



**Report on
Workshop on Measurement of Non-market Output in Education and
Health, London, 3-5 October 2006**

Workshop jointly organised by the OECD, the United Kingdom Office for National Statistics
and the Government of Norway

**Peter C Smith and Andrew Street
Centre for Health Economics
University of York**

Acknowledgements

The authors were commissioned by the Workshop organisers to produce an independent report on the proceedings. Whilst we have benefited greatly from the numerous comments of participants, this is our report and it reflects our personal views.

Measurement of Non-market Output in Education and Health

Contents

Summary	3
Introduction.....	7
Moving from Output=Input to Direct Volume Measurement.....	8
Progress in implementing Direct Volume Measurement.....	10
Health	13
Education.....	19
General themes emerging from the Workshop	24
Conclusions and priorities for the future.....	27
References	30

Summary

In recent years considerable progress has been made in developing improved methodologies to measure non-market output in the National Accounts. Most EU Member States have supported the introduction of a legal framework to implement these methodologies and have introduced current best practice methods to measure output of health and education services. This report summarises contributions at a Workshop held in October 2006 that focussed on building on this foundation and further improving the measurement of non-market output in the National Accounts. The Workshop supports a project intended to provide detailed international guidelines for the further development of volume measures of non-market outputs, in particular for education and health.

The Workshop had three main objectives:

1. to improve *temporal analysis*: by moving from input to direct volume measurement in the health and education sectors;
2. to improve *international comparisons*: by standardising international definitions and processes, and developing Purchasing Power Parities (PPPs), to allow for better cross-country comparisons both of the *levels* and *growth rates* of final consumption in the health and education sectors;
3. to take account not only of the quantity of output (volume of services) but also of the *quality* of output.

There are three fundamental challenges in applying the Direct Volume Measurement approach to non-market services:

1. The output of the non-market sector is often difficult to describe or measure.
2. It is difficult to measure the quality of non-market services.
3. Some means of weighting different services is required in order to aggregate them into a single output index.

For the health service, the Workshop indicated the current state of play on the measurement of health outputs:

- While Eurostat defines health care output in terms of ‘completed treatments’, it is difficult to measure these for the majority of patients. For many patients, this is because there may be no obvious treatment ‘endpoint’. For others, computer systems are unable to track patients that require care across multiple settings.
- Measurement of health care activities has become more comprehensive and disaggregated, particularly as more countries have adopted DRG classification systems.
- There is increasing standardisation across countries in many activity categories, partly in the hospital sector, but there remain considerable challenges in counting non-hospital services.
- Sometimes ‘activities’ are really inputs – most notably the number of prescriptions provided. This raises the question of whether more activities are better.
- Health outputs often reflect contributions made by other sectors, and there is a question of where to attribute these efforts. This is most evident for pharmaceutical inputs and for those provided by the social care sector.
- There are challenges with incorporating new products/technologies in the National Accounts, with tension between reflecting changes in practice and ensuring consistency over time. This is a common problem across economic sectors.
- Costs are seen as the most convenient means to weight different activities, particularly because of their availability. However, cost weights imply the presence of allocative efficiency and constant returns to scale. If these conditions do not hold, costs will be a poor reflection of societal value.
- Taking account of the quality of health services is a major challenge. To meet this, there needs to be debate about what health care is for, how to measure the dimensions of quality, how to isolate the specific contribution of the health sector from other influences on the quality domain, and how to attach weights to each domain.

In the education sector, consideration of output measurement yielded the following insights:

- There is a lack of consensus about the primary purpose of education, with suggestions including equipping recipients with qualifications, making them rounded citizens, or enhancing their earning potential. This lack of consensus makes it difficult to define education ‘outcomes’.
- Output measurement in education is more straightforward than in health, because students are reasonably homogenous within each educational stratum and because they can be tracked over time. That said, the output of higher education institutions is not confined to teaching activities, and their research and other undertakings need to be incorporated into their measure of output.
- Quality adjustment in education is challenging, and countries are adopting different methods. Some make adjustment for class size, recognising that countries with smaller classes appear ‘less productive’ unless the enhanced student experience is accounted for. However, it is not a simple matter to put a value on this experience.
- Unlike in the health sector, it is possible to measure the ‘value added’ of education quite securely by calculating the difference test results at entry and at points along the educational pathway, offering considerable scope for isolating the specific contribution of the sector to levels of educational attainment.
- International comparison of educational quality requires a standardised basis of evaluation. The PISA exercise provides a means to this end, by comparing students in terms of core competencies. The exercise also allows analysis of variation in attainment among students, which can provide an indication of fairness in delivering education among the population. As PISA is repeated, it will be possible to assess changes in educational attainment over time.

In conclusion, we recommend:

- Active engagement of all relevant stakeholders in taking this project forward;
- That the project should continue to develop implementation guidelines and provide a key resource for sharing of best international practice;
- In health, as a long term ambition, that the national accounts should recognise the key purpose of health care as being to enhance health status. Efforts should

be made to measure ‘complete treatments’ and the health gain arising from treatment;

- While recognising costs are convenient means to aggregate different activities, that the project should examine the scope for developing value weights;
- That the project considers how to measure and incorporate quality changes in indices of health output;
- In education, the lack of clarity about primary educational outcomes makes it difficult to establish value weights for different types of output. But it may be that costs are a reasonable reflection of value in this sector. The project may seek to confirm this contention;
- Standardised methods of quality adjustment need to be developed, because existing diversity in national practice does not facilitate international comparisons;
- The PISA exercise is a fruitful basis for future development, offering the promise of standardised international comparisons, assessment of the fairness of each country’s education sector, and changes in educational achievement over time. The project should determine how to make best use of the PISA exercise in the national accounts;
- More generally, the project should explore the scope for measuring levels as well as growth in outputs, thereby improving the basis for inter-country comparability of national accounts.

