

Session Number: Session 7a
Time: Friday 29th August, Am

*Paper Prepared for the 30th General Conference of
The International Association for Research in Income and Wealth*

Portoroz, Slovenia, August 24-30, 2008

Title of Paper

Developments in the measurement of government output and productivity in the UK

Author(s)

Emma Edworthy

For additional information please contact:

Name: Emma Edworthy

Affiliation: Office for National Statistics

Full mailing address: Office for National Statistics
Government Buildings
Cardiff Road
Newport
Wales
NP10 8XG

Email address: emma.edworthy@ons.gsi.gov.uk

This paper is posted on the following website: <http://www.iariw.org>

Developments in the measurement of government output and productivity in the UK

Emma Edworthy¹

Office for National Statistics
emma.edworthy@ons.gsi.gov.uk

Abstract

In July 2007, the UK's Centre for the Measurement of Government Activity (UKCeMGA) published its strategy to measure the quality of public services as a fundamental part of measuring non-market output and productivity. It outlined a quality measurement framework to which public services succeed in delivering intended outcomes, and how they respond to users' needs. Since then, UKCeMGA has carried out an ambitious work programme to improve the quality of measuring government output and productivity. This paper will report some recent developments in the measurement of non-market output in the UK, both in general and in some specific cases. The paper then focuses on the development of two possible models for measuring the quality of social security administration output, and discusses developments of quality measures in other service areas. It then goes on to discuss the use of capital services in general government accounts, and developments on the other input components.

*This paper presents the current stage of an ongoing work stream. As such its content is work in progress and we would welcome comments and suggestions.

¹ I am grateful to contributions from Victor Barzey, Emily Carless, Mark Chandler, Kato Kimbugwe, Helen Patterson, Alison Spence, Maria Tortoriello, Mike G Phelps and Mark Pont.

1. Introduction

The measurement of non-market output raises distinct challenges, compared with the measurement of market output. Given that there is no market, it means that we are unable to directly observe the value that people in society place on particular public service activities. Prices provide this mechanism for the market sector.

Because of the inherent difficulties in measuring the growth in public services output, the traditional approach was to assume that the output of this sector was equal to the amount spent on producing the output, referred to as the outputs=inputs measure. The economic justification behind this was that ‘rational’ governments should spend up to the point that the extra output produced from spending an extra unit was equal to the cost of that spending (marginal benefit from spending is equal to the marginal cost of spending).

However, there are clearly a number of issues attached to the outputs=inputs convention, not least because it assumes that productivity is unchanged. Additionally, it is unlikely that governments act ‘rationally’ or that they indeed have the fiscal capabilities to spend up to the point where marginal benefits equal marginal costs.

As a result of these issues, the UN System of National Accounts (SNA) and European System of National Accounts (ESA) ruled in favour of the direct measurement of public service output. In 2002 ESA95 required all member states to move to direct measurement of output for individual services by the time the accounts covered 2006.

From 1998, Office for National Statistics (ONS) started to measure public service output by direct methods. Public services in the UK account for over a fifth of gross domestic product (GDP). Putting this into perspective, this is larger than the share of manufacturing in the UK economy. Given the size of this sector and the growing emphasis on the calculation of direct output measures, the ONS commissioned Sir Tony Atkinson to conduct a review of the measurement of government activity. The *Atkinson Review of the measurement of Government Output and Productivity for the National Accounts (Atkinson 2005)* produced nine principles (see Annex A) that covered direct measurement of government output, measurement of inputs and measurement of public services productivity. One of the nine principles refers to the issue of measuring the quality of public service output. Principle B states ‘the output of the government sector should in principle be measured in a way that is adjusted for quality, taking account of the attributable incremental service to the outcome. This paper will concentrate on the issue of quality measurement.

The UK Centre for the Measurement of Government Activity (UKCeMGA) was established in 2005, within the ONS, to take forward this work. Measuring output and productivity of public services, such as health care and education, is important for public accountability. Taxpayers, users and providers of public services have an interest in how government spends our money on these services and in whether services are good value for money. UKCeMGA work feeds directly into UK National Accounts (NA). However, information used in the productivity articles may differ

from NA, by using more ‘developmental’ methods. This paper will seek to explain some of the developmental work taking place within UKCeMGA.

The paper concentrates the discussion on possible ways for measuring quality as part of output in social security administration, but also discusses developments made in health care and education. It presents two possible ways for going about a quality adjustment for social security administration, using two quality domains. We look at the issues involved with combining several quality domains into one quality measure. In addition, on the input side, we focus some of our discussion on the production of capital services for the non market sector.

Section 1 provides a general overview. Section two describes the work being carried out on measuring quantity of output. Section 3 looks at the work being undertaken to take in to account the quality of public services. Section 4 reviews the work UKCeMGA has driven forward on inputs. And finally Section 5 concludes.

1.1. Background

1.1.1 The Atkinson Review

The final report of the Atkinson Review (2005) proposed a number of principles, two of which are of particular relevance to this paper:

- Principle A: the measurement of government non-market output should, as far as possible, follow a procedure parallel to that adopted in the National Accounts for market output.
- Principle B: the output of the government sector should in principle be measured in a way that is adjusted for quality, taking account of the attributable incremental contribution of the service to the outcome.

GDP aims to represent the cumulative final value generated and added by the economy in the interactive and successive processes of production. If we apply principle A, it suggests that the output of the public sector should also be measured with reference to the value that it generates i.e. a measure of the attributable incremental contribution of output² to outcome³.

It is important to consider what the ‘unit of output’ for a public service actually is. Units of output could be defined in a way that reflects movements in outcomes. This can be complex in an area like criminal justice and health care where the most desirable state of society is ‘very little activity required’. Thus, if good policing could prevent crime, there would be very little need for prosecution, court and offender management services. It is therefore tempting to argue that an increase in an

² Outputs are the direct product of activities that contribute to outcome

³ Outcomes are the anticipated or actual effects of activities/outputs (among other things)

activity like prosecutions or prison nights shows a worsening outcome for society and so should be seen as reduced output, rather than increased (ONS 2008c).

The Atkinson review placed a lot of importance on developing well based measures of public service quality. Users of public services in the UK are likely to have a keen interest in the quality of the services they receive, be it the time they wait for an operation or their treatment by the police following reporting an incident.

Measuring quality change in the market sector also poses problems. It is easier to measure quality change for tangible goods like cars, however, intangible goods and services also pose issues with quality measurement e.g. how do we measure the change in quality of financial services or R&D services?

The Atkinson Review and the 'Handbook of Price and Volume Measure of National Accounts' (Eurostat 2001) recommended that public service output be measured in a way that adjusts for quality change. *Measuring Quality as part of Public Service Output* (ONS 2007) set out some guidance on this. It proposes that dimensions of quality should be defined for each service within an overall framework of:

- The extent to which the service succeeds in delivering intended outcome;
- The extent to which the service is responsive to users's needs

This paper will discuss the work carried out within this framework for social security administration.

1.1.2 Eurostat Guidance

The development of direct output measures for public services discussed in this paper have been carried out within the Eurostat framework. Eurostat (2001:4) set out criteria for assessing proposed methods:

- A methods are those that approximate the ideal as closely as possible
- B methods are acceptable alternatives: they are further away from the ideal but still provide an acceptable approximation
- C methods are too far away from the ideal to be acceptable. They would generate too great a bias or would simply measure the wrong thing

The criteria set by Eurostat differ depending on whether a public service is 'individual' or 'collective'. Individual services are those consumed by individual households. For example an eye operation is an individual service. It is specific to the individual who consumes it. Whereas collective services are those provided simultaneously to society as a whole, for example, defence. National defence is a classic public good in that its benefits are universally and unavoidably available to all its citizens in the home country, i.e. its effects are non-rival, non-excludable and non-

rejectable⁴. This presents a particular challenge for the measurement of collective goods.

For individual services an output indicator is an A-method if the following criteria are satisfied:

- All services that are provided to the users are covered
- The outputs are weighted by the costs of each type of output in the base year
- The indicator is defined in as much detail as possible
- The indicator is quality adjusted.

If these criteria are not fully satisfied the method becomes a B method. Methods which simply assume inputs are equal to outputs are C methods.

For collective services, classification of methods is broadly the same as for individual services with the following exceptions:

- Methods which simply assume inputs are equal to outputs are B methods for collective services.
- The use of volume indicators of activity is a B method.

It is not always easy to say that a particular public service is classified as ‘individual’ whilst another is classified as collective. Several public services that UKCeMGA measure are hard to classify specifically as either individual or collective. Some services combine elements of both collective and individual services, especially those in the Public Order and Safety classification of government functions (COFOG). For example, the aim of the police service is to provide the public with universal and continuous protection. But, in the event of a failure of this protection, the police also provide a service in clearing up crime, and this is delivered to individual victims (Atkinson, 2005).

