New Measures of Public Service Productivity: Lessons and Results from the United Kingdom

Richard Heys[1]
Office for National Statistics

Abstract

Since the early 2000s, governments, including that of the United Kingdom, have increasingly focused on measuring and improving public sector productivity. In response, the Office for National Statistics (ONS) has been tasked with developing and refining statistical methods to reflect ongoing reforms in public service delivery. In 2025, a new review concluded, which addressed the evolving landscape of public services, with particular reference to the impact of the Covid-19 pandemic. Building on the Atkinson principles, the review introduced innovative methodologies which will be of international interest. These methods provide stronger evidence that public services can achieve both productivity gains and losses, depending on capital investment and funding stability. Applying the latest methods developed under the review suggests that UK GDP growth could have been 0.1 percentage points higher annually since 1997, driven by higher public sector output growth of around 0.5 percentage points per annum. This challenges the long-standing assumption, rooted in Baumol's Cost Disease theory, that public services are inherently non-progressive. The Review's findings are particularly timely given the 2025 revision of the System of National Accounts (SNA), which allows for quality adjustments in measuring public service output. This article highlights the importance of adopting these improved methodologies internationally, as part of the upcoming SNA implementation cycle, to better capture the true value and performance of public services.

Since at least the early 2000s, governments around the world, and particularly in the United Kingdom have been focused on the productivity performance of the public sector. In the United Kingdom, around 20 per cent of Gross Domestic

1 Richard Heys is Acting Chief Economist, Macroeconomic and Environmental Statistics and Analysis Directorate at the Office for National Statistics (ONS). This article is the work of the author and does not necessarily reflect the opinions of the ONS. The author would like to thank all those who worked on the review team led by Sir Ian Diamond, Debra Prestwood and Katherine Kent alongside the author. The author would also like to thank all those who contributed to the review as well as the IPM editorial team and anonymous reviewers. All errors remain the author's own. Email: richard.heys@ons.gov.uk

Product (GDP) is accounted for by the outputs of public services, comparable to most other western economies. How best to measure this substantial part of the economy, particularly in real terms, has been a long-term question for national accountants. The government's desire to find innovative ways to improve UK public services without increasing spending or taxes as well as the recent experience of Covid-19, have brought this back into focus.

In 2023, the then Chancellor of the Exchequer commissioned the National Statistician to undertake a general refresh of methods taking into account the impact of the Covid-19 pandemic on the design and delivery of the public services. In March 2025, the UK Statistics Authority published the outcomes of its work in the Independent Review of the Measurement of Public Services Productivity (UK Statistics Authority, 2025), also referred to as the *Public Sector Productivity Review* (hereafter call the Review).

This article will begin by reviewing the existing literature and key debates surrounding public service productivity. It will then outline the core Atkinson methodology, which underpins much of the current analytical framework. The discussion will move on to the key outcomes of the ONS Review, including outstanding methodological questions. It will then present the main results, and conclude with a summary of findings and implications for future research and policy.

The Existing Literature and Key Debates

Productivity is a long-standing topic of interest to economists and policy makers as productivity is ultimately a driving determinant of standards of living and economic progress. When reflecting on service provision, the classic consideration which has underpinned much academic debate is Baumol's Cost Disease (Baumol, 1967). This remains a seminal theorem in economics: it argues that the real costs of labour-intensive industries (predominantly services) tend to rise faster than those in industries driven by innovation and technological advancements. This is because labour costs in these industries must keep pace with other sectors, without the equivalent increases in productivity as observed in predominantly manufacturing industries. This story is inherently pessimistic in terms of its underlying expectations for productivity gains in the public services, and other similar services; it essentially argues that productivity growth in these sectors will always lag that of other sectors of the economy.

Other studies have started from the more optimistic perspective that there is productivity growth possible in these sectors and observed weakness are, at least partially, driven by measurement techniques, particularly when the sum of real costs is used as a proxy for the volume of output. Under this model, widely used by almost all countries, and frequently still used in national accounts estimation around the world, the common movement in output and inputs necessarily deliver productivity growth of zero, by assumption.

Many governments and academics have found this approach deeply unsatisfactory, either as a reflection of reality or as a metric, which can be sold to the general public.

