micro and macro level frameworks in a systems approach to social protection monitoring and evaluation. Lastly, it addresses the role of institutional arrangements and sets out principles to foster capacity building.

Acknowledgements

The “Monitoring and evaluating social protection systems” paper was prepared by the Social Cohesion Unit of the OECD Development Centre as part of the European Union Social Protection Systems Programme.

The team was led by Alexandre Kolev, Head of the Social Cohesion Unit, and Ji-Yeun Rim, Co-ordinator of the European Union Social Protection Systems Programme, under the guidance of Mario Pezzini, Director of the OECD Development Centre and Special Advisor to the OECD Secretary-General on Development. The report was drafted by Caroline Tassot, based on background papers prepared by the Economic Policy Research Institute. Laura Copete provided assistance at every stage.

The report benefited from valuable inputs and comments by Jürgen Hohmann from the European Commission, Alicia Spengler from GIZ, Markku Malkamäki from THL and Michael Samson from EPRI.

The OECD Development Centre’s publication team, led by Delphine Grandrieux, produced the publication. Justina La assisted with production. The cover was designed by Aida Buendía.

The European Union Social Protection Systems Programme is co-financed by the European Union, the OECD and the Government of Finland.

The contents of this publication are the sole responsibility of the OECD and can in no way be taken to reflect the views of the European Union or the Government of Finland.

Executive summary

Robust monitoring and evaluation (M&E) frameworks support the growth of social protection systems and offer the most productive tools to simultaneously assess programme's effectiveness and provide guidance for improvements. Consequently, M&E systems are increasingly playing a central role in national social protection strategies formulated throughout the world. It provides evidence on social protection's core impacts, which reinforces the political will to scale-up and sustain social protection systems, while exploring why and how the interventions work, contributing insight into how best to strengthen the value-for-money that systems produce.

The Millennium Development Goals demonstrated the importance of identifying and tracking performance indicators, and the need for policy-makers to understand development goals as an element of a complex system in order to effectively measure progress. The Sustainable Development Goals not only include social protection within targets, but also offer the kind of global coordination mechanism that can drive the development of comprehensive M&E systems.

The systems approaches that influence both social protection objectives and the Sustainable Development Goals increasingly require more integrated and effective monitoring and evaluation frameworks, which require strengthening both macro and micro tools. The micro- and macro-evaluation frameworks mutually reinforce each other. Innovative evidence-building approaches inform policy-makers how comprehensive and integrated interventions interact to strengthen cross-sectoral policy outcomes. The macro framework consolidates the results from micro assessments into a planning framework which describes how investments across sectors create developmental synergies.

At a micro-level several tools for impact evaluation exist, each with a unique set of benefits and drawbacks. These, vary to some extent from rigorous-but-expensive to less-expensive-but-less-reliable methods. However, progress towards more integrated and comprehensive evaluation requires innovations, as micro level evidence-building mechanisms that robustly structure learning-by-doing exercises are needed. Moreover, global experience with social protection evaluation highlights the importance of rigorous mixed-methods approach, where quantitative and qualitative techniques complement each other. Comprehensive assessments are rooted in an in-depth understanding of the programme and the poverty context it seeks to address.

The increased prevalence of randomised control trials as the golden standard for evaluations has shifted focus away from national systems as countries cannot be randomised. The macro evaluation framework thus requires tools that link multiple inter-related interventions to cross-sectoral outcomes, and can assess both "intra-sectoral" and

“inter-sectoral” linkages. Innovations in estimating policy production functions have the potential to identify synergies that multiply impact and maximise value-for-money.

An M&E system clearly articulates the institutional arrangements and data sources needed to facilitate a reliable flow of information. Institutional arrangements such as integrated delivery systems and Management Information Systems (MIS) are critical for programme improvement in terms of design and implementations as they enable large-scale and frequent reviews of the social protection programmes. Indicators and targets formulated and regularly collected enable the assessment of programmes, their synergies, and their gaps at various stages of the process, ranging from inputs to activities, outputs, and impacts. Additionally, consolidating effective M&E systems requires capacity building that is not only demand-driven and government owned, but that is both integrated and comprehensive and reflects the national context

This guidance notes aim to provide insight to how countries can build a sound and independent M&E system to monitor and evaluate their expanding social protection systems. It highlights the importance of M&E, evaluates existing methodological approaches, and addresses political economy and institutional issues.

Acronyms and abbreviations

| | |
|----------|--|
| 3iE | International Initiative for Impact Evaluation |
| ATE | Average Treatment Effect |
| ATT | Treatment-on-the-treated |
| BAPPENAS | Indonesian Ministry of National Development Planning- Badan Perencanaan Pembangunan Nasional |
| BDT | Basis Data Terpadu |
| BPS | Statistics Indonesia- Badan Pusat Statistik |
| CCT | Conditional Cash Transfer |
| CEF | Caixa Economica Federal |
| CfW | Cash for Work |
| CGE | Computable General Equilibrium |
| CONEVAL | National Council for the Evaluation of Social Policy |
| CSG | Child Support Grant |
| DiD | Difference-in-differences |
| DSD | Department of Social Development |
| FfW | Food for Work |
| FGD | Focus group discussion |
| GIZ | German Ministry for Economic Cooperation and Development- Deutsche Gesellschaft für Internationale Zusammenarbeit |
| HEF | Health Equity fund |
| IAM | Integrated Assessment Model |
| IFPRI | International Food Policy Research Institute |
| ILO | International Labour Organization |
| ITT | Intention to treat |
| IVs | Instrumental Variables |
| KIIS | Key Informant interviews |

| | |
|--------|--|
| LEAP | Livelihoods Empowerment Against Poverty |
| LEWIE | Local Economy-wide Impact Evaluation |
| LGE | Local General Equilibrium |
| M&E | Monitoring and Evaluation |
| MAMS | Maquette for MDG Simulations |
| MIS | Management Information Systems |
| NGO | Non-Governmental Organisation |
| OSS | One-Stop-Shop |
| PACES | Programme for Extending the Coverage of Secondary School -Programa de Ampliación de Cobertura de la Educación Secundaria |
| PMT | Proxy Means Testing |
| PO | Participant observation |
| PODES | Village Potential |
| PSM | Propensity score matching |
| RCT | Randomised Controlled Trial |
| RDD | Regression discontinuity design |
| SASSA | South African Social Security Agency |
| SDGs | Sustainable Development Goals |
| SHGs | Self-help groups |
| SWS | Single Window Services |
| TN2K | National Team for Accelerated Poverty Reduction |
| ToC | Theories of Change |
| TVET | Technical Vocational Education and Training |
| UDB | Unified Database |
| UNICEF | United Nations Children's Fund |
| VRG | Village Representative Group |
| WFC | Worker Facilitation Centres |

Table of contents

| | |
|--|-----------|
| Executive summary | 5 |
| Acronyms and abbreviations..... | 7 |
| 1. Introduction | 11 |
| 1.1. What is monitoring and evaluation? | 12 |
| 2. Established tools supporting social protection M&E frameworks | 14 |
| 2.1. Taking stock on the evidence-building discourse | 14 |
| 2.2. The evidence-building palate..... | 17 |
| 2.3. Towards comprehensive evaluation..... | 22 |
| Notes | 23 |
| 3. Mixed methods and triangulating evidence-building..... | 24 |
| 3.1. Administrative and monitoring data | 26 |
| Notes | 30 |
| 4. Building effective M&E frameworks for social protection systems..... | 31 |
| 4.1. National systems approaches to M&E and evidence-building | 31 |
| 4.2. Comprehensive and integrated programme-level frameworks for evaluation | 37 |
| 5. Monitoring of social protection in the SDG framework | 39 |
| 6. Institutional arrangements for information and M&E systems..... | 42 |
| 6.1. Management information systems | 43 |
| 6.2. Integrated Delivery Systems | 48 |
| 7. Building national capacity for M&E | 51 |
| 7.1. Key design principles..... | 51 |
| 8. Conclusions and recommendations..... | 55 |
| References | 58 |

Figures

| | |
|---|----|
| Figure 2.1. Rube Goldberg Pencil Sharpener | 15 |
| Figure 4.1. Returns to social protection framework | 32 |
| Figure 4.2. A development planning framework for social protection..... | 34 |
| Figure 6.1. Cadastro Unico..... | 44 |

Boxes

| | |
|--|----|
| Box 1.1. Mexico's case of building an evidence-based social protection system | 13 |
| Box 2.1. Do flying kites cause sharper pencils? The challenges of external validity | 15 |
| Box 2.2. The ethical risks of RCT rigidity | 16 |
| Box 2.3. The importance of a credible comparison group | 18 |
| Box 2.4. Correcting failures of the randomization process of an RCT | 20 |
| Box 2.5. Combining real with quasi-experimental designs in Colombia..... | 21 |
| Box 3.1. A mixed methods evaluation of self-help groups in India | 25 |
| Box 3.2. How South Africa's Child Support Grant Impact Evaluation employed integration | 27 |
| Box 3.3. Transitional poverty analysis using the IDPoor Programme in Cambodia..... | 28 |
| Box 4.1. Calculating the rates of return to social protection using dynamic microsimulation in Cambodia..... | 32 |
| Box 4.2. Policy production function approaches to comprehensive evaluation..... | 36 |
| Box 5.1. Progress towards the Sustainable Development Goal Target 1.3: Secretary-General's Report | 40 |
| Box 6.1. South Africa's institutional framework on M&E | 42 |
| Box 6.2. Brazil's Cadastro Unico..... | 44 |
| Box 6.3. Indonesia's Unified Database..... | 46 |
| Box 6.4. Cambodia's IDPoor | 47 |
| Box 6.5. Single Window Services..... | 49 |
| Box 7.1. South-south examples of social protection capacity building: Brazil's and South Africa's technical assistance to social protection systems in Africa | 54 |

1. Introduction

Robust monitoring and evaluation (M&E) frameworks characterise and support the growth of social protection systems around the world. M&E offers the most productive tools for assessing whether social protection programmes achieve their main goals, while providing guidance for improving programme and systems performance. M&E serves broader learning objectives - building a knowledge base that serves as a global public good. It provides evidence on social protection's core impacts, while exploring why and how the interventions work, contributing insight into how best to strengthen the value-for-money that systems produce. M&E can also provide evidence that reinforces the political will to scale-up and sustain social protection systems.

Around the world, an increasing number of countries have formulated national social protection strategies in which M&E systems play a central role. Social protection can be defined as “a system of formal (and sometimes informal), public (and sometimes private) instruments that aim to reduce social and economic risks, vulnerabilities, exclusions and deprivations for all people and facilitate inclusive social development and equitable economic growth” (Samson and Taylor, 2015^[1]; Devereux and Sabates-Wheeler, 2004^[2]). Within the current policy landscape, framed by the Sustainable Development Goals (SDGs), countries are taking the next step of transitioning from a focus on individual and often fragmented social protection programmes towards a systems approach.

The central question explored in this paper is how countries can build an M&E framework to monitor and evaluate their expanding social protection systems. Systems approaches to monitoring and evaluation require three major initiatives:

- Comprehensive approaches that respond effectively to policy-makers requirements for the evidence required to inform scale-up and sustainability;
- Macro-level evaluation frameworks that measure inter-sectoral synergies and quantify relative rates of return for relevant combinations of interventions; and
- Micro-level evidence-building mechanisms that robustly structure learning-by-doing exercises that guide the development of solutions to complex challenges.

These innovations enable dynamic monitoring and evaluation systems that encompass conventional tools and expand M&E capabilities to support systemic progress.

Comprehensive M&E systems require high-level political will, capacity-building, and careful implementation with continuous cycles of monitoring and improvement. This paper outlines several key areas of a sound and independent M&E system, including methodological approaches as well as institutional and political economy issues.

1.1. What is monitoring and evaluation?

Monitoring describes a process of identifying and tracking performance indicators and reviewing the programme's implementation. It is an ongoing process over the lifetime of the programme, involving the routine observation and recording of activities taking place in a project or programme. Usually, monitoring requires a system that can collect data on aspects of programmes that can be matched against the pre-specified goals and outputs. It aims to assess the progress of project or programme activities. Monitoring systems enable the information collected to be used in making decisions for improving project or programme performance. The systematic and purposeful observation can identify and analyse any unintended (positive or negative) effects of a project or its activities. Monitoring also involves giving feedback about the progress and/or problems of the project or programme to the key principals: government, donors, implementers and beneficiaries.

Evaluation links inputs and activities to the resulting most important outcomes through the objective and systematic assessment of the programme's impacts. Around the world, governments committed to social protection are building evaluation mechanisms into their institutional frameworks for social protection design and implementation. For example, the Government of Brazil established the Secretariat of Evaluation and Information Management to provide monitoring and evaluation for 21 different programmes - employing a harmonised approach and facilitating the integration of complementary programmes supporting both social protection and broader development objectives. Colombia's National Planning Department helps to implement programme impact assessments while simultaneously managing responsibility for policy design. Likewise, South Africa's Performance Monitoring and Evaluation unit in the Office of the Presidency coordinates impact assessments and other reviews nationally, including evaluations of cross-ministerial initiatives such as the nation's public works programmes. Chile's Ministry of Social Development, however, adopted an internal approach, creating the position of Under-Secretary of Social Evaluation to comply with legal obligations to support M&E and improve its quality (Cecchini, Robles and Vargas, 2012^[3]). Box 1.1 discusses Mexico's commitment to evaluations supporting evidence-informed policy development.

