Measuring non-market output and productivity

Paper prepared for Economic Measurement Group meeting
Sydney, December 2008

Phillip Lee
Office for National Statistics. UK

On secondment to Statistics New Zealand
P O Box 2922
Wellington, New Zealand

info@stats.govt.nz
www.stats.govt.nz
Acknowledgements, disclaimer, and notes on confidentiality

The opinions, findings, recommendations and conclusions expressed in this report are those of the author. They do not purport to represent those of either the UK’s Office for National Statistics or Statistics New Zealand or any other organisation. Statistics NZ takes no responsibility for any omissions or errors in the information contained here.

Any table or other material published in this report may be reproduced and published without further licence, provided that it does not purport to be published under government authority and that acknowledgement is made of this source.

Abstract

This paper shares some information that may be useful for those interested in compiling estimates of non-market output and productivity. It presents very short summaries of recent, relevant publications; it sets out current thinking on what are considered to be the ‘ideal’ measures for health and education services; and it discusses a number of issues that have come to the fore during development and compilation in the United Kingdom.

Keywords
Non-market, output, productivity, government, health, education.
Introduction

In recent years, much progress has been made in the field of measurement of non-market output, with several countries moving away from the ‘output=inputs’ method, whereby (change in) output was equated with (change in) total inputs.

Non-market services typically make up a large part of an economy, and account for an even larger part of total government expenditure. Amongst other important users, taxpayers have a right to understand how well their money is being spent, and policymakers need good quality information on economic output in setting monetary policy. National statistical offices have been responding to demands for better information on government output and productivity, principally as part of, or closely connected with, the National Accounts.

This paper focuses on the practical aspects of measuring non-market output, and thus provides relevant background for Statistics New Zealand’s feasibility study into the measurement of non-market output and productivity. This non-market feasibility study is part of a wider project to carry out enhancements to New Zealand productivity measures. The other priorities for the wider productivity project are to:

- introduce a quality-adjusted labour productivity series;
- expand the measured sector to include industries that are currently excluded; and
- develop industry-level measures of productivity.

In regards to the measurement of non-market output and productivity, this paper includes:

- a recent potted history of developments
- the current consensus on methodology
- the progress of other countries
- lessons learned in the last five years or so during the development of non-market output and productivity estimates for the United Kingdom.
Overview of major publications

This section sets out a potted recent history of developments in the measurement of non-market output and productivity, in terms of major publications.

United Nations' System of National Accounts 1993

The *System of National Accounts 1993* (web version, United Nations, 1993) sets out a comprehensive framework for the construction of economic accounts. Although its focus is on concept, it does also provide some practical guidance on compilation. It is particularly concerned with production and accounting for the accompanying flows of money. It therefore does not delve too deeply into the compilation of estimates of non-market output, which are not bought and sold on markets.

What *System of National Accounts 1993* does do is set the scene by suggesting strongly that it is preferable to measure inputs and output independently on the grounds that “Changes in productivity may occur in all fields of production, including the production of non-market services” (para 16.133).

It discourages use of the ‘output=inputs’ method: “When it is not possible to avoid using an input measure as a proxy for an output measure, the input measure should be a comprehensive one and not confined to labour inputs” (para 16.139).

OECD Manual – Measuring Productivity

Known as the *OECD Manual*, the publication *Measuring productivity: Measurement of aggregate and industry-level productivity growth* (Organisation for Economic Co-operation and Development, 2001), is a fairly comprehensive guide to measuring industry-level productivity, especially for the market sector. As such, it does not cover in much detail the measurement of non-market output.

It does explore the main variants of the productivity equation, and explores both the numerator of those equations – output – and the denominator – inputs (capital, labour, intermediate consumption). It is considered to be the main reference manual by compilers of productivity measures, and most publishers of productivity estimates highlight consistency with the *OECD Manual* as a main indicator of the quality of the estimates.

Eurostat's Handbook on price and volume measures in national accounts

The *Handbook on price and volume measures in national accounts* (Eurostat, 2001) expands on the guidance set out in the *European System of National Accounts 1995* (which is fully consistent with *System of National Accounts 1993* and which is more focused on circumstances and data needs of the European Union). In particular, the Eurostat Handbook distinguishes between activities, output and outcomes and introduces A/B/C scores (see below) for methods of European Union member states.

It has since become a legal requirement for EU member states to provide economic data according to the recommendations set out in the Eurostat Handbook. This includes a requirement for Member States to avoid using the ‘output=inputs’ method.