For the areas covered in this paper, their National Account classifications are as follows:

- Social security administration: Individual
- Adult and Children’s Social Care: Individual
- Education: Individual
- Health care: Individual
- Public Order and Safety: Collective

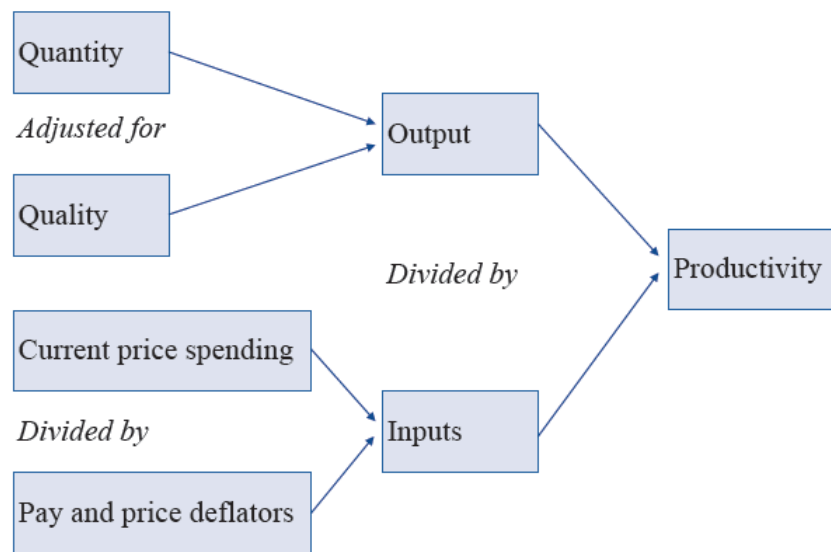
⁴ non-rivalry means that one person’s use of them does not deprive others from using them. Non-excludable means that if one person consumes them it is impossible to restrict others from consuming them. Non-rejectable means that individuals cannot abstain from their consumption even if they want to.

1.2 Productivity and Components of Productivity

1.2.1 Productivity

One of UKCeMGA's primary aims is to develop better measures of government input, output and productivity. Inputs (labour, goods and services and capital) are the resources used in activities which lead to outputs that are directly attributable to those activities.

Figure 1: Components of Productivity Change



1.2.2 Quantity

It is important to carefully consider what the 'unit of output' for a service actually is. As stated in section 1.1, ideally the output of the public service should be measured in a way that takes account of the attributable incremental contribution of the service to the outcome. Most of the current measures used for public service outputs are quantities of 'activities' – for example a pupil attendance at school for a day or the number of income support claims processed. These measures of activities tell us that something has happened, but they do not tell us how successful it was in achieving its intended purpose, or whether the individual claiming income support, for instance, found the experience satisfactory. The ideal way would be to define a unit of output that took into account the quality of that unit.

1.2.3 Quality

Establishing the Principles (ONS, 2006b) emphasised the need for a framework for quality measurement, to be used consistently for different services. UKCeMGA's work on productivity analysis is particularly focused on aspects of quality that may

change over time, and the techniques for incorporating measures of quality into time series. There are a number of issues with incorporating quality which will be addressed in this paper. For instance, is a 1 per cent increase in quality equal to a 1 per cent increase in quantity and how do we combine different measures of quality (ONSg) – for example, how do we weight together a timeliness and an accuracy index for the administration of social security benefits.

1.2.4 Current Price Spending

Current price spending comprises of labour, goods and services and capital. Principle F in the Atkinson Review Final Report (2005) states:

- The measurement of inputs should be as comprehensive as possible, and in particular should include capital services; labour inputs should be compiled using both direct and indirect methods, compared and reconciled.

The simplest direct measure of labour input would be a count of full time equivalents but this approach is problematic as it will hide changes in average hours worked and will mask differences in quality of labour. ‘Hours worked’ is the most appropriate variable for labour input in productivity measurement because it bears a closer relation to the amount of productive services provided by workers. However, measurement of hours actually worked is difficult and so hours paid and full-time equivalent persons can provide reasonable alternatives.

Indirect approaches to the measurement of labour input involve deflating current price expenditure on labour by pay indices. Indirect measures tend to be easier to calculate, with no need for additional data, provided that appropriate deflators exist.

Expenditure on goods and services covers things like NHS drugs, class room equipment, consultation services. These are measured indirectly, using appropriate deflators (see section 1.3.4).

Given the durability of capital (assets last a number of years), the way capital is measured as an input into production cannot be based on annual expenditure on capital assets. The use of capital assets is estimated using a capital consumption measure provided by National Accounts. Capital consumption is a measure of the extent to which the capital stock is used up from year to year. As stated in principle F, the ideal measure of capital should be capital services⁵. However, whereas, there has been extensive work on measuring capital services for the market sector (Dey-Choudhury and Wallis, 2007), work on measuring it in the government sector is relatively new and ongoing. Section 4 discusses UKCeMGA developments in measuring capital services for the government sector.

⁵ Capital services refer to the flow of productive services from the stock of capital. Capital services recognises that the same stock of capital may be used more or less intensively (capacity utilisation).

1.2.5 Pay and Price deflators

Principle G of the Atkinson review final report states:

- Criteria should be established for the quality of pay and price deflators to be applied to the input spending series; they should be sufficiently disaggregated to take account of changes in the mix of inputs; and should reflect full and actual costs.

It is important to develop deflators specific for each area i.e. health, education, public order and social security administration. Work on deflators is much more developed in some areas than others. For example in education UKCeMGA has developed pay deflators specific to support staff in schools. However, in social security administration we still rely on the more general Index of Labour Costs per hour (ILCH) for public administration personnel⁶. Work in this area is ongoing. Given the variety of expenditure on goods and services it is also important that we develop price deflators in this area at a reasonable level of disaggregation e.g. computer software, consultants etc. Finally, deflators for capital are obtained from National Accounts.

1.2.6 Productivity:

UKCeMGA aims to produce a standard Total Factor Productivity measure (TFP). The main difference from the equivalent estimation of market sector TFP is that we estimate the volume of output directly. Output, not adjusted for quality, is the sum of cost weighted activities (Cost Weighted Activity Index). The volume of inputs are weighted by the shares of total input spending.

The following three sections look at quantity, quality and inputs.

⁶ ILCH goes beyond existing earnings indicators to include non-wage costs.

2 Quantity

This section describes the progress made in measuring output in the following areas:

- Health Care
- Adult Social Care
- Education
- Children's Social Care
- Social security administration
- Criminal justice system
- Fire and Rescue services
- Defence

2.1 Health care

Health care is classified as an individual service and output is currently measured in the National Accounts through a cost-weighted activity index covering most types of National Health Service (NHS) activity in England. Since June 2005, data from Northern Ireland have been included.

The index captures 2,500 NHS activity types in England. Activity includes hospital inpatient, day case and outpatient episodes, distinguished by health care resource group; General Practice (GP) and practice nurse consultations and prescriptions; dental treatment, sight tests, and ambulance journeys. This covers more than three quarters of all NHS activity, by value. Coverage is around 80 per cent by expenditure for both England and Northern Ireland.

The last release of the *Public Service Health Care Productivity article* (ONS 2008a) included improvements to the method for measuring change in the quantity of health care output:

- a new approach to measure GP prescribed drugs in order to take proper account of the way drug prices fall when they come off patent, and alternatives are available;
- more reliable underlying data
- more appropriate geographic coverage and an unrestricted collection period
- separate analysis of output and productivity for Hospital and Community Health Services (HCHS) and Family Health Services (FHS)
- the use of FHS Drugs deflators.

2.1 Adult Social Care

The output of Adult Social Care (ASC) is based on the quantity of social care activities measured either in terms of time (number of weeks or residential care) or number of items (number of meals provided). The level of each activity is adjusted to allow for the proportion paid by local government. The index covers 23 activities in

England and 17 in Scotland. The addition of Scotland into the measure since 2002 brought the total coverage of the measure to around 85 per cent of net expenditure on adult social care across the UK.

The index covers a variety of services: assessments of need, day care, home care (personal care at home, provision of meals and, in England also provision of equipment) and provision of residential care home places.

2.3 Education

Education is also classified as an individual service and is currently measured by using a cost weighted index incorporating the following activities for each of the 4 countries of the UK:

- Pupil attendance at primary schools, secondary schools (including Academies and City Technology Colleges), and special schools;
- Full time equivalent pupils at pre schools including those funded by government in the private, voluntary and independent sector;
- Full time equivalent students undertaking initial teacher training and health professional courses;
- Full time equivalent students aged under 19 undertaking further education including education in sixth form colleges. (this has been included for the first time in the public service education article published in April 2008).

The measure of the quantity of education will be developed over the future to incorporate:

- Time spent in schools outside teaching hours
- Higher education

2.4 Children's Social Care

The quantity of children's social care is partly a direct measure and partly outputs = inputs for all 4 countries in the UK. Social care services for 'Children looked after' (a legal term) by the state are measured directly using the number of days spent in different residential settings⁷. The remainder of children's social care is currently measured by assuming outputs = inputs.

- Further research will be undertaken to investigate whether more of social care services can be measured directly.