Introduction

Most governments have an extensive role in the provision and financing of education and health services. These account for a significant part of Gross Domestic Product (GDP) so it is important that their output is measured accurately, not least so that governments can allocate expenditure on an informed basis (Aldin, 2006). Governments also recognise the importance of demonstrating to taxpayers that government expenditure on these services represents value for money.

In October 2006 a Workshop was organised jointly by the OECD, the United Kingdom Office for National Statistics and the Norwegian Government to launch a project for improving the measurement of non-market output in the National Accounts. The project is intended to provide detailed international guidelines for the development of volume measures of non-market output, in particular for education and health.

The Workshop had three main objectives (Konijn *et al.*, 2006; Lequiller, 2006):

1. to improve *temporal analysis*: by moving from input to output or outcome measures in the health and education sectors and, thereby, improve the measurement of non-market output growth and productivity over time;
2. to improve *international comparisons*: by standardising international definitions and processes, and developing Purchasing Power Parities (PPPs), to allow for better cross-country comparisons both of the *levels* and *growth rates* of final consumption in the health and education sectors;
3. to take account not only of the quantity of output but also of the *quality* of that output (the notion of the “volume” of services embraces both concepts).

This report summarises the main issues arising from the Workshop. It first outlines the key conceptual achievements and challenges involved in developing direct volume measures and accounting for quality that were reported at the Workshop. It then summarises the specific issues arising from experience of measuring output in the health and education sectors. Finally, the report reviews the key themes emerging from the Workshop and highlights what we see to be the most important issues to be tackled in future work.

The intention of this report is to highlight achievements, interesting new departures, key challenges, and priorities for the future. It must be emphasised that this report represents only the views of the authors, and is not intended as an official report of the Workshop.

Moving from Output=Input to Direct Volume Measurement

The motivation for the Workshop is the strong international drive towards replacing the traditional treatment of non-market output in the National Accounts. From the early 1960s to the mid-1990s, the output of the public sector in all countries was valued simply by adding up expenditure on inputs, an approach termed the ‘output=input’ convention. The attraction of this approach is that it avoids the need both to measure and to place valuations on non-market outputs. But there are three main drawbacks to this convention:

1. It is circular and self-justifying. The value of output is however much the government chooses to spend on producing or purchasing it. By definition, the higher the level of spending, the better. Taxpayers may disagree.
2. It implies no change in productivity over time, as ‘outputs’ are not measured directly;
3. Reductions in expenditure brought about by technological improvements appear to reduce ‘output’, when in reality only inputs might have been reduced.

The inadequacy of the output=input approach led to recommendations from international bodies such as the United Nations and Eurostat for the development of measures of non-market output using methods that are independent of expenditure on inputs (Atkinson, 2006; Eurostat, 1995). Both the System of National Accounts (SNA93) (United Nations Statistics Division, 1993) and European System of Accounts (ESA95) (Eurostat, 1995) include recommendations to move toward Direct Volume Measurement (DVM) of non-market outputs for many services, including health and education. So, in education for instance, instead of reporting teachers’ salaries, the accounts should measure how many pupils were taught. The output=input

approach remains the recommended method for measuring collective services, such as defence and public order, which have classical ‘public good’ characteristics (Eurostat, 2001). Even so, preliminary attempts are being made to identify outputs for some collective services (Cullinane, 2006; Heikkinen *et al.*, 2006).

All member states of the European Union (with the exception of Denmark, which has secured a postponement) are legally required to implement the DVM approach in time for the 2006 National Accounts (Atkinson, 2006). Considerable progress has been made by member states in meeting these requirements, and these achievements are summarised in box 1 below

Box 1: Progress in implementing Direct Volume Measurement

- The publication by Eurostat of a methodological Manual on Prices and Volumes to provide guidance to Member States
- The support by most EU Member States for the introduction of a legal framework requiring each Member State to measure non-market outputs (as opposed to inputs) by 2006.
- The introduction by most EU countries who have reported of output methods for much of health services; about half of these countries are using methods classified by Eurostat as ‘best practice’. There are examples of countries outside EU doing the same.
- The introduction by nearly all EU countries who have reported of output methods for non-market education services; nearly all of these countries are using methods classified by Eurostat as ‘best practice’. There are examples of countries outside EU doing the same.
- The Workshop papers showed that many countries have been working seriously on developing new methodologies that go well beyond those formally reported to date.

Beyond the EU, the recommendations to move to a DVM approach are not being applied universally. In particular, the US and Canada are likely to continue to use the output=input convention for the foreseeable future (Christian, 2006c). Changing the accounting basis is likely to be politically sensitive, particularly in contexts where input based measures suggest higher levels of ‘output’ growth than alternative volume measures. For example, this appears to be the case for the US and Norwegian educational sectors (Christian, 2006b; Dam *et al.*, 2006) and is likely to hold when the public sector is enjoying a period of expenditure growth because the returns on this expenditure may not be realised immediately.

Direct Volume Measurement avoids many of the deficiencies of the output=input convention and has already secured rapid improvement in the usefulness of the National Accounts. However, there are three fundamental challenges in applying the Direct Volume Measurement approach:

1. The *output* of the public sector is often difficult to describe or measure.
2. It is difficult to measure the *quality* of public sector output.
3. Some means of *weighting* different goods and services is required in order to aggregate them into a single output index.

For goods and services exchanged in the private sector, market prices provide an indication of their relative value to consumers. But such prices do not exist for non-market outputs so some other means must be adopted to assess their relative value. The absence of prices that reflect the true marginal social value is taken to be the fundamental defining feature of the ‘non-market’ sector. Moreover, the role of government in the provision and financing of health and education services varies across countries and over time, so that in most contexts there is a blurred boundary between the public and private sectors (Quintela, 2006). Direct volume measurement seeks to measure volumes of non-market outputs in the National Accounts in an analogous fashion to that employed for the traded sector. This implies the need to infer values for each of the services under scrutiny. These are required so that the different non-market services produced can be aggregated with each other and also aggregated with market services.