The defining study which resulted from this need for better measures was the in-depth investigation from 2003 to 2005 by Sir Tony Atkinson in his independent review of the measurement of government output in the National Accounts, (Atkinson, 2005). He makes the case for direct measurement of output, augmented by adjustment of the volume to reflect changes in quality. This seminal text informed the development of the System of National Accounts (SNA) 2008 in how to conceptualize and then empirically measure the outputs of the public services contained in GDP and established a set of key principles which informed future work.

However, the European System of Accounts (ESA) (2010) which generally follows the SNA as its guiding principles took a flatly opposite view and banned quality adjustment within its 2010 iteration on the basis that more methodological work was required to deliver consistent methods which could be applied across Europe.

The exclusion of these elements from ESA 2010, primarily for reasons of practical consistency, was often misinterpreted, leading to the mistaken belief that the Atkinson Review diverged from core national accounts measurement principles, such as those outlined in the SNA or ESA. In fact, the opposite is true: the Atkinson Review applied core valuation methodologies used in the market sector and considered their application to the non-market sector, specifically public services.

The United Kingdom, along with countries such as Canada and the United States (see below), moved forward with further investigating or implementing these methods, particularly in the areas of health and education, despite the position taken in ESA 2010. Operating under the banner of 'public service productivity', the ONS developed a dataset parallel to the national accounts. Whilst this dataset drew on national accounts data, it also incorporated quality adjustments. As a result, this extended dataset, delivered outside the national accounts, offered valuable insights for policymakers, especially in cases where public service reforms affected the quality rather than the quantity of outputs. This work in the years immediately post-Atkinson's report managed to address the largest parts of the public services, but gaps remained. Dawson *et al.* (2005) further developed these methods for application.

The end-result was a difficult statistic to translate into policy application, being (using cost weights, which can change year-by-year dependent on public spending plans) just under 40 per cent of the public services being quality-adjusted and just over 60 per cent not quality adjusted. By 2016, the Bean Review (Bean, 2016) concluded that the evolution of government data sources and the depth of experience in running the health and education methods gained over the previous decade meant that a new investment in this area could be justified. This led to new quality adjustments being delivered in three policy areas:

- The Criminal Justice System (excluding Policing) was introduced in 2018.
- Adult Social Care was introduced in 2018.
- Children's Social Care was introduced in 2019.

Together these three form around 10 per

cent by weight of all public services and moved the overall balance to just under 50 per cent quality adjusted versus just over 50 per cent not quality adjusted, with a substantive share of the latter representing collective services which are generally perceived as more difficult to measure.

Work on this topic was not restricted to the United Kingdom. There is equally a parallel stream of research in the United States,starting with Fisk *et al.*, (1997), which takes account of the different models of provision utilized in that country. Hall (2017) reviews the literature on adjusting medical sector output for quality in the United States and Cutler *et al.* (2022) provides a US health account. Gu and Wong (2015) equally undertake similar work on Canada's education sector.

Of most significance to the development of this agenda are Schreyer (2010) which broadly mirrored the Atkinson approach in health and education, considering these across OECD countries, and Schreyer (2012) which proposed alternative models, with less reliance on quality adjustment. Following these developments, Diewert (2018) analysed the development of imputed output prices, effectively solving the challenge through quality adjusting prices rather than directly adjusting quantities, as per Atkinson. Meanwhile Foxton *et al.* (2019) and Martin and Riley (2024) undertook reviews of the key lessons learnt through the application of the Atkinson model. Martin and Riley (2024) provides a general review of recent work, particularly in relation to the health sector, including Bojke *et al.* (2018) who suggests current practice could be strengthened by adding further characteristics of the quality of healthcare, including additional National Health Service (NHS) Outcomes Framework indicators. Davies (2020) considers a range of indicators that might be used to draw international comparisons.

Foxton *et al.* (2019) in particular highlighted a set of outstanding questions which merited further investigation around how different aspects of quality, as well as different quality adjusted services, should be weighed against one another in estimating public service output and hence productivity measures. These can be summarized as seven substantive issues:

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

While Atkinson's approach remains the main guide for practical applications, several key imperatives drive the need for further work. These resulted in the 2023-2025 Review led by Sir Ian Diamond as the then-UK National Statistician (UK Statistics Authority, 2025). These included methodological questions, such as those raised above, the ongoing changes observed in public services, the potential for Artificial Intelligence and other innovations to transform public service delivery, as well

as Atkinson's call for periodic reviews in his report. All of these, however, were dwarfed by the impact of Covid-19 from 2020 onwards, and the necessary policy responses which transformed public service delivery.