⁷ Residential settings include children's homes, secure welfare accommodation, fostering services and other looked after setting (which may include family members)

2.5 Social Security Administration

The output of Social Security Administration is measured using a cost weighted activity index. The index developed in the latest *Public Service Social Security Administration Productivity article* (ONS 2008f), covers the administration of twenty five activities⁸. This article included a number of improvements to the output measure:

- It provided ‘claims’ and ‘load’ information for child benefit. In the previous article (ONS 2006) data was only available for ‘load’
- Child support data from the Child Support Agency was included in the cost weighted activity index

2.6 Criminal Justice System

Unlike the other areas discussed above, the criminal justice system contains areas that have both individual and collective elements. In the UK National Accounts, all Criminal Justice System output that is considered as an individual service (Legal services commission, crown prosecution service, crown courts, magistrates courts, prison and probation services) is currently estimated using activity indicators without quality adjustment. Police is considered a collective service and is still measured as outputs=inputs. The method of output measurement for each CJS agency is briefly outlined below:

- Police: Currently measured using deflated expenditure
- Legal Services Commission: Measured directly using a measure of cases in which legal aid is administered broken down into three categories (lower standard fees, higher standard fees and non standard fees plus exempt). These are weighted together to form one index using base year expenditure weights.
- Crown Prosecution Service (CPS): Measured directly using a measure of the number of defendants for which the CPS provide prosecutors. These are then divided between cases in the Magistrates’ and Crown Courts, and are weighted together using base year expenditure weights.
- Crown Courts: Measured directly using a measure of Crown Court cases broken down into three categories (committals for trials, cases for sentence, appeals). These are weighted together to form one index using base year expenditure weights.

⁸ Social security administration covers the following activities: Income support claims and load, Social Fund grants and loans, Job seekers allowance claims and load, retirement pension claims and load, pension credit claims and load, future pension forecasts, international pension credit and load, Incapacity benefit claims and load, Disability living allowance claims and load, Attendance Allowance claims and load, Carers allowance claims and load, other benefits claims, child support , child benefit claims and load and housing benefit claims and load.

- Magistrates' Courts: This is measured directly. A weighted caseload is available from which an output index can be calculated. Completed proceedings are counted in 14 case types. Weightings (reflecting the average time to complete each type of case) are then applied to each case type to provide an overall unitary value of caseload.
- Probation: Measured directly using a cost-weighted activity index based on a number of activities (probation starts, community service, licences etc.). The output measure uses weighted implied workload hours for these activities based on activity sampling exercises carried out in 1997 and 1998.
- Prison: Measured directly using a simple count of the number of nights spent in prison. This output measure is not weighted in any way.

There are a number of limitations with the current output measures. Several examples are given below (ONS 2008c):

- When police productivity is considered, the current output measure (outputs=inputs) assumes change in police productivity to be zero.
- Regarding CPS output it would be desirable to have activities and unit costs further subdivided (e.g. between guilty and non-guilty pleas) in ways which reflect variations between relatively cheap and relatively expensive cases.
- The current probation output measure is extrapolated from a short time series and based on activity sampling exercises conducted ten years ago. A lot has changed in terms of offender management in this time.
- The current measures take no account of a change in the quality of service delivered.

There is a plan to tackle some of these issues and take these measures forward. These include:

- developing output measures with relevant, properly defined 'units of output'
- exploring the use of Home Office Police Activity Based Costing data to produce a direct measure of police output.
- Assessing the viability of using a more disaggregated measure of CPS output.
- exploring the use of newly available unit cost data from the Ministry of Justice (MoJ) and the probation and prison services to produce a cost-weighted activity index for these services.

2.7 Fire and Rescue Services

The output of the Fire and Rescue Service (FRS) has been measured in the National Accounts using a cost-weighted activity index since 2001. Although FRS is classified as an individual service, it can be argued that it has collective elements. However, these are currently not identified separately in National Accounts. Data from three broad categories: fire response, fire prevention and special service incidents are used. The first is a measure of how many fires the FRS has attended during the period in question; the second is a measure of the prevention work⁹ carried out; and the last is a measure of how many non-fire incidents¹⁰ the FRS attends. As one would expect, fire response has the largest weight of the three, followed by special service incidents.

The key issue with the measurement of the FRS work is the perverse relationship between fire response activities and fire prevention activities; given the current methodology, an increase in the effectiveness of the latter leads to a reduction in the volume of the former. Since fire response work has a large weight in the index, even a small fall significantly reduces the output of the service overall despite the fact that this is attributable to an improvement in the FRS work.

In the UK the FRS now has responsibility for a large proportion of the work into National Resilience against major incidents such as terrorist attacks and natural disasters. At present, this work is going unmeasured; developing measures in this area is difficult. Firstly, it would require the measurement of the counterfactual. Secondly, (and this can be applied to the service as a whole) it requires measurement of the capability of the service to respond to incidents: as with the counterfactual this too can vary depending on the assumptions made. Since the aim is to measure the service to the UK as a whole, the assumptions made should match the preferences and values of the population as closely as possible.

2.8 Defence

While significant developments have been made in measuring other areas of public sector productivity, such as Education and Health, Defence is one of the least developed areas, and one of the most difficult conceptually. Defence Activities can be broadly described as the administration, supervision and operation of military affairs and land, sea, air and space defence forces. According to the COFOG, the defence division is made up of 5 groups:

- Military defence;
- Civil defence;
- Foreign military aid;
- Research & development related with defence; and
- Defence not elsewhere classified.

⁹ Namely, community fire safety where firefighters visit local schools and community groups to give fire safety lectures; and fire inspections where firefighters visit homes and businesses to assess the risk of fire and install smoke alarms.

¹⁰ Non-fire incidents include road traffic accidents, animal rescue, flood rescue and rescue from height among other things.

In the UK National Accounts Defence is currently measured using the output = input convention. This satisfies European requirements as Defence is a collective service.

Defence inputs are split into three sections:

- Labour – Soldiers, RAF pilots, Royal Marines and administrative staff.
- Goods and services – Medical supplies (for field hospitals), munitions and uniforms.
- Capital – Air bases, computers and naval docks.

The current measure of defence labour is a complicated one; it is a pseudo-direct measure in that it shows the expenditure on labour in a given year assuming each member of staff is paid the average wage as calculated in the year 2000. There are a number of problems with the current labour input measure:

- It takes no account of changes in pay over time and assumes all staff are paid the same.
- The number of FTE staff is counted but no account is taken of skill. Measuring workers' skill is crucial to the measure of inputs since all other things being equal workers that are more skilled contribute more to production than their less skilled counterparts do.

Based on the data available from Defence Analytical Services Agency (DASA) and the MoD there are a number of improvements which could be made to the current labour input measure:

- A direct labour measure could be composed using data on the number of FTE staff broken down by Service vs. civilian personnel, rank, and trained vs. untrained staff. These measures could be weighted using civilian and service personnel pay scales.
- An indirect measure could be developed where expenditure on staff is deflated using a Service specific pay deflator (based on military salary indices) and a civilian deflator based on the Index of Labour Costs per Hour (ILCH)¹¹ for public sector.

As with labour, goods and services are broken down into the same two categories in the National Accounts: defence and 'other'. Both of these are measured indirectly by deflating current price expenditure. There are a number of limitations with the current measures which UKCeMGA are seeking to improve. The major ones being that the deflators used are not specific enough and that the classification used in the SNA are problematic when it comes to the distinction between defence goods and services and defence capital.

¹¹ The Index of Labour Costs per Hour is an experimental statistic compiled by the ONS. It goes beyond existing earnings indicators (for example the AEI) to include non-wage costs, e.g. the costs of benefits in kind.

The 1993 SNA draws a distinction between two types of durable goods used by the military:

- Those “that are used in much the same way as in any other type of production”. For instance, durable goods such as airfields, docks, or other facilities used as bases. These are durable goods that can potentially be used for civilian purposes and are currently treated as fixed assets (capital); and
- “Destructive military weapons designed for combat”. These goods are also referred to as Single Use Military Equipment (SUME) because they cannot be used by civilian organisations. Because they are considered destructive, they are not treated as fixed assets. They are instead, treated as goods and services.

Under this classification system all SUME items are classified as goods and services and so a number of items which we would usually classify as capital, due to the fact that they are not used up in a production cycle, are classified as goods and services, for example warships and tanks. There is currently a proposal to change the treatment of defence assets with proposed revisions to SNA93 – SNA2008 as follows:

- The acquisition of weapons systems should be recorded as gross fixed capital formation, whilst the acquisition of ammunition is to be recorded as an expense that would result in a change in goods and services.
- Treating weapons systems as fixed assets implies the need to estimate their expected service lives, and capital services.
- Gross fixed capital formation for defence equipment should be presented separately from other types of gross fixed capital formation.
- If weapon systems classified as fixed assets are destroyed in combat, their disappearance should be recorded in the “other changes in volume of assets account,” in the same manner as other fixed assets destroyed in war.

The changes to SNA93 will affect defence inputs (goods and services and capital) leading to an increase in capital stocks on the one side and a reduction in expenditure on goods and services on the other and will require the revision of current capital inputs measures

While the output = input convention satisfies European requirements it is still valid to seek out direct output measures for at least some parts of the service (however, some areas such as counter-terrorism are likely to remain as deflated inputs measures).

There have been a number of suggestions as to how a more direct output measure could be created to replace the current output = input line. The output provided by the military could be thought of as the capabilities the Armed Forces provide, which in turn are determined by the standing commitments and objectives of the MoD, the nature and level of threats to national security and the capabilities of both adversaries and allies. A number of measures could be used to assess capabilities, the most suitable of which is the readiness of the Armed Forces to respond to emerging operations, as reported by the MoD.

Activities lead to outputs (and capabilities) and so where it is not possible to collect output data, it is important to consider collecting data on activities because they reflect what the non-market units are actually doing with their inputs and will

therefore be closer to output, than deflated inputs. As activity data may be difficult to collect activity can be proxied by training data for which measures tend to be more readily available.