Table 1 summarises the A/B/C scores, as shown in the *Atkinson Review: Final report* (Atkinson, 2005).
Table 1

<table>
<thead>
<tr>
<th>Type of service</th>
<th>A/B/C methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Individual services (such as Education, Health,</td>
<td>A methods – output indicator approach where the indicators satisfy the following criteria:</td>
</tr>
<tr>
<td>Social Security, Recreation and Cultural Services)</td>
<td>a) they should cover all services provided; b) they should be weighted by the cost of each type of output in the base year; c) they should be as detailed as possible; and d) they should be quality adjusted.</td>
</tr>
<tr>
<td></td>
<td>B methods – output indicator approach where the criteria are not fully satisfied: eg the level of detail could be improved or the measure does not take into account changes in quality.</td>
</tr>
<tr>
<td></td>
<td>C methods – if input, activity or outcome is used (unless outcome can be interpreted as quality-adjusted output) or if coverage of output method is not representative.</td>
</tr>
<tr>
<td>Collective services (such as General Public</td>
<td>Broadly the same as for individual services but:</td>
</tr>
<tr>
<td>Administration, Defence, Police, and Research and</td>
<td>B methods – input methods are B methods, as are the use of volume indicators of activity. If input methods are used they should estimate the volume of each indicator separately, taking quality changes of inputs into account. Applying productivity or quality adjustments to the sum of the volume of inputs is not recommended.</td>
</tr>
<tr>
<td>Development)</td>
<td>C methods – the use of a single input volume indicator is not a B method.</td>
</tr>
</tbody>
</table>

Source: Atkinson Review: Final report, 2005


The final report of Sir Tony Atkinson’s 2005 review of the measurement of government output and productivity for the national accounts was published in 2005 (Atkinson, 2005). It built on the guidance already available in the field, from OECD, Eurostat as well as within the UK and from other countries. It sets out a number of principles and recommendations for taking development work on non-market output and productivity forward in the UK, and proposed detailed plans for taking work forward for a number of particular areas of non-market output: health, education, public order and safety, and social protection.

It formed the basis for discussions of how to improve non-market output estimates at two OECD/Eurostat workshops in 2005 and 2006, and continues to be the basis on which UK development work is conducted.

OECD Handbook: Towards measuring the volume of health and education services

This handbook (OECD, in press) continues along the same vein as the Eurostat Handbook on price and volume measures in national accounts and the Atkinson Review: Final report, and discusses development from the wider perspective of the 29 countries of the OECD. As well as covering compilation of estimates of change over time, it also covers compilation of estimates across countries.
The current consensus on ‘best’ methods

A consensus on what constitutes a ‘best’ method for measuring non-market output is emerging amongst national statistics offices and international organisations. In particular, there is agreement that:

- The distinction between individual and collective services is important from an output measurement perspective: individual services are acquired by individual consumers, and are used up such that less of the service is available to other consumers; collective services are those which are delivered simultaneously to many consumers and are not used up by any individual consumer.

- It is acceptable to measure collective services using the ‘output=inputs’ method.

- For individual services, the ‘output=inputs’ method should be avoided, and used only as a last resort.

- Output consists of two components: quantity and quality.

- The ‘best’ unit of output quantity is:
  - in health, the ‘health care pathway’ (all of the parts of treatment provided to a person with a specific diagnosis). Without joined-up information, individual health care activities can be used.
  - in education, the number of pupil-hours in pre-primary, primary and secondary education, and the number of full-time equivalent students in tertiary education.

- Differentiating between different types of quantity in an output measure does pick up some aspects of quality:
  - in health, appropriate differentiation could be between different types of health activity at a detailed level (by, for example, Diagnosis Related Groups for hospital care, different kinds of appointments in primary care).
  - in education, appropriate differentiation could be between different types of schooling (by, for example, at least primary, secondary, tertiary).

- The weights to be used for combining together changes in the different quantities should be total costs.

- There is currently no consensus on how to incorporate into an output quantity measure the other aspects of quality that are not, or cannot be, picked up by differentiation, for a number of reasons:
  - there is little agreement on which of the multiple aspects of quality should be part of a total quality measure (which for example in health may include health status, pleasantness of stay in hospital, speed of access, convenience and timing of an appointment);
  - there is little agreement on how to determine the relative importance of the different aspects of quality; and
  - there is typically no systematically available information to support quality adjustment.
Measures of **quality** change that are based on outcome, such as health status, are influenced by factors other than the output of the service of interest. There is currently no means of separating out the individual contributions from these different factors. For example, healthy eating, smoking cessation, taking part in physical activities and so on all play a role alongside the health system in maintaining health status.