Whilst this article will summarize in the main the latest methodological advances delivered by the Diamond Review, a key question remains outstanding. How do the best efforts of statistical offices to apply Atkinson prove or disprove Baumol's contention that productivity growth in such sectors is inherently lower than observed in other sectors of the economy? The concluding section returns to this question in the light of the latest UK estimates to evaluate whether we should be inherently pessimistic or optimistic around the potential for productivity growth in the public services.

The Core Atkinson Methodology

The essence of Atkinson's argument was that in a competitive market, the value society places on a good, service, input or asset is reflected by the market price. This market price should reflect and equilibrate supply and demand. On the supply side, it should reflect the costs of production, including an appropriate margin. On the demand side, it should represent the discounted sum of benefits the consumers believes they will accrue from the product at the time of purchase. The essential logic is that, under the presence of meaningful competitive forces, if there is a rival producer who can deliver the product at a lower cost, or by accepting a lower margin, this will mean they can bid the price lower. Equally, and again in a competitive model, a consumer who believes they will secure more value from the product will be willing to bid up the price. Where the price balances it must be the case that:

Sum of Costs plus margin = Market Price = Discounted sum of Expected Benefits

As such, under conditions where a market (or exchange equivalent) price cannot be observed directly, such as in public services which are often provided free-at-the-point of consumption, this offers two proxy methods which can be used to understand the value of the final output of such public services:

- A sum of costs methodology (referred to as 'inputs = outputs') where the value of the output is set equal to the value of the inputs which go into its production, and given the government does not make profits, no margins applied , or

- A methodology which looks to proxy the discounted sum of expected benefits by splitting the analysis into two computable components: a direct measure of the volume of the relevant output (e.g. the number of hospital operations) and a direct measure of the quality of the final output (e.g. an operation that results in greater average improvements in patients' quality of life is valued more highly than one with lesser improvements.).

In both cases, comparing the aggregate volume growth of all inputs with the volume growth of outputs would deliver an estimate consistent with the concept of Gross Value Added (GVA) growth as applied in the private sector. Dividing the

volume of output by the volume of inputs would present a productivity measure. For the 'inputs = outputs' approach to measuring output, this always gives an implicit result that productivity is assumed to be constant. Hence this method is considered inferior as without being able to apply an appropriate proxy for the margin discussed above, this method is a weak proxy for the exchange equivalent price.

The second method requires a direct measure of quality improvements and rests on two assumptions, which are worth making explicit:

- The quality adjustments act in a way synonymous with a market price, so the better the quality of the product, the higher the quality adjustment factor, in the same way that this would normally be reflected in a higher market price.

- The Government is able to act as a rational 'social planner', making optimum decisions relating to the quantities of each public service to deliver, such that costs are spent up until the quantity where consumers would no longer desire additional units of output.

The first of these assumptions have two corollaries:

- As a quality improvement is implicitly equivalent to an output improvement in volume terms a 1 per cent increase in quality is equivalent to a 1 per cent increase in output.
- This is clearly equivalent to market transactions – if a factory makes one good brick in one time period and then makes in the next time period two broken bricks which cannot be sold for a positive price and can only be given away (price = £0), simply because there are two bricks quantity has not doubled, instead it has fallen one hundred percent.

It is this recognition, that one needs to adjust measures of output to deliver a closer analogy of the methods used in the market sector, which marks Atkinson's Review as the landmark which it is. Regarding the social planner assumption above, this is clearly impacted by budget constraints: whilst with an unlimited budget the social planner may be able to achieve the desired outcome, it may well be that within a fixed budget it is not feasible to take all optimal decisions and achieve:

Sum of Costs = Discount sum of Expected Benefits

As such, assuming the budget available to government is insufficient to deliver this equality, one can still assume the government would choose those interventions where the downstream benefits on average exceed the sum of costs. As such, the sum of costs would inherently be expected to under-estimate the value produced. One would expect governments to draw from the top of the distribution of projects in terms of benefit-to-cost ratios. Indeed, UK fiscal policy, governed by the HMT Green Book (HM Treasury, 2022) is explicit this is the case. For this reason, capturing a direct measure of output and applying a quality adjustment is clearly essential to accurately reflect the value created by the public sector in a form comparable to those observed within the market.