If a direct output measure cannot be developed the current output = input approach could be improved by using a more direct proxy for the value of output, that is the operating costs MoD report against their 3 objectives set out in the PSA for SR2004:

Achieving success in tasks we undertake;
Being ready to respond to the tasks that might arise; and
Building for the future.

However, this measure would require some quality adjustment to take into account to what extent the objectives have been met.

3 Quality

As stressed by principle C of the Atkinson report (see Annex A) and the international guidance governing National Accounts (SNA93), quality change should be treated as an integral dimension of output. In section 2 we discussed the progress made by UKCeMGA in improving direct output measures. This section discusses the quality adjustments that have been applied to output. Given that there has been extensive discussion in quality adjusting health care and education output, we will be focusing our discussion on an area that has not been so well discussed, namely social security administration. However, we will conclude with a brief description of the progress made in health care and education quality adjustments in the UK.

Measuring Quality as part of Public Service Output (ONS 2005) detailed further development work to be done to improve the way we measure quality. The development programme is a long term one. The work described in this section builds on the work done to date and goes in to detail about the most recent piece of developmental work, the proposed method for the measurement of the quality of social security administration within the UK. This section will discuss the quality measure used in the three areas above, but it will concentrate on the methods developed for social security administration.

3.1 Social Security Administration

In section 1 we touched on the ONS quality measurement strategy (ONS 2007). It proposes that dimensions of quality should be defined for each service within an overall framework of:

- the extent to which the service succeeds in delivering intended outcomes; and
- the extent to which the service is responsive to users' needs.

While there is much interest in measuring aspects of quality of service in different public services, and improving quality for users, there is much less experience of constructing a single measure of output combining several dimensions of quality and volume, for a service. Without such a measure, it is not possible to say how value for money has changed over time.

Where public services are delivered to individual users, measurement of quality as part of output is akin to considering the way quality affects price. We may not know to what extent different aspects of a luxury holiday contribute to the price (for example, comfort of bedroom fittings, amenities in hotel and grounds, free taxi from airport, and range of choice at breakfast), but we recognise that individuals get 'utility' from these features and make a choice.

A consumer may take separate account of different aspects of quality. They may put a very high weight on 'extensive hotel grounds', and a lower weight on 'range of choice at breakfast'. Similarly, measurement of public service output needs to consider

whether all aspects of quality are equally important to users. This is compounded by the fact that (as with holidays), different aspects of quality matter to different people at different times.

For social security administration, quality can be considered for ‘impact on intended outcome’ (payments reach eligible recipients on time, without error or fraud) and for ‘user responsiveness’ (ease of making claims, getting information, resolving problems).

Social security administration aims to ensure all those eligible for benefits:

- get paid exactly what they are entitled to, as soon as they are entitled to benefit payment;
- should be satisfied with the information provided, ease with which claims are made, and with the operation of The Appeals System against incorrect payments (overpayment and underpayment) .

These correspond to the two general dimensions of quality set out in the strategy. ‘Impact on intended outcome’ is understood, for social security administration, to mean payments reaching eligible recipients on time, without error or fraud. ‘User responsiveness’ is understood as ease of making claims, getting information, resolving problems etc.

An alternative view is that ‘impact on intended outcome’ is the effectiveness of social security policies in achieving outcomes like reducing poverty (absolute or relative), maximising participation in the workforce and incentivising desired behaviours including personal saving. In other words, it may be regarded as a collective rather than an individual service. Such policy objectives can also be achieved through other means, including the structure of taxation, supply side labour market policies and some aspects of other public services including education, healthcare and personal social care. For this article, the emphasis is on the ‘administration’ of social security rather than the way in which the system as a whole contributes to socially desired outcomes.

Accounting for Quality Change in Estimates of Social Security Administration Output and Productivity (ONS 2008h) considered only the first of the two quality dimensions: impact on intended outcome. Data are available that can be used to measure quality in such dimensions. Impact on intended outcome can be subdivided into accuracy and timeliness of payment. Some data are available about each of these, and the remainder of this paper sets out the theory of how the data can be used to construct a quality adjustment. A more complete approach would also include evidence on user responsiveness. Usable data are currently unavailable on responsiveness to users’ needs. Ad hoc surveys have been undertaken as part of performance management (identifying such things as proportion of phone calls answered within a certain number of rings, and testing customer satisfaction with services provided), but these surveys are generally incomparable over time, and unsuitable for use as a quality adjustment.

3.1.1 Domains of Quality

3.1.1.1 Accuracy

For most of the activities in the current output measure, accuracy describes the extent to which benefit recipients receive the correct amount of money to which they are entitled. Inaccuracy can arise for several reasons:

- a) deliberate and successful attempt to defraud the benefits system;
- b) mistakes made by a claimant in providing information;
- c) failure by a claimant to provide information;
- d) failure by a claimant to update information in the light of changing circumstances,
- e) or errors by staff in processing information provided; and
- f) clerical error and other human errors (e.g. medical judgements in assessing disability claims)

Inaccuracy is split between three causes: fraud; claimant error and official error, (DWP 2007) each of which is detailed below and referenced to one of the reasons listed above:

Fraud (a): This includes all detected cases where:

- the basic conditions for receipt of benefit, or the rate of benefit in payment, are not being met;
- the claimant can reasonably be expected to be aware of the effect on entitlement; and
- benefit stops or reduces as a result of the review.

The fraud category comprises cases where claimants deliberately misrepresent their circumstances, or fail to notify changes in their circumstances, with the intent of obtaining benefit to which they are not entitled. It includes cases where fraudulent activity is substantiated through third party verification or an admission is obtained from the claimant.

Claimant (or Customer) Error (b,c,d): The claimant has provided inaccurate or incomplete information, or failed to report a change in their circumstances, but there is no fraudulent intent on the claimant's part.

Official Error (d,e): Benefit has been paid incorrectly due to inaction, delay or an incorrect assessment by the administering department.

Inaccuracies can lead to either underpayment or overpayment of benefits. Underpayment clearly leads to losses for recipients, both in terms of money received as well as time and effort, if they realise the problem and try to resolve it. Overpayment, if undetected, leads to losses to the benefits system but also potential difficulty for benefit recipients if detected and reclaimed. In addition, dealing with known inaccuracies adds to the administration costs of the relevant departments.

Fraud and error affect the credibility and reputation of the service overall, so quality can be seen as an issue for wider society, not just for recipients.

With regards to departmental errors, in using this type of measure one has to assume that the checking procedures from which the measure comes are applied consistently, both across types of benefit and over time.

There are five areas of accuracy that are considered:

- underpayments caused by official error;
- overpayments caused by official error;
- underpayments caused by claimant error;
- overpayments caused by claimant error; and
- overpayments caused by fraud (note that there is no underpayment due to fraud).

3.1.1.2 Timeliness

Timeliness refers to the time it takes to process a benefit application or to process new information about a claim. Timeliness can be measured either as the average processing time or by the number of cases dealt with in a threshold limit.

There are pros and cons associated with both threshold and average clearance time measure. Threshold measures are prone to discontinuity in the time series as thresholds are moved and do not capture the experience of the full distribution of users. Average processing times will mask vastly different customer experiences. However, they are seen as preferable because they capture improvements and do not have the same discontinuity issues as threshold measures.

Given that benefits are paid from the date of entitlement, the actual impact on claimants for waiting may not be severe. Additionally, small improvements in timeliness may not have much impact on claimants. The additional wait caused by small differences in waiting times will not always cause a claimant to take action (eg to take out loan), and a 'Social Protection Fund is provided by DWP for people needing to access immediate financial assistance.

3.1.2 Proposed methods

When considering two different domains for quality an issue arises with how to combine the domains. Where there are multiple attributes, these attributes need to be weighted to form an overall quality adjustment. *Adjusting Measures of Public Service Output of Quality Service* (ONS 2008g) suggests there are a number of sources we could consider when choosing our weights:

1. User preferences – In some cases the preference of users might provide the most useful indication of what dimensions of quality are most important. This would be the type of weighting that would be most akin to market sector quality where prices reflect the perspective of the producer. This may be suitable for services where the user instigates the use of the service and therefore has the most idea about its utility.
2. Provider preferences – In some cases the preference of providers might best reflect the relative importance of domains of quality. This might be the case where the ‘user’ has no control over the service, or where the user lacks expert knowledge of the characteristics of quality of the service and relies on the service provider as agent.
3. Population preferences – In some cases the relative importance of domains to the general population might best reflect quality. For example, the relative importance of aspects of quality for the prison service should reflect the views of the population as actual or potential victims of crime, not just those of prisoners. Preference studies used in health and social care have generally used population preference rather than those with a particular illness or condition, since those individuals adapt to their condition; population views on importance of avoiding illness or disability are regarded as more relevant. The QMF study is now identifying population preferences for the relative importance of the domains of adult social care, using a population group rather than actual social care users.
4. Expert review – an alternative source might be through expert advice where neither the users nor providers have the full perspective on how the service meets its aims and where the population would be unlikely to be able to provide a view on the relative importance of different dimensions of service.
5. Cost/ Resource – In some cases it might be possible to identify the cost associated with changing quality in each of the domains or the amount of resource that is put into different dimensions of a service. This comes close to a provider perspective on relative importance of different aspects of quality – providers are likely to adopt quality changes they perceive as important and cheap. However they may also not adopt changes that they consider are important if they are expensive. Data on cost of quality change, based on those adopted, could be regarded as a proxy for the preferences of the provider (or budget controller). An example could be the introduction of a new computer system to facilitate and speed up bookings of appointments. Alternatively, if data is available on the amount of staff time that is put into different dimensions of a service then this could be used to weight the dimensions. For example, the amount of teacher contact time that is used for lessons in subjects to increase attainment versus the amount that is spent on personal and social issues relating to the other domains of the Every Child Matters framework. This type of weighting could be affected by increased costs of improvements as the quality comes closer to ‘perfect’.

