

The Welfare Implications of Public Goods: Lessons from 10 years of Atkinson in the UK

Fred Foxton, Joe Grice, Richard Heys & James Lewis

September 2018

Executive Summary

When considering issues of measuring welfare beyond GDP, a key ongoing but unfinished agenda is to how to measure the outputs of goods and services which are free at the point of delivery, for the purposes of national accounts. Public services such as schools and health services are the major example of this kind. Over a decade ago, Sir Tony Atkinson provided a principled framework for this end. Consistent with the basic principles of national accounting, he advocated an approach by which this output should be measured as the value added by the services concerned. This value, in turn, equated to the improvement in outcomes directly attributable to the activities of the public services concerned. Implementing this approach, as he recognised, is by no means straightforward, but the UK experience recounted above shows that strong progress can be made. Working with experts and practitioners, quantity and quality measures can be identified and used to give a good approximation to the value added by the key public services, and thus their contribution to GDP. New data and intelligent use of existing data means this can be done at low cost and in a way which maximises stakeholder understanding and acceptance.

But national statistical institutes are also now grappling with a second task; measuring changes in welfare or wellbeing more generally, regardless of how they are generated. Health outcomes - life expectancy, for example, or healthy life expectancy – are influenced by a variety of factors besides publicly funded health services: diet, smoking prevalence and other lifestyle choices are obvious determinants. So, the central tasks under this agenda become first the identification of appropriate measures of outcome changes and then to determine how much value our societies place on those changes. We illustrate this using a widely recognised measure of well-being, life expectancy, and readily available information on the shares attributable to health services and hence non-health service factors, alongside existing estimates of the value of a quality adjusted life year have enabled us to create estimates of welfare gains alongside new quality adjustments which could be used to both feed into GDP under the SNA and into a wider welfare measure.

The Welfare Implications of Public Goods: Lessons from 10 years of Atkinson in the UK

Fred Foxtton, Joe Grice, Richard Heys & James Lewis¹

Introduction

Current debates about measuring the impact of the digital economy, specifically free digital goods which deliver welfare gains to consumers, even if their exact treatment in the National Accounts is under debate, need to be seen in the context of a larger group of transactions which are also free, or nearly free, to consumers. These are mainly the public services. In the UK, around 20% of Gross Domestic Product (GDP) is accounted for by the outputs of public services. Other G7 countries exhibit similar magnitudes ranging from 19% to 24%, with the one exception being the United States at around 14%. Measurement of these free goods is a common issue affecting almost all countries.

The UK has had an interest in this question since 2003 when the then National Statistician, Len Cook, asked Sir Tony Atkinson to conduct an independent review of the measurement of government output in the National Accounts, with a Final Report produced in 2005. The resultant publication was a seminal text which informed the development of System of National Accounts (SNA) 2008 in how to conceptualise and then empirically measure the outputs of the public services contained in GDP. The UK, alongside several other countries, pressed ahead with implementing these methods. This work managed to address the largest parts of the public services, but gaps remained.

The Bean Review (2016) commended the Office for National Statistics for this work but identified that renewed efforts were needed to update the methods being applied where quality adjustments were in place and to create new adjustments where these were not.

This paper makes four contributions to this debate. First, it draws attention to the importance of these issues both in terms of economic activity and more widely to consumer welfare: whilst the impact of changes in digital technology over the last 20-30 years are important, considering the life-saving and life-enhancing improvements in medical care over the same period gives important context. Secondly, to re-iterate a commonly missed Atkinson recommendation: the fast pace of change in public service² delivery *and usage* means that methodologies need to be kept under regular monitoring and updated as required. Thirdly, the paper draws out key lessons the UK has learnt over this period which the authors hope might contribute to the process of mutual learning. Finally, it highlights how a better understanding of the public sector's contribution can only enhance efforts to measure economic welfare.

The paper is structured as follows:

- A brief account of the historical context of measuring public services in the UK, in the wake of the 1993 SNA, and the problems that were encountered that led to the Atkinson Review,
- An account of the Atkinson Review and its implementation in the UK.
- A discussion of the current methods to calculate quality adjustments in the UK.
- A summary of the most significant issues identified in the UK in measuring public service outputs and outcomes, and how these have been addressed over the last decade

¹ With thanks to Katherine Kent, Heather Bovill, Jonathan Athow Richard Smith and an anonymous referee for their comments. All errors remain the authors'. The views expressed within are the personal views of the authors and do not represent, or claim to represent the views of the Office for National Statistics.

² This paper focuses on 'public service' rather than 'public sector' simply because mainly public services are now delivered via both the public and private sectors in many countries.

- A discussion of the challenges in capturing welfare gains related to public services alongside other non-GDP welfare gains in any new metric.
- Conclusions.

Part One: Measuring Public Service Output and Productivity: The Historical Context

The treatment and measurement of public service output and, by extension, public service productivity, has long been known to raise tricky but important issues. Quite clearly, their measurement is not straightforward. Most transactions included within Gross Domestic Product are measured at their market or exchange value. But most outputs provided by the public sector – health services or public provision of education, for example – are non-market services. So, while such services clearly have value, there is no observable price to guide the valuation. The value, therefore, must be imputed and this may not be simple to do.

The founding fathers of national accounting wrestled with how public service outputs should be treated in the accounts and indeed some, like Kuznets, proposed excluding them entirely. Hicks changed his mind at least twice on this question. In the event, the consensus was to adopt a convention – the so-called ‘output equals inputs’ convention – whereby these non-market outputs were deemed to be equal to the observable value of the inputs used to produce them. The implication of ‘outputs equals inputs’ is that public service productivity is always constant, with its growth rate, by definition, zero.

Leaving aside for a moment the complexity of measurement, there are important reasons for taking public service output and productivity seriously. One is the sheer scale of the transactions involved. In the UK, for example, non-market public service output accounts for around a fifth of UK GDP³; the sector is well over twice the size of manufacturing. So, omitting them from the National Accounts would be to ignore a major part of the value which the economy generates. Similarly, to do so would be to overlook a material contribution to the overall productivity of the economy. Nor does such productivity performance simply mirror that of the rest of the economy. In recent years public service productivity in the UK has been rising while the productivity performance of the rest of the economy has been stagnant.

A second reason why public service productivity is important relates to fiscal policy. Finance ministries are continuously in the horns of a dilemma, though one whose acuteness varies over time. On the one hand, the political pressure for improved public services is strong. Citizens as users have rising expectations of what they receive from health services, from publicly provided education, by way of social care and so on – no less than they have had rising expectations of the economy overall. Where many public services are key to tackling inequality and improving life chances, as these issues are important in public debate and amenable to improved public services, understanding the output of the public sector helps users understand governments’ steps to tackle inequality. But citizens as taxpayers are also reluctant to pay the rising taxes that might finance the improving public services. The only way to square this circle is to improve the efficiency and effectiveness of how taxpayers’ funds are used, so that through increased productivity, more output is produced by the same amount of taxpayers’ money. Accordingly monitoring public service productivity is of policy importance over and above the sector’s (sizeable) contribution to productivity performance overall.

Third, the performance and efficiency of the public services conditions the productivity of the rest of the economy. A well performing legal system, for example, is a vital underpinning of a well-functioning commercial sector. An efficient and well performing health service is a major contributor to a healthy and productive workforce, while the outputs of the publicly provided education system make a direct contribution to the nation’s human capital. Arguably, the same outputs also feed into social capital and thus again underpin a well-performing economy overall.

³ This can vary marginally by year selected.

Given the importance of these issues for economic commentary and policymaking, the balance of opinion in the national accounting community increasingly moved towards thinking that the ‘outputs equals inputs’ convention was untenable. There was no reason to suppose that it gave an accurate view of how the outputs and productivity of this growing sector were behaving within the overall economy. Since, by definition, it implied necessarily unchanging productivity within the sector, it could give no useful information regarding the other two issues: how well public services were making use of taxpayers’ funds or how productively the public services condition the performance of the rest of the economy. These drawbacks from ‘outputs equal inputs’ were substantial.

Accordingly, the 1993 System of National Accounts (SNA) recommended that, in future, countries should move away from the previous convention and instead adopt methodologies which measured the output of public services directly, using observable information relating to these services. This would mean of course that there was no reason why the estimated outputs from such methodologies would equate to the observed inputs. Accordingly, it would also be possible to estimate how productivity in these various sectors was changing over time.

ONS was one of the early movers, together with a handful of other NSIs, in taking forward this new agenda. By the late 1990s, measured by value, some two-thirds of public service outputs were measured directly. The remaining one-third or so continued to be measured by ‘outputs equals inputs’; so-called collective services, in particular defence, were the main part of this residuum. However, not long after the new methodologies were put in place, the estimated productivity series began to demonstrate paradoxical behaviour. Having been rising at fairly steady rates up to 1997, the estimated productivity of the directly measured sectors fell by over 20 per cent in the four or five years after 1997. It was hard to understand why the estimates were showing such declines. Nor was there any corroborating evidence to suggest that such declines had occurred. Accordingly, users’ confidence in the validity of the estimates became increasingly strained. Since the output-driven estimates also now fed into the UK National Accounts overall, confidence in those, too, was also in question.

In these circumstances, at the end of 2003, the then UK National Statistician, Len Cook, asked Sir Tony Atkinson to conduct an independent review of methodologies to measure public service output and productivity. His terms of reference also included looking at the way ONS had approached the new SNA agenda and its implementation of direct measurement methodologies. The Atkinson Review lasted for just over a year and Sir Tony published a report in January 2005 setting out his conclusions.

The Atkinson Review and its Legacy

The Atkinson Review was a milestone in this agenda. The report clarified many issues and proposed a model for measuring public service outputs. On the basis of its recommendations, it outlined a research and implementation programme in the various main public service areas. Len Cook accepted Atkinson’s conclusions, subject to underlining that their full implementation would take time and be conditioned by availability of resources.

Fundamentally, Atkinson agreed wholeheartedly that the SNA had been right to counsel direct measurement of non-market public services. The drawbacks of the traditional ‘outputs equal inputs’ convention were just too great to be acceptable, for the reasons set out earlier in this paper. By the same token, ONS had been right to take up this agenda. The problems observed in the UK data were real ones but were rooted in the way the agenda had been implemented, as discussed further below, not because the overall agenda was problematic.

Atkinson’s report saw the problem as being ONS’s failure to base its methodologies and estimates on a clear set of explicit principles. Not unnaturally, when faced with a difficult task, in many cases ONS statisticians had sometimes used readily available indicators or other data sources but which did not necessarily relate directly to what was needed, and/or stop-gap methodologies to measure public service outputs, in the hope that this would be better than nothing. Experience, however, showed these hopes

were not always been realised: it can be argued, in some cases, that the procedures had led to estimates which were worse than not having anything.

The complete set of Atkinson's principles are shown in annex A. One superficial reaction to them is that many look like common sense. Who would not be able to agree to them? On the other hand, their usefulness and power comes from employing them as a yardstick against which to compare the actual procedures which were in place. They quickly highlighted areas where ONS's existing procedures did not measure up. This gave a clear indication of where remedial action was required as well as helping guide the nature of the remedial action and revised procedures.

One particularly important principle related to what should, in theory, be included in a country's national accounts and therefore what the methodologies should be striving to capture. Atkinson contended that the key consideration in national accounts was value; thus, GDP could be considered as the cumulative value added from the economy, going through the various stages of production. It was therefore essential to avoid measuring public service output solely by what were essentially activities – say, the number of medical procedures performed or the number of pupils taught, particularly where such measures may incentivise perverse outcomes; such as fire services being measured using the number of fires they put out, where increasing fire prevention activity would lead to a reduction in output, rather than a growth.

The problem he saw was that such activities may or may not have value. His private sector analogy was production of broken bricks. A factory which produced only broken bricks would find its output next to negligible since the broken bricks would have little or no value, as opposed to well produced whole bricks which, of course, would have value. In the public services, the equivalent issue was to establish whether the hospital procedures carried out or the number of pupils taught were adding value or otherwise; what was the quality of the 'bricks' they represented.

A key principle was therefore that the estimates of public service output should be quality adjusted. At a common-sense level, the value of a health care intervention clearly depends upon its quality. The procedure is of value only to the extent that it leads to a health outcome superior to a counterfactual where the procedure had not been carried out. This leads inevitably to the question as to how outcomes should relate to the estimates. Traditionally, national accountants had been reluctant to consider outcomes as relevant and with some good reason. In most countries, life expectancies and healthy life expectancies have risen significantly over time. While improving health services have played a part in this, the broad evidence is that this has been a minority contributor with factors such as improving diets, falling smoking and healthier environments being much more important. It would therefore be quite wrong to ascribe the whole value of the improved health outcomes to the output of healthcare sectors. On the other hand, to the extent that an improved health outcome can be directly attributed to the activities of healthcare systems, then that should be taken into account in the estimated output.