Progress in implementing Direct Volume Measurement

Considerable progress has also been made in many countries across Europe seeking to implement the European Union National Accounts directive, particularly in health care and education (Malherbe *et al.*, 2006a, 2006b). Key definitional terms for the health and education sectors are reproduced from the Eurostat handbook in box 2 below (Eurostat, 2001). This reveals differences in applying the definitional concepts to the two sectors.

Box 2: Key definitional terms

	Health	Education
Input	what the health system uses in order to provide its output	what the education sector uses in order to provide its output
Output	the quantity of health care received by patients, in terms of complete treatments, adjusted to allow for the qualities of the services provided	the quantity of teaching received by students, adjusted to allow for the qualities of the services provided
Activities	the individual actions carried out by the health sector in delivering a completed treatment	n/a
Outcome	the change in health status due to health sector interventions	lack of consensus over what constitute educational outcomes

There is a shared definition of *inputs* between the two sectors. Although not an explicit focus of the Workshop, comments were made about the need to improve input measures. Issues that were raised in discussion include:

- Accurate measurement of capital inputs;
- How best to take account of differences the *quality* of inputs, which requires some means of assessing how quality changes of the inputs are related to quality changes of the output (Brathaug, 2006);
- Appropriate apportionment of shared inputs, when contributions to output are made by institutions that are not defined as part of the health or education sectors. This is particularly challenging where the institutional divide may vary across countries. For instance, the provision of long-term care for elderly people may be the primary responsibility of the health sector in some countries but of the social care sector in others (Huber, 2006);
- How best to take account of the contribution to current output of inputs utilised in previous periods. The inputs to education, in particular, are provided over long periods of time, whereas output may be measured only at a single point of time.
- In health care in particular, how to treat improvements in technologies, such as pharmaceuticals, that are derived from developments in other (market) sectors. Moreover, it is not obvious whether the contributions to output arising from the use of pharmaceuticals should be attributed to the health sector or to the pharmaceutical sector, which is treated separately in the National Accounts.

Measuring the quantity of educational *output* is less challenging than it is in the health sector, even though Eurostat's broad definitions of 'output' are similar. In education, there is usually accurate information about the number of pupils taught at each stage of their education and, often, it is possible to track the educational attainment of individual pupils over time. Furthermore, pupils are a relatively homogenous set of service users.

In the health sector there is much more heterogeneity among service users in terms of the nature of their contact with the health system and what this contact is designed to achieve. It is often difficult to arrive at a precise definition of what constitutes a 'completed treatment' and to measure this accurately, particularly for the large volume of treatments for patients with chronic conditions (Brathaug, 2006). The fallback position in the health sector, therefore, is to count the number of *activities* undertaken by the various institutions that comprise the health sector. This is discussed at greater length in the next section of the report.

Conversely, there is greater consensus about how to define the desirable *outcomes* of health care than there is of education. The primary, although not exclusive, aim of health care is to improve health status. While it is a challenge to measure this improvement, it is becoming both technically and practically more feasible.

In contrast, participants at the Workshop did not share a consensus about what constitutes the primary purpose of education. Education may be designed to ensure that pupils have more qualifications, are 'more rounded' citizens, or are better able to command higher wages. The Workshop featured presentations that assumed different conceptions of educational outcomes. We summarise this debate in more detail later in the report.

The SNA requires each country to measure output volume *growth* and most discussion was, not surprisingly, about this measure (Konijn *et al.*, 2006). Output volume growth in one or more countries can be measured in a comparable way without having to measure the actual volume *levels* in each country.

But inter-country comparisons are also important (Konijn *et al.*, 2006). Measures of volume levels indicate, in a comparable way, the output or consumption of different goods and services by each country in a particular time period. Such comparisons of the volume of education or health services produced or consumed in different countries - perhaps expressed per pupil or per head of population - are likely to be of considerable policy interest. Although challenging, constructing this type of measure is, in principle, feasible for education and health services, as shown by the contributions made at the Workshop (Pritchard *et al.*, 2006a, 2006b, Huber 2006). Some recent developments have helped make success more likely. Volume levels of different goods and services produced in different countries are now regularly estimated and compared with the help of purchasing power parities (which remove the differences in national price levels from the money value of each country's production). An extension of the PPP methodology could result in the development of comparable measures of the volume levels of education and health services in different countries.

Health

As Chessa and Kleima point out, “finding an adequate measure for output volume is not straightforward. This task involves the partitioning of total output into a set of individual products, quantifying their volume and finding weights in order to aggregate the volumes into a value representing total output volume” (Chessa *et al.*, 2006). If the health system produced only a single output (x) the change in output from one year to the next can be expressed as:

$$I^1 = \frac{\text{output}_{yr2}}{\text{output}_{yr1}} = \frac{x_2}{x_1}$$

But this oversimplifies matters considerably, as encapsulated by the Eurostat definition of health care output. “The health output is the quantity of health care received by patients, adjusted to allow for the qualities of services provided, for each type of health care. The quantity of health care received by patients should be measured in terms of complete treatments” (Eurostat, 2001).

Putting this definition into practice is not straightforward because it is a challenge to measure ‘complete treatments’ (Brathaug, 2006; Pritchard *et al.*, 2006b). The majority of patients receive a range of interventions from different providers, in different settings, and, in the case of patients with chronic conditions, over a long period of time. It is not always possible to define when the treatment is complete and most countries lack the informational capability to track patients across different settings. This means that counting the number of patients who have completed their treatment is not currently possible.