Monitoring and evaluation systems serve multiple functions. Broadly, it is important that the institutional set-up encourages large-scale and frequent reviews of the social protection system, with a healthy debate within ministries and nationally regarding issues such as targeting and the benefit level. South Africa is an example of a country where there have been frequent reviews and healthy debates, leading to an expansion of programme coverage, increases in benefit amounts and significant reforms in cases where the programme has not operated effectively. For example, an evaluation of the social grant system's targeting process led to substantial reforms of the associated means test. For evaluations to yield substantial gains, policymakers must be willing to "fine-tune" or adjust a programme's operations whenever necessary in order to improve its performance.

Box 1.1. Mexico's case of building an evidence-based social protection system

Mexico established the global standard of linking evidence-building to social protection policy. The country's flagship programme is a conditional cash transfer, now called Prospera, but originally known as Progresá and subsequently as Oportunidades. The programme was implemented in 1997 to ameliorate extreme poverty in which a quarter of Mexico's population had been living for the previous two decades. Importantly, the government, specifically Congress, mandated that the program had to be evaluated annually by external evaluators. In 2004, the Social Development Law and the creation of the National Council for the Evaluation of Social Policy (CONEVAL) institutionalized evaluation – an autonomous institution mandated to measure poverty reduction nationally and coordinate evaluation of social protection programmes by the federal government. This has led to the construction of a social sector M&E system.

Since its implementation, Prospera has been evaluated through several a randomised controlled trails (RCTs). The evaluations have systematically demonstrated direct effects on health and nutrition outcomes, and on important social determinants such as education. The monitoring of Prospera's implementation allowed policy-makers to identify windows of opportunity to improve its design and operation. One of the main indicators used to monitor its performance is the percentage of people living in extreme poverty. This has been regularly measured every two years through the country's periodically updated national household income and expenditure survey.

Evaluators have linked the resulting evidence to a strengthened political will to expand and sustain the cash transfer programme (if not the name) as political dispensations evolved. The designers of the programme understood that impact assessment is an essential tool to strengthen the operations of the programmes, as well as to generate credible evidence and rigorous proof of the programmes' achievements. As such, the programme is regarded as setting international standards for social policy evaluation.

Source: Valle, A.M. (2016^[4]), "The Mexican experience in monitoring and evaluation of public policies addressing social determinants of health", *Global Health Action*, <http://dx.doi.org/10.3402/gha.v9.29030>.

Global experience with social protection evaluation highlights the importance of a comprehensive and rigorous mixed-methods approach. Good assessments are rooted in a thorough understanding of the programme and the context of poverty it addresses. The groundwork for monitoring and evaluation begins with the initial design of the programme. Quantitative and qualitative techniques complement each other, and rigorous analysis must grapple with the omnipresence of the potential for error.

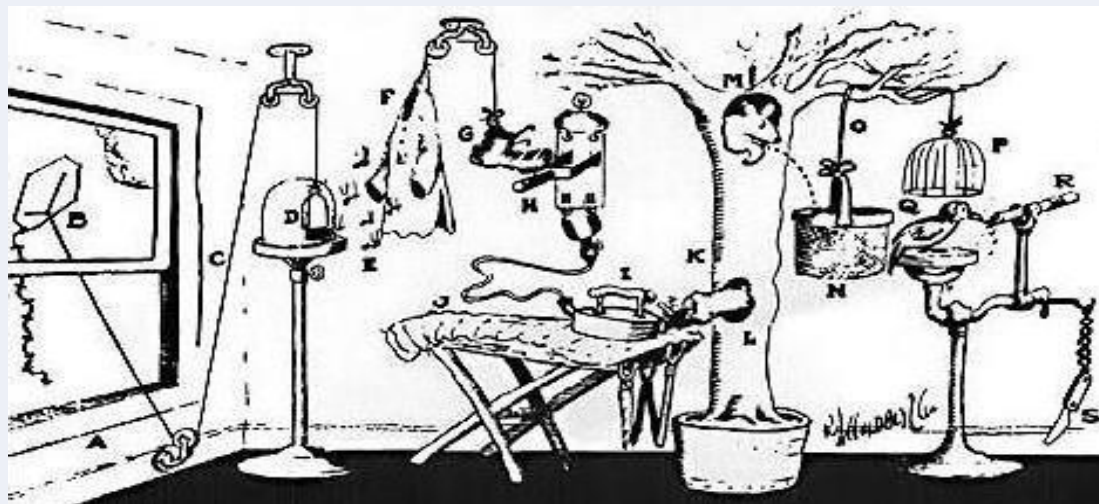
2. Established tools supporting social protection M&E frameworks

2.1. Taking stock on the evidence-building discourse

A policy or programme is only as good as its ability to attain its intended goals; the degree to which it ‘works’. Robust, independent evaluation of social protection programmes provides information about what works, what does not, and—very importantly—why (UNDP, 2009^[5]). Over the past decade, the world has seen a sea of change in the role of evidence informing social policy design and implementation. Particularly in the social protection sector, rigorous quantitative impact assessments have changed how policy analysts and implementers measure success. A range of impact evaluation tools exist, each with a unique set of pros and cons. These pros and cons, to some extent, fall along a continuum of rigorous-but-expensive to less-expensive-but-less-reliable methods.

First, a growing emphasis on methodological rigour has crowded out a more comprehensive approach to evidence-building, thereby creating gaps in the evidence base policy-makers require to design and implement more effective strategies. Notably, while rigour versus cost is a much-debated trade-off in the world of impact evaluation, other issues need not be overlooked. First, the ‘type’ of research question should guide which evaluation tool should be applied. Some evaluators view experimental methods as a gold standard for ‘what works’, but they are not necessarily effective for answering ‘why it works’ (Deaton and Cartwright, 2017^[6]).

Second, a pilot programme can only be scaled up if the adopted model can succeed in different locations. Often, programme impacts are enabled through context-dependent factors. Deaton and Cartwright offer the example of a conditional cash transfer (CCT) programme for which parents are required to take their children to clinics to receive their benefit. The scheme will fail to achieve outcomes if clinics are not available. Such findings sparked policy analysts and development organisations such as UNICEF to consider the provision of combined social policy packages, known as “Cash Plus Care” (Roelen et al., 2017^[7]). The Rube Goldberg machine (Figure 2.1) provides an example that illustrates the point: a flying kite uniquely rigged to simultaneously sharpen a pencil generates an underlying channel of cause-and-effect that is so specific that in no other context would a flying kite be expected to achieve the same outcome (Deaton and Cartwright, 2017^[6]; Peters, Langbein and Roberts, 2018^[8]; Banerjee et al., 2017^[9]). Sound monitoring and evaluation (M&E) frameworks should therefore reflect such context-dependencies and other factors that enable and disable outcomes. If M&E frameworks do not fully explain ‘why’ a programme works, the evidence cannot ensure success in scaling-up.

Box 2.1. Do flying kites cause sharper pencils? The challenges of external validity**Figure 2.1. Rube Goldberg Pencil Sharpener**

Note: Open window (A) and fly kite (B). String (C) lifts small door (D) allowing moths (E) to escape and eat red flannel shirt (F). As weight of shirt becomes less, shoe (G) steps on switch (H) which heats electric iron (I) and burns hole in pants (J). Smoke (K) enters hole in tree (L), smoking out opossum (M) which jumps into basket (N), pulling rope (O) and lifting cage (P), allowing woodpecker (Q) to chew wood from pencil (R), exposing lead. Emergency knife (S) is always handy in case opossum or the woodpecker gets sick and can't work

Source: <http://msl.cs.uiuc.edu/~lavalle/cs397/goldberg.html>

The illustration above illustrates the limitations of external validity – the degree to which a conclusion holds for other persons in other places and at other times. Causal processes often require highly specific economic, cultural, or social structures to enable them to work. The Rube Goldberg Pencil Sharpener is a highly exaggerated example showing the absurdity of assuming that the effect of the kite on the pencil will work when the context is different. Policy makers continually attempt to design causal relations between programmes and outcomes.

While the internal validity – the capability to test causality – is attributed as a key advantage of RCTs, which is equally contrasted by its shortcomings in external validity. Some critics even argue that establishing external validity becomes harder the more the social context is explicitly manipulated by the researcher – which is a central characteristics of an RCT where researchers manipulate treatment assignment. Moreover, scaling an RCT pilot often changes (reduces) the treatment effects because the scaled programme is typically implemented by resource-constrained governments, while the original pilot RCT is often implemented by better-resourced NGOs or the researchers themselves.

A recent systematic review of all RCTs conducted in developing countries and published

in leading economic journals between 2009 and 2014 found that the majority of published RCTs do not provide a comprehensive presentation of how the experiment was implemented and that important limitations to external validity were not addressed. A recent paper on challenges and solutions to “scalable policies” drawing on RCT-based M&E suggest an approach of “multiple iterations of experimentation” where a successful pilot is seen as an effective answer to a policy problem where the “why” has to be pinned down and tested elsewhere. Besides this requiring a multi-year approach which is not always possible, the lesson is that programme impacts have to be viewed in their context, where the context is constantly further refined and understood.

Source: Deaton and Cartwright (2017^[6]), Understanding and misunderstanding randomized controlled trials, *Social Science and Medicine*, Vol. 28, pp. 2-21; Banerjee et al. (2017^[9]), “From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application”, *Journal of Economic Perspectives* 31/4, pp. 73-102; Peters, Langbein, and Roberts (2018^[8]), *Generalization in the Tropics - Development Policy, Randomized Controlled Trials, and External Validity*, Oxford University Press, Oxford.

A third factor that can improve the effectiveness of evidence in building social protection systems is ensuring learning-by-doing over time, which can be guided by expert observation and subsequent adaptation and adjustment. Some evaluation experts view Randomised Control Trials (RCTs) as theory-neutral and reflect scepticism of expertise (Deaton and Cartwright, 2017^[6]). RCTs thrive for evaluation problems where theories are ample but common ground is sparse. However, their rigour can make them unnecessarily rigid (see Box 2.2).

Box 2.2. The ethical risks of RCT rigidity

A treatment for new-borns with pulmonary hypertension – a life-threatening form of increased blood pressure – was developed in the 1970s by expert-driven trial-and-error, drawing from a well-understood theory of the disease. This led in early experimentation to a reduction in mortality from 80 to 20%. The investigators felt compelled to conduct an RCT. One baby in the control group received conventional therapy and died, while 11 received the newer treatment and lived. Nonetheless, the control group was deemed necessary. With a stopping rule of four deaths, four more babies (out of ten) in the control group died, while those in the experimental group all lived.

Source: Deaton and Cartwright (2017^[6]), Understanding and misunderstanding randomized controlled trials, *Social Science and Medicine*, Vol. 28, pp. 2-21.

Fourth, the mass effort of finding ways to mimic the controlled setting of the laboratory environment fails to take account of ‘spill over effects’. Analysing the path from cause to effect should not disregard what this path does to the system at large. Cash transfers randomised at the household level often generate spill over effects that affect non-beneficiaries, including the comparison group, because the aggregate cash injection into

the local economy can stimulate economic activity with widespread effects. RCTs may not estimate what would happen if the project were scaled up to the larger community (Banerjee and Duflo, 2011_[10]). One approach for assessing the broader set of outcomes resulting from social programmes involves the Local Economy-wide Impact Evaluations (LEWIE) methodology. These models have demonstrated that cash transfers in sub-Saharan Africa benefit non-eligible households and largely rural economies by stimulating production activities within the community (Taylor, 2013_[11]).

An appropriate choice of evaluation methodologies therefore requires: (1) carefully constructed Theories of Change (ToCs) that explain how an intervention will deliver the required impacts and (2) a selection of evaluation methods that enable testing the causal links that drive the theory of change.

2.2. The evidence-building palate

Evaluators can choose from a palate of evaluation methods, each with their own statistical strengths and weaknesses as discussed above. Furthermore, a carefully constructed ToC might lead to a preferred method to test its causal links. They are discussed below.

2.2.1. Randomization

RCTs offer a powerful evaluation option that largely solves some of the most vexing problems in impact assessment. If they are implemented properly, they can provide the most robust indication of programme impact. And once an evaluation survey has been designed and the data collected, the empirical exercises to infer impacts from RCTs are straightforward (Khandker, Koolwal and Samad, 2010_[12]). The most convenient phase to start thinking of an RCT is at the inception of a programme, during the pilot phase. This phase offers a natural opportunity to introduce randomization before the programme is scaled up. It presents an opportunity for the implementation partner to rigorously assess the effectiveness of the programme. Equally, it can provide a chance to improve the programme's design to increase its impact.