The Atkinson Report was widely debated in the years following its publication. Its approach was largely accepted and helped shape the revised 2008 System of National Accounts (SNA). The principle of allowing for quality adjustment in estimates of output was accepted and emphasised, as part of a wider trend of economists becoming increasingly comfortable in addressing social welfare function issues. The European System of Accounts (ESA) which generally follows the SNA, as its guiding principles, surprisingly, and somewhat regrettably, took a flatly opposite view and banned quality adjustment within its 2010 iteration. This illustrates how contentious this topic remains. This decision was purportedly in the interests of international comparability but the authors would argue there seems to have been some muddled thinking at work. Imposing arbitrary comparability in methods does not necessarily serve the interests of comparability of the realities. With quality adjustment not allowed, those countries where public services have improved in quality are estimated with outputs below the reality and conversely for those where quality improvement has been relatively low. This can only

prejudice rather than help international comparability. Eurostat has been organizing work to review this issue so, hopefully, this is on the way to being resolved.

Since the Atkinson Review, the UK has delivered public service output and productivity estimates, with varying degrees of success, differing both between and within service areas as shown in Figure 1. The approaches are categorised broadly into three types.

‘Output equals inputs’ – Accounting for around 37.8% of public service output in 2015, this approach assumes that the volume of output is equivalent to the volume of inputs used to create them. Typically capturing what are referred to as “collective services” (such as defence), this convention is used when the output of a service area is conceptually difficult to define and/or measure. As a result, productivity is assumed to remain constant and growth will always be zero. This is the least satisfactory method.

Quantity output - Representing around 17.7% of public service output in 2015, this approach utilises long-standing indicators of activities known as a cost-weighted activity index (CWAI). Here an index is constructed as the weighted sum of change in the level of different activities from one year to the next. As most public services do not have a market price to use as a weight, given they are not sold on a market, the cost of producing a unit of activity (unit cost) are used as a proxy. Although this cost weighting occurs, the use of measured outputs is believed to be an improvement on the previous input based methodology and is used as a measure of output in the UK estimates of total public service output. More detail about this approach, and the steps involved, can be found in Annex B.

It is, however, recognised as the second-best approach. While some elements of quality change can be captured (e.g. through the differentiation of activities), a CWAI will fail to capture all quality improvements. An example of this would be a technique where improvement lead to a lower cost of treatment. This would be recorded as a reduction in output, when no change has occurred and outcomes have perhaps improved.

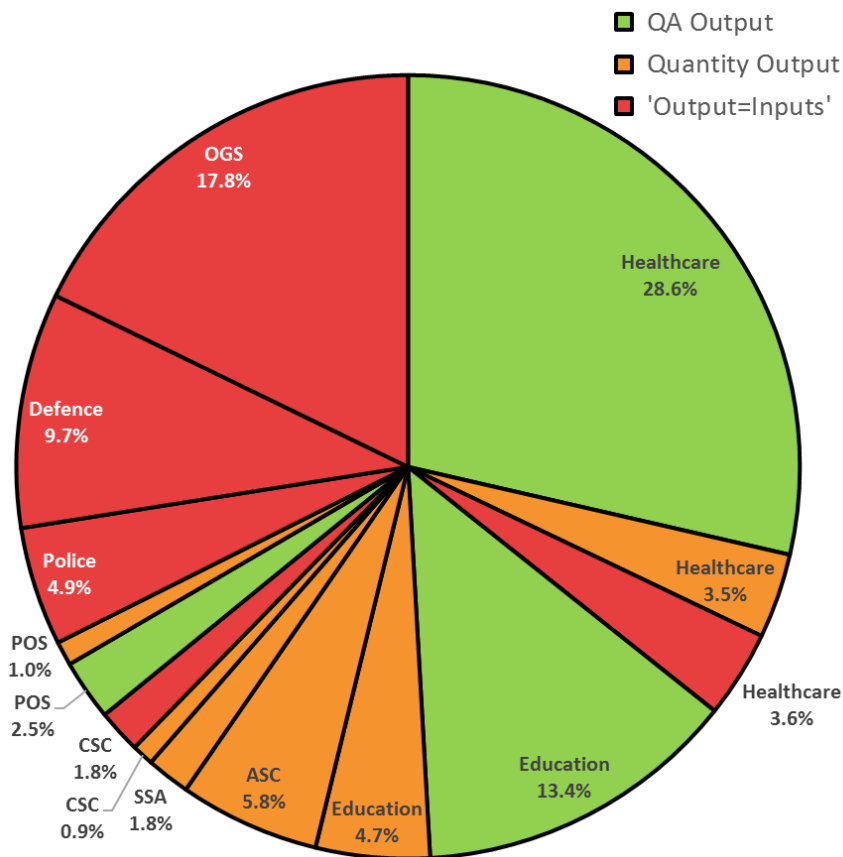
Quality adjusted output – The third category then accounts for the remaining 44.5% of public service output in 2015. Looking to address the short-coming with using CWAI, elements of public services output are adjusted to take account of changes in quality, in line with the recommendations of the Atkinson Review, reflecting improvements in outcomes that can be attributed directly to public service activity.

Within the market sector, higher-quality variants of outputs can be picked out. As higher-quality outputs sell for more than the lower-quality, the change in quality is accounted for by the price differential. This is, of course, much harder (but no less important) to do for public services because service users do not pay directly for the services, and thus there is no user-driven differential to use. Where they can be identified quality metrics are, therefore, used to augment volume data, based on how far outcomes can be attributed to public services, to give a well-based measure of the public service output concerned.

It is important to note that such quality adjustments are explicitly excluded from the measurement of output in the National Accounts Framework by ESA10, and are not part of the output series used in other Office for National Statistics (ONS) measures of productivity.

More details about the general methodology can be found in Annex C, while more specific details are provided later in Part Two of this paper.

Figure One: Output-type share by service area, 2015, UK



Source: ONS (2018a)

Notes: Shares may not sum to 100 due to rounding. QA refers to quality adjusted. Public order and safety (POS) includes courts and probation services, prison service and fire service. OGS refers to other government services which includes services such as economic affairs, recreation, and housing.

Atkinson made a further important recommendation to ‘triangulate’ estimates with corroborating evidence when assessing public service output. Such evidence might be subjective or objective. The sharp downturn in ONS’ estimates of public service output that had led to the Atkinson Review being set up turned out to be largely illusory and due to problems with data sources and methodologies. A subjective source of evidence that might have shown up the problems earlier would have been talking to practitioners and expert commentators. When, during the Review, they were asked what might have caused the sharp downturn in productivity, their invariable response was that they were not aware there had been such a downturn.

Such evidence would not have been conclusive but would, at least, have rung alarm bells. Such subjective evidence can, moreover, be supplemented by objective evidence. In the hospital sector, one of the key factors affecting efficiency is average length of stay. With relatively fixed hospital capacity in the short term, a shorter length of stay allows more patients to be treated. So, if there had been a sharp downturn in health sector productivity, it would have been reasonable to expect average length of stay to have increased. But, in fact, it had not behaved out of the ordinary. Again, not conclusive but another possible alarm bell to give reason to question the estimates.

A third precept from the Atkinson Review stemmed from the nature of what it saw to be the task. Assessing changes in the quality of various public services, or for that matter collecting and assessing triangulation evidence, may well not be within most National Statistical Institutes’ core competences.

Fortunately, such issues are the core business of other communities. Assessing, for example, the quality of teaching and the contribution that schools make is what many education experts spend much of their time doing. Similarly, such issues in the healthcare sector are a central preoccupation of public health experts, epidemiologists, health economists and so on. Practitioners, by definition, are a complementary source of expertise. So, too, are government departments and other public authorities, who will have much greater expertise and experience of the services in their fields than an NSI could ever hope to muster. Atkinson therefore recommended that ONS (and other NSIs) should form networks with such experts to allow them to tap into the expertise that would be needed to compile authoritative estimates of public services output.

One important purpose that such networks could serve would be to feed into periodic reviews of the ways that public services are delivered and whether intervening changes mean that the original data sources and methodologies for compiling output and productivity estimates remain valid or whether changes are necessary. Models for delivering public services change no less quickly than the business models underpinning private sector activity. So, without such periodic reviews, there would be the possibility of maintaining methodologies that no longer corresponded to the real world. Of course, in principle, detecting such changes should be an ongoing concern. But the periodic reviews should serve as a safety net to ensure that relevant changes are picked up.

The Atkinson Report set out a principled approach which it recommended as the general model for measuring public service output and productivity. It also had chapters with suggested agendas for applying the approach in four key areas:

- Healthcare services
- Public sector education services
- Public order and safety (specifically the criminal justice system)
- Services relating to adult social care

It recognised that completion of the work programme to fulfil these agendas would take a number of years. The rest of this paper discusses UK experience in taking this work forward and some of the principal lessons learned. Then, in light of this work, it considers both the welfare implications of public services but also how we treat those welfare gains which are not attributable to changes in public service provision.

Part Two: The Current Methods of Calculating Atkinson Quality Adjustments

We begin by briefly explaining the methods used to derive the four quality adjustments currently in use in the UK.

Healthcare

In the United Kingdom, health care is primarily a public service under the Atkinson Review definitions. Nearly 80% of UK health care expenditure is publicly-funded with much of this public expenditure funding free-at-the-point-of-use care through the National Health Service (NHS)⁴.

The task of valuing the output and measuring the productivity of a free-at-the-point-of-use service, without insurers or other intermediaries negotiating prices from care providers, therefore faces the same challenges Atkinson sought to address across other public services.

⁴ The NHS provides healthcare in Great Britain. In Northern Ireland, the Health and Social Care Service provides similar free-at-the-point-of-use care. For brevity, 'NHS' is used to describe all UK public health services.

But mitigating the considerable challenge of measuring the productivity of a service for which a price does not exist, the NHS provides the advantage of a wealth of data, collected on a uniform basis from all NHS providers.

Measuring health care output

As with other public service sectors, quantity output is measured through a cost-weighted activity index (CWAI)⁵. The data for this come from detailed published management information. NHS provider organisations responsible for hospital, community and mental health care report detailed data on activity and unit costs as part of the process of setting reimbursement rates for the thousands of different activity types carried out across these sectors, as well as for use as a management information resource. For this purpose, activity and expenditure are analysed by Health Resource Group (HRG) and by care setting. The HRG system provides a more detailed and precise treatment-classification system as an alternative to the internationally-used Diagnosis-Related Group (DRG) system, with over 25,000 individual activity types in the most recent years.

For other elements of publicly-funded health care outside of NHS hospital, community and mental health care provision and drug prescriptions, data are scarcer. Data availability is particularly problematic for general practice, where output is currently measured using modelled estimates based on historical and demographic data, and for the rapidly growing component of NHS-funded services that are outsourced to independent sector providers.

By spending, NHS hospital, community and mental health provision accounts for 64% in the most recent data, with a further 10% from prescription drugs. Other family health services, of which General Practice is the largest component, along with the more easily measurable dental and ophthalmological services, account for 15% and services purchased from non-NHS providers a further 11%.

UK public service health care output therefore combines a large element of some of the most precise output measures available for UK public services, with estimations needed for some of the other elements of the service. But as with other service sectors, the limits of cost-weighted activity in determining the value of public service provided still hold across all service elements. Hence a quality adjustment is required.

Measuring health care performance and outcomes

The comparative wealth of data available for health care extends to data on the quality of services. Here, a large variety of measures are available – NHS performance statistics provide monthly measures of performance against targets for a range of activities, while outcomes data from life expectancy to cancer survival rates provide indicators of the ultimate goals of the health service.

However, this trove of data does not automatically translate into the quality adjustments envisaged by Atkinson for output and productivity.

Consider the use of NHS performance indicators as quality adjustments and, as an example, Accident and Emergency (A&E) department waiting times, which are one of the NHS's highest-profile headline performance measures.

We can track the percentage of A&E patients who are seen within the NHS's national four-hour waiting time target. But it is not clear how a change in a quality adjustment incorporating the proportion of patients seen in four hours should affect the value of A&E output. Should we give equivalence to the

⁵Produced by chain-linked Laspeyres indices. Estimates of quality-adjusted output are produced in a similar manner as explained in Annex B.

volume and quality measure such that a 1% increase in activity and a 1% decrease in patients seen within the target roughly the same as a stable value of output?