In view of these difficulties, it is common practice to define output in the health sector by counting the number of activities undertaken – for instance, the number of patients treated in hospital. If only a single ‘activity’ (x) is undertaken, the index is rewritten as:

$$I^2 = \frac{\text{activity}_{yr2}}{\text{activity}_{yr1}} = \frac{x_2}{x_1}$$

But, of course, the health sector performs many different activities – patients visit their general practitioners, are treated in hospital, are provided with medicine by pharmacists, and so on. To be able to assess the output of the health system it is necessary both to count all of these activities ($x_j, j=1\dots J$) and to attach a relative value to each type of activity ($v_j, j=1\dots J$). Thus, in Laspeyres form, where activities are valued in the base period, the index becomes:

$$I^3 = \frac{(\text{number_of_activities}_{yr2}) \times (\text{value_per_activity}_{yr1})}{(\text{number_of_activities}_{yr1}) \times (\text{value_per_activity}_{yr1})} = \frac{\sum_{j=1}^J x_{j2} v_{j1}}{\sum_{j=1}^J x_{j1} v_{j1}}$$

Even counting activities can be difficult. In the hospital sector, activities are counted reasonably accurately. As well as being able to count the number of patients admitted to (or discharged from) hospital, there are good ways of distinguishing between one type of patient and another, the most common classification system being diagnosis

related groups (DRGs) (Brathaug, 2006; Huber, 2006; Pritchard *et al.*, 2006b) and this is the approach recommended by Eurostat (Eurostat, 2001). Malherbe and Gallais report that eight EU countries, as well as Australia and New Zealand, use some form of DRGs as the basis for quantifying hospital activity (Malherbe *et al.*, 2006b).

While recognising that DRGs represent a considerable improvement to highly aggregated descriptions of hospital activity (eg inpatient admissions, outpatient visits), participants at the Workshop pointed out the lag between the appearance of new technologies and their inclusion in the classification system. It was noted, however, that this was a similar issue to the quality change problem in the measurement of the Consumer Price Index.

Outside hospitals, counting activities is more difficult (Jedlickova *et al.*, 2006). Often data are unavailable. In the UK, for instance, no routine data are collected on the number of consultations patients have with general practitioners. Instead the UK National Accounts rely on estimates derived from surveys of the general population (Lee *et al.*, 2006). Reliance on survey data may give a misleading impression both of the volume of activity, because this depends on accuracy of recall and how representative the sample is of the general population, and of changes in volume over time, particularly if a new sample is surveyed in each period.

The method adopted to counting activities varies according to institutional stratification and by country. A survey of the range of approaches currently in place across European and OECD countries was reported at the conference (Malherbe *et al.*, 2006b). A summary of the findings is reported in box 3 below.

Box 3: Methods to count activities

Stratification	Most common method	Other methods
Hospital services		
In-patient services	Number of treatments by DRG	Occupant bed days
Hospital psychiatric services	Occupant bed days	Number of treatments by DRG
Rehabilitation services provided in rehabilitation centres / hospitals	Occupant bed days	Number of treatments by DRG
Nursing services	Occupant bed days by level of care	Number of treatments by DRG
Medical practice services		
Services provided by medical specialists	Number of consultations (by type of treatment)	Number of treatments
Services provided by GPs	Number of consultations (by type of treatment)	Number of treatments
Dental practice services	Number of consultations (by type of treatment)	Number of treatments
Other human health services	Number of treatments	

While counting activities is difficult, it is even more problematic to attach a value (v_j) to these activities, as required to calculate total output. The ‘value’ of hospital treatment following cardiac arrest will be different to the ‘value’ of a consultation with a GP about back pain. But how should we determine these relative values? By definition, for non-market services, there are no market prices to indicate the consumer’s marginal willingness to pay for them. Instead, the recent convention in the National Accounts has been to use cost to reflect the value of non-market outputs. Thus the index becomes:

$$I^4 = \frac{(\text{number_of_activities}_{yr2}) \times (\text{cost_per_activity}_{yr1})}{(\text{number_of_activities}_{yr1}) \times (\text{cost_per_activity}_{yr1})} = \frac{\sum_{j=1}^J x_{j2} c_{j1}}{\sum_{j=1}^J x_{j1} c_{j1}}$$

where c_j reflects the cost of activity j . This practice is consistent with the SNA and the recommendations of the European Union (Eurostat, 2001). However it implies that costs reflect the marginal value that society places on these activities. So, in the UK, a cochlear implant to treat deafness (at £23,889) is assumed to be fifteen times more ‘valuable’ than a normal delivery in maternity care (at £1,598). The use of costs to reflect value also rests on strong implicit assumptions which are unlikely to be valid, especially that health care resources are allocated in line with societal preferences (ie the health system is allocatively efficient).

Nevertheless, cost-weights have the advantage that they are reasonably easy to obtain and incorporate into the index. In general, participants at the Workshop indicated little support at the current time for an index in which cost weights are replaced by value weights, because of the conceptual and practical challenges involved. However some did take the view that this was the right direction for the future.

One way to capture aspects of quality is to incorporate them as adjustments in a cost-weighted output index. An example of this approach has been developed by the University of York and National Institute of Economic and Social Research (Dawson *et al.*, 2005). The English Department of Health has used some partial measures of quality (waiting times, 30 day survival after hospital admission, blood pressure control in general practice) to demonstrate the method (Simkins, 2006).