The main goal of impact evaluations is to construct a reliable counterfactual by finding an appropriate comparison for the treatment group (see Box 2.3). Proponents of 'experimental' rather than 'non-experimental' methodologies understand RCTs to require fewer assumptions to establish a rigorous path from cause to effect, referred to by evaluators as 'internal validity'. This is primarily due to its outstanding ability to randomize away differences between groups of households – one group which is treated and the other not (Khandker, Koolwal and Samad, 2010_[12]). Like tossing a coin, each person participating has as much chance of being selected into treatment as into the comparison groups. Some evaluation experts even suggest that this is also a fair and transparent way to assign scarce programme services (Khandker, Koolwal and Samad, 2010_[12]).

Once participants are randomly assigned and the programme has run for some time, simply comparing the two groups on various outcomes gives the 'average treatment effect' (ATE). However, in practice, not every person assigned to be treated will necessarily comply, willingly or unwillingly. A person who is eligible for unemployment

training may not show up on the first day, while an ineligible person may find a way to attend the course anyway. Such a ‘selection bias’ can create distortions that undermine the internal validity of the evaluation. To address this common risk, evaluators often use the assignment or eligibility for the programme rather than the actual participation to estimate the ‘intention-to-treat’ effect (ITT). Randomly assigned eligibility is usually less vulnerable to selection bias than is actual participation. The ITT estimator is more appropriate when policy-makers and programme implementers mainly “offer” a programme opportunity but cannot (or should not) require participation.

Box 2.3. The importance of a credible comparison group

The International Food Policy Research Institute (IFPRI) conducted an impact assessment of conditional cash transfers in Nicaragua from 2000 to 2002. Households benefiting from the cash transfers received an average of 3 500 Córdobas in the first year, and 3 800 in the second year. Annual household expenditure for these beneficiaries, however, increased only by 404 Córdobas over the two years – an insignificant impact, particularly given the cost of the programme. Can policy analysts conclude the programme was a failure?

No, the problem is not that simple. Simply measuring the change in a welfare measure over time for beneficiary households ignores other factors affecting households. In this study, IFPRI also measured the change in household expenditure for a similar group of households that did not have the benefit of the conditional cash transfer programme. Their average expenditure fell significantly over the two years. It is likely that the average expenditure of the participating households would have fallen also in the absence of the programme. The significant expenditure stabilising impact of the conditional cash transfers is a mark of programme success.

This example demonstrates the importance of a “control group” in evaluating social transfer programmes. Comparing the living standards of households after the introduction of a social transfer programme to their status before the intervention is a “before-after” comparison and ignores the impact of all the other factors that affect the beneficiaries. To separate the impact of the social transfer programme from all the other effects, it is useful to identify a group of households as similar to the beneficiary households as possible – and to monitor their living standards simultaneously with those of the beneficiaries. This group is called a “control group” because it enables the study to control for all the other factors affecting the households.

Source: Samson, Van Niekerk, and Mac Quene (2010_[13]), *Designing and Implementing Social Transfer Programmes*, Economic Policy Research Institute, Cape Town.

It must be noted that even when an RCT is implemented, difficulties with randomization can occur that confound the validity of the impact assessment (Box 2.2). It is also not uncommon that programme implementation is announced on short notice and randomization is not feasible.

2.2.2. *Quasi-experimental methods*

The main quantitative alternative to RCTs is a set of quasi-experimental methods. These empirical methods aim to overcome the lack of an observed counterfactual by drawing on sophisticated statistical analysis to synthetically construct a credible comparison group (Deaton and Cartwright, 2017_[6]). The lack of the RCTs essential instrument, randomization, creates risks that programme participants and non-participants differ in ways other than just whether or not they participate in the programme. For instance, a targeted cash transfer will identify those who are particularly poor and separate them from those who are better off. As a result, participants may be poorer than non-participants even after the intervention. The intervention may still work, but the different starting points complicate the analysis of results. This phenomenon is known as ‘selection bias’: the very reason some households are selected into a programme is because of their background characteristics, be it poverty, number of children, unemployment, lack of education, or other factors. Deaton and Cartwright argue that quasi-experimental approaches can be just as powerful as evaluation tools when they are applied with excellence (Deaton and Cartwright, 2017_[6]).

A common quasi-experimental method, called propensity score matching (PSM), constructs a statistical rather than real-world comparison group that is based on predicting the likelihood that an individual is participating in the programme (treatment), based on observed characteristics.¹ Matching individuals with similar likelihoods across the range of likelihood scores imitates – to the best extent possible given the circumstances and data – a randomization process. The average treatment effect is then calculated, just like in an RCT, by comparing the two groups. PSM estimation often requires a larger sample size than that for a comparable RCT. It assumes that no unobserved characteristics influence both programme participation and the outcomes of interest.

Cash transfer programmes with continuous targeting criteria, such as proxy means testing (PMT), can be used to compare individuals right above and below the cut-off score. As such, they behave as ‘locally randomized experiments’. This evaluation method, known as regression discontinuity design (RDD), exploits a ‘sharp’ cut-off in household or individual characteristics on which they themselves have no direct control. This programme-driven discontinuity between individuals creates an experiment of sorts because there are no systematic differences that drive programme participation. Sudden spikes in household’s income above the score will likely be directly attributable to the cash transfer. RDDs’ specific ability to act as ‘locally randomized experiments’ gives it a distinct edge, allowing it to highlight finer programme impacts, that can sometimes be overlooked when using RTCs - which evaluate a broader ‘average treatment affect’.

Box 2.4. Correcting failures of the randomization process of an RCT

Behrman and Hoddinott (2005) examined the impact of food supplements on children from Prospera (formerly known as Progres and Oportunidades, a government social assistance programme in Mexico) which involved the distribution of food supplements to children. Although the programme was randomized geographically, a shortage in one nutritional supplement provided to preschool children led local administrators to exercise discretion in how they allocated this supplement, favouring children with poorer nutritional status. As a result, when average outcomes between treatment and control groups were compared, the effect of the programme diminished and seemed non-existent.

Behrman and Hoddinott examined a sample of about 320 children in project and control households (for a total sample of about 640). Introducing child-specific fixed-effects regressions revealed a positive programme impact on health outcomes for children. It was predicted that this effect alone could potentially increase lifetime earnings for these children by about 3%. The fixed-effects estimates controlled for unobserved heterogeneity that were also correlated with access to the nutritional supplement.

Source: Behrman and Hoddinott (2005^[14]), “Programme Evaluation with Unobserved Heterogeneity and Selective Implementation: The Mexican PROGRESA Impact on Child Nutrition”, Oxford Bulletin of Economics and Statistics, pp. 547-69.

In some cases, beneficiary-status can be indirectly guided by randomness. Such is the case when a programme is assigned randomly such as in an experimental design. However, the lottery does not perfectly predict programme participation, as discussed above, as participants can opt-out or find ways to opt-in. Through a method called two-stage least squares (2SLS) in which instrumental variables (IVs) are estimated, only the random part of programme participation can be isolated and used to evaluate the ‘average treatment effect on the treated’ estimator (Box 2.5).

The example narrated in the box below shows that quasi-experimental methods can also be used to complement and enhance experimental methods. A very common method that is used in this regard is the difference-in-differences (DiD) estimation. DiD estimates combine with-and-without comparisons (randomized or not) with before-and-after comparisons. This implies that ‘panel data’ is needed - a survey must be carried out before the programme commences and a survey after a designated period of time has elapsed - in order to measure impact. DiD compares treatment and comparison groups in terms of changes over time. Anything that is unobserved but remains constant over time is cancelled out by the double-differencing. The approach assumes that both treatment and comparison groups have no important characteristics that evolve differently over time.

Box 2.5. Combining real with quasi-experimental designs in Colombia

The Programme for Extending the Coverage of Secondary School (Programa de Ampliación de Cobertura de la Educación Secundaria [PACES]) in Colombia, provided more than 125 000 students with vouchers that covered over half the cost of attending private secondary school. Because of the limited PACES budget, the vouchers were given through a lottery. Angrist et al. (2002_[15]) took advantage of this randomly assigned treatment to determine the effect of the voucher programme on educational and social outcomes.

They found that voucher beneficiaries were 10% % more likely to complete the 8th grade and scored higher on standardized tests three years after the initial lottery. They also found that the educational impacts of PACES were greater for girls than boys. They also looked at the impact on several social outcomes and found that those receiving vouchers were less likely to be married and worked fewer hours per week.

However, there was some non-compliance with the randomized design. Only 90% % of the initial lottery winners had actually used the voucher or another form of scholarship, and 24% % of the lottery losers had actually received scholarships. Angrist and colleagues therefore also used intention-to-treat – or a student’s lottery win or loss status – as an instrumental variable for programme participation, creating an effect called the ‘treatment-on-the-treated’ (ATT). This estimate separated the pool of voucher beneficiaries who had won the voucher from those who lost but still participated, creating a truly random treatment group.

Finally, the researchers were able to calculate a cost-benefit analysis to better understand the impact of the voucher programme on both household and government expenditures. They concluded that the total social costs of the programme are small and outweighed by the expected returns to participants and their families. This suggested that demand-side programmes such as PACES can be a cost-effective way to increase educational attainment.

Source: Angrist et al. (2002_[15]), Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment, *American Economic Review* 92/5, pp. 1535-58.

2.2.3. Micro-simulation for ex ante evaluation

An alternative to the more traditional ex-post methods is **micro-simulation modelling**. This ex-ante evaluation tool enables researchers to analyse predictive scenarios of policy impacts. These models are built on a foundation of large-scale household surveys and use a range of rules and assumptions to model ‘first-order’ effects such as income, poverty, and inequality. By adjusting benefit sizes in the case of a cash transfer programme, these models can effectively calculate programme costs at different levels of impact, serving as a simple cost-benefit analysis.

Micro-simulation models analyse at the level of individual units, such as persons, households, firms or other micro-economic agents. As such, it estimates impacts beyond averages and is able to examine heterogeneity across agents – particularly the distributional impact of a particular set of rules. The model is built around a database in which each unit is represented by a data record containing an information set including at minimum a unique unit identifier, a weight that enables generalization to a larger population, and a set of associated attributes of interest to the modeller. For example, a typical social policy micro-simulation model will include a database of individuals with information on age, sex, educational attainment, family structure, income, employment, livelihoods indicators and other attributes.

Because microsimulation is driven by a modelling of behaviour and interactions of agents, its robustness depends on the validity and reliability of its underlying set of computational rules and assumptions. Robustness can be improved by modelling additional layers of complexity. Simplicity is not bad, but it creates a narrower range of validity. For example, deterministic models apply rules based on the various attributes of the agent, such as a cash transfer added to an individual's income based on age. However, real life is not that predictable. Stochastic models include probabilistic rules. For example, the probability of receiving a cash transfer might depend on household attributes as well as a random factor.

Another such layer is its 'temporal' or time dimension. Static models provide a snapshot of immediate programme impact and model the direct relationship between policy inputs and outcomes. Dynamic models incorporate time-effects such as economic and demographic growth. Importantly, these models can also include the behavioural response, or 'second-order', impacts of a policy. As such, they can model complex theories of change, with various impacts and feedback loops.

Micro-simulation models can be powerful tools for policy-makers to understand the effects of policies, particularly taxes and benefits, and reforms to them. And they are incredibly cost-efficient. But their robustness depends on model complexities and the degree to which models approximate the real-world within a range of validity. Its results should however not be over read, especially when the economic and social context is not clear cut, which it seldom is. Yet analysing and comparing different policy alternatives to a policy case is where microsimulation excels.

2.3. Towards comprehensive evaluation

Policy-makers wish to know more than whether a programme works. If a programme is to be scaled-up, adjusted and/or made more cost-efficient, one wishes to understand exactly what it is in the intervention that is driving impacts and why. Many complex research questions are relevant. Will the impacts carry over to other contexts? What will the long-term impacts be? Are there heterogeneous impacts for different population groups? How do policies interact and do these interactions compound or weaken impacts? How does the programme affect the macro-framework?

The tools discussed above, if designed appropriately, can contribute to answering these questions. These quantitative methodologies, however, are insufficient. Additional tools are required for more comprehensive programme evaluation:

- By expanding surveys with a range of **community-level questions**, programme effects on the local economy can be examined. Programmes can spark behavioural changes in its recipients, which cause multipliers to other areas of their lives as well as externalities and spill over to the rest of the community. As cash transfer recipients spend their cash, these households unleash ‘local general equilibrium’ (LGE) effects that transmit programme impacts to others in the economy, including non-beneficiaries. Experimental economists often ignore the effects of programmes on ineligible groups, instead focusing on the average effects of treatments on the treated. Ignoring general-equilibrium effects can give an incomplete and often biased picture of how cash transfers affect local economies, including production activities.
- **Qualitative methods** can often better explain causal pathways. Key informant interviews (KIIs), focus group discussions (FGDs), and participant observation (PO) illuminate the programme experience from the beneficiary’s viewpoint. Qualitative approaches are increasingly understood as playing a pivotal role in sound monitoring and evaluation systems. They have been given centre stage in areas such as human-centred design and adopted both within the public and private sphere. The abovementioned qualitative instruments (KIIs, FDGs, POs) highlight that qualitative methods have unique and important advantages. For example, quantitative evaluations struggle to pick up low prevalence or underreported outcomes, such as child abuse. In-depth interviews in such cases provide a more effective tool (Glass, Gajwani and Halliday, 2016^[16]). Complex programmes such as training and mentoring programmes likewise benefit from qualitative evaluation approaches. Focus groups and in-depth interviews can observe and identify a vastly wider range of variables as well as highlight how these interact with each other and create feedback loops.