This would imply that the value of providing A&E services to patients after the four-hour target is near-zero. However, given that patients counted in the activity data after a wait of four hours have endured the loss of their valuable time in surroundings not of their choice to receive care, it appears clear, even to a logic-seeking economist, that the value patients place on receiving emergency medicine services is greater than zero. So, such a simple solution would clearly be inadequate.

And the problems of how to apply such a performance measure to output do not stop with only the question of how such expenditure is scaled. Such a performance indicator only reflects one aspect of quality and research shows a tendency of providers to modify their behaviour to meet the minimum requirements, but not necessarily the spirit, of performance targets (Kings Fund (2017)). For instance, the four-hour waiting time target may encourage A&E department to prioritise seeing patients who are approaching the four-hour mark, but improvements in performance against the four-hour target may not reflect shorter waiting times for patients in other parts of the waiting times distribution.

Therefore, robust quality adjustments cannot simply be drawn from the NHS performance targets. Instead, they should inform the effect of health provision on the outcomes they are trying to achieve.

One alternative measure from the health economics literature does provide a conceptual framework which fits the criteria for quality adjustment far more closely, the Quality-Adjusted Life Year (QALY). The QALY is a tool for evaluating health care interventions that was first developed in the 1960s and 1970s and is now used globally (Mackillop & Sheard (2018)). The QALY is particularly prominent in the evaluation of health care in the UK, where the National Institute of Clinical Excellence (NICE) uses it to make recommendations on what treatments should be funded on the NHS.

While there is no single definition of a QALY, NICE uses the definition that a QALY is ‘*a measure of the state of health of a person or group in which the benefits, in terms of length of life, are adjusted to reflect the quality of life.*’

The QALY thus has two elements, a health-related quality of life element and a temporal element; and can therefore combine the effect of improvements in health-related quality of life and increases in the length of life resulting from treatment.

Health-related quality of life is measured on a scale between zero and one, with zero being a state equivalent to death and one representing perfect health. For the evaluation of health care, the gain in health-related quality of life from an intervention is then measured across time to produce a measure of QALY gain, such that a gain of one QALY represents one additional year of life in perfect health following the intervention.

However, while the QALY serves to provide much of the theoretical grounding for a quality adjustment, the quality adjustment used by ONS cannot simply consist of a change in measured QALY both for the practical reasons no systematic regular data collection on patients’ health-related quality of life before and after treatment exists, nor consistent data available on the increase in patients’ life expectancy resulting from treatment, but also for the conceptual reason that changes in health states are not just caused by health provision, but also by an array of other factors.

Quality-adjusting health care output

Unlike other adjustments where ONS and the relevant government department generally undertook the relevant work, given the challenges of constructing a quality adjustment to meet the principles from the

Atkinson Review, the current health care quality adjustment was designed through a rigorous process, which set out the measurement framework, involving an expert group of health economists⁶.

The construction of the measure incorporated a range of relevant factors, while taking care to minimise combining metrics which would overlap and thus record the same quality drivers multiple times. For instance, the elective inpatient care adjustment combines health gain, survival, waiting times and patient satisfaction, thereby covering the main aspects of care quality. The quality adjustment used by ONS for healthcare output continues to be based on this research.

The measure produced can be divided into three components:

- Hospital procedures adjustment
- Primary care outcomes adjustment
- Patient experience adjustment

While we will discuss each of these in turn, of the three, the hospital procedures adjustment is by a large margin the most significant in terms of its effect on the measure, while also being by far the most complex. The hospital procedures adjustment continues to be produced by the Centre for Health Economics at the University of York and is used both to quality adjust healthcare output in the ONS measure and a separate productivity analysis carried out by the Centre for Health Economics (see Dawson et al., 2005).

Hospital procedures adjustment

The quality adjustment utilises the Hospital Episode Statistics (HES) dataset, whilst incorporating other data from various sources. The HES dataset is a highly detailed administrative dataset recording details of all patients receiving hospital treatment on the NHS in England, and observations in HES are coded to appropriate activity types using the aforementioned HRG system which is used to produce cost-weighted output.

The team did not try to value welfare gain on a QALY basis, although the concept of QALY is central to the hospital procedures quality adjustment, for the following reasons:

- There is no certain value to one QALY – NICE does not specify a single value, though their treatment recommendations imply a value for one QALY of £20,000 - £30,000. However, other bodies use different approaches to value health, with the Department for Transport using a single figure per life lost to evaluate road safety interventions (Glover and Henderson, 2010).
- The value of a QALY should vary over time with average incomes and the marginal utility of income, but an increase in incomes should not be attributed to a quality adjustment for the NHS. However, holding the value of a QALY constant could result in the quality adjustment effect declining over time as cost inflation affects the ratio between quality and quantity value.
- Other factors beyond pure QALY gain may be important for the quality adjustment, such as patient experience.
- As discussed in part three, below, QALY gain could be influenced by other factors outside of the influence of the NHS, such as the output of other services, environmental factors or changes in patients' behaviour.

The hospital procedures adjustment which was instead developed can itself be broken down into three sub-components:

⁶ Funded by the UK Department of Health, the project team consisted of several economists from the Centre for Health Economics at the University of York and the National Institute of Economic and Social Research, along with the input of other involved bodies, including the Department of Health and ONS.

- Estimate of health gain
- Short-term survival
- Waiting times

The health gain estimate is an attempt to proxy for the gain in health-related quality of life on an equivalent zero-to-one scale as is used in the calculations for QALYs. While, as previously mentioned, there is no systematic collection of health-related quality of life for all patients across the NHS, patient reported outcome measures (PROMs) are collected from patients across two high-volume treatment groups – hip replacements and knee replacements, and until 2017, were also collected for groin hernia and varicose veins procedures. The PROMs give measures on health-related quality of life gain before to after treatment using the EQ-5D scale, a widely used assessment framework, which uses patient responses to questions on ability to pursue usual activities, anxiety/depression, pain, mobility and ability to self-care to produce a health-related quality of life score on a scale of zero to one.

However, these PROMS measures cover a tiny fraction of the total number of patients receiving hospital treatment on the NHS. The health-related quality of life gain for the majority of patients therefore needs to be estimated. The Centre for Health Economics produced a single ‘rough estimate’⁷ for all remaining elective treatments (procedures scheduled in advance) and a single ‘rough estimate’ for all non-elective treatments (urgent unscheduled procedures). These estimates assume a greater health gain for non-elective procedures as patients generally arrive in a worse health state than elective patients, and so experience a greater health gain⁸.

The gain in health-related quality of life is spread across remaining life expectancy as derived from the ONS data on life expectancy by age, although due to a lack of data, no adjustment is made to counter the effect that treatment may extend life expectancy or that patients may have a lower life expectancy than the general population due to their pre-existing health issues. Gains in health-related quality of life across the future are discounted using the social time preference rate.

Post-operative short-term survival rates are then incorporated in this measure to reflect health-related quality of life falling to zero for patients who do not survive the procedure or die before being discharged.

A waiting times factor then incorporates the forgone potential health gain for treatment being delayed, with waiting times at the 80th-percentile taken to reflect the importance of uncertainty and the risk of long waiting times to patient well-being.

Application of the hospital procedures adjustment

The overall adjustment for hospital procedures is then calculated from the change in the health-related quality of life element (health gain/survival rate factor) multiplied by the change in the temporal element (life expectancy/waiting time factor). This adjustment is calculated individually for each HRG and then applied to the output data which is also calculated at the HRG level, meaning the quality adjustment is not simply applied as an aggregate of all procedures carried out, but incorporates the same cost-weighting as non-quality adjusted output.

However, the complexity of the calculation means that the drivers of this quality adjustment can be difficult to discern and the direction of effect not always immediately intuitive. Table One explains the effect of these changes:

⁷ The CHE’s assessment, not that of the authors.

⁸ Some further adjustments to these rates are applied to procedures with high mortality rates to avoid the quality adjustment giving a negative valuation to these procedures, but for brevity we will not explore the details of these further adjustments here.

Table One: Effect of changes in components of hospital procedures quality adjustment on output

An increase in ...	Effect on quality-adjusted output	Mechanism of effect
Health-related quality of life (HRQoL) gain reported in PROMs	Increases	Changes HRQoL gain
Proportion of treatments that are elective ⁹	Decreases	Changes HRQoL gain
Post-operative survival rate	Increases	Changes HRQoL gain
Average age of patients being treated (<i>life expectancy at birth unchanged</i>)	Decreases	Changes length of period gain experienced over
Life expectancy at birth (<i>age at treatment unchanged</i>)	Increases	Changes length of period gain experienced over
Waiting times (80 th percentile)	Decreases	Changes length of period gain experienced over

Primary care output adjustment

While the hospital procedures adjustment provides a quality adjustment for a large proportion of spending, the NHS also comprises many other smaller services. As previously mentioned, the relative paucity of detailed output data for other NHS services also applies to quality data.

For general practice, the largest component of primary care, a quality adjustment has been built using a selection of appropriate outcomes data from the Quality and Outcomes Framework (QOF), an incentive scheme for general practitioners (GPs). These measures relate to the extent to which patients' health risk factors fall above or below risk thresholds, thus incentivising GPs to monitor, medicate and promote behaviours for healthy outcomes. For example, the proportion of patients with coronary heart disease who have blood pressure and cholesterol readings above a threshold. The quality adjustment is scaled down to reflect the fact that only a small proportion of the population has the relevant risk factors.

The GP outcomes adjustment is a demonstration of the fact that the collection of quality and outcomes data can vary as policy changes. QOF outcomes improved rapidly after their introduction as an incentive scheme in the early 2000s, but the scale of improvements decreased in subsequent years as GPs moved closer to the maximum achievable scores. As the QOF system has matured and the gains in outcomes have become more marginal, the number of measures collected has fallen further reducing the proportion of GP activity that the quality adjustment covers.

Patient experience adjustment

The patient experience adjustment covers a range of NHS services and was included to account for the issue that the other aspects of the quality adjustment do not incorporate non-clinical aspects of care quality which may also be valued by patients, such as being well-informed and involved in decision-making, having good relationships with staff and having a comfortable environment.

⁹ CHE assign a greater health gain factor to non-elective treatments than elective treatments (see paragraph on pp11-12.)

The patient experience adjustment is calculated using data from the Overall Patient Experience Scores, which are based on national surveys covering inpatients, outpatients, emergency care, community mental health services and primary care, although the patient survey for primary care has been discontinued since the quality adjustment was designed.

Generally, the patient experience scores experience only minor variation over time, and as with the GP outcomes adjustment, result in a relatively minor effect on the overall quality adjustment.

Education

Education services comprise eight publicly funded sectors ranging from pre-school to higher education training of teachers (Initial Teacher Training (ITT)) and is captured using data on student numbers for the respective sectors. Unsurprisingly then, the two largest sectors are primary schools and secondary schools (including academies) both in terms of spend and actives. Like other individual services, the output of education services is measured directly, reflecting changes in the aggregated activities delivered. However, looking deeper into this, the demography of the UK will mean that using pupil numbers alone gives only a very low or constant growth in education output over time, implying the actual output of the public education system may not have significantly improved during this period.

However, it is highly likely that factors like the quality of care pupils receives; the depth to which a syllabus is taught; the individual attention afforded to them by teachers; and the skill sets which they develop will have changed. Therefore, if these changes in quality are accounted for, it would be expected that the volume of education service output would change, even if demographic factors hold student numbers constant. Therefore, following Atkinson (2005), additional steps were incorporated to explicitly account for changes in the quality of provision in estimates of education output.

In general, the output of the education sectors is quality adjusted in two stages. Firstly, student numbers in sectors are adjusted by attendance rate. In line with specific recommendations outlined by Atkinson, rather than using pure registered pupil numbers, adjusting by absence aims to provide a more accurate measure of the amount of teaching activity received by pupils, so absence (both authorised and unauthorised) are captured.

Secondly, metrics of “high-level” attainment, using information about examination results, measure changes in the overall quality of services provided.

In Foxton (2018a), output associated with both primary and secondary schools is adjusted using the average point score (APS) per student in GCSE level or equivalent examinations, which are normally taken during the student’s eleventh year of schooling. The average point score (APS) for GCSE and equivalent qualifications relates to the attainment of pupils aged 15/16 at the end of Year 11. It is the best current measure for the annual change in the quality of output. It rests on the assumptions that the change in the APS can be used to approximate quality, and:

- should be applied to all pupils in primary and secondary schools¹⁰ (from reception class to the end of the sixth form) in the UK,
- is an adequate approximation for all educational outcomes, for example attainment after 16 and development of wider outcomes such as citizenship.