The following section discusses two specific proposals for combining accuracy and timeliness in to a single quality adjustment.

3.1.2.1 Method 1:

Method one is a multiplicative model. Although this makes the combination of the two domains, accuracy and timeliness, relatively straight forward, it does have one major weakness in that it forces us to make assumptions about the relative importance of each domain.

At present definitive weights for the relative importance of timeliness and accuracy are not available. Whereas one benefit recipient may feel that an accurate payment is more important than a timely one, someone else may view a timely payment as more important than an accurate one. This example highlights the subjective nature of assigning weights.

There are other weighting issues with this method. Inaccuracies can lead to either underpayment or overpayment of benefits. We can make a case that both forms of inaccuracies inconvenience recipients. But how do we value which is more inconvenient? Whereas underpayment is clearly a loss to the individual, overpayment can be seen as an inconvenience to the individual if it is detected and recouped from the administering department. There are also wider issues for society and tax payers if overpayments are a large percentage of inaccuracies. Since there is no research to suggest which type of mispayment should have the highest weight in our quality measure, we assume they carry an equal weight. Additionally, inaccuracy is classified into three types (explained in more detail below) fraud, customer error and official error. Once again in the light of no evidence to suggest one type of inaccuracy should be weighted more highly or lower than another we assume they each carry an equal weight.

The following approach assumes a particular relationship between quantity and quality, namely that a 1 per cent rise in one is equivalent, in output terms, to a 1 per cent increase in the other. This relationship is open to some debate, and is being considered further as part of UKCeMGA's wider project on adjusting output for quality change (ONS 2008).

This model starts with the standard Laspeyres volume measure (cost weighted activity index) that is used for measuring government output:

$$L_{0t} = 100 \frac{\sum_i p_{0i} q_{ti}}{\sum_i p_{0i} q_{0i}}$$

where:

i represents a type of activity

t represents the time period

p_{0i} represents the average unit cost of activity i in the base period, 0

q_{ti} represents the quantity of activity i in time t

Quality is then included through adjusting the quantity in time t by the change in quality since the base period. That is:

$$L_{0t}^* = 100 \frac{\sum_i p_{0i} q_{ti} K_{0ti}}{\sum_i p_{0i} q_{0i}}$$

where $K_{0ti} = \frac{k_{ti}}{k_{0i}}$

and k_{ti} represents the quality in time t for activity i.

The paper now considers the two dimensions in turn.

Accuracy Measure

For accuracy, we need data to reflect the extent to which the service delivers accurate payments. Accuracy can be considered in two ways – the number of accurate payments, or the total value of payments that is made accurately. Using simply the number of inaccurate claims does not reveal the extent to which payments were in error - differing levels of inaccuracy may not be thought to represent different quality levels. Initially, we consider that any data that cover any of these aspects could be used to develop a measure of quality over time.

We can take the proportion of money that is either overpaid or underpaid according to the five types of error described earlier in this article, and calculate a weighted sum of mispayments, m_{it} as follows:

$$m_{it} = \sum_{j=1}^5 w_{ij} e_{ij}$$

where:

i represents type of claim

t represents time period

j represents the 5 types of mispayments described in Section 3.1.1

e_{ij} represents proportion of money mispaid

$w_{ij} \in [0,1]$ is the weight, as described in the previous section, for each of the mispayment types

A similar formulation would be valid either on ‘number of payments’ or ‘value of payments’. Clearly if the proportion of overpayments and underpayments goes up this represents a decreasing quality of service. We would want an index for accuracy that increases as quality of service increases. One such index is to use the percentage ‘accurate’ i.e. not known to be a mispayment rather than the raw overpayment or underpayment percentage:

$$1 - m_{ti}$$

Change in quality between period 0 and period t is therefore:

$$\frac{1 - m_{ti}}{1 - m_{0i}}$$

Clearly if the proportion of overpayments and underpayments goes up then the quality of the service is decreasing. We want to develop an index for accuracy that increases as the quality of the service increases and decreases as quality falls. Therefore we chose to use the percentage 'accurate' i.e. not known to be a mispayment rather than the raw overpayment or underpayment percentage:

Timeliness Measure

Timeliness data can be in one of two forms – some kind of average clearance time for processing claims, or in the form of the number/proportion of claims that are not dealt with within a specific time (which could be relative to a target/deadline). Further considerations here include whether a payment that is one day late is as poor quality as one which is ten days late. Such threshold measures disguise the distribution of actual times much more than average measures.

The longer it takes a benefit to clear, i.e. be processed and paid, the longer the claimant is waiting. An increase in the average clearance time therefore represents a fall in quality, so we want an index that increases as clearance time drops. In a similar way for accuracy, we choose the reciprocal of the average clearance time:

$$\frac{1}{\bar{c}_{ti}} \text{ or } \bar{c}_{ti}^{-1}$$

where \bar{c}_{ti} represents the average clearance time in days of benefit i in time t

The change in quality between period 0 and period t is therefore:

$$\left(\frac{\bar{c}_{ti}}{\bar{c}_{0i}} \right)^{-1} = \frac{\bar{c}_{0i}}{\bar{c}_{ti}}$$

The index is constructed so that a fall in clearance time counts as an increase in the quality index. All benefits have been weighted equally. This could be questioned since recipients of certain benefits will suffer more than others if they have to wait too

long. Ideally an index could use different weights to reflect the relative hardship and inconvenience caused to benefit recipients in different circumstances.

Timeliness measures of quality are most relevant to claims, though they could in principle be developed also for processing times on changes in circumstances, requests for information, and indeed the physical transfer of money into claimants' bank accounts.

Accuracy measures are for load, but could in principle be applied to claims, since payment may be wrong from the start. Data are currently collected by sampling benefits already in payment so are relevant to the load measures. It should be recognised that mispayment may have arisen at the 'claims' stage.

These arguments point towards an index where claims activities are adjusted for timeliness and load activities are adjusted for accuracy, assuming evidence is available. In other words, in quality adjusting the output indices, for claims we assume that timeliness is fully weighted and accuracy has no weight, and vice versa for load.

In more general terms, it is important to recognise the need to choose and justify separate weights for timeliness and accuracy to be applied to both claims and load. We have adopted the simple initial approach as described in the previous paragraph.

In this context we consider only one quality domain for each of claim and load. The quality-adjusted output indices for each are then defined as:

$$L(c)_{0t}^* = 100 \frac{\sum_{i=1}^n p_{0i} q_{ti} \left(\frac{\bar{c}_{ti}}{\bar{c}_{0i}} \right)^{-1}}{\sum_{i=1}^n p_{0i} q_{0i}}$$

for claims, and

$$L(l)_{0t}^* = 100 \frac{\sum_{i=1}^n p_{0i} q_{ti} \left(\frac{1 - m_{ti}}{1 - m_{0i}} \right)}{\sum_{i=1}^n p_{0i} q_{0i}}$$

for load.

As explained above, this model assumes equal weights within each of the following set of quality domains:

1. the relative balance of the importance of overpayment and underpayment; and
2. the relative importance of each cause of error.
3. the relative importance of timeliness and accuracy
4. the relative importance of quality and quantity.

One way to test these assumptions would be the use of sensitivity analysis. Sensitivity analysis involves evaluating the impact of using different sets of weights. If results are quite different based on different weighting assumptions, we would need to better justify the assumption we made about weights. Conversely, if it was shown that the results were insensitive to the particular choice of weights, it would matter less what we assumed about the weights. This is an area of further development.

3.1.2.2 Method 2

A second way to regard quality adjustment is to differentiate between the different levels of quality that a service may have and consider only those claims that are processed accurately and in a timely manner as part of output. In the formation of the index number, the output is distinguished into cases that have a particularly quality level. The quantity of output in each of those groups is then adjusted for the level of quality in that group.

If we regard k_{it} as being the quality value in time t of output i , and adjust each of the quantity measures q_{it} we reformulate (1) as follows:

$$L_{0t} = \frac{\sum_i p_{0i} q_{it} k_{it}}{\sum_i p_{0i} q_{0i} k_{0i}}$$

Then, the index increases if more of the output moves from lower valued categories to higher valued categories and vice versa.

Atkinson (2005) discusses this approach, drawing an analogy with the case of defining the quality of the output of postal services by taking account not just of letters handled but also whether the letters were delivered to the correct address and on time. Defining output as letters delivered to the correct address on time combines both measures of quality that we're considering (timeliness and accuracy) with the original quantity unit (namely "letters") to get a single measure. It gives a natural way of weighting the quality elements together. It also provides an reason why a one per cent increase in the quantity of letters handled should be of equivalent value to a one per cent increase in letters delivered to the right address: namely that both represent a one per cent increase in effective units of output.