In practical terms, this raises the question of how to define quality in health care. Ultimately quality adjustment requires deciding upon the domains of quality in which people are interested, which might include the following (Atkinson, 2005; Brathaug, 2006; Dawson *et al.*, 2005; Huber, 2006; Simkins, 2006):

- the contribution made to improving health
- having a choice about when and where care is delivered
- the delay (waiting time) before receipt of care
- patient satisfaction or patient experience
- the environment in which care is delivered

Quality-adjustment is difficult most fundamentally because people do not demand health care for its own sake, but because of the contribution it makes to their health status. This requires a means of measuring the health outcome of treatment. At the Workshop, Kelley provided the following definition: “The degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge” (Kelley, 2006).

Ideally, the measure of health outcome should indicate the ‘value added’ to health as a result of contact with the health system. This reflects the need to isolate the specific

contribution of health services to health outcome – the problem of attribution. Value added measures have proved difficult to make operational in the health sector, mainly because the counterfactual – what health status would have been in the absence of intervention – is rarely observed. Instead health status measurement tends to rely on comparisons of health states before and (sometime) after intervention. For the purposes of the measuring output growth in the National Accounts, before-and-after measurements can be considered reasonable approximations to with- and without-treatment comparisons (Dawson *et al.*, 2005).

However, currently available data on quality are too limited in virtually every OECD country for such an adjustment to be made adequately. Many countries are able to provide a description of efforts to improve quality of care, but few are able to quantify these efforts (Kelley, 2006). Even then, quantification is usually confined to areas where efforts have been made to improve quality, without commensurate attention to where quality might have declined. This may induce biased assessment of progress (Kelley, 2006). One suggestion is to select a subset of tracer or sentinel procedures and indicators in those areas (Huber, 2006; Kelley, 2006).

In summary, the Workshop indicated the current state of play on the measurement of health outputs and outcomes:

- While Eurostat defines health care output in terms of ‘completed treatments’, it is difficult to measure these for the majority of patients. For many patients, this is because there may be no obvious treatment ‘endpoint’. For others, computer systems are unable to track patients that require care across multiple settings.
- Measurement of health care activities has become more comprehensive and disaggregated, particularly as more countries have adopted DRG classification systems.
- There is increasing standardisation across countries in many activity categories, partly in the hospital sector, but there remain considerable challenges in counting non-hospital services.

- Sometimes ‘activities’ are really inputs – most notably the number of prescriptions provided. This raises the question of whether more activities are better.
- Health outputs often reflect contributions made by other sectors, and there is a question of where to attribute these efforts. This is most evident for pharmaceutical inputs and for those provided by the social care sector.
- There are challenges with incorporating new products/technologies in the National Accounts, with tension between reflecting changes in practice and ensuring consistency over time. This is a common problem across economic sectors.
- Costs are seen as the most convenient means to weight different activities, particularly because of their availability. However, cost weights imply the presence of allocative efficiency and constant returns to scale. If these conditions do not hold, costs will be a poor reflection of societal value.
- Taking account of the quality of health services is a major challenge. To meet this, there needs to be debate about what health care is for, how to measure the dimensions of quality, how to isolate the specific contribution of the health sector from other influences on the quality domain, and how to attach weights to each domain.

Education

Participants at the Workshop suggested that measurement of output was more manageable in the educational sector than in health: education is less complex, there is more stability over time, there is greater homogeneity among service users, and there is greater availability of data. But there are other challenges, particularly in reaching agreement about the primary purpose of education.

This challenge can be re-formulated as a question: “what does the education system aim to produce?” Garniss argues that the focus of measurement should be on the contribution that the education system makes to social and economic welfare (Garniss, 2006). But operationalising this concept requires intermediate measures. Various possibilities were alluded to at the Workshop including giving pupils qualifications,

moulding them into ‘well-rounded citizens’ or enhancing their expected lifetime earnings. These are not easily measured, so the current focus in the National Accounts is to count the volume of education delivered, and then make adjustments for educational quality.

The general form of an output index in education is as follows.

$$I^6 = \frac{(\text{number_of_outputs}_{\text{yr2}}) \times (\text{cost_per_output}_{\text{yr1}})}{(\text{number_of_outputs}_{\text{yr1}}) \times (\text{cost_per_output}_{\text{yr1}})} = \frac{\sum_{j=1}^J x_{j2} c_{j1}}{\sum_{j=1}^J x_{j1} c_{j1}}$$

Here outputs are weighted to reflect the differential costs of teaching pupils at different stages of the educational pathway. Eurostat suggests an institutional stratification comprising six levels (Eurostat, 2001):

- preschool
- primary
- lower secondary
- upper secondary
- higher (tertiary) education
- other education

This institutional stratification is not always applied, with some countries finding it difficult to separate lower and upper secondary education in particular.

Eurostat’s recommended approach to how best to count ‘output’ depends on the educational level (Eurostat, 2001):

- Number of student hours is recommended except in higher education, though it is acceptable to use number of students if hours per pupil is considered stable;
- Number of students is considered a better reflection of output in higher education.

Workshop participants recognised, however, that hours of teaching may not be highly correlated with the quality of teaching (Christian, 2006a). This was demonstrated by Davidson in his presentation of the PISA exercise (see box 4 for summary) (Davidson,

2006). In international comparisons Finnish children (7-15 year olds) receive among the lowest number of student instruction hours but achieve the best PISA scores. This suggests that input (hours of instruction) does not necessarily equate with outcome in terms of test results.

Box 4: Key features of PISA exercise

- Aim to be a measure of quality, focusing on measuring student competence in reading, mathematics and science *not* what students have learned (so it is not a test of national curricula)
- Rigorous sampling criteria, with international sampling referee, to minimise selection bias
- Undertaken on a three-yearly cycle with a set of core questions allowing (future) construction of a time series
- Captures range of background school and student characteristics
- Presentation of results allows comparison across countries of both *mean* levels of achievement and *variation* among students in achievement: ie captures both quality and equity

Countries have been considering how to account for the quality of their educational services, drawing on such things as national inspection regimes (Caplan, 2006; Garniss, 2006). While these might be politically important measures within countries, incorporating them into the National Accounts raises problems of cross-country comparability. There is also a concern with consistency across time.