As these examinations vary across geographical areas, the APS quality-adjustment is applied to Primary and Secondary school output in each country separately. The APS at GCSE level for England and Wales are provided by the Department for Education and the Welsh Government respectively, while the APS associated with the Standard exams in Scotland are provided by the Scottish Government. For reasons of data comparability and availability, the level of education quantity in Primary and Secondary schools in Northern Ireland is quality-adjusted using the APS of English schools. Initial Teacher Training (ITT)

¹⁰ Including, in England, Academies and City Technology Colleges (CTCs).

quantity in each geographical area of the UK is adjusted using the QTS award rate for England, which is also provided by the Department for Education. Here the implicit assumption is made that changes in quality in ITT in Wales, Scotland and Northern Ireland follows the trend in England.¹¹

The Criminal Justice System

Introduced as part of Foxton (2018b), when measuring the output and productivity of public order and safety (POS), explicit adjustments are made to the measure of output from the Criminal Justice System (CJS) to take account of changes in quality and improvements in associated outcomes. The basic activity measures, common to both public service productivity estimates and national accounts, consist of cost-weighted aggregates of services provided (such as prison bed-days or cases processed per court) which are paid for by the UK government. This is covered in greater detail both in Part 1 of this paper and Annex B. The quality adjustments applied then consider some of the aspects of quality not already captured by the simple activity measure of output for POS.

Within the POS service area there are four main components: fire-protection, courts (which itself has five further sub-components), probation and prisons. The quality adjustments are applied to a sub-set of these components, as shown in Table Two, which are identified as forming part of the CJS, alongside an indication of the weightings used. A quality adjustment is not applied to fire-protection or county courts services, which deliver civil cases¹².

The criminal justice quality adjustment has four components:

- recidivism (re-offending) adjustment
- prison safety adjustment
- custody escapes adjustment
- courts' timeliness adjustment

The first relates to achieving an overall outcome - reducing re-offending - of the whole criminal justice system and therefore treats the CJS as one interlinking system that allocates and providing appropriate disposals¹³ and rehabilitation services. It can, however, be argued that the associated sub-components may have specific target outcomes, in addition to reducing recidivism. Therefore, the remaining three adjustments relate to specific target outcomes for sub-components of the CJS.

¹¹ This is a key issue in relation to geographical comparability – is it better to quality adjust all geographies, even when no quality adjustment data is available in that area, or is it better to present unadjusted data in these areas, even if this introduces a variation of its own.

¹² UK court cases are divided into 'civil' and 'criminal'. Civil cases covering areas of family and contract law do not address 'criminal offences' and are therefore out of scope of a quality adjustment designed around addressing criminal behaviour. Similarly, firefighting services are exempted from quality adjustment.

¹³ A disposal can be thought of as an appropriate sentence for the crime and mitigating factors, such as repetition, aggravation, or factors which make the case more severe (i.e. assault with a weapon, as opposed to assault without a weapon).

Table Two: Quality adjustment weights by output component

Component	Quality adjusted	Recidivism	Prison safety	Custody escapes	Courts' timeliness
Fire-protection Services	No				
Magistrates Courts ¹	Yes	50.0%			50.0%
County Courts ¹	No				
Crown Courts ¹	Yes	50.0%			50.0%
Crown Prosecution Service ¹	Yes	100.0%			
Legal Aid ¹	Yes	100.0%			
Probation	Yes	100.0%			
Prisons ²	Yes	29.2%	37.5%	33.3%	

Source: Office for National Statistics

Notes:

1. Subcomponent of UK courts and related activities
2. Weights for 'Prisons' quality adjustments are taken from Prison and probation performance statistics 2014 to 2015. Further details on sources and methods used can be found in Foxton (2018b)

Recidivism adjustment

The recidivism adjustment is applied across all output associated with the Criminal Justice System (CJS). It approximates the effect the CJS has on reducing the volume and severity of further crimes being committed by those who have gone through it – this being an important social outcome for the system. The ONS measure works by adjusting the cost-weighted activity indices of the service areas identified in Table 2 by a severity-adjusted rate of recidivism.

This adjustment itself is composed of three parts, the first being the change in the number of proven re-offences committed by adult and juvenile offenders categorised between crime types. These include such categories as Violence Against the Person, Robbery and Fraud. Secondly, an adjustment is made to offenders, to account for differences between cohort characteristics and their likelihood to re-offend. The final adjustment made provides a weighting by which to aggregate together all re-offences. This weighting is based upon the relative severity of the re-offence and is derived from ONS (2016). More information on this source, as well as others used, can be found in Foxton (2018b).

Prisons safety adjustment

The prisons safety adjustment relates to the number of incidents of assaults, self-harm and deaths that occur in prison custody. The purpose of this being to reflect that safety of prisons is an important component of the quality in the activity and services provided, as set out in the Prison Safety and Reform White Paper (MoJ, 2016).

We measure the number of incidents per 1,000 prisoners, which are grouped into “Severe”, “Less severe” and “Those resulting in a death”. These groups are subsequently weighted and aggregated together based on their relative cost. This is achieved by using the total cost to society of workplace injuries as a proxy, taken from the Health and Safety Executive.

Custody escapes adjustment

The escape adjustment relates to ensuring prisons fulfil the role of public protection and is applied to activities used to measure the output of the prison service.

The measure is based on changes in the difference between the number of escapes and a baseline of 0.05% of the England and Wales prison population – a historic target used by the Ministry of Justice. The purpose of this being that as the absolute number of escapes approaches zero, the relative change year-on-year would have a disproportionate effect on a non-baselined quality adjustment index.

Courts' timeliness adjustment

The courts' timeliness adjustment relates to the average time taken for criminal cases to be taken to completion, on the basis that the delivery of a sentence in a timely manner is favourable. However, there is currently no adjustment made to reflect whether there has been fair treatment of the suspect or victims or to allow the appropriate time for preparations of criminal cases with differing levels of severity or complexity.

For magistrate courts, the measure is based on the mean average time of charge and laying of information to completion. For Crown Courts, the measure captures the average waiting times experienced by all defendants and the mean time from main hearing to completion. As implemented, the measure accounts for changes in the average time taken to completion by criminal courts because increases in volume may reflect a worsening.

Adult Social Care

Adult social care (ASC) services comprise care and support provided to older people, adults with learning or physical disabilities, adults with mental health problems, drug and alcohol misusers, and carers. By spending, the largest two client groups are older adults and adults with learning disabilities. The care services covered by ASC include placements in residential and nursing care homes, home visits by carers, day care services and supported living arrangements in accommodation adapted to users' needs.

Unlike the National Health Service (NHS), which has provided free-at-the-point-of-use health care to patients since it was formed 70 years ago, ASC services have not undergone the same funding and policy unification. While the NHS is a single public body in each of England, Scotland and Wales, responsibility for the provision of ASC services remains with local authorities^{14,15}. The funding of ASC comprises a number of streams, with the main sources:

- Local authorities' own funds.
- Fees charged to clients, which for many services is subject to means tests based on clients' wealth and income.
- Payments from the NHS. There are a number of schemes under which NHS bodies transfer funds to ASC services and the funding transferred through these schemes has grown in recent years at the behest of government policy. The transfers are intended to financially support local authorities to relieve pressures on the NHS by reducing unnecessary hospital attendances by care clients (so-called 'bed-blocking') and promote co-operation between NHS and ASC service providers.

The provision arrangements are further complicated by most ASC services being contracted out by local authorities to independent sector (typically private firms and charities), while a minority of services continue to be provided directly by local authorities.

There is also a substantial private sector, where clients purchase ASC services directly from private providers without necessarily involving local authorities.

As explained above, following the Atkinson Review guidelines, the remit of public service output and productivity is delineated by public spending as opposed to public provision. Therefore, the proportion

¹⁴ Local government in England is divided between 152 'top-tier' local authorities, with fewer in the smaller nations of Scotland and Wales.

¹⁵ The Health and Social Care Service in Northern Ireland is a public body with responsibility for both health care and adult social care, although funding arrangements are similar to the rest of the UK, with health care being available free at the point of use, but certain ASC services means-testing and charging clients.

of ASC services funded by local authorities and payments from the NHS is within public service output and productivity, whether provided by local authority or independent sector providers; while client-funded activity is excluded from the measures.

Measuring the output of ASC services

Activity and expenditure for ASC services are measured using data collected by the NHS from local authorities, enabling the construction of a cost-weighted activity index. When the Atkinson Review was published, activity data were available for residential care, nursing care, home care, day care, the provision of equipment and home adaptations, meal deliveries and referrals and care assessments undertaken. Residential, nursing and day care were further split by client group to reflect differences in the costs of providing care to different groups.

However, this data collection was ended in 2013/14 and the collection that replaced it in 2014/15 covered a reduced set of activities, causing the proportion of ASC expenditure covered by the cost-weighted activity index to fall from 76% to just 36%. As a result, a new methodology has been developed (see Lewis, 2018b) which uses activity data, where available, to generate a cost-weighted activity index, and where it is not, calculates volume output using the ‘outputs equals inputs’ convention.

While this is the best measure available for output, the loss of such a large proportion of activity data limits our ability to measure productivity across the whole ASC service sector, although separate productivity measures covering the service elements for which activity data remain (residential and nursing care) are produced to analyse these services specifically.

Developing quality indicators for ASC services

While activity data matching the requirements of the Atkinson Review were readily available in the 2000s, suitable quality measures were not.

The absence of available quality measures for ASC was not only a problem for the implementation of the Atkinson Review guidelines but also a problem for policy analysts trying to understand the performance of the ASC sector, whose main data source was the inspection reports of care homes carried out by the Care Quality Commission.

As a result, the ONS organised a cross-body programme, Measuring Outcomes for Public Service Users (MOPSU) to develop a toolkit for measuring ASC outcomes, along with other strands on building quality measures for early years education and measuring the third sector (see ONS, 2010). The MOPSU project on ASC outcomes was led by the Personal Social Services Research Unit (PSSRU) at the University of Kent providing sector-specific research and economic expertise.

At first glance, health care and adult social care may appear to be similar services, with both involving the care of individuals with health problems. However, while the main element of the health care quality adjustment measures the gain in health resulting from a hospital procedure performed at a point in time, the primary benefit from social care is an improvement in quality of life over the period social care is being received.

The project considered several approaches to measuring the outcomes of social care (see ONS, 2007), including:

- The extra-welfarist approach, where the desired outcome is pre-determined by the researcher and achievement against this outcome measured on a scale.
- The hedonic psychology approach, which involves studying clients’ spontaneous approach/avoid, continue/desist and good/bad reactions at various moments in time as they use services.

- The capabilities and functioning approach, first developed by Sen (1985), which measures clients' opportunities or potential to obtain desirable 'functionings' such as being fed or having meaningful social relationships.

The approach taken followed the capabilities and functioning approach, and applied it to form a measure on a QALY-style zero-to-one scale, known as social care-related quality of life (SCRQoL). This is used as a quality adjustment on ASC output.

As described in the section on health care, there are several alternative questionnaire forms for measuring the quality of life element of QALYs, with the EQ-5D used in the NHS patient-reported outcome measures. The MOPSU project therefore needed to design a questionnaire for eliciting SCRQoL, based loosely on the capabilities described by Sen as essential elements of well-being.

An analysis of existing literature revealed eight broad domains which, with minimal overlap, appear to determine quality of life:

- Personal cleanliness and comfort
- Accommodation cleanliness and comfort
- Safety
- Food and nutrition
- Control over daily life
- Occupation
- Social participation and involvement
- Dignity

However, simply surveying care clients to rate their satisfaction on each of the eight domains against four possible responses for each creates two problems. Firstly, there is no reason to assume that each of the domains is of equal value to care clients – some may be more important to overall well-being than others. Secondly, it is not certain that the levels of responses that clients give against their experience (such as needs fully met, mainly met, partly met or not met) should be allocated a set of equally-spaced utility values, such as 1, 0.67, 0.33 and 0.

To deal with these issues, the PSSRU project worked with RAND Europe on a study to determine the relative importance of each of the domains, and various 'levels' of experience within the domains, by asking care clients to rank the best and worst outcomes of a range of possible 'levels' of the above categories.

This study¹⁶ enabled the construction of 'weights' for the preferences such that each 'level' of experience for each domain is attributed a utility value. Table Three shows an example with two of the domains. The weights demonstrate that a difference in utility value between the top and bottom level responses for the control over daily life domain (the client having as much control over their daily life as they want and the client having no control over their daily life) is greater than the difference in utility value between the top and bottom responses for the social participation domain (the client having as much social contact as they want with people they like and the client having little social contact with people and feeling socially isolated). Of the eight domains, control over daily life had the greatest range in utility between the highest and lowest response, and this was bounded between zero and one. However, the utility weighting study also revealed that the difference between the first and second level response of each domain was lower for the control domain than for the social participation domain.