Notes

¹ Importantly, these are observed characteristics measured before the programme begins—pre-treatment – so that they are not influenced by the programme itself, which would confound impact assessment.

3. Mixed methods and triangulating evidence-building

Today, global best-practice evaluation approaches require ‘mixed methods’, where both quantitative and qualitative methods are applied in complementary ways (Devereux and Sabates-Wheeler, 2004^[2]). Whereas quantitative methods focus on averages and generalizable effects, qualitative methods identify outliers and very context-specific effects. An evaluation of a social protection programme in Latin America utilised both quantitative surveys and ethnographic models. The South African Child Support Grant Impact Assessment¹ employed non-experimental quantitative approaches combined with a multi-method qualitative component (DSD/SASSA/UNICEF, 2012^[17]). While these and similar studies have been identified as cases of good practice (Devereux and Sabates-Wheeler, 2004^[2]), leveraging the full potential of mixed method impact evaluation remains an area of ongoing progress.

Mixed methods often employ ‘triangulation’ to obtain different types of data addressing a common set of evaluation questions (Morse, 1990^[18]). Complementary designs consolidate the differing strengths and address the weaknesses of quantitative and qualitative methods. Qualitative methods provide analytical insights into the complex ‘missing middle’ between interventions and impacts (Garbarino and Holland, 2009^[19]). Without such insights, researchers and policy analysts tend to make ‘interpretative leaps’ based on what is measured. A pitfall here is that what is unquantifiable becomes irrelevant while ‘what is measurable and measured then becomes what is real and what matters’ (Chambers, 1995^[20]). Conversely, if qualitative methods inductively throw up surprising explanations, quantitative methods are then able to ask, ‘how common is this’? And establish how confident we can be in these explanations. This iterative relationship between describing and explaining provides the key to effective combination of methods and data (Box 2.5) (Garbarino and Holland, 2009^[19]).

Box 3.1. A mixed methods evaluation of self-help groups in India

The government of the Indian state of Andhra Pradesh has stimulated the formation of women-only grassroots organisations at the village-level called Self-Help Groups (SHGs). By 2007, more than 700 000 such groups had formed with help of a range of development partners. An impact evaluation conducted in the same year found a puzzling result. Its village survey showed a continued rise in the number of SHGs as compared with two years prior. However, the individual-level survey ran parallel showed a reduction in participation of the groups. Why were there more and more groups when participation was on the decline?

The apparent discrepancy was explained with help of qualitative data collected alongside the quantitative surveys. These revealed that there was an increasing number of non-functioning SHGs that only existed on paper but had dissolved due to a lack of skills of staff and non-payment of microloans. Yet they had remained on the books.

Worryingly, SHG dropout had impacted the poor most. The qualitative fieldwork informed several policy responses to this challenge, including support to illiterate groups in record keeping and adoption of simpler bookkeeping methods, as well as finding alternative payment arrangements. Further policy implications came from the quantitative analysis. These indicated that households with multiple eligible female members were not receiving additional loans and so these women were not participating. By triangulating the results, a range of in-depth policy options were identified to turn the decrease in participation around.

Source: White (2008_[21]), *Making impact evaluation matter*. International Initiative for Impact Evaluation, International Initiative for Impact Evaluation, New Delhi; Garbarino and Holland, (2009_[19]), *Quantitative and Qualitative Methods in Impact Evaluation and Measuring Results*, Social Development Direct, London.

Integrated evaluation, defined as the interaction of qualitative and quantitative approaches for strengthening evidence building, often offers the most effective way to build relevant evidence. Qualitative-quantitative integration offers seven major contributions to comprehensive evaluation:

- **Consistency:** Ensuring that both quantitative and qualitative components are aligned at the strategic level, particularly with regard to their mutual alignment in the overall strategy and consistency across the design documentation.
- **Technical design in terms of models:** Results and findings from the qualitative exercise are used to inform and validate key elements of the overall strategy, particularly in the case of quasi-experimental design in terms of modelling the participation of beneficiaries in the programme.

- **Technical design in terms of instruments:** Results and findings from the qualitative component directly inform the development of the survey instruments.
- **Explanatory scope and capacity:** Using the findings from the qualitative study to explain motivation and reasoning for interesting or anomalous results obtained from the quantitative study, and to provide further insight into the “how” and “why” of impact and operational lessons.
- **Quantifying qualitative results:** Testing and quantifying the qualitative results to identify which are purely anecdotal and which represent statistically significant findings across the broader population.
- **Practical implementation lessons:** Ensuring that the practical lessons learnt from the preparatory qualitative component (design, fieldwork, analysis and reporting, but particularly sampling and finding respondents) are consciously noted and used in the quantitative component.
- **Assumptions:** Using the qualitative component to identify and check underlying assumptions that need to be recognised and which may impact in one way or another on the project.

Box 3.2 discusses how South Africa’s Child Support Grant impact evaluation benefitted from all seven of these major contributions. Integrated approaches not only link quantitative and qualitative approaches, but also ex ante (micro-simulation) and ex post (programme) evaluation, monitoring evidence and information from other sources.

3.1. Administrative and monitoring data

Evaluations usually employ data specifically collected for the impact assessment. Increasingly, evaluators employ existing databases of administrative and/or monitoring data. This can expand the scope and improve the efficiency of the evaluation. Box 3.3 exhibits the usefulness of Cambodia’s administrative data supporting a social protection systems evaluation.

Box 3.2. How South Africa's Child Support Grant Impact Evaluation employed integration

South Africa's Child Support Grant impact evaluation benefitted from all seven of the major contributions of qualitative-quantitative integration:

1. Consistency: In South Africa's Child Support Grant impact evaluation, both quantitative and qualitative teams worked together to ensure consistency, with the reference team interrogating design documents intensively with the support of global experts.
2. Technical design in terms of models: In South Africa's Child Support Grant impact evaluation, focus group topics were identified to specifically address key issues affecting the quantitative analysis of impacts, such as modelling early versus late enrolments and adjusting for potential selection bias associated with variation in CSG take-up and the gradual roll-out of the CSG to teenagers.
3. Technical design in terms of instruments: In South Africa's Child Support Grant impact evaluation, the results and findings from the qualitative component informed the development of each survey instrument module prior to and during pre-testing.
4. Explanatory scope and capacity: In South Africa's Child Support Grant impact evaluation, findings from the qualitative study were used in the final report to help interpret and explain programme impact results obtained through the quantitative analysis.
5. Quantifying qualitative results: In South Africa's Child Support Grant impact evaluation, due to the integration, the questionnaires allowed for a test and quantification of the qualitative results (with larger sample sizes), which enabled the study to distinguish which effects or relationships were purely anecdotal and which represent statistically significant findings across a broader population. For example, expenditures on "beauty products" were virtually nil.
6. Practical implementation lessons: In South Africa's Child Support Grant impact evaluation, many lessons were learned in the qualitative research (e.g. sampling and finding respondents) that have informed survey design and implementation. For example, the qualitative team identified a problem with SOCPEN addresses.
7. Assumptions: In South Africa's Child Support Grant impact evaluation, the qualitative component was used to identify and check assumptions for the quantitative models.

Source: Samson et al. (2016^[22]), "The Impact of a Promise Realized: South Africa's Child Support grant", From Evidence to Action: The Story of Cash Transfers and Impact Evaluation in Sub-Saharan Africa, Oxford University Press, Oxford.

Box 3.3. Transitional poverty analysis using the IDPoor Programme in Cambodia

IDPoor is a panel survey and database employed by the Government of Cambodia and its development partners to identify poor households and to improve the targeting and distribution of social protection services. The survey captures proxy poverty indicators in a scoring system, as well as community input, in order to assign households a level of poverty. From these processes, officials assign each household IDPoor level 1 (very poor), IDPoor level 2 (poor), or other (non-poor). Information on households identified as poor (either level 1 or 2) have additional data entered into the IDPoor database, and these households then receive an IDPoor “Equity Card”, which enables them to access a variety of social protection services. A study used the IDPoor database to analyse the dimensions of poverty and the nature of transitions in and out of poverty and across levels of poverty intensity in Cambodia. The analysis included the construction of poverty transition matrices and the estimate of probit regressions to assess the covariates of poverty transitions.

Methodology

The IDPoor database consists of eight rounds of interviews, conducted since 2007, each targeting a different subset of provinces. The survey re-interviews each household in the same provinces every three years, allowing for the panel analysis of poverty transitions. The individual households were identifiable across the various rounds through their unique region codes and household codes. Using these unique identifiers to merge the dataset to create a panel data made it possible to track the poverty status of these households up to three periods. This panel dataset allowed the study to explore poverty dynamics over two different periods, enabling the probit analysis to incorporate the explanatory impact of prior transitions on current changes in poverty status.

In total, there were 584 105 households that appeared across all three waves of the IDPoor database, and these households formed the core analysis group. Since not all households are covered in each wave, an analysis of selection bias provided a better understanding of potential selection bias that might affect the outcomes in the transition matrices. The analysis was split into three samples: The first matrix for sub-sample 1, S1, presented poverty transitions for households in provinces that the data verification process identified as having a high likelihood of exhibiting selection bias (the provinces for which the increases in the included population were greatest from wave to wave). The second matrix for sub-sample 2, S2, presented poverty transitions for households in provinces that the data verification process identified as having a low likelihood of exhibiting selection bias (provinces for which the IDPoor teams consistently sampled nearly the entire rural population). The last matrix, S3, presented poverty transitions for all households regardless of province. An analysis of the material differences and the magnitudes of the transition matrix differences between the identified sub-samples suggested that the potential for selection bias was limited.

Results

Poverty Transition Matrices

The poverty transition matrices demonstrated that households in the first sub-sample (S1) are much more likely to be poor compared to households in the second sub-sample (S2). Although, households in the first sub-sample were more likely to be poor on all IDPoor levels than households in sub-sample 2, transitions in and out of poverty for both sub-samples took place at similar rates, varying by a few percentage points at the most. The third matrix, S3, which covered all households matched between waves 2 and 3 of the IDPoor dataset demonstrated the resilience of households that are not poor and the prevalence of positive poverty transitions amongst poor households. The results showed that most households that are non-poor do not fall into poverty between waves 2 and 3 and a majority of households that are either poor or very poor experience upward transition between waves 2 and 3.

Conditional Transition Matrices

Aimed at identifying the determinants of poverty transition across periods, the conditional transition matrix analysis found that “newcomers to poverty” defined as those who fell into poverty between waves 1 and 2, are very likely to move out of poverty in wave 3. Households that graduated out of poverty between waves 1 and 2 were more likely to fall back into poverty between waves 2 and 3 than the full national sample; “newcomers to non-poverty” were more vulnerable to falling back into poverty in S1, S2 and S3. The study also found that female-headed households make more frequent transitions into poverty and then make less frequent transitions out.

Probit Analysis

The probit analysis explored the marginal effects for the probability of transitioning out of poverty between wave 2 and wave 3, separately for three samples – S1, S2 and S3. In all three specifications, a set of household characteristics recorded in wave 2 and having experienced an economic shock between wave 1 and wave 2 have a detrimental effect on the probability of moving from poverty to non-poverty between wave 2 and wave 3. Across the three samples, having a disabled head or spouse, having older persons (above the age of 60), and the household head being divorced or widowed decreased the probability of graduating from poverty. Similarly, households that had sold land or where not all members were generating income were also less likely to graduate out of poverty.

Source: MoP (n.d.^[23]), “IDPoor” (*website*), Ministry of Planning, Royal Government of Cambodia, Phnom Penh.

Notes

¹ Commissioned and funded by the Department of Social Development (DSD), the South African Social Security Agency (SASSA) and the United Nations Children's Fund (UNICEF) South Africa.

4. Building effective M&E frameworks for social protection systems

The increased prevalence of randomised control trials as “gold standard” evaluations has shifted focus away from national systems: countries cannot be randomised. The systems approaches that influence both social protection objectives and the larger Sustainable Development Goals increasingly require more integrated and comprehensive monitoring and evaluation frameworks, which require strengthening both macro and micro tools. Several countries—particularly those with planning institutions—provide important lessons for building these monitoring and evaluation (M&E) systems. Comprehensive M&E systems that strengthen inclusive social development and equitable economic growth tackle complex challenges, build bridges across policy sectors and coordinate interventions within a larger planning framework (OECD, 2013^[24]). The following section assesses first the national systems level framework and then discusses programme-level approaches.