A simple way of applying this principle in the case of Social Security Administration is to regard accurate, timely payments as having value, and those that are late or incorrect as having no value. In other words, we consider only those claims that are processed accurately and in a timely manner as part of output.

This gives a revised output measure as follows.

$$q_{ii}^* = q_{ii} (1 - m_{ii} - v_{ii} + a_{ii})$$

Where

v_{ii} is the proportion of untimely payments

m_{ii} is the proportion of mispayments and

a_{ii} is the proportion of claims that are both mispaid and untimely.

Note that the unit of output is still claims - but now only timely accurate claims count.

This approach may be thought extreme because a claim either contributes entirely or not at all to the quantity measure. We may choose to consider degree of lateness or degree of accuracy as meriting partial inclusion of some claims in the quantity measure. The simplest way to achieve this based on the formulation above is to simply subtract some proportion of the “bad” claims rather than the full amount. However, exactly how much to subtract becomes a matter for judgement and discussion.

For mispaid claims the simplest adjustment is just to take a proportion of the claims. For untimely claims the best approach if there were data would be to weight claims according to the degree of lateness, with a zero weight for claims processed within a certain number of days and weights increasing towards and eventually reaching one for very late payments.

The above expression is then modified as follows:

$$q_{ii}^* = q_{ii} \left(1 - \alpha_{ii} m_{ii} - \sum_{j=1}^n \beta_{ij} v_{ij} + \sum_{j=1}^n \min[\alpha_{ii}, \beta_{ij}] a_{ij} \right)$$

$\alpha_{ii}, \beta_{iid} \in [0,1]$ are the factors to represent the degree of exclusion of a particular measure from the quantity measure. d is the number of days taken to process the claim.

Note that claims that are both untimely and inaccurate need to be added back in, to avoid double counting, but with the lower of the relevant weights. For example, if inaccuracy is weighted more highly than untimeliness then claims that are both are added back in as if they were untimely.

If α_{ij}, β_{ijd} for all j are all 0, then this measure is simply the original quantity measure. Likewise, if α_{ij}, β_{ijd} for all d are all 1 then the quality adjusted measure is simply the number of timely accurate claims processed. Intermediate cases allow for the possibility that different types of mistake can be counted to some degree as output. But the overall measure remains a measure of “effective” claims processed. In the absence of any evidence upon which to base weights, sensitivity analysis may be carried out by varying α_{ij}, β_{ijd} between 0 and 1 to investigate the impact.

A particularly simple version of this measure arises when there is a target cut-off date for claims to be processed such that all claims after that date are defined as untimely with no differentiation between claims which are more or less late. In this case the above equation becomes:

$$q_{ii}^* = q_{ii} (1 - \alpha_{ii} m_{ii} - \beta_{ii} v_{ii} + \min[\alpha_{ii}, \beta_{ii}] a_{ii})$$

This measure will equal the multiplicative measure discussed in section 4.2, which

can be reformulated as $q_{ii} = (1 - \alpha m_{ii})(1 - \beta v_{ii})$ if $a = \frac{\alpha m_{ii} \cdot \beta v_{ii}}{\min[\alpha, \beta]}$. However, in general there is no guarantee that this condition will be fulfilled.

The measure proposed so far takes no account of the value or the type of mispayment. These could be incorporated by using the proportion of the value of mispayments to total payments instead of the proportion of the quantity of mispaid claims to total claims, and by distinguishing overpayments and underpayments and weighting them differently. So the general expression, still measured in units of effective forms, would be

$$q_{ii}^* = q_{ii} \left(1 - \alpha_{1ii} o_{ii} - \alpha_{2ii} u_{ii} - \sum_{j=1}^n \beta_{ij} v_{ij} + \sum_{j=1}^n \min[\alpha_{1ii}, \alpha_{2ii}, \beta_{ij}] a_{ij} \right)$$

Where o_{ii} is the proportion of the value of claims overpaid, and u_{ii} is the proportion of the value of claims underpaid.

The problems with this method are data availability and what to do about including any additional quality domains. For instance, claimant recipients may care a lot about the ‘smiliness’ of the administrator dealing with their claim. If we wanted to include this in to this method, we go back to the same problem in method 1 i.e. how to weight the different domains.

3.2 Health Care and Adult Social care

‘*Public Service Productivity: Health Care*’ (ONS 2008) included quality adjustments to take account of the following aspects:

- short term survival after hospital admission
- health gain following treatment in hospital
- impact on health gain as a result of change in waiting times for health treatment
- outcomes from primary health care
- patient experience

The original work on the impact of survival, health gain and waiting times was done by a joint project between the Centre for Health Economics and NIESR (York/NIESR 2005). Separate work by the Department for Health (DH) added a quality adjustment for primary health care and patient experience (DH 2005). ‘*Further Developments in Measuring Quality Adjusted Healthcare Output*’ (DH 2007), funded by DH, updated results in all five areas and provided adjustments to output growth for 2000/01 – 2005/06. The overall impact on annual health care output growth from these five adjustments together was 0.5 per cent.

UKCeMGA has commissioned further research to develop a quality adjustment for adult social care. The ‘Quality Measurement Framework’ project¹² is working with academics at the Personal Social Services Research Unit to provide new metrics of adult social care quality, and aims to run a population preference study to estimate weights which to combine quality metrics into the overall output index. For health care work is continuing in cooperation with the Department of Health and the Centre for Health Economics to investigate issues arising from the decline of mortality amenable to medical intervention and to look at the use of patient reported outcome measures to provide quality indicators.

3.3 Education and children’s social care

The quality measure of education is based on the annual change in the Average Point Score (APS) of GCSE and equivalent exams taken at age 16. Details of the underlying methodology are given in *Methods for Public Service Productivity: Quality adjusting School Education Output* (ONS 2007), with the latest estimates available in *Public Service Productivity: Education* (ONS 2007). The adjustment is applied to the pupil attendance index. This measure is subject to error for a number of reasons including:

- Only one year of exam results are taken into account at age 16 while children are at school for 11 years. Any improvements in earlier years education will only be picked up with a time lag.

¹² Financed by the Treasury as part of the Invest to Save programme.

- The focus of the measure is narrow as it considers only one outcome measure for schools. Schools in England have a duty to improve the well-being of pupils based on the five 'Every Child Matters' outcomes. These are:
 - enjoy and achieve
 - achieve economic well-being
 - be healthy
 - stay safe
 - Make a positive contribution to society

Future development plans include:

- Extending the attainment measure to include key stage results and post-16 results;
- Produce a quality measure for pre-schools;
- Extend the quality measure to incorporate the wider school outcomes based on the 'Every Child Matters' outcomes.

There is currently no quality adjustment for children's social care.

4. Inputs

4.1 Capital inputs

Following the Atkinson review (2005), the UK government set out to improve its measure of inputs associated with government final consumption expenditure (GFCE) on public services. The review addressed a number of issues on the measurement of capital and made specific recommendations on how they should be improved (Box 1). This section of the paper provides a review of the current work within ONS to improve capital inputs measures and how these impact or influence UKCeMGA's work on government productivity analysis.

Box 1: Atkinson review on capital inputs

Rec: 5.3 - ONS and HMT should work together with CLG and the devolved administrations to improve the accuracy of data classification for government spending on public services in the National Accounts
--

Rec: 5.5 - Develop estimates of capital services, while increasing the level of detail presented to distinguish between functions and public and private sectors, to assist in analysis of productivity of public service spending
--

Rec: 5.6 - Move towards use of the accounts of departments and other public bodies as a basis for estimating capital consumption, rather than its own Perpetual Inventory Model

Source: Atkinson (2005: 68 – 70).

4.1.1 Capital in productivity analysis

Productivity measures the efficiency with which organisations produce products or deliver services by measuring the relationship between the outputs and the inputs used to produce them. Estimating public sector productivity requires accurate measurement of both inputs and outputs. Given the nature of public sector expenditure and the services provided, obtaining such a measure is not as straight forward as one would expect. However, measuring inputs in public services should not present a greater challenge compared to a similar process in the private sector, given that data collection and measures of inputs (labour, intermediate consumption and capital) would not differ greatly. Nevertheless, dealing with a durable input such as capital is more difficult since it lasts for more than one accounting period. Similarly, when a capital good delivers a service to its owner, no market transaction is recorded. As such measuring these implicit transactions - whose quantities are the services drawn from the capital stock during a period - is one of the challenges of capital measurement for productivity analysis.

4.1.2 Capital input measures

There is an extensive literature on both types of capital inputs measures – capital consumption and capital services. Capital consumption is the decline in value of an asset (depreciation) expressed in the replacement value of the asset in the production of a service. Capital services on the other hand, is the input that flows to production from a capital asset used in the productive process of production. Hence simply defined as:

Capital service = consumption of fixed capital (CFC) + return on capital / opportunity cost.

The two OECD manuals on *Measuring Productivity* (2001a) and *Measuring Capital* (2001b) provide an in-depth discussion on both the traditional measure of gross and net capital stocks using the perpetual inventory method (PIM) and the concept and measurement of capital services using the volume index of capital services (VICS). From the literature, there is unanimity on the importance of capital inputs and recognition that capital services are the more appropriate measure of capital inputs in activity and production analysis. The conceptual issues and the methods on both measures are dealt with by a number of authors with ONS (2008e) providing a summarised overview.