¹⁶ While the MOPSU project established the principles of weighting different domains of quality of life, the actual weights used in the quality adjustment are derived from a later study based on a number of specific surveys (see Netten et al., 2012).

Table Three: Utility weights for two example domains

Domain level	Utility weight
<i>Control over daily life</i>	
1. I have as much control over my daily life as I want	1.000
2. I have adequate control over my daily life	0.919
3. I have some control over my daily life, but not enough	0.541
4. I have no control over my daily life	0
<i>Social participation and involvement</i>	
1. I have as much social contact as I want with people I like	0.873
2. I have adequate social contact with people	0.748
3. I have some social contact with people, but not enough	0.497
4. I have little social contact with people and feel socially isolated	0.241

Source: Netten et al. (2012)

Implementation of the adult social care quality adjustment

To collect the social care-related quality of life (SCRQoL) data needed to measure the performance of local authority social care services, the Adult Social Care Survey was introduced in April 2010 and now interviews over 10% of adult social care clients in England annually. The measure of SCRQoL, along with other outcome measures from the Adult Social Care Survey, form the Adult Social Care Outcomes Framework, a set of indicators used to evaluate the performance of local authority ASC services across England.

While a change in the measure of SCRQoL gives a good indication of changes in the well-being of the care population, the measure does not give a definitive answer on whether a change in SCRQoL can be attributed to social care services or results from changes in the underlying care population or their wider environment. For instance, an improvement in the average response to the control over daily life question in Table Three could result from improvements to the quality of care which result in clients being more involved in decisions about their care, but could also result from a change to the care population to include more lower-need clients whose health status may afford them more independence than other clients.

To produce an attributable quality adjustment, it is therefore necessary to develop a measure which isolates the effect of service quality on outcomes from the other factors which may also influence these outcomes. Adjusted social care-related quality of life (Adjusted-SCRQoL), was developed by the Quality and Outcomes of Person-Centred Care Research Unit (QORU) from the earlier work on SCRQoL to provide such a measure for the ASCOF and was introduced into the 2016/17 indicator set.

The Adjusted-SCRQoL measure controls for a range of factors outside the control of social care providers which may affect SCRQoL including age, health status, the suitability of clients' home for meeting their needs and clients' ease of travelling around outside in their local environment through using regression analysis to derive an estimate for the expected effect of these factors on SCRQoL (Forder et al., 2016).

While the Adjusted-SCRQoL measure has only been published in 2016/17 and for community care clients, the ASC output quality adjustment used by ONS for community care is produced using the same parameters using Adult Social Care Survey data for the period 2010/11-2016/17.

For residential and nursing care, the quality adjustment is derived from a similar regression analysis informed by Yang, Forder and Nizalova (2017) and controls for:

- gender
- ethnicity
- age

- self-reported health status, level of pain and level of anxiety
- the number of basic activities of daily living (ADLs) the client needs support with
- whether the client can deal with their finances and paperwork

The ASC quality adjustment therefore provides an estimate of the change in the key ASC outcome attributable to social care services.

Part Three: The big issues in measuring public service outcomes

As outlined above, generating quality adjustments for a variety of public services is both feasible and capable of significantly improving the quality of the statistics being produced. In addition, where methodologies have been developed the authors believe there is a strong potential for other countries to use these as a substantive foundation upon which to build methodologies tailored to their countries service design.

However, in line with Atkinson's recommendations we do not believe it is feasible in the UK, or any country to '*stand still*'. Continuous improvement and development of these methods is required, and to make the most of this opportunity it is sensible to consider the key lessons which the current practices provide, particularly if other countries wish to learn from these examples.

Similarly, before discussing the welfare implications of outcomes which the public services contribute towards, this section addresses the key issues and 'lessons learnt' from a decade of attempting to apply the Atkinson principles in the UK context. These are:

- How should various aspects of quality change be valued and weighted?
- How should different quality-adjusted services be weighted together?
- How do we keep pace with the rate of technological change?
- Should we following individuals or use aggregate data?
- What do we do when a change in policy affects our measure?
- Where do we source objective weights?
- How do we trade off consistency of estimates with different needs for data in relation to devolved matters?

How should various aspects of quality change be valued and weighted?

There are two symmetric problems in relation to the valuation and weighting of quality change:

- What to do when a public service delivers multiple outcomes, which could all contribute towards the quality adjustment we calculate for that service, and
- What to do when a single outcome is impacted by multiple parts of the public services?

Clearly for the first of these, when a common metric exists which can be applied to multiple outcomes, such as QALYs in health, this appears a trivial question: Once the value of a single QALY is established healthcare interventions can be theoretically evaluated by comparing their cost to the value of the number of QALYs they deliver per course of treatment.

Complexities can, however, still emerge. As explained above, the QALY has two elements, a health-related quality of life element and a temporal element; and can therefore combine the effect of improvements in health-related quality of life and increases in the length of life resulting from treatment. Data is required on both dimensions for the whole population to derive a quality adjustment, which is a non-trivial investment. Equally, while the increase in QALY following an intervention would be far

closer to a measure of the quality of health service provision than NHS performance indicators or broader outcome measures such as life expectancy, which would require attribution factors to be generated, the use of a quality adjustment solely based on QALY itself would still face the problems described in Part Two.

So, given that in the UK we do not use QALYs, the health quality adjustment is applied to output as a simple scalar variable, such that a 1% increase in the quality adjustment results in a 1% increase in quality adjusted output.

But changes in quality may reflect changes in the value of the service that are less or greater than this simple scalar imposes. Deriving accurate valuations to either weight contributions to the quality adjustment, or to weight the quality adjustment vis-à-vis the outputs remains a formidable challenge, as illustrated by research by Ryan et al. (2014) on the case of valuing patient satisfaction. Whilst a range of methods are available, eliciting firm reliable values is at present almost impossible, and relies on an ability to calculate objective weights. This can make it difficult to diagnose the exact causes of quality change over time, a non-trivial complaint in a measure regularly used to inform public policy analysis.

The second scenario is perhaps most easily explained in relation to the way health and social care interact to support improved health outcomes, or the interactions between the various agencies within the criminal justice system (CJS). The CJS is a collection of agencies working in partnership towards a common goal. The effective functioning of the CJS requires the processing of offenders from arrest to prosecution, to the delivery of justice – whether punishment or acquittal. An accurate measure of the increment to collective welfare from the CJS should reflect this. This implies that one cannot treat the police, the courts, and the prisons as entirely individual, stand-alone entities: The effectiveness of each agent within the CJS depends, to varying degrees, on the effectiveness of the others. For example, the quality of prosecutions undertaken by the Crown Prosecution Service will depend on the quality of the investigative work undertaken by the police. In describing the CJS we have sometimes used the analogy of a car engine. Subsequently, the question becomes one of attributing a system-level outcome (reducing re-offending) across the various component parts of the criminal justice system, when some of these, particularly prisons, have their specific quality measures (e.g. safety and decency). This can lead to some parts of the system having a lower weight on reducing re-offending when they may be assumed to have a higher weight than others.

However, we continue to need to weight together the activities which count towards output, and in doing so we need to weight their output by the quality measures, returning us to the first bullet above. This brings together the different outcome measures when different parts of the one system have different outcomes relating to them. In principle, it is most desirable to weight together different quality metrics or indicators based on the value placed by individuals and society on services and their various attributes. Such an approach was taken with the adult social care quality adjustment, using data from separate studies specifically commissioned to understand how social care clients valued different aspects of well-being. The stated preference approaches provided differential weights for different aspects of well-being and enabled different levels of responses to be assigned relative values.

Extending such an approach across all service sectors requires the conducting of a range of studies to identify such preferences. In each case, three key questions would need to be answered before extending this approach: whose preferences are used, how to reflect changes in preferences over time, and how to derive an ‘average’ valuation?

Another possibility may be to weight domains of quality according to the relative costs associated with them. This assumes that the revealed preference of service managers on what they consider important reflect social preferences. However, costs do not necessarily correlate with quality – this is where the Atkinson Review came into the story.

In the absence of adequate data then a default solution may be to tend towards equal weights. However, this method is clearly sub-optimal: the long-term robustness of equal weighting, lacking in empirical support, could undermine the validity of the associated measure produced. In relation to prisons, where re-offending, prison safety and preventing escapes come together as three clear outcomes we need to recognise, we used the approach of taking a weighted average of prison performance measures, as used by HM Inspectorate of Prisons, on the basis that these were set by Ministers in Parliament, and as such could be argued to represent social preferences. This relies on assumptions of government efficacy in delivering this role, but appeared justifiable over any other arbitrary set of weights. Such ‘social preferences’, however, cannot be observed in all areas of public services, although the experiment of using such performance structures may be replicable in other ‘inspected services’.

How should different quality adjusted services be weighted together?

The challenge described above in relation to a single service, become an even more complex matter when one begins to combine services together. The classic method used under Atkinson is to do this using cost weights. These are objective, and if one believes the marginal pound is efficiently allocated by government then should be reflecting equal value. However, if one considers the question; would a 10% reduction in the quality of £1,000 of spending on health be worth more or less than a 10% reduction in the quality of £1,000 of spending on forestry, one can immediately see that, by dint of the sheer differential in the volume of funding devoted to these two services, this would likely deliver very different marginal impacts and public reactions. Therefore, are cost weights appropriate if we cannot break down changes in costs between changes in the prices of output and changes in the quality of the outcomes delivered by these outputs? Diewert and Fox (2017) present arguments for alternative weighting approaches based upon the relative value to users, which the authors consider intuitively strong and worthy of significant further consideration.

Keeping pace with new technology, systems and data

An additional recommendation raised by Atkinson was the need to maintain and continue to develop quality adjustments through time. Quality measures which were identified initially, particularly in periods of rapid technological change, may no longer be fit for purpose and may, with the passing of time, fail to continue to measure the key underlying principle you are trying to measure. Principled measures are key, but must reflect change.

For example, the health quality adjustment itself remains largely unchanged since its introduction in 2005, but the range of metrics health policy analysts study has not. In 2010, a new set of indicators for measuring health care performance, the NHS Outcomes Framework (NHSOF), was introduced and has become the central source for analysts measuring health care outcomes. Having been produced prior to the NHSOF, the results of the current health care quality adjustment do not always triangulate with the story that health care policy analysts derive from the NHSOF. In 2016, the Centre for Health Economics at the University of York convened a workshop to bring together policy analysts and health economists to consider the criteria that should be used for selecting NHSOF indicators, and this was followed up with a paper, Bojke et al. (2018), applying the criteria to these indicators. The challenges of adopting NHSOF indicators within a quality adjustment are considerable, as they are not drawn from a single data source, and so are published at different times and often variable frequencies. However, such a review demonstrates the need to regularly review quality measures to ensure their continued relevance as policies and data sources change and may lead to a quality adjustment with relevance to users.

The *quid pro quo* here is the allocation of development time by NSIs. In the authors’ experience the trade-off between investing in updating existing quality adjustments versus the creation of new measures covering new service areas has regularly required consideration. In recent years, new service areas have been prioritised where developments in reporting and data sources have opened the door to creating a new adjustment at relatively low costs. The continued pace of change in health and education,

our two largest services, however probably make the need to revert to revising their measures inevitable in the coming years.

In relation to new data, where it is difficult to forge a link through to an individual's experience, which is the approach followed in health and adult social care, we have found that the most practicable application is through the use of published data sources, which are generally aggregated at the population level to track the movement in group or average performance. Criminal Justice is a prime example of this. Efforts to focus on individual offenders, or the 'offender journey' resulted in a failure to deliver a quality adjustment in this space until 2017, when the ONS changed tack to focus on aggregate performance data. The key to unlocking this was the delivery by ONS of an experimental dataset on the relative severity of crime (capable answering questions like 'how many burglaries equal a murder?'), which provided a set of objective weights to adjust raw re-offending data and provide a consistent measure of whether outcomes were improving or weakening through time.

This approach brought the additional benefit of allowing service providers to better engage with the productivity statistics, as they were grounded in concepts and measures they were currently working with and understood. Providing objective insights where these had been missing previously appears to the author's to be a positive direction of travel, particularly when this work could be delivered at little additional cost.

What do we do when a change in policy affects our measure?