4.1. National systems approaches to M&E and evidence-building

Over the past decade, several countries have developed evidence-building approaches that support national systems. Brazil, Cambodia, Indonesia, Nepal, South Africa and Uganda, among others, all enabled macro-level planning frameworks that supported inter-ministerial initiatives. Brazil innovated integrating cross-sectoral complementary interventions into the *Bolsa Familia* programme more than decade ago and has led South-South policy diffusion efforts in many countries (see Box 6.5).

Countries like Indonesia and Nepal, for more than a decade, have centred social protection policy development within their respective national planning agencies. In Indonesia, BAPPENAS ensured that integrated social protection systems supported both poverty reduction and larger development objectives. For example, cash transfers enabled the government to reduce reliance on regressive and economy-distorting price subsidies while expanding human capital-building investments in health, education and nutrition for poor families. Similarly, Nepal’s National Planning Commission authored the nation’s first social protection strategy, building a life-cycle system that tackles poverty and vulnerability while building social cohesion and strengthening developmental outcomes.

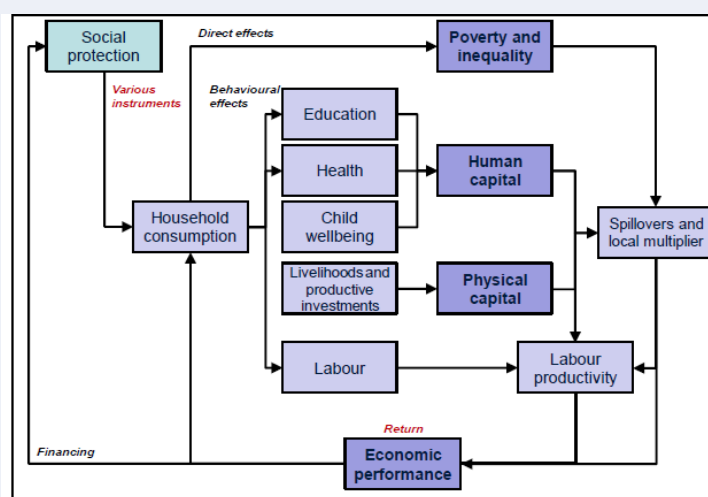
Box 4.1. Calculating the rates of return to social protection using dynamic microsimulation in Cambodia

The Government of Cambodia and UNICEF commissioned an evaluation of the rates of return to social protection which assessed the impacts of a range of social protection instruments – a cash transfer for children, a social pension, a scholarship for lower secondary education, and a public works programme – on a range of household economic and social outcomes, including poverty, inequality, school attendance, nutrition, adult labour force participation and productivity.

The analytical framework analysed the effects of social assistance instruments on socio-economic outcomes, incorporating feedback loops and multiple order impacts. The results show that changes in household disposable income affect household behaviour, including investments in health, education, child well-being and livelihoods and productive activities. Improving health status, education and child well-being increase the level of human capital, while livelihoods and productive investments increase physical capital, increasing labour productivity increase and raising disposable incomes, creating a virtuous circle of inclusive social development and equitable economic growth.

The study also concluded, “The effects, benefits and returns of social protection can be higher if it is implemented together with complementary policies such as improving health and education coverage and quality, enhancing sanitation conditions and infrastructure, fostering economic productivity and the formal labour market, and promoting industrialization, innovation and technical change.”

Figure 4.1. Returns to social protection framework



Source: Mideros, Gassmann and Mohmen (2013^[25]), Estimation of rates of return on social protection: Making the case for non-contributory social transfers in Cambodia, United Nations University.

The development planning approach is strengthening multi-sectoral interventions and reinforcing multidimensional impacts in Bangladesh, Cambodia, Indonesia, Nepal, Rwanda, South Africa, Tanzania, Uganda and other countries. The framework supports both planning and impact assessment—recognizing the importance of both *ex ante* and *ex post* evaluation. Planning agents within governments employ the framework to balance social and economic investment priorities to jointly achieve the primary policy objectives. These can include the promotion of equity, reduction of poverty, increasing employment, and the strengthening of pro-poor and inclusive economic growth and development. The planning process builds linkages within specific sectors (such as the mutual strengthening of cash transfers and social health insurance within the social protection sector, as in Ghana) and across sectors (such as designing social protection programmes to maximize education, health, nutrition and livelihoods impacts). Uganda employed a development planning process for its National Development Plan 1 in 2010 that assessed the importance of intra- and inter-sectoral linkages.

A development planning framework reinterprets the economic input–output matrices employed by development planning models decades ago, adapting them as policy production models. “Inputs” are defined as the set of public (and sometimes private) instruments, programmes and policies that enable the government to achieve priority “outputs”, which are defined in this approach as the achievement of national policy objectives. The framework emphasises the importance of “intra-sectoral” and “inter-sectoral” linkages. For example, intra-sectoral linkages within social protection reflect the mutual strengthening that results when cash transfers finance the contributions of otherwise destitute households for social health insurance. This in turn protects the members from the catastrophic health shocks from which social transfers alone provide inadequate protection. The combination of cash transfers and social health insurance works much better to reduce poverty and vulnerability than the individual instruments on their own.

The approach also analyses inter-sectoral linkages. For example, social protection instruments strengthen outcomes outside their own sector, deepening human capital, strengthening livelihoods development and broadly promoting pro-poor and inclusive economic growth. Many social protection instruments work by expanding poor people’s access to markets. Cash transfers expand the effective demand for market goods and services, enabling people to meet their basic needs while strengthening livelihoods engagement and stimulating economic activity.

Likewise, a full range of policy sectors can effectively strengthen the achievement of social protection objectives: education builds human capital and effectively tackles the intergenerational transmission of poverty, livelihoods programmes strengthen household resilience and reduce vulnerability, and economic reforms can create opportunities that enable poor households to lift themselves out of poverty.

Figure 4.2. A development planning framework for social protection

Type the subtitle here. If you do not need a subtitle, please delete this line.

| Policy Instruments (Inputs) | | | | | | | Development planning matrix | Policy objectives (Outputs) | |
|-----------------------------|------------------|--------------------|-----------------|---------------------|--------------------|----------------|-----------------------------|-----------------------------|-------------------|
| Social Protection | | Other sectors | | | | | | | |
| Cash transfers | Health Insurance | Education/training | Social services | Financial inclusion | Livelihood support | Infrastructure | | | |
| | | | | | | | Poverty reduction | | Social protection |
| | | | | | | | Risk management | | |
| | | | | | | | Social inclusion | | |
| | | | | | | | Human capital development | | Other |
| | | | | | | | Increased employment | | |
| | | | | | | | Economic growth | | |

Source: Samson and Taylor (2015^[1]), “Supporting Social Protection Systems”, Concept Paper N°4, Tools and Methods Series, European Commission, Luxembourg.

The essential element of the development planning framework for social protection is a national coordinating mechanism that plans, prioritizes and integrates the set of public policies and practices (including those of the social protection sector). This mechanism increasingly includes a national development plan coordinated by a national planning institution (such as a National Planning Commission). These comprehensive and integrated planning approaches strengthen both impact and efficiency by increasing the likelihood of achieving the priority policy objectives while reducing costs and risks. In addition, practical development plans reinforce credibility in the government’s strategy, enabling the government to expand its policy options.

The main gap in this macro evaluation framework is the measurement of conditional rates of returns. Consider the investment of a billion dollars in a social cash transfer programme. In a single-sector/single-outcome evaluation, reflecting the optimal assignment of cash transfers to the poverty reduction objective, a culmination evaluation would quantify the future impact on poverty reduction.

A multi-sectoral approach may be more effective and more efficient. Combining cash transfers with social health interventions improves the effectiveness of both interventions. Cash transfer initiatives better enable poor households to access social health insurance programmes, and the health benefits protect poor households from the catastrophic shocks which are too great for cash transfer interventions to provide adequate protection. The result is a greater impact on both of the priority outcomes. The rate of return to a cash transfer programme depends on the level of investment in the social health insurance initiative, as well as on the pattern of investments in the entire range of the government's social and economic investments. The same is true for the rate of return to a social health insurance investment. In order to analyse a manageable problem, conventional evaluation approaches generally assume the investments in complementary interventions are unchanged. As a result, these frameworks are usually unable to provide evidence enabling the government to optimize the mix of investments and benefit from synergies across sectors.

Policy makers demand evidence on the trade-offs in returns to multi-sectoral complex interventions, understanding that the incremental return to an additional investment in one sector depends on the pattern of investment in every other sector. Simple culmination evaluations cannot answer this question, since they usually focus on a single intervention in one sector, or at most a few complementary or alternative treatments. Policy makers require evidence on the impact of different states of a socioeconomic system, and this defies RCT: one cannot randomize countries. The benefits to complex evaluation are twofold: (i) when complex interventions benefit from substantial interaction effects across sectors, only a complex evaluation can measure the additional synergy benefits and illuminate the complex trade-offs in combining the multiple initiatives required to generate the optimal impact, and (ii) complex evaluations capture the entire spectrum of benefits, avoiding the undercounting associated with single-sector/single-outcome trials.

Box 4.2. Policy production function approaches to comprehensive evaluation

Economists decades ago pioneered the use of production functions to model how governments produced social and economic policy outputs—health, education, poverty reduction and other development goals. Computable General Equilibrium (CGE) models provided economy-wide assessments that explicitly incorporated development goal production functions. One of the more sophisticated policy models, the World Bank’s Maquette for MDG Simulations (MAMS) supported economy-wide assessments, incorporating cross-sectoral interactions and macro-economic forces. CGE models like MAMS usually use restrictive Cobb-Douglas production functions to predict development goal outcomes, limiting policy technologies to constant returns to scale and similarly restricting opportunities for cross-sectoral synergies.

Recent work adopts an innovative approach. The translog production function specification enables the estimation of complex relationships among policy sectors, with investments in one set of interventions directly influencing the productivity of government spending in other areas. For example, investments in health can accelerate achievement of education goals, and spending that strengthens gender equity initiatives can improve livelihoods. In addition, these models can better analyse non-linear relationships—the high returns to investing in policy initiatives that harvest “low-hanging fruit”, but the challenges of diminishing returns as policies scale up further and aim to achieve universal success—grappling with the challenges of extending progress to the most vulnerable and marginalised groups.

The availability of better fiscal datasets enables this new analysis. Historically, economists employed cross-country datasets with highly aggregated sectoral data. The aggregate data and contextual differences across countries obscured important synergies. Today, governments increasingly invest in public finance systems that precisely track programme-level expenditures at local (e.g. district) levels. Each district or local unit makes fiscal decisions allocating resources to health, education, social protection, livelihoods and other sectors. Household surveys enable measurement of district-level outcomes for these sectors, which can be linked to the fiscal data. These comprehensive datasets enable more powerful estimations of sophisticated policy production functions (e.g. translog specifications), identifying synergies and quantifying efficiencies that document substantial value-for-money.

Source: Samson, M. (2018_[26]), *Financing the child-centred SDGs in Ethiopia*, UNICEF Ethiopia, Addis Ababa.

The box above discusses the policy production function methodological approach for quantifying this development planning framework. Other options include *Integrated Assessment Models* (IAMs), which provide a flexible approach for assessing cross-sectoral dynamic interactions over long horizons. However, IAMs generally do not support goal-based budgeting with concrete estimates of required investment resources.

4.2. Comprehensive and integrated programme-level frameworks for evaluation

The rising dominance of “gold standard” randomised evaluations has provided government policy-makers with a wealth of programme- and project-level impact assessments, many of which provide robust and credible policy guidance on important practical questions. These tools, however, provide increasingly limited insights as policy options become more complicated.

The systems approaches that influence both social protection objectives shift toward more integrated and comprehensive evaluation also requires methodological change at a micro level. The conventional micro- evaluation approach of an RCT—or, in social policy terms, an experimental pilot—tends to bias toward single-sector/single-outcome trials. The conventional evaluation approach is best suited to trials where implementers know the precise parameters of the interventions in advance, and these can be implemented homogenously across the treatment group(s). RCTs often fail if the treatment must be adjusted during the course of the trial. RCTs are better suited to the objectives of culmination evaluation rather than the broader process of evidence-building that often requires a learning-by-doing approach. In contrast to an experimental evaluation, an evidence-building evaluation allows for greater flexibility in both design and implementation. The evidence-building framework sacrifices some risk of internal validity for greater potential variability in the intervention and stronger policy relevance. Both types of evaluation share common tools. For example, both can use randomization techniques to map out the counterfactual. However, experimental evaluations are generally designed to maximize internal validity, often constraining the intervention to a limited and constant set of outcomes and treatments, given the typical resource constraints that limit any evaluation initiative.

Evidence-building evaluations aim not only to test whether an initiative works, but also to help design the optimal combination of interventions. Evidence-building evaluations tend to have more broadly represented samples, to increase the external validity of the assessment. The treatment in an evidence-building pilot can change over time, as learning-by-doing leads to an improvement in the technology for delivering impact. The result is an increased variability in the effectiveness of the intervention, which may reduce statistical power.