The main types of **quality** adjustment for health are considered to be:

- health status and survival rates
- process quality indicators (for example, vaccination take-up rates, at-risk patient monitoring in primary care)
- patient experience indicators (for example, timing and convenience of appointment, cleanliness of hospital)

The main types of **quality** adjustment for education are considered to be:

- exam scores
- school inspection results
- class size.
How far have countries progressed in moving away from the ‘output = inputs’ method?

OECD and Eurostat asked countries in a 2006 survey about the way they compiled estimates of output for health and education services in their National Accounts. The following tables summarise the responses.

Table 2
Summarised results of 2006 Eurostat and OECD surveys of methods for education services

<table>
<thead>
<tr>
<th>Method Description</th>
<th>Countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality adjusted quantity measure</td>
<td>Austria, France, Hungary, Italy, Latvia, Lithuania, Malta, Spain, Sweden, UK</td>
</tr>
<tr>
<td>Quality measure only, no quality adjustment</td>
<td>Australia, Belgium, Czech Republic, Finland, Germany, Greece, New Zealand</td>
</tr>
<tr>
<td>Output=inputs</td>
<td>Canada, Denmark, Japan, Korea, Luxembourg, Switzerland, US</td>
</tr>
</tbody>
</table>

Source: OECD Handbook Towards Measuring the volume of health and education services

Table 3
Summarised results of 2006 Eurostat and OECD surveys of methods for health services

<table>
<thead>
<tr>
<th>Method Description</th>
<th>Countries</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality adjusted quantity measure for both primary and secondary care</td>
<td>No countries</td>
</tr>
<tr>
<td>Quantity measure only, no quality adjustment for both primary and secondary care</td>
<td>Australia?, Czech Republic, Finland, Sweden, UK</td>
</tr>
<tr>
<td>Quantity measure only, no quality adjustment for secondary care; output=inputs for primary care</td>
<td>Austria?, France?, Germany?, Greece, Hungary?, Italy?, Netherlands, Norway, Slovak Republic?</td>
</tr>
<tr>
<td>Output=inputs for primary and secondary care</td>
<td>Belgium?, Canada?, Denmark, Iceland?, Ireland?, Japan, Korea, Luxembourg?, Poland?, Portugal?, Switzerland?, US</td>
</tr>
</tbody>
</table>

Source: OECD Handbook towards measuring the volume of health and education services
Thoughts on lessons learned from the UK experience

This section lists and discusses a number of issues that came up during the Atkinson Review and the subsequent development of the UK’s Office for National Statistics’ (ONS) non-market output and productivity estimates.

(i) Influence

One of the first lessons learned relates to the extent to which ONS had sufficient influence and power to take the development work forward. ONS had been developing direct output methods for measuring non-market output for a number of years. In 1998, the first direct methods were published as part of the UK’s National Accounts. Work continued unabated for a number of years, with a team of up to four people, and a series of articles presenting new methods and data. However, whilst progress was being made, swifter progress was sought.

There were a number of skills that were perceived to need fortifying:

- The ability to gain access to expertise and information across government.
- The ability to influence key stakeholders across government and convince them of the benefits of better productivity measurement.
- The ability to draw on a wide range of expertise, including statistics, economics, administration and accountancy.

In 2003, the approach to developing non-market output estimates was revamped. A commitment from the UK’s National Statistician lead to the setting up of an independent project headed by Sir Tony Atkinson, with significant resources. At its height, the team supporting Sir Tony counted 14 people, some of whom were ONS staff, some of whom were seconded in from other government departments and the Bank of England. The deputy head of the UK Government Economic Service, Joe Grice, was brought in from HM Treasury as director of the Review, and a senior civil servant from the Department of Health, Aileen Simkins, was brought in as co-director.

The combination of greater numbers, but perhaps more importantly, the greater seniority of the directors, their wider set of contacts, and their experience in working across government, brought the required progress.

(ii) Collaboration

While greater numbers for the Review team was helpful, much of the work of the Review was carried out by colleagues in other government departments. These colleagues brought a much better understanding of the services of interest and of the relevant data sources and accompanying limitations. In some cases, the value added by these colleagues covered advice on what the service involved (and therefore what needed to be measured). In other cases, the value added included the initial specification of, and construction of, the basic datasets and even the required indices for initial estimates.