There are instances where the measure itself is subject to policy decisions, and is directly affected by policy change, not just in terms of the level, but also in terms of the definition of the measure itself. This is, broadly, always the case, but in some areas it is more pertinent than others. This is particularly the case when there are fears of 'gaming', that is where the definition of the measurement itself leads to undesired outcomes. Education is a prime example of where government policy has been shaped by the need to address a set of interlocking concerns.

Whilst the ONS has historically used GCSE APS attainment as a quality metric, the actual application has changed noticeably over time, in response to three key issues, where there was a perceived threat that the measure had been corrupted.¹⁷ Firstly, there is the question of whether attainment through time has been consistently measured or suffered from 'grade inflation'¹⁸, secondly have schools making greater use of 'easier' or more vocational courses to artificially inflate APS scores, and thirdly have schools improved marks by '*teaching to the test*' rather than giving a rounded education.

In light of this, the Department for Education established a review which found evidence of improvement in pupil's attainment in England over the period. However, when similar analysis was carried out on other measures and systems of pupil attainment used in the UK and within the OECD, they found, in contrast to the APS, little overall improvement in the level of pupil attainment. It is, however, worth noting that these findings were based on less timely data based on much smaller sample sizes than national performance data, but they called into question the validity of GCSE APS as a proxy either of educational attainment at that age, or as a proxy for the whole system, as the current quality adjustment implies.

¹⁷ Notwithstanding the fact that our experience of using a single measure (age 16 GCSE test results) as a proxy for performance across the age spectrum is that this model makes it difficult to reflect differential performance in one part of the education system (e.g. primary or early years) against another (e.g. secondary), when the quality measure simply does not capture more than one of these.

¹⁸ The converse argument is that, in the face of increasing tuition fees, and low wage growth in low skilled jobs, students have responded to market forces by investing more heavily in their own development whilst education is free, resulting in improving performance.

To address these worries, reforms to GCSE grades were introduced by the Department for Education in 2014, following the Wolf Report (2011). This changed the qualifications eligible to count towards APS, particularly in relation to vocational qualifications on school performance measures in England¹⁹. To reflect this an alternative approach to quality adjusting UK public service education output was proposed (ONS, 2015a) and adopted (ONS, 2015b).

The method replaced the use of APS data for England with Level 2 attainment at age 16 for years 2008 to 2013. Level 2 attainment at age 16 (or L2) equated to achievement of 5 or more GCSEs at grades A*-C or an eligible Level 2 vocational qualification of equivalent size. This is a threshold measure of the percentage of students achieving a particular level of attainment, compared to the APS which takes into account the full distribution of attainment data, making Level 2 attainment less susceptible to changes in the education system and pupil behaviour. This is the current method used for quality-adjustment in England. However, alongside these changes, in 2017 a further revision to GCSE grading was introduced which presented a more fundamental challenge. In a further effort to address perceptions of grade inflation, a new grading structure was introduced. This deliberately did not enable a one-to-one matching with the old banding structure, introducing computational challenges in preventing a discontinuity in the series.

Table Four: Old and New GCSE band equivalences

Old structure	New structure
A*	9
A	8
	7
B	6
	5
C	4
D	3
E	2
F	1
G	
U	U

Clearly, a measure which is the subject of frequent change is not a stable base upon which to build a long-term quality adjustment.

How do we trade off consistency of estimates with different needs for data in relation to devolved matters?

As mentioned above in relation to the education quality adjustment, for reasons of data comparability and availability, the level of education quantity in Primary and Secondary schools in Northern Ireland is quality adjusted in line with that applied to English schools.

Similarly, while current measures and methodologies to reflect quality change in the CJS are applied to the output of the UK as a whole, the associated metrics reflect but a subset – covering England and Wales. Here the implicit assumption is made that changes in quality in CJS in Scotland and Northern Ireland follows the trend observed in England and Wales.

¹⁹ The significant increase in APS between 2008 to 2009 and 2011 to 2012 that could partly be attributed to increases in the number of non-GCSE examinations taken because of changes in the type of examinations, which counted towards performance.

Whilst only UK level estimates are produced we can, to some degree duck these issues, but in light of a growing need to provide statistics for devolved administrations and lower-level geographies it is clearly problematic to either attempt to compare an area whose quantity of output is quality adjusted with one which is not, or to quality adjust two areas by an adjustment factor derived in only one of them. In a world where decision-making powers in relation to these services have been devolved to administrations in each of the component countries of the UK it is clearly problematic for decision-makers in Northern Ireland or Scotland to have to view the productivity of services they have responsibility for through a lens which can be argued to be distorting their view of their system relative to the other nations of the UK. This issue was explicitly recognised in Atkinson's Principle E.

This is exacerbated where different administrations or legal systems have resulted in long-standing differences between the model of services provided by the constituent countries, their methods of delivery and the machinery of government. These differences are set to become potentially more important because of devolution. Likewise, by applying common factors, we may well fail to reflect variations in priorities/desired outcomes, particularly as quality metrics and their associated weightings become more granular.

Given that many public services in the current model are not quality adjusted and at the aggregate level we are therefore regularly comparing quality-adjusted and non-quality adjusted sectors, ONS is exploring removing non-native quality adjustments where these are currently applied.

Part Four: The treatment of non-attributable outcomes as welfare gains

Atkinson and the first three parts of this paper focus on the outcomes which are directly attributable to the activities and outputs of the public services, however there is merit in stepping back to consider some fundamental questions about the exact scope under consideration and the implications of that scope on the object of interest: welfare gains.

At its simplest, the Atkinson framework conceives that the volume of activity is not adequately measured by the outputs of that sector if insufficient attention is paid to quality change. For products in the market-sector this is captured through adjustments made to the deflator to de-compose price changes into those caused by changes in the general price level and those caused by changes in the quality of the product. The relative price of a product should increase as its quality increases. This is obviously a more complex exercise when prices cannot be observed, and the public services, where such prices do not exist exemplify this. The Atkinson Review therefore argued for the application of quality adjustments derived from directly observable data.

Are then these quality adjustments equivalent to the welfare gains received by consumers? The answer here from Atkinson is unequivocally 'no'. The Atkinson quality adjustments only captures that aspect of the welfare gain which is directly attributable to the service.

However, in terms of the debates about measuring the modern economy and the need to understand why citizen's increasingly view GDP as a poor proxy for welfare measures, this is a key point for two reasons:

- If public services are a significant fraction of GDP, and quality adjustments have a noticeable impact on volume growth in relation of these services, and this is not being taken into account, as it isn't in the UK and Europe then this may introduce a wedge between GDP and welfare growth even if GDP should share a common growth rate with welfare.
- Welfare gains from outcomes which relate to, but are not attributable to public services, may be a significant driver of this perceived difference in the behaviour of GDP and welfare, but may not currently be captured in our welfare measures.

Should we then consider moving away from the Atkinson model of bottom-up estimates of the directly attributable component and instead focus on a top-down apportionment of a more universally estimated outcome? There are three arguments for doing this:

- We could then gain additional information on both GDP and welfare effects through the element attributable to the public services, and the part apportioned elsewhere.
- As fears about how very different models of public service delivery could be consistently quality adjusted by multiple countries was an important reason for not including quality adjustment in ESA10, it is possible that using a more standardised method based on readily available data may be more attractive. Such data is already being produced on consistent bases for the Sustainable Development Goals, and it is possible that developing countries may be better able to make use of this approach within limited approaches.
- We have identified in part three that whilst by no means insurmountable, identifying methods to derive only attributable quality adjustments is not methodological simple, and very possibly no simpler than a top-down method.

This is not to argue, however, that this issue does not come without its own challenges to overcome. Three methodological issues emerge:

- **How to calculate the value of the improvement in outcomes?** Much of the work undertaken by the ONS is in terms of growth indices: so, if re-offending falls from a base level by 1%, this is applied as an uplift to the volume of output produced by the criminal justice system. However, this does not require an estimate of the absolute value of that re-offending to be produced, merely for the growth rate to be captured and applied equally to the ‘public services part’ and the ‘non-public services part’, on the assumption that these both have the same impact on growth through time. To capture the value caused by factors outside the productive boundary this method would need to be adapted to monetise rather than just scale the outcomes data.
- **How to estimate weights to apportion this value between the public services and other factors?** Atkinson and his methodologies absolutely side-step this issue by only using measures, particularly survival rates, weighting times and patient satisfaction, which directly relate to the services in scope. Taking health, for example, this means simpler, more direct measures, such as life expectancy are currently considered an unsuitable quality adjustment given that it is affected by factors outside of the control of the health service.
- **How to keep apportionment weights up to date?** Whilst it is possible to make the simplifying assumption that the two components have the same impact on growth through time, in which case it does not matter what proportion you attribute to the public sector²⁰, this is clearly sub-optimal as it does not allow any change in relative effectiveness or productivity to emerge. To gain an accurate understanding of the impact of public services, we need to keep this apportionment up to date, because of the difference of technical and allocative efficiency. If public services become more efficient at delivering outcomes than non-public service drivers this may affect citizen’s allocation decisions. Let us imagine that in the following year medicine has become more effective and is an increasingly better substitute for giving up smoking or gym attendance. The technical efficiency of healthcare has changed, but it may make sense for there to be a change in resources by the citizen to achieve allocative efficiency, by swapping out going to the gym and increasing use of medication. In this instance, the apportionment factors described above would need to be altered, ascribing more weight to the activity of the public service, as opposed to that of the individual.

²⁰ If you estimate re-offending has fallen by 1%, and this is composed of two elements, A & B, with weights (x) and (1-x), respectively, where the weights sum to unity, then if one assumes both A & B fall by 1% also, then x can take any value from 0 to 1.

So, let us consider the key steps which one may take to meet these three methods challenges, or at least to understand whether they would so onerous as to make such an approach unfeasible.

Clearly in the first instance we would require international agreement on the outcomes which should be in scope for the exercise. Whilst frameworks such as the Sustainable Development Goals provide us with a start, there are two issues: firstly, the SDGs contain an exceedingly large number of metrics, so one imagines that these would need to be short-listed, but secondly there is the potential that some of these outcomes may in conflict with one another. Let us consider the headline outcomes of improving life expectancy and providing national security to citizens. It is easy to imagine a situation where providing national security involves sending troops into battle and large number of casualties being incurred. In such instances, would we see a positive impact on the security outcome measure but a negative on the life expectancy.

Assuming we can derive an agreed set of outcome measures, the second stage is producing monetised values of changes in these outcomes, both positive and negative, to generate metrics in the same denomination as the National Accounts to allow comparison. Whilst in some areas this appears feasible, in others we may reach the same challenges discovered in the Atkinson bottom-up method. Theoretically government cost-benefit analyses should provide key insights in instances where these are available.

Next, we will need a parallel set of analyses to be run, or values to be available from elsewhere, to apportion impacts to public services and other factors. There is a question of granularity here. If three public services contribute to an outcome, do we need to be able to produce granular quality adjustments for each outcome, to allow us to observe the relative productivity and efficacy of different services?

Whilst it is beyond the scope of this paper to provide a comprehensive assessment of the ability to deliver each of these requirements for each of an un-defined set of outcomes, we can consider a thought experiment in relation to the UK's largest public service. Let us imagine we wished to us a single top-down measure which could be used in place of the complex collection of measures currently used and which related to both public service and non-public service driver of health.

The most obvious outcome is life expectancy. Life expectancy has three clearly relevant attributes as a quality adjustment measure: a) that is an overarching indicator rather than one that only applies to certain services with the health service, b) it is clearly very important, and c) if one wished to undertake these types of quality adjustments in multiple countries in a consistent fashion, then this metric is one that almost all countries would be capable of implementing²¹.

To commence, a quality adjustment index from life expectancy statistics requires us to control for population change, essential as any output measure to which the quality adjustment would be applied would implicitly account for population growth. The index would also need to control for small changes in the gender ratio at birth, to account for differences in male and female life expectancy. Table Four shows an example of such an index.

To apply the adjustment to healthcare output and incorporate it into GDP, it would be desirable to monetise the gains in life expectancy attributable to healthcare. For this we use the value of a quality adjusted life year (QALY), as used by NICE in healthcare economic evaluations. In terms of the value of life expectancy, this provides a value of £30,000 for each additional quality-adjusted life year, the upper end of the range used by the economic evaluations conducted by NICE²². As such, in addition to

²¹ Under the simplifying assumption that gains in life expectancy in any given year are a result of improvements in health care and other factors in that year, as opposed to those made in previous years.

²² As shown by Ryen and Svensson (2014), studies undertaken on willingness to pay per QALY yield a wide range of results, with a mean value greater than the NICE range.

the example life expectancy quality adjustment index, Table Four also shows some example figures for the value of healthcare quality. The values shown are built from a range of assumptions.