There are also differences in ethical approaches that experimental evaluations and evidence-building evaluations adopt. Howard White, Executive Director of the International Initiative for Impact Evaluation (3iE), makes the case that “the really unethical thing is not the withholding of the program, perhaps temporarily, from some group. The really unethical thing is the spending billions of dollars each year on programs that don’t work. And without rigorous impact studies, including RCTs, we won’t know if they work or not. The sacrifice of having a control group is a small one compared to the benefits of building an evidence base about effective development programs.” Essentially, it is unethical not to experiment if one does not have rigorous evidence on programme effectiveness. For example, in a 3ie/IFPRI Seminar Series on Impact Evaluation, Renos Vakis of the World Bank reported evidence on the impact of a cash transfer programme on early childhood cognitive development, based on an RCT that randomized benefits to a sample of infants (3ie and IFPRI, 2012^[27]).

Evidence-building pilots generally adopt a “do no harm” ethical approach, adopting quasi-experimental methodologies (such as PSM or regression discontinuity techniques) or else providing treatments with unambiguous net benefits (such as cash transfers) to both treatment and control groups while randomizing treatments with uncertain costs and benefits. For example, evaluations of South Africa’s Child Support Grant, a rights-based cash transfer programme, have successfully employed PSM techniques that have not required the withholding of benefits to any child. An RCT of savings and investment linkages to the Child Support Grant employed treatment and control groups of adolescents who all received the cash transfer, but randomized the assignment of treatments that involved costs and benefits to participating households and an uncertain net benefit (in clinical terms, satisfying the condition of “equipoise”).

5. Monitoring of social protection in the SDG framework

Experience with the Millennium Development Goals has demonstrated that to effectively monitor the progress of a nation towards a development goal, policy-makers must understand it as an element of a “complex system”, in which numerous disparate factors interact in multifaceted ways to affect relevant indicators. If they instead study policies and results as if they exist in a vacuum, only analysing factors that may initially seem most directly related to the goal, they will be ignoring much of the data and mechanisms which truly underlie progress. This limited understanding leads to policy strategies based on an incomplete view of the issue to be addressed, and the nation will likely underperform. Conversely, if policy-makers tackle goals holistically and draw from a deeper and more informed understanding of the complex interactions at play, they will be equipped to leverage the potential of rapidly advancing data analysis tools to construct a policy approach which will lead their nation on the best path available given the evidence at their disposal (IIED, 2016^[28]; 2016^[29]).

The first SDG aims to “end poverty in all its forms everywhere.” The social protection target associated with this goal (1.3) plans to “implement nationally appropriate social protection systems and measures for all, including floors, and by 2030 achieve substantial coverage of the poor and the vulnerable.” Two main indicators for this target include:

- Proportion of population covered by social protection floors/systems, by sex, distinguishing children, unemployed persons, older persons, persons with disabilities, pregnant women, newborns, work-injury victims and the poor and the vulnerable
- Losses from natural disasters, by climate and non-climate-related events (including both lives lost and financial cost in US dollars)

Sustainable Development Goal 1.3, which proposes comprehensive social protection for vulnerable groups, is uniquely suited to the advantages of a holistic approach to evaluation. It is easy to imagine how a wide array of factors, such as technology, culture, infrastructure and macroeconomic trends, might affect the ability of countries to finance and organize social protection programmes, and how in turn such programmes would be vital to the UN’s overall goals of equitable and sustainable economic development and the eradication of extreme poverty. Furthermore, the various variables of a social protection programme, such as its eligibility requirements, benefit structure, and the mechanisms by which benefits are distributed, provide myriad ways to tailor a programme for the unique needs and challenges of a given nation or subnational area, and underline how interconnected such programs are with other policy factors.

Box 5.1. Progress towards the Sustainable Development Goal Target 1.3: Secretary-General's Report

United Nations General Assembly resolution 70/1 mandates the Secretary-General, in cooperation with the United Nations system, to submit a report on global progress towards the Sustainable Development Goals. The following bullet points, drawn from the report, provides a global overview of the current situation of Target 1.3 (social protection) for Goal 1 (“end poverty in all its forms everywhere”), on the basis of the latest available data for indicators in the global indicator framework.

- Social protection systems are fundamental to preventing and reducing poverty and inequality at every stage of people’s lives, through benefits for children, mothers with new-borns, persons with disabilities, older persons and those persons who are poor and without jobs. Preliminary data show that in 2016, only 45% of the world’s population was effectively protected by a social protection system and that coverage varied widely across countries and regions.
- In 2016, 68% of people above retirement age received a pension. However, that global average masks large regional differences. In Oceania, excluding Australia and New Zealand, and in sub-Saharan Africa, only 10% and 22%, respectively, of people above retirement age received a pension in 2016.
- Other vulnerable groups lack social protections as well. In 2016, only 28% of people with severe disabilities collected disability benefits, only 22% of unemployed individuals worldwide received unemployment benefits and only 41% of women giving birth received maternity benefits.
- Building the resilience of the poor and strengthening disaster risk reduction is a core development strategy for ending extreme poverty in the most afflicted countries. Economic losses from disasters are now reaching an average of USD 250 billion to USD 300 billion a year. Disaster risk globally is highly concentrated in low- and lower-middle-income countries. In relation to the size of their economies, small island developing States have borne a disproportionate impact.

Source: UNSD (2017_[30]), *The Sustainable Development Goals Report 2017*, United Nations Statistics Division, New York.

Global trends in climate change, for example, affect the severity and nature of natural disasters. A country’s educational system will affect not only whether potential beneficiaries are aware of available government programs, but also the degree to which members of vulnerable groups are able to use government assistance as stepping-stones to higher-pay jobs. Changes in culture might affect the family environment in which a child is raised, or the willingness of people to take advantage of available programs.

Infrastructure development could transform the ability of the government to reach potential beneficiaries or open them up to a wider geographic range of employment opportunities. It is clear that there are numerous potential ways that efforts to institute social protection programs are affected by wide-ranging factors, and nations can use the interactions they find to inform more effective and holistic policies, which in turn can be adapted to the changing contexts in which they are implemented.

Policy-makers must therefore look not only at simply the expenditure allocated to each programme and the overall coverage rate achieved, but rather study, in addition, a wide range of other data points that could point to interactions which could inform changes to policy structure and financing. For example, difficulties in expanding reliable coverage to certain rural areas might traditionally only present themselves as higher cost-per-capita, indicating a need to allocate more money to those areas. Holistic research could suggest much more efficient solutions, such as a recommendation of infrastructure investment. By viewing programs within the context of other government investments, policy-makers can conceivably construct evidence-based arguments for holistic budgetary approaches to the problem of social insurance, so that money can be spent wherever it will have the most impact towards the goal of complete coverage. It is beneficial to analyse spending in other areas as it affects social protection coverage, and to study how other economic and development indicators are affecting efforts to expand social protection. The discussion in the previous section provides a framework supporting this approach to monitoring and evaluating the SDGs.

6. Institutional arrangements for information and M&E systems

Appropriate monitoring and evaluation (M&E) systems are essential for ensuring the success of social protection programmes. Monitoring – the process of identifying and tracking performance indicators and reviewing the programme’s implementation at regular intervals over the lifetime of the programme – is critical as to provide accessible details on various facets of the programme. Programme evaluations are instrumental in assessing whether programmes achieve their core objectives and conducting a systematic assessment of the programme’s impacts. These together are critical to enable improvements in programme design and implementation, strengthen political will for expansion and sustainability and contribute to the global evidence base (Cecchini, Robles and Vargas, 2012_[3]). A higher quality of decision-making is the most prized outcome of M&E that is put in practice.

Box 6.1. South Africa’s institutional framework on M&E

South Africa’s Performance Monitoring and Evaluation unit in the Office of the Presidency coordinates impact assessments and other reviews nationally (most recently including a twenty year of social protection in South Africa). Additionally, South Africa is an example of a country where there have been frequent reviews and healthy debate, leading to an expansion of programme coverage, increases in benefit amounts and significant reforms in cases where the programme has not operated effectively. For example, an evaluation of the social grant system’s targeting process led to substantial reforms of the associated means test.

Source: Samson, Van Niekerk, and Mac Quene (2010_[13]), *Designing and Implementing Social Transfer Programmes*, Economic Policy Research Institute, Cape Town.

Institutional arrangements such as Management Information Systems (MIS) and integrated delivery systems including Single Window Services that encourage and enable large-scale and frequent reviews of the social protection programmes are critical to enabling improvements in programme design and implementation, strengthening political will for expansion and sustainability and contributing to the global evidence base (Cecchini, Robles and Vargas, 2012_[3]).

6.1. Management information systems

A decade ago, MIS were often paper-based, and payment mechanisms relied on physical delivery of cash at pay points – making delivery undependable and monitoring and evaluation challenging. Today, technology has transformed the social protection space. Integrated systems such as the Single Registries have been institutionalized in varying forms to improve the ease of delivering, monitoring and evaluating social protection programmes across the globe. Nonetheless, few countries have optimised the use of the gathered data for the development of efficient and effective social protection systems. In recent years, there has been positive development on this front in East Asia, with Indonesia and Cambodia investing in the development and utilisation of Single Registries to enhance their social protection systems.

Brazil pioneered the now common single registry, creating a central beneficiary database, containing *all* information of *all* the beneficiaries for *all* nation-wide social protection programmes called the *Cadastro Unico* (see Box 6.2).

Box 6.2. Brazil's Cadastro Único

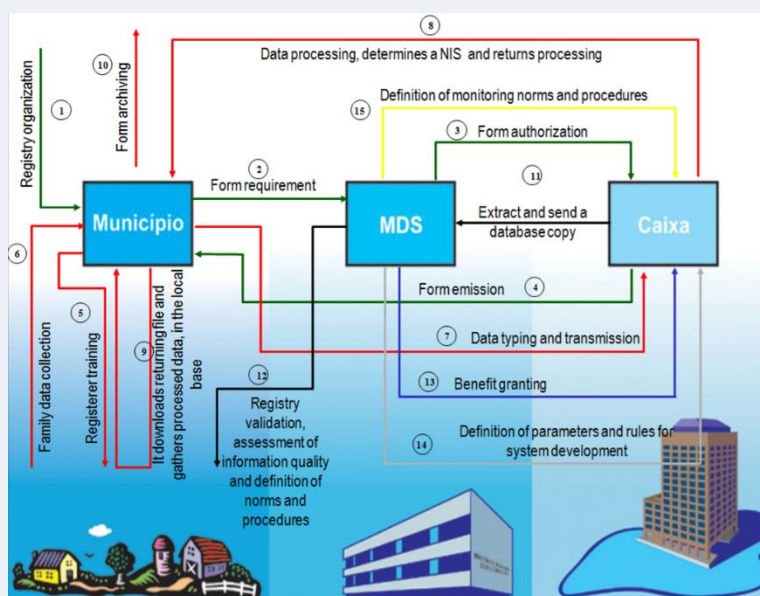
In 2001, the federal government of Brazil launched a major effort to construct a single beneficiary registry database and a system of unique social identification numbers (NIS) called the Cadastro Único. Data collection and beneficiary registry are still decentralised to the municipalities, but operation and maintenance of the database are centralised at the federal level, with oversight provided by the Ministry of Social Development, and system management and operation conducted by the Caixa Econômica Federal (CEF).

Brazil's Cadastro Único management information system (MIS), which supports one of the world's largest and longest running cash transfer programmes (Bolsa Familia), offers a particularly important role model for single registry systems globally. Brazil's public development bank Caixa Econômica Federal administers Cadastro Único, providing four important sets of benefits:

Reduced fiduciary risk: Housing the single registry within an autonomous public institution reduces the risks inherent with single registries, including data manipulation, fraud, and clientelism. The independent role of the public bank offers a system of checks-and-balances that reduces fiduciary risk while sharing the technical systems expertise that resides within the public banking system.

Linkages between payment services and social protection: Brazil's linking Cadastro Único to Caixa strengthens the synergies between payments and management information systems, enabling success delivery of 14 million monthly payments.

Figure 6.1. Cadastro Único



Facilitating municipal/national linkages: As the figure above illustrates, Brazil's Cadastro Único requires iterative processes between municipalities and the national stakeholders—the Ministry of Social Development and the Caixa. The federal bank provides the information infrastructure and management expertise to maintain national standards while enabling municipalities to customise the single registry to meet local requirements.

Developmental linkages: The Ministry of Social Development employs the Cadastro Único to construct an Index of Family Development, which can be used to identify complementary programmes to support human capital and livelihoods opportunities at the household, community and municipal levels. The index enables the ministry to track progress over time at aggregate and dis-aggregated geographic levels.

Source: Hellman and Gazola (2015^[31]), “How does Bolsa Familia work? Best Practices in the Implementation of Conditional Cash Transfer Programs in Latin America and the Caribbean”, IDB Technical Note; Inter-American Development Bank, Washington, DC; ILO (2014^[32]), *Cadastro Único: Operating a registry through a national public bank*, International Labour Organization, Geneva.

Despite its existence, Brazil failed to reap all the benefits of a single registry as (i) not all programmes in the country use the *Cadastro Unico*, and (ii) the data is primarily used at the municipal level and by the Ministry of Social Development for programme targeting and monitoring but at the federal level, the government still relies on statistical survey data, which is updated annually and considered more reliable.

