

FEATURE

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

The measurement of non-market output in education and health

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 article summarises contributions at a workshop held in October 2006 that focused 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.

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. Governments also recognise the importance of demonstrating to taxpayers that government expenditure on these services represents value for money.

In October 2006, the Office for National Statistics (ONS) organised a workshop jointly with the Organisation for Economic Co-operation and Development (OECD) 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:

- 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
- 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, and

- to take account not only of the quantity of output but also the quality of that output (the notion of the volume of services embraces both concepts)

This article 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 article describes key themes emerging from the workshop and highlights what are considered to be the most important issues to be tackled in future work.

The intention of this article is to highlight achievements, interesting new departures, key challenges, and priorities for the future. It must be emphasised that the article represents only the views of the authors, and is not intended as an official report of the workshop. The article is an abridgement of the formal report of the workshop, which is available from the OECD website at www.oecd.org/dataoecd/61/9/37903961.pdf

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:

- 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
- it implies no change in productivity over time, as outputs are not measured directly, and
- 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. Both the System of National Accounts (SNA93) and European System of Accounts (ESA95) 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. Even so, preliminary

attempts are being made to identify outputs for some collective services.

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. Considerable progress has been made by Member States in meeting these requirements, and these achievements are summarised in **Box 1**.

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. 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.

DVM 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 DVM approach:

- the output of the public sector is often difficult to describe or measure
- it is difficult to measure the quality of public sector output, and
- 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. DVM 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 DVM

Considerable progress has also been made in many countries across Europe seeking to implement the European Union National Accounts directive, particularly in healthcare and education. Key definitional terms for the health and education sectors are reproduced from the Eurostat handbook in **Box 2**. This reveals differences in applying the definitional concepts to the two sectors.

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

Box 1

Progress in implementing direct volume measurement

Progress has been achieved in the following areas:

- 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 the EU doing the same
- many countries have been working seriously on developing new methodologies that go well beyond those formally reported to date, as evidenced by the workshop papers

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 healthcare 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

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. 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 article.

Conversely, there is greater consensus about how to define the desirable outcomes of healthcare than there is of education. The primary, although not exclusive, aim of healthcare 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. The debate is summarised in more detail later in the article.

The SNA requires each country to measure output volume growth and most discussion was, not surprisingly, about this measure. 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. 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. 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 PPPs (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

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:

$$I^1 = \frac{output_{yr2}}{output_{yr1}} = \frac{x_2}{x_1}$$

But this oversimplifies matters considerably, as encapsulated by the Eurostat definition of healthcare 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. 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:

$$I^2 = \frac{activity_{yr2}}{activity_{yr1}} = \frac{x_2}{x_1}$$

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 ($x_j, j=1...J$) and to attach a relative value to each type of activity ($v_j, j=1...J$). Thus, in Laspeyres form, where activities are valued in the base period, the index becomes:

$$P^3 = \frac{(\text{number_of_activities}_{y_{t2}}) \times (\text{value_per_activity}_{y_{t1}})}{(\text{number_of_activities}_{y_{t1}}) \times (\text{value_per_activity}_{y_{t1}})} = \frac{\sum_{j=1}^J x_{jt}^2 v_{jt}}{\sum_{j=1}^J x_{jt} v_{jt}}$$

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) and this is the approach recommended by Eurostat. Eight EU countries, as well as Australia and New Zealand, use some form of DRGs as the basis for quantifying hospital activity.

While recognising that DRGs represent a considerable improvement to highly aggregated descriptions of hospital activity, for example, in-patient admissions and out-patient 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 prices index.

Outside hospitals, counting activities is more difficult. 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. 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. A summary of the findings is reported in **Box 3**.

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 from the value of a consultation with a GP about back pain. But how should these relative values be determined? 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:

$$P^4 = \frac{(\text{number_of_activities}_{y_{t2}}) \times (\text{cost_per_activity}_{y_{t1}})}{(\text{number_of_activities}_{y_{t1}}) \times (\text{cost_per_activity}_{y_{t1}})} = \frac{\sum_{j=1}^J x_{jt}^2 c_{jt}}{\sum_{j=1}^J x_{jt} c_{jt}}$$

where c_j reflects the cost of activity j . This practice is consistent with the SNA and the recommendations of the EU. 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 healthcare resources are allocated in line with societal preferences, that is, 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.

In practical terms, this raises the question of how to define quality in healthcare. Ultimately, quality adjustment requires deciding upon the domains of quality in which people are interested, which might include the following:

- the contribution made to improving health
- having a choice about when and where care is delivered

Box 3

Methods to count activities

Stratification

Hospital services

In-patient services

Hospital psychiatric services

Rehabilitation services provided in rehabilitation centres/hospitals

Nursing services

Medical practice services

Services provided by medical specialists

Services provided by GPs

Dental practice services

Other human health services

Most common method

Number of treatments by DRG

Occupant bed days

Occupant bed days

Occupant bed days by level of care

Number of consultations (by type of treatment)

Number of consultations (by type of treatment)

Number of consultations (by type of treatment)

Number of treatments

Other methods

Occupant bed days

Number of treatments by DRG

Number of treatments by DRG

Number of treatments by DRG

Number of treatments

Number of treatments

Number of treatments

- 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 healthcare 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, the following definition was provided: ‘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 measuring output growth in the National Accounts, before and after measurements can be considered reasonable approximations to with- and without-treatment comparisons. However, currently available data on quality are too limited in virtually every OECD country for such an adjustment to be made adequately.