Table Four: Example index of quality adjustment for life expectancy

	Est. average lifetime QALY		Est. increase in the stock of lifetime QALY attributable to healthcare (£bn) (4)	Est. annual valuation of lifetime QALY attributable to healthcare (£bn) (5)	Quality adjustment index for health output based on est. value of lifetime QALY (6)	Existing Atkinson output quality adjustment for est. QALY gain from healthcare (7)
	Male (2)	Female (3)				
2001	65.56	69.04		184.3	100.0	100.0
2002	65.73	69.17	96.8	185.5	100.2	101.1
2003	65.97	69.29	109.2	186.8	100.5	101.4
2004	66.26	69.51	138.2	188.5	100.9	102.1
2005	66.60	69.81	189.1	190.9	101.3	103.0
2006	66.75	69.90	133.9	192.5	101.5	104.0
2007	67.14	70.28	214.0	195.2	102.1	104.3
2008	67.39	70.53	187.6	197.5	102.5	104.7
2009	67.69	70.82	182.0	199.7	102.9	105.0
2010	68.04	71.06	199.7	202.2	103.4	105.3
2011	68.35	71.30	201.7	204.7	103.8	105.9
2012	68.56	71.38	145.6	206.5	104.0	105.9
2013	68.73	71.47	137.6	208.2	104.2	106.3
2014	68.85	71.56	153.6	210.1	104.3	106.9
2015	68.85	71.54	133.2	211.7	104.3	107.2

Source: Author's calculations using ONS data and assumptions sourced from McGinnis et al (2002) and NICE.

Columns two and three in Table Four present ONS estimates of average lifetime quality adjusted life years (QALY). As covered in section two, the quality adjusted life years are measured using the level of health-related quality of life a person has (bounded between zero and one) over each year they live.

Health-related quality of life varies over the population depending on health status, and therefore to get a better estimate of population quality adjusted life years, we therefore need to draw on more than a simple series of life expectancy. The General Practice Patient Survey (GPPS) records health-related quality of life²³ for respondents with and without a long-term health condition and by age group, as contained in Table Five.

²³ The GPPS uses the same EQ-5D scale to record health-related quality of life as is used in the existing healthcare output quality adjustment, as covered in section 2. The GPPS data are taken from a point in time (Q1 2017), whereas mortality and healthy life expectancy data are taken by year.

Table Five: Health-related quality of life status²⁴

Age	No long-term health condition	Long-term health condition
aged 16 to 24	0.928	0.761
aged 25 to 34	0.927	0.770
aged 35 to 44	0.921	0.756
aged 45 to 54	0.913	0.730
aged 55 to 64	0.910	0.725
aged 65 to 74	0.911	0.742
aged 75 to 84	0.869	0.689
aged 85 or over	0.775	0.563

In addition to the GPPS, the number of years an individual is expected to live in good health is estimated in the ONS Healthy Life Expectancy statistics. Thus, a quality adjustment can be derived from combining data on health-related quality of life with life expectancy and healthy life expectancy statistics.

This measure could be made more sophisticated by undertaking the analysis at an individual level using mortality data, to derive estimates for the average quality-adjusted life years experienced by the population over the course of their life. Such a measure is more sensitive to reductions in the proportion of the population dying at a young age, when more years would have been expected to be lived in good health, than to gains in years lived in older age groups.

Using this data, we then derive three options for a QALY-driven adjustment or a direct estimate of outcomes in columns four to six. Column seven presents the current health quality adjustment applied to health service output in the existing public service productivity estimates.

The methods behind these different models is described below:

Column four presents the estimated increase in the stock of lifetime QALY attributable to healthcare, is calculated by multiplying lifetime QALY for each gender by the unit value of a QALY, as drawn from NICE, and an apportionment factor as changes in the quality adjusted life years the population experiences cannot be wholly, or even mostly attributed to health care. Various studies have looked at the proportion of population health improvements can be attributed to health care relative to other sources. The King's Fund (2015) provides a short literature review, covering three studies, giving values of 10-15%, up to 25% and 43%, respectively.

Taking the mid-point of the most conservative of these, McGinnis et al (2002), would give us a value of the proportion of population health attributable to healthcare of 12.5%.

Applying this factor to the marginal annual change in the total value of the stock of lifetime QALYs for the current UK population generates the estimates in column 4 of Table 4. Applying this factor to the total annual stock and dividing through by the average (raw) life expectancy in the latest year delivers column 5, which can be considered an annual flow of life-expectancy benefits attributable to healthcare.

With these assumptions in place, columns 4 and 5 in table 4 suggest an order of magnitude in 2015 for the annual increase in the value of the stock of total quality-adjusted life expectancy of £133bn and an annual flow of life-expectancy benefits attributable to healthcare was £212bn. But as these figures

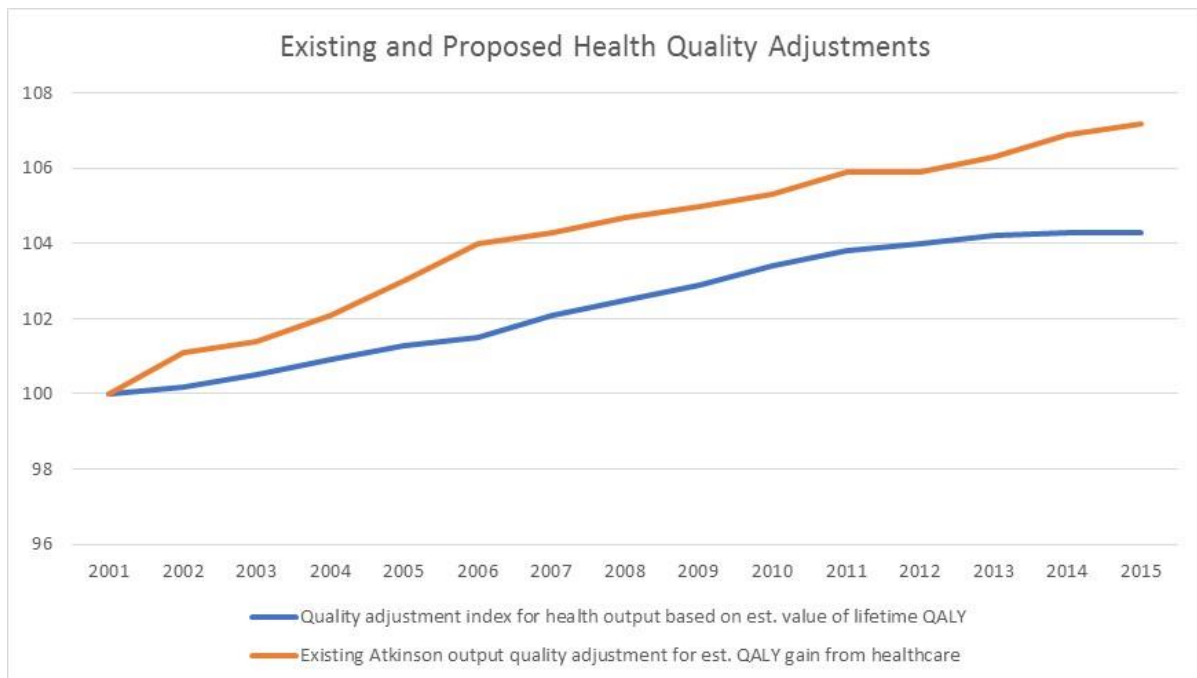
²⁴ These figures are based on the 2017 GPPS. In theory, it would be appropriate to include changes in health-related quality of life over time, although data from the GPPS do not suggest a material change over time.

represent the 12.5% difference in life expectancy that can be attributed to healthcare, there is a remaining value of life expectancy, some £1.5tn, not accounted for in GDP.

Much like healthcare output, both these measures are influenced by changes in the population. Both Columns 4 and 5 increase with both increases in average QALY per head and with population. Whilst these provide good valuations of total value of health in the UK, there is a good case to be asked in terms of whether population change should be controlled for. In an example where the population grew at a faster pace than a fall in QALYs per head, we would see these measures increase even though the average quality adjusted life expectancy is falling. Whilst it can be argued that health services would be reaching a larger population, this is an important debate for which we do not have space in this article. We merely conclude that columns 4 and 5 provide two alternative methods on estimating directly the value accruing to the UK from observable high-level health outcomes resulting from health services. Taking these back to the Atkinson methodology, such estimates would not act as a ‘quality adjustment’ applied to health service output, but would instead, by being a direct measure of the monetary value of the desired outcome, would entirely replace output based measures within the National Accounts. Conversely, if one stripped out the effect of population growth one would instead be able to create an index which could serve as a traditional ‘quality adjustment factor’ for public service output, as is presented in column 6 of Table 4, as a “Quality adjustment index for health output based on est. value of lifetime QALY”.

Column 7 provides the current ‘bottom-up’ quality adjustment which can be compared to these ‘top-down’ measures. Columns 6 and 7 can be compared and when one focuses only on QALY-measured health outcomes, rather than the basket of measures described in Part Two, that the quality adjustment currently used grows at a consistently faster pace. Importantly, the Atkinson measure may be expected to differ from the life expectancy adjustment as it also includes improvements in satisfaction and primary care service performance, alongside measures of QALY gain from hospital services.

Figure Two: Comparing alternative quality adjustment indices



Stepping away from the quality adjustment to GDP estimates, however, and returning to the wider benefits of improvements in QALYs not caused by public health services, whilst clearly both partial and simplistic the £1.5bn value of the welfare gains attributable to health improvements which are not caused by public health services is an obvious scalar for other aspects of the modern economy and the

relative importance of further research into different aspects of change in the modern citizen's experience. Considering this is only one aspect of 17 high level SDG domains, the scope for this agenda to make a substantive difference to our perceptions of the societies we live in is clear. This does not imply the author's do not recognise the impact of, say, free at the point of delivery digital goods. We merely ask if these gains are in the same ball park as those we estimate above and where the marginal investment is most productively employed.

Part Five: Conclusions

This paper discusses possible ways forward in two related but different areas. It draws upon UK experience for this purpose.

One relates to the ongoing but unfinished agenda as to how to measure the outputs of goods and services which are free at the point of delivery, for the purposes of national accounts. Public services such as schools and health services are the major example of this kind. Over a decade ago, Sir Tony Atkinson provided a principled framework for this end. Consistent with the basic principles of national accounting, he advocated an approach by which this output should be measured as the value added by the services concerned. This value, in turn, equated to the improvement in outcomes directly attributable to the activities of the public services concerned.

Implementing this approach, as he recognised, is by no means straightforward, but the UK experience recounted above shows that strong progress can be made. Working with experts and practitioners, quantity and quality measures can be identified and used to give a good approximation to the value added by the key public services, and thus their contribution to GDP. New data and intelligent use of existing data means this can be done at low cost and in a way which maximises stakeholder understanding and acceptance.

But national statistical institutes are also now grappling with a second task; measuring changes in welfare or wellbeing more generally, regardless of how they are generated. Health outcomes - life expectancy, for example, or healthy life expectancy – are influenced by a variety of factors besides publicly funded health services: diet, smoking prevalence and other lifestyle choices are obvious determinants. So, the central tasks under this agenda become first the identification of appropriate measures of outcome changes and then to determine how much value our societies place on those changes.

Adopting approaches based on clear principles, as Atkinson advocated, looks important for both agendas. For one thing, the outcomes used for the purposes of measuring the output of public services should be consistent with those used for measuring welfare more widely. Secondly, a principled approach helps to ensure intellectual rigour. Third, international comparability is important. The specific circumstances and institutions of particular countries will vary and methodologies need to take this into account. But provided methodologies are all based on the same underlying principles, comparability can be safeguarded, particularly if they make use of commonly accepted and produced high level outcome measures.

Our illustration of using a widely recognised measure of well-being, life expectancy, and readily available information on the shares attributable to health services and hence non-health service factors, alongside existing estimates of the value of a quality adjusted life year have enabled us to create estimates of welfare gains alongside new quality adjustments which could be used to feed into GDP under the SNA. Similarly, in education, it would be possible to use human capital as a similar high-level outcome. Improved skills might be expected to lead to higher wages and salaries which equate to further value for those who earn them. There may also be an inter-temporal impact. Improved human capital might be expected to lead not just to higher wages and salaries now but over a period of time. Applying necessary discounting would enable us to quality adjust the value of educational services in

GDP and the impact of non-educational drivers of human capital in a welfare estimate or welfare account.