The topic was considered to be of such importance to the English health ministry, that they funded a project on the measurement of health output by the University of York and the National Institute of Economic and Social Research. Equally, the education ministry published it own study on the measurement of education output.
(iii) Independence from political interference

The productivity of government services is an inherently political topic. At no point in time has there seemed to be inappropriate pressure from colleagues working in political ministries: there has been a great deal of propriety, with analysts and other contributors offering only their best to the development work. However, the perception that ONS is just another arm of government and therefore under the thumb of ministers is strong (and hence led in part to the introduction of the Statistics Act earlier this year, 2008). Prior to the Act coming into law, the way ONS has bolstered (at least the perception of) statistical independence has been by appointing an independent Advisory Group.

This Advisory Group is chaired by a leading external figure and comprises people who are well respected in their various fields. The Advisory Group has thus far had two chairs: the first, Peter Smith, Professor of Economics and Director of the Centre For Health Economics at the University of York, England, and the current chair, Martin Weale CBE, Director of the National Institute of Economic and Social Research, London. Members include (and have included): the Professor of Economics at the University of Oxford, the Director of the Institute for Fiscal Studies, the Chair of the Health Statistics Users Group, and the Professor of Economics at the University of Bristol.

As well as reinforcing (the perception of) statistical independence, the Advisory Group advises on the general strategy of the productivity work and reviews the detailed content of each of the articles published on non-market productivity. It should be noted that final decisions on strategy and content of articles lies with the UK’s National Statistician.

Another key way in which we have managed user perceptions has involved being completely transparent with our data sources, methods and assumptions. Accompanying each of the productivity articles is the comprehensive Sources & Methods (ONS, 2008). This publication allows users to replicate the calculations for the productivity estimates, and change data sources, methods and assumptions. Simplified spreadsheets replicating the calculations and providing the raw and calculated data are also available.

(iv) Managing tension between different interested parties

Early on in the Atkinson Review, it was clear that different parties had different requirements relating to the National Accounts. Some favour time series continuity while others favour the latest and best estimates. Colleagues compiling the National Accounts are necessarily cautious about making changes to important data series and have rigorous quality assurance procedures in place for judging whether or not to accept changes to data sources and methods. Making sure that change, and proposals for change, were managed in the interests of all paid dividends.

A key element to managing these particular interests was to make only those improvements to the National Accounts that came out of consensus: all other improvements are exposed in a series of articles which use the National Accounts as a base. When the output estimates in the National Accounts and in productivity articles are not the same, care is taken to explain differences.

A key part of the implementation plan was to consult with users on the requirements, as well as their thoughts on developments in different methods and concepts. These consultation exercises were informed by short, comprehensive papers and were held around the country. They brought together many different people – economists, statisticians, academics, researchers, accountants, service providers and so on. These discussions have contributed to the work strategy and plan, and decisions taken on
proposals for change in output and productivity measures have been influenced strongly by these consultations.

(v) Perfection is the enemy of the good

Measurement of non-market output is as much an art as it is a science. As much relies on good judgement as it does on sound methods and good data sources. For example, there is a lack of systematic data on quality change of health services, and an absence of a consensus on what the relationship is between different aspects of quality change and quantity change.

Striving to get a perfect answer before publication would mean that no information – even reasonably good information – would be released. Consistent with the UK’s National Statistics Code of Practice on Methodology and Quality, the Atkinson Review considered it better to put less-than-perfect information into the public domain, so long as the information was fully explained.

This is the case with the quantity measures of non-market output. Perfection would require that quality was taken into account, but this is not practicable. It is considered good enough to include only the quantity measures of health output in the National Accounts in the UK (and not quality adjustments).

In order to make sure that the best decisions on publication are made, balancing the desire to inform with the need to not misinform, material in articles is quality assured by both expert and lay, internal and external audiences.

(vi) ‘Absurdity’

The flipside of a policy to publish less than absolutely perfect information is that some commentators think that the data may be misleading. ONS’s first Public Service Productivity: Health article (October 2004) only included quantity estimates of output and productivity: no quality adjustments were available at the time, and this was explained clearly. The article showed that, measured without taking into account quality change, health service productivity was falling. The reaction from the minister of health was that it was ‘absurd’ to try to measure health service productivity without taking into account quality change.