4.1.3 Why is capital service a better measure?

The traditional approach uses the PIM to measure gross investment (new assets acquired) and the price of that new investment while making some assumptions about how the quantity and value of older assets change over time (arithmetic depreciation). This approach, does not take into account the interest tied up in the purchase of the asset as the true economic cost; rather it tends to regard the interest as intermediate consumption. Diewert (2007) and others have argued that the interest on the cost of the purchase is productive, given that it is the cost of foregoing intermediate consumption.

The capital service measure on the other hand, addresses this by taking into account the opportunity cost of tying up resources in capital, as opposed to a summation of the value of the stock in a given period¹³. Capital service estimates therefore, measure the actual contribution of the asset in the productive process and therefore are the best measure of capital inputs for productivity purposes (for a detailed exposition, see Oulton and Srinivasan 2003 and ONS 2008).

The value of capital services is the productive stock of the asset multiplied by the price of those services. This price is the unit cost for the use of the capital asset for one period and is sometimes referred to as the “rental price” of a capital good or the “user cost of capital”. By definition the capital services are observable but it is the rental price that is not observable, given that firms own their assets as opposed to renting them. As such, this also means that the rental price is not always evident hence has to be modelled. For example, if all fixed assets were leased on the market then it would be plausible that the rental values would be directly observable. In practice however, many fixed assets are owned by the users and consequently there is no rental transaction that can be observed. The computation of these unobserved values raises conceptual and empirical questions that have been addressed by current work within ONS (Wallis 2005 & 2007, Dey-Chowdhury & Wallis 2007)

¹³ The formula for calculating capital services comprises of four principle terms: (i) a return to capital equal to the real interest rate multiplied by the value of the asset at the beginning of the period (ii) less the anticipated real holding gain or loss of owning the asset (opportunity cost) (iii) plus consumption of fixed capital (iv) multiplied by a discount factor.

4.1.4 Matters arising from practical application

The ONS (2008e) paper deals at great length with the issues arising from the practical application of capital inputs in UKCeMGA's work on public sector productivity analysis. In summary, it covers issues such as; (i) the appropriate rate of return for the public sector investment; (ii) how this is modelled in the VICS; (iii) the level of input GFCF data disaggregation that feeds into both the PIM and VICS; (iv) and the lack of specific guidance on the treatment of Private Finance Initiatives (PFI) capital assets.

4.1.4.1 An appropriate rate of return for the public sector

Of all the issues raised, the question of the appropriate rate of return for the public sector has been widely debated (Diewert 1980 and 2001, Harper et al 1989) and is at the fore of the next development of the UK VICS. For UKCeMGAs work, the current rate of return used in the calculation of capital services is not representative of the public sector. In the model, the rate of return is equal across all assets and industries and is one that exhausts the gross operating surplus (GOS) of the UK economy. The concern, however, is with how the rate of return is modelled.

Essentially the argument is the following: GOS is a market-based concept but the rate of return is modelled assuming that this is completely exhausted across all assets (or industries). This includes government and yet Diewert et al (2004) argue that this endogenous method is inapplicable when it comes to those institutional units for which the National Accounts do not generate an independent measure of GOS – the non-market producers. This follows the current SNA93 convention that takes the view that there is zero net operating surplus for non-market producers, which in turn implies that no return to capital on investments in assets in this sector. There are some merits to this interpretation; however, a major flaw is that similar assets will be valued differently when used in the market and non-market sectors. ONS currently gets round this problem by making an assumption that the rates of return in the non-market sector is equal to that in the market sector thus eliminating the differential valuation of similar assets across market / non-market boundary.

Schreyer (2004) suggests an exogenous rate of return, which among its advantages deals with the question of non-market production for GOS is zero as defined by SNA93. In this case, a return to capital is estimated for non-market producers. Harrison (2008) notes that the Canberra II Group has considered the issue and proposed a number of changes to the System of National Accounts; (i) A return on capital should be estimated for non-financial assets of non-market producers; (ii) The return should reflect the value of the asset (see Harrison 2008 for more detail on the proposal to revise this section of SNA93). Therefore, as indicated earlier the debate returns to what the appropriate rate of return on government assets would be. The ONS is looking at the feasibility of using the test discount rate¹⁴, as defined in the Treasury's Green Book, given that it is used currently, to assess returns accruing to the public sector and reflects the opportunity cost of public sector investment, hence, ideally suited for the estimation of the rate of return for the public sector. The debate

¹⁴ The test discount is the rate used to assess public sector projects. It is related to the fall in the social value of consumption over time. It is currently 3.5 per cent, although it has changed several times over the last twenty years.

on whether we should have different rates of return is on going and we look forward to SNA 2008 to provide guidance.

4.1.5 Current improvements to capital input measures in UK

A number of initiatives are underway that address some of the issues highlighted in ONS (2008) and other related papers. These are addressed below:

The ONS includes five year life lengths for computer software and hardware, and in 2007 implemented revised estimates for own account ICT investment for the private sector following OECD and Eurostat recommendations (see Ahmad 2003 and Chamberlin et al 2007). No further work on proposed research into asset type classifications has been possible due to the ongoing National Accounts Re-Engineering project. Wilson et al (2007) observe that the new own account ICT methodology has resulted in upward revisions to current price investment, gross and net capital stocks, and capital investment in intangible fixed assets. Thus for 2006, revisions were £8.6 billion for capital investment, £7.6 billion in capital consumption, and changes in gross and net capital stocks of £45.6 billion and £29.7 billion respectively, attributed to changes in the treatment of service lives in computer software investment (Wilson et al 2007). To improve the level of detail in the PIM results, National Accounts Group aims to increase the level of detail of GFCF input data which then means that ONS should be able to use results at a higher level of disaggregation (4digit). This would also address the VICS data disaggregation problem.

The most recent developments on the VICS model have been focused on achieving a broader asset breakdown / definition in an effort to isolate the trends and impact in the capital investment series. Dey-Chowdhury and Wallis (2007) assessed the potential impact of treating software as a separate asset when estimating capital services. They have extended the current capital services asset definition to eight assets by separating own-account software copyrights and mineral exploration from the National Accounts definition of intangible assets¹⁵ in the model. With better own-account software data, it was possible to split out National Account intangibles. For instance purchased software is separated out from plant and machinery. Their results show strong growth in capital services from computers and purchased software and much stronger growth in the service industries than production industries over the years.

¹⁵ National Accounts definition of ‘intangibles’ consists of own-account software, mineral exploration and copyrights.

4.1.6 Way forward on capital input measurement in the UK

While the issues highlighted in the ONS (2008) publication on capital consumption and capital service measures are valid in the interim, the current focus within ONS is on the production of reliable estimates of capital services (consumption of fixed capital plus return on capital) given the deficiency of the capital consumption measure in taking into account the opportunity cost of capital. This can be achieved by using a reliable tool in the derivation of the consumption of fixed capital – Whole of Government Accounts (WGA) and a measure of the return on capital or opportunity cost of capital that is applicable to the public sector (the test discount rate in the Treasury Green Book). As highlighted in ONS (2008), the WGA provides a more direct and accurate measure of government depreciation and hence consumption of fixed capital than the traditional PIM would. In addressing the current limits of the existing methodology, the move to WGA would provide an improved capital consumption measure and a better level of data disaggregation. As such, ONS is working with Treasury to achieve a timely delivery of this data. As regards the opportunity cost of capital, the test discount rate offers a sensible convention to adopt given that it is a measure of the social time preference and reflects the opportunity cost of public sector investment. The ONS (2008) paper makes the case for its inclusion for the public sector and how it can be modelled exogenously. UKCeMGA and the VICS team at the ONS-National Account Group (NAG) will work to include the 3.5% discount rate in the next capital service estimates for the public sector.

4.2 Other areas of development in inputs

As stated in section 2, UKCeMGA is also seeking to produce both direct and indirect measures for labour and to develop deflators at a practically disaggregated level. The areas which have seen the most work have been:

1. Health Care
2. Education
3. Social security administration

This section now looks at labour, goods and services and capital for each of the services listed above. Work is also underway to investigate the development of input measures for criminal justice system, fire and rescue services and defence, but these are not discussed in this paper.

4.2.1 Health Care

Input volume is mainly derived indirectly from deflation of current price expenditure. It is currently based solely on data from England. The main exception to this is the volume of labour in the NHS, which we estimate via the ‘direct’ method. The direct method means that we take information on the volume of different categories of staff within the NHS and combine them into a total labour volume index using average salaries as weights. We derive the volume of goods and services in health care by deflating expenditure divided into ten categories by a category specific deflator. In adult social care expenditure is divided into labour, independent care inputs and other goods and services, each with its own deflator. For both health care and adult social

care we use the capital consumption series available from national accounts for health and adult social care in aggregate.

The main thrust of development work on inputs is to provide a better reconciliation of the various estimates of NHS expenditure available. Within that we will particularly focus on purchases of health care from non-NHS providers. Other work will seek to incorporate input data from the devolved administrations. There may also be scope for improvement in the deflators, particularly for intermediate consumption in family health services.

4.2.2 Education

Currently an indirect method for inputs is used for labour. The labour input is expenditure on wages and salaries adjusted by a pay deflator. The pay deflator is a composite deflator based on a teacher specific deflator and an hourly wage rates by occupation estimated in the Annual Survey of Hours and Earnings (ASHE) for support staff.