In recent years, Indonesia has achieved great strides in developing their single registry – the Unified Database (UDB). As the country undertakes a large-scale transformation of the fragmented social protection system; the single registry electronic database containing the social, economic and demographic information of approximately 96 million individuals belonging to the lowest welfare bracket demonstrates remarkable progress. In the latest wave of transformations, Indonesia hopes to centralise the targeting, monitoring and evaluation of all social protection programmes through the UDB, as opposed to the current arrangement of separate mechanisms adopted by the various implementing ministries (Box 6.3).

Box 6.3. Indonesia' Unified Database

In 2011, the government began investing in the development of a Unified Database (UDB), known as the Basis Data Terpadu (BDT), as a single registry electronic database containing social, economic and demographic information on an estimated 24.5 million households belonging to the lowest welfare bracket using data from the Pendataan Program Perlindungan Sosial or Data Collection on Social Protection Programmes (PPLS 2011). PPLS 2011 recorded information for approximately 40% % of households across Indonesia belonging to the lowest socio-economic bracket, who were identified through poverty mapping the results of the 2010 population census, 2010 social economy survey (Survei Sosial Ekonomi, or SUSENAS) and village Potential (Potensi Desa, or PODES).

In addition, PPLS 2011 officials also employed the ratchet method and recorded information on other households suspected of being poor and excluded from the census, SUSENAS or PODES data based on information from other poor households through consultations and from direct observations. In February 2012, the results of PPLS 2011 were submitted by Statistics Indonesia (BPS) to The National Team for Accelerated Poverty Reduction (TNP2K) for the creation of the Unified Database. Household data in the Unified Database is listed by welfare rank using proxy means testing (PMT). PMT was used to estimate the socio-economic condition of households using data on household characteristics, such as the number of family members, education status, housing condition, asset ownership, etc.

Households listed in the Unified Database can be ranked and used for programme planning and identifying the names and addresses of prospective social assistance recipients, whether they target households, families or individuals. Ideally, the database should contain welfare information (including the names and addresses) of Indonesia's entire population; however, when the Unified Database was developed, the 40 % coverage was considered sufficient to meet the targeting needs of social protection and poverty alleviation programmes.

When all Indonesian households are grouped by wealth deciles (1 – poorest 10 % and 10 – richest 10%); the Unified Database contains information on households belonging to Decile 1, Decile 2, Decile 3 and Decile 4. Moreover, the UDB also identifies individuals by demographic groups (e.g. children, senior citizens, working aged people, widows, etc.), school children and those who do not go to school, individuals working in several pre-identified areas, disability status, status of home/residential ownership, drinking water source in households, and source of cooking fuel in households. The database is used for carrying out analysis or planning for poverty alleviation programmes to be used by government institutions and NGOs as well as targeting beneficiaries of social protection programmes.

Source: TNP2K (2018^[33]), *Data Terpadu PPFM*, Tim Nasional Percepatan Penanggulangan Kemiskinan, Government of Indonesia, Jakarta.

Box 6.4. Cambodia's IDPoor

Since its inception in 2005, the IDPoor programme has expanded to cover all rural households. Before that, Cambodia's efforts at establishing a concerted social protection information system were fragmented. Individual programmes, such as the Health Equity Fund (HEF), improving access to basic health care services, or labour market-related Cash for Work (CfW), Food for Work (FfW) and Technical Vocational Education and Training (TVET), have operated in Cambodia for over two decades. Their decentralised databases, all applying different systems for targeting, have stymied their impact and maintained costs at unnecessarily high levels.

Because of its transparent ten-step targeting, monitoring, evaluation and implementation protocol that includes comprehensive grievance and redressal mechanisms, IDPoor has become a widely trusted system throughout rural Cambodia. The rigorous multi-dimensional targeting mechanism and the willingness of local communities to provide feedback, raise complaints and regularly update the platform provides Cambodia's government with a centralised targeting database.

The targeting process includes interviews with scoring and non-scoring sections, whose responses are in turn validated by a Village Representative Group (VRG). Following assessment, the VRG displays a list of all eligible households in the village, which passes through a consultative verification meeting that involves a community-based feedback loop. A commune council subsequently revises the list and endorses its results upon which villagers can lodge complaints. The reviewed and, if necessary amended, application returns to the VRG for final verification. The VRG enters the list into the IDPoor database. Ultimately the selected households receive a so-called Equity Card. The database is updated every three years.

Targeting therefore includes comparatively expensive proxy means testing. However, considering IDPoor's strong performance regarding accurate identification, keeping the cost of inclusion and exclusion errors in monetary terms minimal, the system displays strong value-for-money. The distribution of Equity Cards that allow poor individuals to access social assistance programmes and facilitate efficient monitoring.

Materialised in the Equity Cards, the ability to utilise IDPoor as a proxy for poverty yields significant reductions in operational costs by reducing future targeting expenditures. A database that updates regularly, providing accurate information from ministerial to village levels through robust monitoring, evaluation and grievance processes represents a remarkable step forward for Cambodia.

The government of Cambodia has now implemented the Urban IDPoor programme to create a centralised M&E system. Through the expansion of the IDPoor to urban areas and enhancing the frequency of database updates, the Cambodian government envisages that the index will gradually transition from a social registry into a single registry which will optimise the efficiency of social protection programmes.

6.2. Integrated Delivery Systems

Service delivery mechanisms have also evolved in ways that can support and improve monitoring and evaluation. Single Window Services (SWS), for example, integrate multiple services – from assessing potential beneficiary eligibility, enrolling beneficiaries into programmes and providing services – to facilitate transparency and overcome the difficulty of gathering critical data from fragmented sources. Such systems enable information sharing horizontally within government institutions and ministries as well as vertically by strengthening sub-national, local authorities.

Single Window Services (SWS) programmes offer a “one-stop shop” for the coordination and delivery of social protection programmes and labour-market initiatives, better enabling access for poor households and contributing to the extension of programme coverage. “Single windows” facilitate transparency and information sharing across various government institutions, fostering horizontal and vertical integration by strengthening sub-national authorities. A “one-stop shop” embedded in the government institutions that deliver multiple services from information dissemination to enrolment facilitation to the provision of services can realise delivery and cost efficiency gains for providers and enhance programme uptake by simplifying the process for the beneficiaries.

Box 6.5. Single Window Services

As a response to the challenge of providing quality social services, Mongolia implemented One-Stop-Shops (OSSs) nation-wide in 2013. OSSs deliver social protection programmes, employment counselling services, and notary and banking services at the provincial and district levels. Mongolia's OSSs group officers from different social protection schemes, employment programmes, and public and private services into one room for better coordination and monitoring. Over 60% of Mongolia's population uses an OSS regularly with a satisfaction rate of 85% for the services. In 2011, 31 established OSSs served more than 1.8 million customers, including over 1.2 million customers in rural areas.

GIZ, on behalf of the German Ministry for Economic Cooperation and Development, implemented a pilot establishing 250 Worker Facilitation Centres (WFC) in 6 districts in Karnataka in India. Designed as a single window service, these WFCs promote awareness, assist with enrolment and employ a management information system to monitor the progress of the programme. The system's database of unorganised workers archives the required documents that facilitate future programme applications. An impact evaluation documented a 13% improvement in awareness of government social security schemes as well as a 15% increase in programme take-up in villages with WFCs. Responding to this evidence, the Government of Karnataka is extending single window services to the remaining 23 districts.

Source: Ekben (2014^[34]), "Single Window Services in Social Protection: rationale and design features in developing country contexts", Discussion Papers on Social Protection, GIZ.

The ease of gathering information that integrated delivery systems allow for policy makers and programme officials to better monitor and evaluate the efficiency and effectiveness of programmes (GIZ, 2016^[35]). These delivery systems not only enable better monitoring by helping the creation of a database but can help realise better efficiency gains and translate to better programme outcomes.

Single-registry management information systems and integrated payments systems create opportunities to move beyond just delivering the benefit and help develop the beneficiaries' capabilities, particularly through the associated access to financial, information and communications services. This progress has required a supporting institutional framework that increasingly supports cross-sectoral co-operation.

A single registry system can help to build support and legitimacy for a given programme. However, the adoption of a single registry must be done with some key considerations such as:

- Clear objectives.

- Regular system updates ensuring dynamism and allowing for demographic and geographic updates and changes in welfare statuses of households
- Auditing and quality control systems for the registry and other management information systems (MIS)
- Access to MIS data for key potential users (programmers or regional government)
- Reliable and user-friendly MIS software systems.
- External help for capacity-building from banks, private consultancy firms or foreign ministerial bodies to cultivate local technical expertise.

7. Building national capacity for M&E

The appropriate design, effective implementation and robust financial management of national social protection policies, programmes and systems require strengthening the government's institutional capacity at national and local levels. This capacity includes the institutions, human resources, systems and other public resources that support the delivery of social protection benefits and services.

Policymakers also require technical support in making decisions at a policy level. Understanding the advantages and disadvantages of cash transfers, public works, school feeding, contributory programmes and other types of social protection interventions informs better policy decisions – a prudence that is particularly important when fiscal reforms require difficult choices.

To achieve programme success, policymakers require the necessary capacity to gather and analyse data to make evidence-based policy decisions; the capacity to design and implement programmes at a national level and to deliver these at the local level.

In addition, governments may require support for the inter-ministerial co-operation often required to build appropriate social protection systems. A number of important topics require capacity development support at both national and sub-national levels, including the strengthening of national systems for targeting, delivery mechanisms (particularly payments), fiduciary risk management and monitoring and evaluation (M&E).

Development partners also often support systems for managing fiduciary risk and effective M&E. Increasingly, rigorous evaluations constitute a global public good, contributing to an international evidence base supporting good practices.

7.1. Key design principles

Four key design principles support the foundation for sustaining the capacity development strategy:

- Demand-driven
- Government and national ownership
- Integrated and comprehensive
- Reflecting the national social and policy context

The capacity development requirements also involve four key characteristics of appropriate and effective capacity development:

- **Policy relevant:** The capacity development must address the key policy and skills areas and answer the questions directly related to the critical issues the government faces.
- **Timely:** The capacity development interventions must be available to government in a timeframe that enables the required policy decisions and programme activities.
- **Credible:** The content delivered must be appropriate and rigorous, providing objective information that meets global standards.
- **Accessible:** Capacity development should be available to all those who require it, and the interventions should be delivered in a broadly accessible style consistent with adult learning requirements.

7.1.1. Demand-driven

A demand-driven approach communicates to participants the vital role of social protection policies in not only cost-effectively tackling poverty, vulnerability and social exclusion but also generating economic returns by promoting livelihoods, employment and meaningful developmental impacts. Both of these dimensions contribute to successful social protection public finance management by expanding budgetary resources and ensuring strong economic returns.

7.1.2. Government and national ownership

Government support and ownership of the capacity-building strategies enables greater sustainability and effectiveness. Key government mechanisms will guide capacity development decisions and resource allocations with support from the key partners. While at times international technical assistance combined with local expertise will be paired with national counterparts in order to ensure a transfer of skills to ensure a national cadre of social protection experts can sustain a national capacity development process, international service providers will always work closely with national counterparts to ensure the development of national capacity for world-class social protection, ensuring the strengthening of a national cadre of professionals who can both support capacity development and the required technical assistance to government otherwise sourced internationally.

7.1.3. Integrated and comprehensive

A national capacity development effort strengthening the thematic areas linked with social protection requires an integrated and comprehensive approach in order to ensure efficiency and value for money. This integrated approach strengthens input cost-effectiveness and enhances the value of the training by fostering collegial interactions across the wide range of active social protection ministries and other stakeholders.

The approach incorporates a comprehensive design that builds linkages across sectors, in order to strengthen fiscal efficiencies and reinforce developmental impacts in managing social protection programmes and projects.

7.1.4. Reflecting the national social and policy context

While the capacity building strategy draws on rich lessons of global experience, it recognises the critical importance of reflecting the national social and policy contexts. Global experts and best practices can contribute valuable skills and insights, but assuring relevance to unique circumstances and policy priorities requires a reliance on the country's robust evidence base on social protection and national expertise. In practice, this approach focuses on peer learning mechanisms which marshal the active involvement of capacity development programme participants. The strategy also incorporates mechanisms to strengthen networks for ongoing learning and evidence sharing, built on a "virtual platform" that harnesses electronic communications and social media to deepen capacity development impacts and build a community of practice.

The peer learning approach leverages the lessons the participants bring to the strategic activities and facilitates the development of a social protection public finance management network that nurtures ongoing capacity development. The resulting e-forum will provide a constructive opportunity for sharing experiences and articulating a unified vision of developmental social protection that delivers value-for-money in a fiscally sustainable manner. The format of the strategic activities promotes meaningful and continuing dialogue, which can reinforce political will and increase the likelihood of successful implementation of these programmes within the context of a broader development planning framework.