UKCeMGA has proposed a 1.5% annual adjustment for the changing value of education, the rationale being that educational qualifications have higher returns for individuals today than they used to have in the past because of increasing earnings in a growing economy (Luciano, 2006). At the Workshop, Weale questioned the economic basis for making this adjustment, arguing that the output index should not incorporate income effects (Weale, 2006). The ONS has recently launched a consultation exercise to explore this controversy.

At present, those countries that have tried to account for educational quality have adopted different strategies, though most make adjustments for class size or test scores attained by pupils (Malherbe *et al.*, 2006a). In Italy, there is a class size adjustment, in the belief that the quality of teaching declines once class become too

large ('congested') (Malherbe *et al.*, 2006a). In practice, this adjustment will have minimal effect if class size does not vary over time (Jedlickova, 2006). However, if class sizes are falling, this would imply productivity decreases if no account is taken of enhanced student experience.

There are a number of ways to measure educational attainment, including examination data, moving up data (proportion of students who pass to higher year) and standardised assessment tests (Christian, 2006a). The measures of attainment available in the education sector have the potential to be substantially more powerful than any equivalent in the health sector. The reason is that, because students can be tested over time, it is possible to calculate the 'value-added' provided by the education sector. This approach seeks to isolate the extent to which educational attainment can be attributed to the education sector.

International comparisons of the quality of education might best be secured by drawing on the PISA exercise, under which a sample of students is tested periodically in key competencies in reading, maths and science. PISA also asks students about their 'soft skills', such as their engagement and motivation (Davidson, 2006). The availability of student-level data also allows country-level comparisons to be made both of the level of educational attainment and the variation in results among students.

It would be difficult and, perhaps, inappropriate to extend PISA to tertiary education, given the more specialised nature of teaching at this level and because institutions have objectives in addition to teaching, including research and broader contributions to public services (Christian, 2006a). Therefore alternative ways of measuring tertiary educational quality are required. Many countries have attempted to do this in order to facilitate performance management and evaluation of higher education at a time when institutions have a large degree of autonomy over management and organisation of their activities. At the Workshop Skodvin *et al* described attempts in Norway to assess the performance of higher education institutions in Norway (Skodvin *et al.*, 2006). This evaluation takes account of teaching and research activities, engagement with the community, and resource management.

Attempts in the US to move from relatively straightforward counts of the number of students taught towards an outcome based measure were described at the Workshop (Christian, 2006a; Jorgenson *et al.*, 2006). This involves applying a human capital approach to measuring educational output, the argument being that this would capture more accurately the likely significant contribution of education to economic growth. Reporting plans to update earlier empirical work, Jorgenson *et al* describe their proposal to investigate the impact of education on lifetime earnings which involves estimating the wage gap between groups of people who have had different levels of education (Jorgenson *et al.*, 2006).

The issue of ‘fairness’ or equity in the provision of education services was raised by audience members at the Workshop as an important political objective. Incorporating equity considerations in the National Accounts is not straightforward, but the PISA exercise may prove valuable for quality-adjustment of volume, especially if testing could be undertaken more frequently. Pritchard and Gallais outline a method for incorporating PISA scores into the national accounts in order to make international comparison of the growth in quality-adjusted volume in education services (Pritchard *et al.*, 2006a). Their illustrative results reveal marked differences across countries in the relationship between spending per student (after taking account of purchasing power parities) and quality-adjusted educational output.

In summary, consideration of Direct Volume Measurement in the education sector yielded the following insights:

- There is a lack of consensus about the primary purpose of education, with suggestions including equipping recipients with qualifications, making them rounded citizens, or enhancing their earning potential. This lack of consensus makes it difficult to define education ‘outcomes’.
- Output measurement in education is more straightforward than in health, because students are reasonably homogenous within each educational stratum and because they can be tracked over time. That said, the output of higher education institutions is not confined to teaching activities, and their research and other undertakings need to be incorporated into their measure of output.

- Quality adjustment in education is challenging, and countries are adopting different methods. Some make adjustment for class size, recognising that countries with smaller classes appear ‘less productive’ unless the enhanced student experience is accounted for. However, it is not a simple matter to put a value on this experience.
- Unlike in the health sector, it is possible to measure the ‘value added’ of education quite securely by calculating the difference test results at entry and at points along the educational pathway, offering considerable scope for isolating the specific contribution of the sector to levels of educational attainment.
- International comparison of educational quality requires a standardised basis of evaluation. The PISA exercise provides a means to this end, by comparing students in terms of core competencies. The exercise also allows analysis of variation in attainment among students, which can provide an indication of fairness in delivering education among the population. As PISA is repeated, it will be possible to assess changes in educational attainment over time.

General themes emerging from the Workshop

In this section we summarise what we perceive to be the main themes emerging from the Workshop.

There was widespread (if not unanimous) agreement that the pursuit of DVM for non-market output in National Accounts is a desirable objective, as it is more likely to offer meaningful information than the input=output convention for resource allocation and accountability in the public services. Considerable progress has already been made in implementing the DVM approach in EU member states and other OECD countries. However, there remain many challenges, and progress towards more satisfactory treatment of non-market services is likely to be a long-term project.

Measurement of activities at quite a fine level of detail has been shown to be feasible across a wide range of health care. In the education sector, outputs in the form of pupils at various stages of education are readily measured. This offers considerable

potential for some very rapid improvement in the estimates of output growth in these sectors.

However, a persistent concern of the Workshop was how to attach ‘values’ to the various activities or outputs. The traditional approach of using costs as a proxy for value has the virtue of feasibility, but assumes that resources within the sector under scrutiny are already being allocated in line with the traditional market rule (marginal benefits = marginal costs). In a non-market setting this is for many good reasons unlikely to be the case.