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 reformulated as a question: ‘What does the education system aim to produce?’ It was argued that the focus of measurement should be on the contribution that the education system makes to social and economic welfare. In addition to educational attainment, outcomes such as generating social skills, making good citizens, producing healthy

people, as well as the purely custodial function of providing childcare could also be considered. 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:

$$F = \frac{(\text{number_of_outputs}_{y2}) \times (\text{cost_per_output}_{y2})}{(\text{number_of_outputs}_{y1}) \times (\text{cost_per_output}_{y1})} = \frac{\sum_{j=1}^J x_{j2} c_{j2}}{\sum_{j=1}^J x_{j1} c_{j1}}$$

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:

- preschool
- primary
- lower secondary
- upper secondary
- higher (tertiary) education, and
- 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:

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

Countries have been considering how to account for the quality of their educational services, drawing on such things as national inspection regimes. 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.

Atkinson (Atkinson 2005) proposed a 1.5 per cent 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. At the workshop, the economic basis for making this adjustment was questioned, it being argued that the output index should not incorporate income effects. 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. In Italy, there is a class size adjustment, in the belief that the quality of teaching declines once classes become too large – ‘congested’. In practice, this adjustment will have minimal effect if class size does not vary over time. However, if class sizes were falling, this would imply productivity decreases if no account were 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. 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 OECD’s Programme for International Student Assessment (PISA) exercise (Davidson, 2006). PISA is a collaborative process among the 30 member countries of the OECD and nearly 30 partner countries, which attempts to measure student achievement. Under PISA, 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. 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. 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. Norway has attempted to assess the performance of its higher education institutions by taking 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. 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. This involves estimating the wage gap between groups of people who have had different levels of education.

General themes emerging from the workshop

In this section, the main themes emerging from the workshop are summarised.

There was widespread (if not unanimous) agreement that the pursuit of DVM for non-market output in the 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 healthcare. 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 equal marginal costs). In a non-market setting, this is unlikely to be the case.

An alternative approach is to seek to attach valuations derived from other sources to activities and outputs. For example, in healthcare, 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 and some aspects of value may not be captured in this measure. However, if measured properly, these weights would address the problem of isolating the specific contribution of the healthcare 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 multidimensional 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 usable 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 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 healthcare are more formidable.

Finally, while the workshop emphasised the issues in healthcare and education, the agenda is also relevant to a range of other collective services, such as criminal justice

and transport. 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

These concluding remarks indicate what the authors feel should be the priorities in the development of detailed international guidelines for improving the measurement of non-market output in the National Accounts. It should be emphasised 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, there is an urgent need to secure the active engagement of all relevant stakeholders. Advances will be made only if they meet the priorities and needs of those stakeholders, including governments, legislatures, national statistical agencies, and other users of national accounts. At an early stage, therefore, participants in this work 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. Documenting these 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 (Eurostat, 2001) and the Atkinson Review Report (Atkinson, 2005). Future work should act as a focus for developing and refining international guidelines on the implementation of these principles. To that end, it is noted that, as a result of the workshop, OECD is preparing a handbook on measuring education and health volume output.

Future work in the area of health 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.

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 research should verify that this is the case.

In the short term, it is likely that healthcare methodology will 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 research should examine the possibility of moving towards measures of completed treatments, at least for a subset of healthcare.

Existing methodology relies almost universally on the use of cost weights to aggregate measures of activity or output in healthcare. 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. Research efforts should 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 healthcare 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 healthcare 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, for example, prescriptions dispensed, and healthcare 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 healthcare, and it is felt 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 healthcare 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, future research 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, research should seek to clarify this issue.

Because of the lack of value weights, measures of quality improvement may be more important in education than in healthcare, 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, further research 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 be a priority.

Considerable progress has already been made across EU Member States and other OECD countries in implementing DVM 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. Future efforts should 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.

ACKNOWLEDGEMENTS

The authors were commissioned by the workshop organisers to produce an independent report on the proceedings. While they have benefited greatly from the numerous comments of participants, this is their report and it reflects their personal views.

CONTACT

✉ elmr@ons.gsi.gov.uk

REFERENCES

Atkinson T (2005) *Atkinson Review: Final Report. Measurement of Government Output and Productivity for the National Accounts*, Palgrave Macmillan: Basingstoke.

Chessa A and 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, 3–5 October 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, 3–5 October 2006.

Dawson D, Gravelle H, O'Mahony M, Street A, Weale M, Castelli A, Jacobs R, Kind P, Loveridge P, Martin S, Stevens P and Stokes L (2005) *Developing new approaches to measuring NHS outputs and productivity, Final Report*, Centre for Health Economic research paper 6: York.

Eurostat (2001) *Handbook on price and volume measures in national accounts* (2001 edition), Luxembourg: Office for Official Publications of the European Communities.

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, 3–5 October 2006.