Box 7.1. South-south examples of social protection capacity building: Brazil’s and South Africa’s technical assistance to social protection systems in Africa

Brazil’s social policy development support to a number of countries influenced policy development in Africa. Brazil and South Africa provided technical assistance in the design of Ghana’s Livelihoods Empowerment Against Poverty (LEAP) programme, which links a cash transfer with access to social health insurance benefits, providing more comprehensive social protection and strengthening the impacts of both interventions. Technical experts from both countries supported a nationally led team designing the system, providing support to build targeting mechanisms, delivery vehicles, management information systems and a monitoring and evaluation framework.

Similar teams from Brazil and South Africa supported Mozambique’s expansion of its social protection system. Mozambique developed its own Bolsa Familia pilot, influenced by the Brazilian namesake and supported by the International Labour Office (ILO). Technical experts from Brazil, South Africa and other countries developed social protection capacity building materials and offered courses in Mozambique, supporting the integration of inter-ministerial initiatives promoting livelihoods and employment by “considering broader macro-economic areas for social investments to raise overall living standards (such as in agriculture, food security and employment-generating activities).” Experts from these countries (and others) also provided technical assistance in the design and implementation of social protection systems.

Source: OECD (2013^[24]), *Agricultural Policy Monitoring and Evaluation*, OECD Publishing, Paris; UNICEF (2010^[36]), *Child poverty and disparities*, UNICEF, Maputo.

8. Conclusions and recommendations

The more robust evidence provided by randomised control trials has built a better global evidence base supporting the design and implementation of effective social protection systems and their integration with other policy sectors. Ironically, the increasing focus on national systems approaches has begun to limit the incremental usefulness of experimental methodologies—one cannot randomise countries. As policy-makers demand more relevant evidence that supports more comprehensive and integrated policy innovations, the limits of project- and programme-oriented evaluation become more apparent.

Experimental and quasi-experimental quantitative methodologies mainly aim to rigorously evaluate core impacts. The methodologies offer powerful tools for rigorously, precisely and credibly attributing outcomes and proving causality, particularly for simple interventions that easily satisfy the criteria for the evaluation approaches. As problems become more complex and the options for effective interventions become murkier, the effectiveness of counter-factual based approaches diminishes. They cannot evaluate an intervention that cannot be precisely defined. Today, policy-makers require comprehensive monitoring and evaluation (M&E) systems that better tackle complex questions and provide evidence on the impact of integrated policy responses.

On the supply side, an M&E system clearly articulates the institutional arrangements and data sources needed to facilitate a reliable flow of information. Indicators and targets are formulated and regularly collected to monitor programmes, their synergies, and their gaps at various stages of the process, ranging between inputs, activities, outputs, and impacts.

On the demand side, the system responds to the main questions policy-makers are asking. The intended users of the system – high-level policy-makers, programme managers, front-line workers – have access to useful information and understand how the system provides them with the evidence they require, which aligns incentives and supports the provision of more reliable high-quality data. Ultimately, this fosters a national policy environment and institutional culture where lessons-learned are part of the conversation and data is used to answer questions (Attah et al., 2015^[37]).

Such comprehensive and integrated evaluation approaches strengthen a complex evidence base that works better for national social protection systems—and for strategies that aim to achieve the Sustainable Development Goals. Comprehensive evaluation can:

- Evaluate the interactions and linkages among different actors, institutions and policies.
- Map both the direct and indirect as well as the static and dynamic pathways that may compound or weaken an intervention’s impacts.
- Assess how the social, economic, cultural and political context influences the comprehensive set of results (including impacts) of the intervention.
- Respond specifically to the key questions policy-makers are asking, and for which answers are required in order to scale up the interventions.
- Recognise that the impact of one sector’s intervention on a specific outcome depends on the inter-related investments across a range of policy sectors.
- Integrate the optimal mix of inter-related interventions in order to achieve a range of joint outcomes.

Progress towards more integrated and comprehensive evaluation requires innovations at both macro and micro levels. The macro evaluation framework requires tools that link multiple inter-related interventions to cross-sectoral outcomes, and can assess both “intra-sectoral” and “inter-sectoral” linkages. Innovations in estimating policy production functions offer promise in identifying synergies that multiply impact and maximise value-for-money. Similarly, the micro evaluation framework requires tools that go further than proving causality. More flexible evidence-building approaches are required that harness learning-by-doing mechanisms to identify and evaluate appropriate and effective combinations of interventions. Complex problems require tools that not only test hypotheses but also evolve hypotheses, contributing to the design of effective interventions.

Some of the world’s most successful social protection programmes—Mexico’s Prospera (formerly Progres a and Oportunidades), South Africa’s Child Support Grant, Bangladesh’s Challenging the Frontiers of Poverty Reduction, and many others—originated as programme failures. These failures led to outstanding success because the responsible policy entrepreneurs intrepidly harnessed the resulting lessons and channelled them into policy improvement. They understood that tackling a complex problem necessarily risks failure, which is at worst a stepping stone to success-achieving innovation.

The micro- and macro-evaluation frameworks mutually reinforce each other. Innovative evidence-building approaches inform policy-makers how comprehensive and integrated interventions interact to strengthen cross-sectoral policy outcomes. The macro framework consolidates the results from micro assessments into a planning framework which describes how investments across sectors create developmental synergies. Integrated and comprehensive systems evaluation offers numerous innovative benefits:

- They measure how investments in one sector depend on the related investments in other sectors, better enabling policy-makers to choose optimal combinations of interventions to optimise impact, efficiency and value-for-money.
- They enable policy designers and implementers to build a complex set of interventions that is not knowable in advance and that will cost-effectively achieve a range of diverse but critical policy outcomes, maximising the likelihood of success while providing a cost-effective solution.
- They facilitate a full benefit analysis across a spectrum of policy sectors, providing a more complete and accurate evaluation, and as a result expanding the range of stakeholders who value the joint intervention, thereby strengthening political support for the necessary reforms and innovations.

These comprehensive M&E approaches are particularly important for development partners. The evidence from comprehensive M&E frameworks represents a global public good. National stakeholders optimise country-level investment decisions. Systems approaches are new, opaque, expensive and risky. In the absence of a global framework, they will suffer under-resourcing, because national decisions usually will not reflect global spill-over benefits. The Sustainable Development Goals offer the kind of global coordination mechanism that can drive the development of comprehensive M&E systems.

References

- 3ie and IFPRI (2012), *Seminar Series on Impact Evaluation : Cash Transfer, Behavioral Changes, and Cognitive Development in Early Childhood*, <http://www.3ieimpact.org/en/events/3ie-ifpri-washington-seminars/> (accessed on 4 December 2018). [27]
- Angrist, J. et al. (2002), “Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment”, *American Economic Review*, Vol. 92/5, pp. 1535-1558. [15]
- Attah, R. et al. (2015), *How to move beyond the impact evaluation trap? Challenges and solutions for the setting up of comprehensive M&E systems for Social Protection Programmes*, Oxford Policy Management, Oxford. [37]
- Banerjee, A. et al. (2017), “From Proof of Concept to Scalable Policies: Challenges and Solutions, with an Application”, *Journal of Economic Perspectives*, Vol. 31/4, pp. 73-102. [9]
- Banerjee, A. and E. Duflo (2011), *Poor Economics: A radical rethinking of the way to fight global poverty*, Public Affairs, New York. [10]
- Behrmann, J. and J. Hoddinott (2005), “Programme Evaluation with Unobserved Heterogeneity and Selective Implementation: The Mexican ‘PROGRESA’ Impact on Child Nutrition”, *Oxford Bulletin of Economics and Statistics*, pp. 547-569. [14]
- Cecchini, S., C. Robles and H. Vargas (2012), *The Expansion of Cash Transfers in Chile and its Challenges: Ethical Family Income*, International Policy Centre for Inclusive Growth, Brasilia. [3]
- Chambers, R. (1995), “Poverty and livelihoods: Whose reality counts?”, *IDS Discussion Paper*, No. 347, Institute of Development Studies, Brighton. [20]
- Deaton, A. and N. Cartwright (2017), “Understanding and misunderstanding randomized controlled trials”, *Social Science & Medicine*. [6]
- Devereux, S. and R. Sabates-Wheeler (2004), “Transformative Social Protection”, *IDS Working Paper* 232, https://www.ids.ac.uk/files/IF1_3.pdf. [2]

- DSD/SASSA/UNICEF (2012), *The South African Child Support Grant Impact Assessment: Evidence from a survey of children*, DSD, SASSA & UNICEF. [17]
- Ekben, C. (2014), “Single Window Services in Social Protection: rationale and design features in developing country contexts”, *Discussion Papers on Social Protection*, No. 19. [34]
- Garbarino, S. and J. Holland (2009), “Quantitative and Qualitative Methods in Impact Evaluation and Measuring Results”, *Issues Paper*, Social Development Direct, London. [19]
- GIZ (2016), *Integrated Public Service Delivery for Social Protection International Experiences with Single Window Services (SWS)*, <https://www.giz.de/expertise/downloads/giz2017-en-integrated-public-service-delivery-study.pdf> (accessed on 19 November 2018). [35]
- Glass, S., R. Gajwani and F. Halliday (2016), “Does Quantitative Research in Child Maltreatment Tell the Whole Story? The Need for Mixed-Methods Approaches to Explore the Effects of Maltreatment in Infancy”, *The Scientific World Journal*. [16]
- Hellman, A. and A. Gazola (2015), “How does Bolsa Familia work? Best Practices in the Implementation of Conditional Cash Transfer Programs in Latin America and the Caribbean”, *IDB Technical Note*, No. 856, Inter-American Development Bank. [31]
- IIED (2016), “Evaluation: a crucial ingredient for SDG success”, *Governance, Policy and planning*, IIED, <http://pubs.iied.org/17357IIED> (accessed on 4 December 2018). [28]
- IIED (2016), *Five considerations for national evaluation agendas informed by the SDGs*, <http://dx.doi.org/10.3138/cjpe.30.3.02./11>. [29]
- ILO (2014), *Cadastró Único: Operating a registry through a national public bank*, ILO, Geneva. [32]
- Khandker, S., G. Koolwal and H. Samad (2010), *Handbook on Impact Evaluation: Quantitative Methods and Practices*, World Bank, Washington. [12]
- Mideros, A., F. Gassmann and P. Mohnen (2013), *Estimation of rates of return on social protection: Making the case for non-contributory social transfers in Cambodia*, United Nations University, Maastricht. [25]
- Ministry of Planning (n.d.), *IDPoor*, <http://www.idpoor.gov.kh/en/home> (accessed on 28 May 2018). [23]
- Morse, J. (1990), *Qualitative Nursing Research: A contemporary dialog*, Sage Publications. [18]
- OECD (2013), *Agricultural Policy Monitoring and Evaluation*, OECD Publishing, Paris. [24]
- OECD (2007), *Employment, International Migration, Social Policies and Health*, OECD Publishing, Paris. [38]

- Peters, J., J. Langbein and G. Roberts (2018), *Generalization in the Tropics - Development Policy, Randomized Controlled Trials, and External Validity*, Oxford University Press, Oxford. [8]
- Roelen, K. et al. (2017), *How to make 'Cash Plus' work: Linking cash transfers to services and sectors*, UNICEF Office of Research, Florence. [7]
- Samson, M. (2018), *Financing the child-centred SDGs in Ethiopia*, UNICEF Ethiopia, Addis Ababa. [26]
- Samson, M. et al. (2016), "The Impact of a Promise Realized: South Africa's Child Support grant", in FAO and UNICEF (ed.), *From Evidence to Action: The Story of Cash Transfers and Impact Evaluation in Sub-Saharan Africa*, Oxford University Press, <http://www.fao.org/3/a-i5157e.pdf> (accessed on 4 December 2018). [22]
- Samson, M. and N. Taylor (2015), "Supporting Social Protection Systems", *Tools and Methods Series*, No. Concept Paper Nr. 4, European Commission, Luxembourg. [1]
- Samson, M., I. Van Niekerk and K. Mac Quene (2010), *Designing and Implementing Social Transfer Programmes*, Economic Policy Research Institute, Cape Town. [13]
- Taylor, E. (2013), *A Methodology for Local Economy-wide Impact Evaluation (LEWIE) of Cash Transfers*, FAO, Rome. [11]
- TNP2K (2018), *Data Terpadu PPFM*, Tim Nasional Percepatan Penanggulangan Kemiskinan, <http://bdt.tnp2k.go.id/> (accessed on 19 November 2018). [33]
- UNDP (2009), *Handbook on Planning, Monitoring and Evaluating for Development Results*, UNDP, New York. [5]
- UNICEF (2010), *Child poverty and disparities*, UNICEF, Maputo. [36]
- UNSD (2017), *The Sustainable Development Goals Report 2017*, United Nations Statistics Division, New York. [30]
- Valle, A. (2016), "The Mexican experience in monitoring and evaluation of public policies addressing social determinants of health", *Global health action*, Vol. 9/1. [4]
- White, H. (2008), *Making impact evaluation matter. International Initiative for Impact Evaluation*, International Initiative for Impact Evaluation, New Delhi. [21]

For more information about the EU Social Protection Systems Programme:

<https://oe.cd/social-protection>