ONS’s second Public Service Productivity: Health article (February 2006) did include some estimates of the impact from introducing quality adjustment. However, the quality adjustments were novel and had not been subjected to rigorous quality assurance processes. Accordingly, the article did not present a single time series of productivity change, but a number of them (actually, three different versions, each of which had an upper and lower bound). Many commentators found this approach difficult to interpret, as the sets of estimates showed that health productivity could have been increasing, could have been flat or could have been falling, depending on the sources, methods and assumptions made.

ONS’s third Public Service Productivity: Health article (January 2008) benefited from the advice and opinion gained during the consultation exercises (see “(iv) Managing tension between different interested parties” above) and allowed a single ‘best’ time series of health productivity change to be presented. The accompanying text made the limitations of the analysis, as well as the impact from using different data sources, methods and assumptions, clear.
(vii) The 80–20 rule

The 80–20 rule goes hand-in-hand with ‘perfection is the enemy of the good’. Getting a first answer to most of the questions raised has not taken up too much time: most of the development time has been taken up with trying to refine early efforts.

For example, initial research into the quality of the expenditure figures, which underpin the estimates of the volume of intermediate consumption (volume change is calculated by deflating expenditure) identified a better source with a much finer disaggregation by type of expenditure. This meant that more specific price deflators could be applied. However, the greater level of detail now available has meant that the quality assurance required has become more resource intensive – a wider network of experts needs to be serviced; a greater degree of understanding of the expenditure data were needed; more independent sources were needed to corroborate the main source and so on.

(viii) Low hanging fruit

There was much that was done without significant resources. For example, once a data source had been identified that provided a much greater degree of disaggregation for health services in hospital, it was a fairly straightforward matter to specify exact requirements and produce a dataset that could be used in the National Accounts.

Easy wins at the early stage of the Review helped to keep up momentum, allowed resources to be concentrated on the more problematic issues, and showed interested parties that the Review would have benefits.

(ix) Inputs are as important as output

Much of the first half or more of the Atkinson Review was concentrated on output: this was what was required for both the National Accounts and for productivity analyses. Inputs, at least in volume terms, were only required for productivity analyses (the UK does not compile, for example, constant price input-output tables).

When we began looking at inputs to production, we found that the potential for error was at least as large as for output. For example, the use of poor price deflators alongside inconsistent expenditure estimates, not taking into account quality change in labour and so on, had as much of an impact on health productivity estimates as did incorporating greater differentiation in output.

(x) Pebble beach

Dealing with the many issues involved in measuring productivity involves much choice. First of all, which of the many do you look at, and in what order? How do you decide which is the first? What do you do when you lift one pebble up, and find that there are more intriguing pebbles underneath? Where do you stop? At which point in time do you say ‘enough is enough – for the moment we have done all that we can.’ There does not seem to be any easy answer to this. The key is in keeping track of where you are, and continuously revisiting the priority list. It would be ideal if the issues could be categorised according to impact on the productivity figures themselves, but unfortunately, this is usually not possible without hindsight.

This makes planning quite difficult. Uncovering significant stones means changes in publication timetables. ONS’s third Public Service Productivity: Health Care article was originally scheduled to be published in July 2007, then September, then November. It was finally published in January 2008. The delay related to uncovering one issue buried under
another in the estimates of expenditure on intermediate consumption (see ‘(vii) The 80–20 rule’ above).

(xi) Personal preferences

Given the large number of avenues to explore in developing measures of non-market output, and the lack of a systematic means for identifying the issues that are going to have the greatest impact on the quality of the results, it was easy to concentrate on doing the things that were of most interest to the individuals doing the work. While this was not necessarily a bad thing, it was important to make sure that no more time than necessary was spent, and other avenues were not cut off because of dwindling time or resources.

Part of the way in which we introduced some rigour was to continue to search for ways to evaluate any potential impact as information and understanding of the issues and potential solutions built up. Reviewing whether it was worth spending (much) further time on issues that were interesting but that did not seem likely to make much difference was useful. Also, it helped to share information about issues with colleagues, so that at least a subjective but shared agreement on how to spend time was reached.
Concluding remarks

This paper has shared some information that may be useful for those interested in compiling estimates of non-market output and productivity, which will inform Statistics NZ’s own feasibility study. It has included:

- short summaries of recent, relevant publications
- an overview of current thinking on what are considered to be the ‘ideal’ measures, particularly for health and education services
- information on countries’ progress
- discussion of a number of issues that came to the fore during development and compilation in the United Kingdom.

The lessons learned from the UK’s non-market output and productivity project are probably relevant to any other development project. It is therefore hoped that the lessons learned are useful for a broader audience that just those interested in non-market output and productivity measurement.
References