The challenges presented by these twin agendas are ones we believe the statistical community needs to take up. We are convinced that whilst implementation raises non-trivial issues, these are not insurmountable. If we chose not to do so, we would have little to say about a fifth or so of our respective economies. We would also miss a vital contributor to measuring the changing well-being of our societies. The cost of such a decision to our reputation would be profound. In a world where digital innovation is offering a stream of new free goods and services which undoubtedly add to welfare, missing flows of value such as described above would cast any measure of welfare into doubt as incomplete and potentially misleading. The need to tackle these issues is both important and pressing. Failing to push on from the start that Atkinson established in this area would be a huge opportunity missed.

Bibliography

Atkinson, A. (2005) 'Atkinson Review: Final report. Measurement of Government Output and Productivity for the National Accounts'. Available at <https://www.ons.gov.uk/ons/guide-method/method-quality/specific/public-sector-methodology/articles/atkinson-review-final-report.pdf>

Bean, Sir C., (2016) 'Independent Review of Economic Statistics: Final Report, 2016.' Available at https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/507081/2904936_Bean_Review_Web_Accessible.pdf

Bojke, C. et al (2018) 'Accounting for the Quality of NHS Output', Centre for Health Economics, University of York: York

Castelli, A., M. Chalkley and I. R. Santana (2018) 'Productivity of the English National Health Service: 2015/16 Update', Centre for Health Economics, University of York: York

Dawson, D. et al (2005) 'Developing new approaches to measuring NHS outputs and productivity', National Institute of Economic and Social Research/Centre for Health Economics, University of York: York

Diewert, E. & Fox K.J. (20x17) 'Productivity Measurement in the Public Sector: Theory and Practice', Vancouver School of Economics, University of British Columbia: Vancouver. Available at <https://pdfs.semanticscholar.org/634a/0bbf731c4269d865dcdc22358bd8e4283be3.pdf>

Ford, G. and J. Lewis (2018) 'UK Health Accounts: 2016', Office for National Statistics. Available at: <https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthcaresystem/bulletins/ukhealthaccounts/2016>

Forder, J. et al (2008) 'Measuring Outcomes in Social Care: Second Interim Report', Personal Social Services Research Unit, University of Kent: Canterbury

Forder, J. et al (2016) 'Identifying the impact of adult social care: Interpreting outcome data for use in the Adult Social Care Outcomes Framework'. Quality and Outcomes of person-centred care Research Unit, University of Kent: Canterbury

Foxton, F. (2018a): 'Public service productivity estimates: total public service, UK: 2015' Office for National Statistics. Available at <https://www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/articles/publicservicesproductivityestimatestotalpublicservices/2015>

Foxton, F. (2018b): 'Quality adjustment of public service public order and safety output: current method' Office for National Statistics. Available at <https://www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/methodologies/qualityadjustmentofpublicservicepublicorderandsafetyoutputcurrentmethod>

Glover, D. and J. Henderson (2010) 'Quantifying health impacts of government policies', Department of Health, Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/216003/dh_120108.pdf

Hicks, Sir J.R. (1941) 'The Rehabilitation of Consumers' Surplus,' Review of Economic Studies.

Lewis, J. (2018a) ‘Public service productivity estimates, healthcare: 2015’, Office for National Statistics. Available at:
<https://www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/articles/publicservicesproductivityestimateshealthcare/healthcare2015>

Lewis, J. (2018b) ‘Measuring adult social care productivity’, Office for National Statistics. Available at:
<https://www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/articles/measuringadultsocialcareproductivity/2018-06-01>

King’s Fund (2015) ‘Inequalities in life expectancy: Changes over time and implications for policy’, retrieved 26/10/2018. Available at
https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/inequalities-in-life-expectancy-kings-fund-aug15.pdf

King’s Fund (2017) ‘What’s going on with A&E waiting times?’, retrieved 2/8/2018
<https://www.kingsfund.org.uk/projects/urgent-emergency-care/urgent-and-emergency-care-mythbusters>

Kuznets, S. (1937) ‘National Income and Capital Formation, 1919–1935’. NBER

Kuznets, S., Epstein, L., & Jenks, E. (1941): ‘National Income and Its Composition, 1919–1938’. NBER.

Mackillop, E., and S. Sheard (2018) ‘Quantifying life: Understanding the history of Quality-Adjusted Life-Years (QALYs)’, *Social Science & Medicine*, Volume 211, August 2018, Pages 359-366

Mason, A., P. Ward and A. Street (2011) ‘England: The Healthcare Resource Group system’ in R. Busse (ed.) ‘Diagnosis-Related Groups in Europe’, European Observatory on Health Systems and Policies

McGinnis, J.M., P. Williams-Russo and J. Knickman (2002) ‘The case for more active policy attention to health promotion’, *Health Affairs*, Volume 21, Number 2, pp. 78-93. Available at
https://www.researchgate.net/publication/11464071_The_Case_For_More_Active_Policy_Attention_To_Health_Promotion

MoJ (2016), ‘Prison safety and reform’, Ministry of Justice, Available at
<https://www.gov.uk/government/publications/prison-safety-and-reform>

Netten, A. et al (2012) ‘Outcomes of social care for adults: developing a preference-weighted measure’. *Health technology assessment*, 16 (16). pp. 1-166.

ONS (2007) ‘Initial report: quality measurement framework project’, Office for National Statistics. Available at
<http://webarchive.nationalarchives.gov.uk/20110110153641/http://www.ons.gov.uk/about-statistics/methodology-and-quality/measuring-outcomes-for-public-service-users/mopsu-reports-and-updates/initial-report.pdf>

ONS (2010) ‘Measuring Outcomes for Public Service Users’, Office for National Statistics. Available at
<http://webarchive.nationalarchives.gov.uk/20110110153641/http://www.ons.gov.uk/about-statistics/methodology-and-quality/measuring-outcomes-for-public-service-users/mopsu-reports-and-updates/mopsu-final-report.pdf>

ONS (2014) 'Health Expectancies at Birth and at Age 65 in the United Kingdom: 2009-11', Office for National Statistics. Available at

<https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/bulletins/healthexpectanciesatbirthandage65intheunitedkingdom/2014-11-18>

ONS (2015a) 'Methods changes in Public Service Productivity Estimates: Education 2013', Office for National Statistics. Available at

<http://webarchive.nationalarchives.gov.uk/20160106223636/http://www.ons.gov.uk/ons/guide-method/method-quality/specific/public-sector-methodology/articles/methods-change-in-public-service-productivity-estimates--education-2013.pdf>

ONS (2015b) 'Public Service Productivity Estimates: Education 2013', Office for National Statistics. Available at

<https://www.ons.gov.uk/peoplepopulationandcommunity/educationandchildcare/articles/publicserviceproductivityestimates/education2013>

ONS (2016) 'Crime Severity Score for England and Wales.' Available at <https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/articles/researchoutputsdevelopingacrimeseverityscoreforenglandandwalesusingdataoncrimesrecordedbythepolice/2016-11-29>

ONS (2017) 'Health State Life Expectancies: 2014 to 2016', Office for National Statistics. Available at

<https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandlifeexpectancies/bulletins/healthstatelifeexpectanciesuk/2014to2016>

ONS (2018) 'Estimates of the population for the UK, England and Wales, Scotland and Northern Ireland', Available at

<https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland>

NHS Digital (2018) 'NHS Outcomes Framework Indicators - August 2018 Release' Retrieved 26/10/2018, available at: <https://digital.nhs.uk/data-and-information/publications/clinical-indicators/nhs-outcomes-framework/current>

Ryan, M. et al (2014) 'Valuing patients' experiences of healthcare processes: Towards broader applications of existing methods', *Social Science & Medicine*, Volume 106, April 2014, Pages 194-203

Ryen, L. and M. Svensson (2014) 'The Willingness to Pay for a Quality Adjusted Life Year: A Review of the Empirical Literature', *Health Economics*, Volume 24, Issue 10, October 2015, Pages 1289-1301

Sanders, S. (2017) 'National life tables, UK: 2014 to 2016', Office for National Statistics. Available at

<https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/bulletins/nationallifetablesunitedkingdom/2014to2016>

Sen, A. (1985). *Commodities and capabilities*. Amsterdam: North-Holland.

van Loon, M. S. et al (2018) 'Quality of life in a broader perspective: Does ASCOT reflect the capability approach?', *Quality of Life Research*, 2018; 27(5): 1181–1189

Wolf, A. (2011) 'Review of Vocational Education' Available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/180504/DFE-00031-2011.pdf

Yang, W., Forder, J., and O. Nizalova (2017) 'Measuring the productivity of residential long-term care in England: methods for quality adjustment and regional comparison', *Eur J Health Econ* (2017) 18: 635.

Annex A: The Atkinson Principles

As drawn from pp55-56 of Atkinson (2005).

Principle A: the measurement of government non-market output should, as far as possible, follow a procedure parallel to that adopted in 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.

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.

Annex B: Estimating public service quantity output

The process is carried out in several steps:

1. Time series data are compiled examining (a) the number of differentiated activities and (b) the level of expenditure in each individual sector, at the available geographic granularity.
2. A chain-linked Laspeyres volume index of output is produced for each educational sector such that:

$$\psi_t = \psi_{t-1} \left(\sum_i \left(\frac{((a_{i,j,k,t}) - (a_{i,j,k,t-1}))}{a_{i,j,k,t-1}} * \frac{x_{i,j,k,t-1}}{\sum_j x_{i,j,k,t-1}} \right) + 1 \right)$$

Where:

i, j, k and t index individual sectors, differentiated activities, geographical area and time respectively

ψ_t is a chain-linked Laspeyres index of quantity output

a_i is the number of activities

x_i is the level of expenditure in current price terms

Output in the initial period (t=0) is set equal to 100

3. A UK-level, chain-linked Laspeyres volume index of output is calculated using the individual sector indices and the relative cost weights, such that:

$$\Psi_t = \Psi_{t-1} \left(\sum_i \left(\frac{(\psi_{i,t} - \psi_{i,t-1})}{\psi_{i,t-1}} * \frac{x_{i,t-1}}{\sum_i x_{i,t-1}} \right) + 1 \right)$$

Where:

i and t index individual sectors and time respectively

Ψ_t is a chain-linked, aggregate UK, Laspeyres index of quantity output

ψ_t is a chain-linked Laspeyres index of individual sector quantity output

x_i is the level of expenditure in current price terms

Output in the initial period (t=0) is set equal to 100

The result of this process is a chain-linked, UK-level, Laspeyres index of quantity output for the respective service area. There are several equivalent methods of generating this result. In particular, this approach is equivalent to first calculating the indices for geographical areas and then aggregating over educational sectors.

Annex C: Estimating public service quality adjusted output

The process is carried out in several steps:

4. The quality adjustment measures are converted into indices such that:

$$q_{i,j,k,z,t} = q_{i,t-1} \left(\frac{\beta_{i,j,k,z,t} - \beta_{i,j,k,z,t-1}}{\beta_{i,j,k,z,t-1}} \right)$$

Where:

i, j, k, z and t index individual sectors, differentiated activities, geographical area, quality measure and time respectively

β_i is respective quality metric

q_i is the level of quality achieved in delivery

$q_{i,t=0}$ equals 1

5. A chain-linked Laspeyres volume index of quality-adjusted output is produced for each individual sector such that:

$$l_t^Q = l_{t-1}^Q \left(\sum_i \left(\frac{((a_{i,j,t}q_{i,t}) - (a_{i,j,t-1}q_{i,t-1}))}{a_{i,j,t-1}q_{i,t-1}} * \frac{x_{i,j,t-1}}{\sum_j x_{i,j,t-1}} \right) + 1 \right)$$

Where:

i, j and t index educational sectors, geographical area and time respectively

l_t^Q is a chain-linked Laspeyres index of quality adjusted output

a_i is the number of activities

q_i is the level of quality achieved in delivery

x_i is the level of expenditure in current price terms

Output in the initial period (t=0) is set equal to 100

For sectors which are not explicitly quality adjusted, $q_{i,t} = q_{i,t-1} = q_{i,t=0} = 1$

6. As before, a UK-level, chain-linked Laspeyres volume index of quality-adjusted output is calculated using the individual sector indices and the relative cost weights, such that:

$$L_t^Q = L_{t-1}^Q \left(\sum_i \left(\frac{(l_{i,t}^Q - l_{i,t-1}^Q)}{l_{i,t-1}^Q} * \frac{x_{i,t-1}}{\sum_i x_{i,t-1}} \right) + 1 \right)$$

Where:

i and t index educational sectors and time respectively

L_t^Q is a chain-linked, aggregate UK, Laspeyres index of quality adjusted output

l_t^Q is a chain-linked Laspeyres index of quality adjusted output for each individual sector