These principles of applying quality ad-

justment to estimate output were accepted as part of a wider trend of economists becoming increasingly comfortable in addressing social welfare function issues.

Questions Raised by the Public Sector Productivity Review

The Public Sector Productivity Review was, to a significant extent, motivated by the Covid 19 pandemic and its aftermath. The pandemic highlighted two key challenges. First, measures of public services productivity can be subject to significant changes in times of crisis which traditional measurement systems often struggle to accommodate. Second, differences in national accounting methodologies across countries can undermine international comparability. Joint research by the ONS and OECD (Mitchell *et al.*, 2022) demonstrated that, in addition to genuine differences in the timing and impact of the pandemic and in policy responses, part of the variation in reported GDP figures stemmed from inconsistent methods used to measure public services. These methodological discrepancies contributed to the observed cross-country differences in economic performance.

In addition to COVID-19, there were other issues which required renewed attention from the perspective of measurement, reinforcing the Atkinson principle that methods require routine updating to continue to meet need. The passage of time delivered two distinct challenges:

- **Changes experienced within public services areas** – in some areas a service may have been re-designed in a fundamental fashion such that the measures no longer reflect the landscape. For example:

o During the Covid pandemic, the Health sector created new Test and Trace capabilities which was outside the existing measurement framework, requiring the creation of new metrics.

o In 2018, a significant change in UK welfare payment policies was made as the introduction of Universal Credit replaced a number of benefits which formed the core of the ONS measurement model until that time. As reacting to Covid-19 was prioritized the ONS reverted to 'inputs = outputs' for the measurement of productivity in social security administration. A new method was therefore a priority for the Review as with other areas affected by changes, which impacted measures including Education.

- **Changes within the measurement of the service areas** – in some areas the service may have remained consistent through time, but the measurement system may have deteriorated. This may have been for a number of reasons:

o Data sources may have ceased to be published by various agencies. For example, during the pandemic, various educational examinations were replaced by teacher-grading as students were unable to attend school and moved to home-teaching. Similarly measures of re-offending which are captured through re-conviction data had to be paused whilst courts were closed during the pandemic, and hence re-conviction patterns exhibited unusual trends.

o Data quality may have deteriorated due to falling sample sizes or other statistical reasons.

o In some areas data is forecast to cover more recent time periods, but the model may need updating and bringing up-to-date. Below we first discuss the changes related to Covid-19 and then other changes which formed part of the Review.

The Direct Impact of Covid-19

The Covid-19 pandemic delivered a fundamental challenge to almost all area of public services, which broadly fell into four categories:

- Doing the same activities in a new way, in ways one could measure.
- Doing the same activities in a new way, in ways one could no longer measure.
- Doing new activities:
- Changing the relative weights between different activities

Both health and education services, alongside numerous others saw the operating model for their services fundamentally transformed by Covid-19. In health services, even setting aside Test and Trace, there were dramatic movements from high-cost in-patient routine operations to medium-cost critical care. Whilst this new task was more labour-intensive and staff in the NHS worked harder than ever before, in cost-weighted volume terms, activity fell during this period, due to the lower value cost-weighting attributed to this provision.

In education, where output is measured via exit qualifications (e.g. GCSEs in England), it is not just the current year's teaching and learning which shape this year's results: the previous ten years of formal education, and pre-primary early years' provision, need to be taken into account. As such the existing methodology pro-rated qualification results back through the cohort's education, using a method called cubic splining. So, for example, a significant fraction of a student's success in Year 11 has been attributed to previous years.

Covid-19 fundamentally disrupted this pattern and caused attainment to behave in fundamentally different ways. However, it would be inappropriate to model that a student sitting their exams in 2021, and whose results suffered due to disruption in 2020, should see their 2017 attribution downgraded: pre-Covid, the student would likely have performed as well as the preceding cohort in that year.