An alternative approach is therefore to seek to attach valuations derived from other sources to activities and outputs. For example, in health care the additional quality-adjusted life years secured per patient offers a good indication of the relative values of different treatments. In practice, pursuit of this value added metric is at a very early stage. Furthermore, there remain some aspects of ‘value’ that may not be captured in the QALY. However, if measured properly, these weights would address the problem of isolating the specific contribution of the health care sector to health outcomes.

Even if acceptable values can be inferred, there remains the question of whether they are sensitive to changes in the ‘quality’ of the service under scrutiny. For example, the health benefits of a treatment may remain unchanged, but the non-health characteristics of the treatment may improve (for example, in the form of reduced waiting times). The multi-dimensional nature of measures of the ‘quality’ of public services becomes a particular challenge as attention shifts from measures of output towards measures of outcome.

The Workshop also highlighted other challenges. There was some concern about the need to maintain useable time series of data with which to assemble estimates of productivity change. The rapid changes in methodologies run the risk of fracturing the continuity of data series, and there is a need to maintain the integrity of series wherever possible.

At present most countries are seeking to enhance their own country’s methodology without too much regard for international standardisation. Yet international

comparison remains one of the most important vehicles for promoting improvement in public services, so the pursuit of comparability remains a high priority. Contributions to the Workshop have demonstrated the feasibility of making progress in this domain. The work of OECD on the PISA initiative offers a promising model in the education sector. However, the challenges in health care are more formidable.

In contrast to market sectors, a particular preoccupation of many public services is the pursuit of 'equity' objectives. For example, many education systems place a particular emphasis on securing improved outcomes for the most disadvantaged pupils. In principle, this implies attaching a higher value to a given outcome for the population groups of concern. However, there is currently little evidence on which such an adjustment could be made.

Careful thought also needs to be given to presentation of the information. It will be worth evaluating the effectiveness of the StatRes project in Norway which attempts to improve the transparency of public sector performance reports (Strømgren, 2006). The StatRes project will provide the public with a website showing key performance indicators for central-government activities. The system implements the 'input-activity-output-outcome' logic and other international standards. The aim is to strengthen incentives for agencies, stimulate analyses and search for improvement areas, trigger improvement in data/indicator quality, and stimulate and support the debate about effective use of public resources. Publication of such data is not without risks, however, particularly if it induces undesirable behavioural change, such as 'gaming' of reported performance (Van Dooren *et al.*, 2006)

Finally, whilst the Workshop emphasised the issues in health care and education, the agenda is also relevant to a range of other collective services, such as criminal justice and transport (Cullinane, 2006; Heikkinen *et al.*, 2006). Although not a central feature of the discussion at the Workshop, it is clear that the challenges in these services, which are generally less measurable in terms of individual activities, are greater than those in the health and education sectors, and any major advances are likely to be in the longer term.

Conclusions and priorities for the future

In conclusion, we indicate what we feel should be the priorities for the proposed project to develop detailed international guidelines for improving the measurement of non-market output in the national accounts. We should emphasise that these are personal observations, developed in the light of the Workshop proceedings. They do not necessarily reflect the views of the sponsors or the participants at the Workshop.

First, we feel there is an urgent need to secure the active engagement in the project of all relevant stakeholders (Grice, 2006). The project will succeed only if it meets the priorities and needs of those stakeholders, including governments, legislatures, national statistical agencies, and other users of national accounts. At an early stage, therefore, we feel that participants in the project should seek out the information needs and priorities of these key stakeholders, and ensure that they are taken into account in the subsequent methodological developments.

There is then a clear need to continue to survey existing practice in participant countries. The Workshop identified a great deal of common practice across many countries, but also marked differences in methodology. The project should create an important resource for sharing existing practice.

Many of the fundamental principles to which all systems of national accounts should seek to adhere in measuring non-market output have been addressed already, notably in the SNA, the Eurostat Manual on Prices and Volumes and the Atkinson Review Report. The project should act as a focus for developing and refining international guidelines on the implementation of these principles. To that end, we note that, as a result of the Workshop, OECD are preparing a Handbook on Measuring Education and Health Volume Output.

We agree that the two priorities for further work should be the health care and educational sectors.

The project should seek confirmation that health gain (in the form for example of quality adjusted life years) should be the fundamental outcome on which

methodology is developed. We believe that system responsiveness unrelated to health gain (in the form for example of waiting times) is likely to be a secondary issue at this early stage of methodological development, but the project should verify that this is the case.

In the short term, it is likely that health care methodology is likely to be driven by counts of activities rather than outputs. However, there is an imperfect relationship between the number of activities delivered and health gain, and we feel that the project should examine the possibility of moving towards measures of completed treatments, at least for a subset of health care.

Existing methodology relies almost universally on the use of cost weights to aggregate measures of activity or output in health care. We believe that in due course measures of activity or output should be aggregated using value weights, based on the relative contribution of the treatment to health gain. In many circumstances these may yield very different measures of volume growth. The project should therefore in our view examine the scope for developing value weights as a matter of urgency, and recommend methodologies for estimating such weights.

It is clearly desirable to seek out indices of the ‘quality’ of health care with which to augment the counts of activities or outputs. However, these indices may merely indicate the extent to which the ‘value’ of the treatment is increasing over time, and a good measure of ‘value’, updated every year, may render such quality measurement redundant. That is, if a system of accounts uses adequate disaggregation with value weights, the need to measure quality change is less urgent.

The treatment of pharmaceutical expenditure in health care volume measurement is an especially challenging problem that deserves urgent attention, given its importance as a percentage of total expenditure. This involves determining the relationship between the amount of input (eg prescriptions dispensed) and health care output; and deciding whether these contributions to output should be attributed to the health or pharmaceutical sector.