Goods and services expenditure is adjusted by a composite deflator based on components of the RPI / PPI indices for different components of procurement spending.

The capital input measure uses an estimate of capital services. Due to the volatility in capital services, the series is smoothed using a 3 year moving average.

The education and children's team are developing a direct measure of labour using a new source of data for teachers hours.

4.2.3 Social Security Administration

Inputs for Social Security Administration are currently measured using an indirect method by deflating current price expenditure. Expenditure data is split into three components: labour, goods and services and capital.

SSA labour input is deflated by the Index of Labour Costs per Hour (ILCH). The ILCH goes beyond existing earnings indicators to include non-wage costs.

Expenditure on goods and services is deflated using the retail price index (excluding the cost of mortgage interest payments) and capital input is measured using capital consumption which involves deflating expenditure on capital using an implied deflator for capital consumption based on the current and constant price series for public administration in National Accounts

There are a number of areas POSSA will work on to improve input measures for SSA in the future. This will include working towards complete consistency between input and output measures, currently it is not possible to include HMRC expenditure on child benefit but child benefit output measures are included. There are a number of

issues with the deflation methods which we will look to improve. The DWP pay deflator is based on data for all the civil service staff and the LA pay deflator is not specific to housing benefit staff. It may be possible to produce more specific deflators in the future, as recommended in the Atkinson Review. The Atkinson Review also recommended that ONS incorporate direct measures of labour input, e.g. number of FTEs broken down by staff grades/job families. Including this sort of measure would be beneficial in analysing changes in the volume of labour input in a way that takes account of implied quality change. It may also be appropriate to include a more specific deflator for goods and services in the future following developments in the Whole Government Accounting Systems.

5. Conclusion

Measures of output (both quantity and quality), inputs and productivity are under going major developmental work to improve the measures by the ONS. This will enable:

- Better estimates of GDP
- Improved estimates in National Accounts
- Better information for Government and the public on how the public sector is performing
- Better information for policy advisers
- Better information for Government and the public on how the quality of public services is changing over time

While development work in some areas has been undertaken for some years, new areas are now being explored, in particular in areas of quality. Looking at measures for the quality of public services is particular challenging, as was highlighted in Section 3.

In this paper, we have explored some of the issues surrounding the development of a specific measure for quality adjusting social security administration output. We have also illustrated, for three benefits, the kind of impact quality adjustment might make under some very general assumptions.

In developing a robust quality adjustment, it is important to understand the different aspects of quality, and how they interact with each other. It is inappropriate to use only one measure without a good understanding about how the various domains interact for we risk misrepresenting overall quality change by the use of only one indicator. Furthermore, we are unlikely to know whether (unknown) measures for other domains move in the same or an opposite way. For example, a decision to spend more on speeding up time taken to process initial claims may mean a worsening of accuracy on load. It is perfectly feasible that the two measures could also move in the same direction, particularly, for example, if a new computer system for example has fundamentally improved the quality of the overall service.

UKCeMGA's wider review (ONS 2008g) of the way output estimates are adjusted for changing quality is addressing the following issues:

- the appropriateness of using continuous or categorical methods for calculating quality adjustments;
- how quantity and quality relate to each other, and how quality adjustments should apply to output series;
- in the context of multiple aspects of quality, how we determine the weights that should be applied to each in order to derive a single measure of quality change;
- how we take account of domains of quality for which we have no measurement of quality; and
- how we apply quality adjustment to series for which we have no measure of quality.

The review aims to determine methods and procedures that can apply to all public services. ONS plans to hold a seminar later this year on the principles and procedures for adjusting a range of public service outputs for quality of service. If you have any comments about the content of this article, either general or specific, please email emma.edworthy@ons.gsi.gov.uk

Annex A

Principles from the Atkinson Review Final Report

Atkinson Review Recommendation 4.1:

‘The direct measurement of the output from government spending, and the measurement of inputs and productivity, should be based on a set of principles, within the framework set by international guidelines.’

The following main principles cover the direct measurement of output, the measurement of inputs, and the measurement of productivity.

- **Principle A:** the measurement of government non-market output should, as far as possible, follow a procedure parallel to that adopted in the national accounts for market output
- **Principle B:** the output of the government sector should in principle be measured in a way that is adjusted for quality, taking account of the attributable incremental contribution of the service of the outcome
- **Principle C:** account should be taken of the complementarity between public and private output, allowing for the increased real value of public services in an economy with rising real GDP.
- **Principle D:** formal criteria should be set in place for the extension of direct output measurement to new functions of government. Specifically, the conditions for introducing a new directly measured output indicator should be that (i) it covers adequately the full range of services for that functional area, (ii) it makes appropriate allowance for quality change, (iii) the effects of its introduction have been tested service by service, (iv) the context in which it will be published has been fully assessed, in particular the implied productivity estimate, and (v) there should be provision for regular statistical review.
- **Principle E:** measures should cover the whole of the United Kingdom; where systems for public service delivery and/or data collection differ across the different countries of the United Kingdom, it is necessary to reflect this variation in the choice of indicators.
- **Principle F:** the measurement of inputs should be as comprehensive as possible, and in particular should include capital services; labour inputs should be compiled using both direct and indirect methods, compared and reconciled.
- **Principle G:** criteria should be established for the quality of pay and price deflators to be applied to the input spending series; they should be sufficiently disaggregated to take account of changes in the mix of inputs; and should reflect full and actual costs.
- **Principle H:** independent corroborative evidence should be sought on government productivity, as part of a process of ‘triangulation’, recognising the limitations in reducing productivity to a single number
- **Principle I:** explicit reference should be made to the margins of error surrounding national accounts estimates’.

References

Ahmed, N., (2003), 'Measuring Investment in Software', *STI Working Paper 2003/6* (OECD, Paris)

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

Centre for Health Economics and NIESR (2005) *Further Developments in Measuring Quality Adjusted Healthcare Output*.

Chamberlin, G., T. Clayton, and S. Farooqui, (2007), 'New Measures of UK Private Sector Software Investment'. *Economic & Labour Market Review*, **1**, 5:17-28.

Dey-Chowdhury, S., and G Wallis, (2007), 'Volume of Capital Services: Estimates for 1950 to 2006', *Economic and Labour Market Review*, **1**, 12:37-47

Department of Health (2007b) *Further Developments in Measuring Quality Adjusted Healthcare Output*.

Diewert, E.W., (1980), 'Aggregation Problems in the Measurement of Capital', in D. Usher (ed.) *The measurement of Capital*, (Chicago: the University of Chicago Press):433-528

Diewert, E.W., (2001), 'Measuring the Price and Quantity as Capital Service Under Alternative Assumptions', *Discussion paper No.01* 24th June; (Department of Economics, University of British Columbia).

Diewert, E.W., A. Harrison, P. Schreyer, (2004) 'Cost of Capital Services in the Production Account'. Available at <http://www.ipeer.ca/workshops/Shreyer%20Harrison%20Diewert,Cost%20of%20Capital%20Services.doc>

Diewert, E.W., (2007), '*Productivity Measurement in National Accounts and its Importance*', A Paper presented at the Asian Productivity Organisation – Keio University Lecture Program October 22, 2007. (Department of Economics, University of British Columbia).

Dey-Chowdhury, S., and G. Wallis, (2007), 'Volume of Capital Services: Estimates for 1950 to 2006', *Economic & Labour Market Review*, **1**, 12:37-47.

European Commission (2001) *Handbook on Price and Volume Measures in National Accounts* (2001 edition), Luxembourg: Office for Official Publications of the European Communities.

Eurostat (1995) *European system of National Accounts*.

Harper, M., E. Berndt, and D Wood, (1989), 'Rates of Return and Capital Aggregation Using Alternative Rental Prices' in Jorgenson, D. and R. Lndau (eds.), *Technology and Capital Formation*, (MIT Press).

OECD (2001a), *Measuring Productivity: Measurement of Aggregate and Industry-Level Productivity Growth – OECD Manual* (Paris: OECD).

OECD (2001b), *Measuring Capital: A Manual on the Measurement of Capital Stocks, Consumption of Fixed Capital, and Capital Services* (Paris: OECD).

ONS (2006a) *Public Service Productivity: Social Security Administration*.

ONS (2006b) *Establishing the Principles Consultation*.

ONS (2007) *Measuring Quality as Part of Public Service Output: Strategy following consultation.*

ONS (2008a) *Public Service Health Care Productivity article.*

ONS (2008b) *Guidelines for measuring Statistical Quality .*

ONS (2008c) *Criminal Justice System: Scoping document*

ONS (2008d) *Framework for measuring the quality of methods used to estimate public service output and productivity.*

ONS (2008e), *Capital Inputs in Public Sector Productivity: Methods, Issues, and Data.*

ONS (2008f), *Public Service Productivity: Social Security Administration.*

ONS (2008g), *Adjusting Measures of Public Service Output for Quality of Service.*

ONS (2008h) *Accounting for Quality Change in Estimates of Social Security Administration Output and Productivity.*

ONS (2008i) *Public Service Output, Input and Productivity: Quantity measure of Children's Social Care Output.*

Oulton, N., and S. Srinivasan, (2003), 'Capital Stocks, Capital Services, and Depreciation: An Integrated Framework', *Bank of England Working Paper No. 192* (London, Bank of England).