Volume measurement is much more straightforward in education than in health care, and we feel that a great deal of progress can be made in developing guidelines for such measurement in national accounts. However, there is less agreement than in health care about the ultimate outcomes of educational services. This diversity of views still permits individual countries to develop measures of growth. However, in the interests of promoting international comparison, we feel that the project should examine the scope for a broader consensus on the measurement of educational outcome.

The lack of clarity on outcomes leads to special challenges on the estimation of weights with which to aggregate measures of educational output. It may be the case that – in education – society has been more successful in aligning expenditure with values, so that the divergence between cost weights and value weights may be less marked in this sector than in other domains of government activity. Nevertheless we feel that the project should seek out some clarification on this issue.

Because of the lack of value weights, measures of quality improvement may be more important in education than in health care, and this deserves careful attention in the project. Here the challenge may be to ensure that the metric for quality remains consistent over time. The PISA initiative may yield a useful resource in this respect.

More generally, we feel that the project should explore the scope for measuring levels of output as well as growth. Inter-country comparisons are one of the most powerful devices for engaging policy interest and prompting the search for service improvements, and developing robust methodology to that end, building on experience with PPP, should in our view be a priority.

Considerable progress has already made across EU member states and other OECD countries in implementing Direct Volume Measurement in the National Accounts. This has made the National Accounts more meaningful as measures of the changes in output over time and for making international comparisons. We hope that the project will build on the foundation that has been laid in order to further enhance international comparison of non-market output in the health and education sectors.

References

- Aldin V. 2006. Measuring government output and productivity: meeting the need for more timely, detailed and better quality data on government expenditure, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Atkinson T. 2005. *Atkinson Review: Final Report. Measurement of Government Output and Productivity for the National Accounts*. Palgrave Macmillan: Basingstoke
- Atkinson T. 2006. Measurement of Government output and productivity. *Journal of the Royal Statistical Society, series A* **169**(4): 659-662
- Brathaug AL. 2006. Measurement of health output - experiences from the Norwegian National Accounts, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Caplan D. 1998. Measuring the output of non-market services. *Economic Trends* **no. 539**(October): 45-79
- Caplan D. 2006. Non-market services: the view from the national accounts, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Chessa A, Kleima F. 2006. The Dutch experience in measuring health output and labour productivity, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Christian M. 2006a. Measuring the education function of government in the United States, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Christian M. 2006b. Measuring the output of health in the United States, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Christian M. 2006c. spoken comment, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Cullinane P. 2006. Can the output of collective services be measured? The case of public order and safety, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Dam KW, Sørensen KØ. 2006. Volume measures of education in the Norwegian National Accounts, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Davidson M. 2006. PISA database for pupils of 15 year old, and similar data for other projects: output and outcome, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Dawson D, Gravelle H, O'Mahony M, Street A, Weale M, Castelli A, Jacobs R, Kind P, Loveridge P, Martin S, Stevens P, Stokes L. 2005. *Developing new*

- approaches to measuring NHS outputs and productivity, Final Report*. Centre for Health Economic research paper 6: York
- Eurostat. 1995. European Systems of Accounts. Eurostat
- Eurostat. 2001. *Handbook on price and volume measures in national accounts* (2001 ed.). Luxembourg: Office for Official Publications of the European Communities
- Garniss D. 2006. Measuring government education output quality in England: an overview of the issues and approaches developed by the Department for Education and Skills, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Grice J. 2006. Stakeholders: the key to success in measuring the productivity of government, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Heikkinen J, Lehmuskoski M. 2006. Productivity in the central government sector in Finland - the case of the Ministry of Transport and Communications 2000-2005, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Huber M. 2006. International comparisons of prices and volumes in health care among OECD countries, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Jedlickova E. 2006. The Czech experimental calculations in education, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Jedlickova E, Sarovcova L. 2006. Health in the Czech republic, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Jorgenson DW, Fraumeni B, Christian M. 2006. A human capital approach to measuring the output of education for the US, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Kelley E. 2006. Measuring health care quality at the international level, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Konijn P, Gallais A. 2006. The national accounts perspective, time and spatial comparisons: the projects of Eurostat and OECD, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Lee P, Little C. 2006. Measuring health output and productivity in the UK: an essential element of public accountability, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Lequiller F. 2006. Introductory comments, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Luciano M. 2006. Measurement of non-market output in education, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006

- Malherbe F, Gallais A. 2006a. Review of European and OECD countries' experiences from questionnaires on education services launched in June 2006, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Malherbe F, Gallais A. 2006b. Review of European and OECD countries' experiences from questionnaires on health services launched in June 2006, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Pritchard A, Gallais A. 2006a. Cross-country comparison of production volumes and prices: non-market education services, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Pritchard A, Gallais A. 2006b. Proposal for an output method for Purchasing Power Parities for (non market) health services, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Quintela I. 2006. The Portuguese experience: estimation of the output of hospitals according to different methodologies, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Simkins A. 2006. Health quality indicators and value of output, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Skodvin O-J, Andreassen GA, Wien Fjell M, Moi A. 2006. Assessing performance within higher education in Norway, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Strømgren M. 2006. The StatRes project - more transparency of public-sector performance, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- United Nations Statistics Division. 1993. *Systems of National Accounts*. United Nations
- Van Dooren W, Malinska J, Manning N, Sterck M, Kraan D-J, Bouckaert G. 2006. Issues in Output Measurement for "Government at a Glance", *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006
- Weale M. 2006. Public sector output: output indicators and welfare indicators, *OECD/ONS/Government of Norway Workshop on "Measurement of non-market output in education and health"*: London, October 3-5, 2006