Preventative Services and Latent Capability

At the heart of the conceptual challenges raised by Covid-19 was the role of preventative services. Following on the discussion above which demonstrated the need to quality-adjust output data to reflect the true value created by a service, the impact of prevention had already been recognized as one of the most challenging issues as such services are generally designed to cost significantly less than the downstream benefits they may unlock. Consider a low-cost tobacco cessation program designed to reduce future demand for costly cancer treatments. In such a case, a cost-based ap-

proach, even if quality-adjusted, may undervalue the program's long-term benefits. This is especially true when using a cost-weighted activity index to aggregate services, as high-cost treatments like cancer operations would disproportionately influence the overall valuation. Weale (2024) describes exploratory work in this area on diabetes prevention.

In a similar fashion, excess capacity which is laid down to help a system cope with periods of peak/surge demand, would generally appear to depress productivity in the years where inputs spent but not used to produce output, even if these investments may be essential in allowing the system to work at peak times. How to account for this latent capacity was another key issue for consideration.

Opportunities Presented by New Data

Moving on to issues beyond Covid-19, there was an opportunity for the Review to benefit from significantly more data across government than in 2005. This allowed consideration of situations where services can now be reliably measured and where data can be disaggregated to better match inputs and outputs at the detailed level. The importance of disaggregation is particularly important when estimating volumes because this process also allows more detailed deflators to be used.

The Challenge of Collective Services

Atkinson, as other authors, divided services into those which were 'individual' and those which were 'collective'. That is those where the service would affect one individual – such as an operation on person x means the same operating theatre and medical staff cannot simultaneously be used for person y – and those which affect us all – for example, no-one can 'opt-out' of the UK-wide nuclear deterrence. How to value this deterrence, and how it changes (would citizens today feel better defended if the UK government had purchased an additional nuclear submarine earlier?) are significant questions, which have not been resolved globally (Smith, 2024).

This challenge remains as fundamentally difficult today as faced by Atkinson and those who have worked on this topic in the interim. Importantly, the Review did not draw a distinction between individual and collective services, a distinction with a long heritage in national accounts, finding this to be an increasingly unhelpful and outdated concept in a period of increasingly personalized services and better data allowing the link between individuals and services to be better understood.

The Challenge of Services with Multiple Outcomes

Measuring the output and the outcomes delivered by a service can be complex even when there is a simple one-to-one relationships (e.g. health services), but some services are characterized by delivering multiple outcomes. Policing is a clear example, with responsibilities to both prevent and solve crime, deliver crowd-control, undertake missing persons investigations, work to reduce re-offending with key partners, attend road-traffic accidents, undertake community policing, and tackle anti-social

behaviour, alongside counter-terrorism activity and addressing organised crime. This raises a number of distinct challenges:

- Mapping inputs to each activity.
- Accessing good quality and consistent activity data, with no double-counting.
- Calculating the relative weights of these activities in the aggregation process based on accurate and timely data.
- Attributing outputs and outcomes to the participating bodies. For example, if police work with local probation staff to manage dangerous offenders upon release, how should this activity be split between police and probation agencies?

This complexity made policing an area which Atkinson was unable to resolve, and so at the start of the Review it was still the second largest individual service (after defence) to be treated as 'inputs = outputs'. Resolving this was therefore a priority. Similarly, in social security administration and taxation the question of how to weight different taxes and benefits directly relate to this. The Review implemented alternative weighting methodologies which may better reflect the value users receive (for example, do citizens place more weight on a benefit which delivers a larger share of benefits disbursed or which costs more to administrate?)

The Outcome of the ONS Review

The Review, developed over two years contains 120 recommendations, divided into all areas of the public services. Healthcare, social security administration, criminal justice and fire, and policing form the bulk of these recommendations by number, even though several of those may be referred to as future developmental work. In general, the Review has refreshed input and output data sources, but in addition the key issues by area considered were:

- Environmental services and local services – Scope and definition of service areas
- Tax administration – What is the output?
- Public order and safety – Data and implementation
- Social security administration – Fundamental change of service design (Universal Credit)
- Healthcare services – Preventative services and equivalization
- Education services – The impact of Covid
- Defence services – Conceptual challenges

Finally, the Review explored accelerating the pace of statistical production to enable more timely measures using nowcasting techniques. A longstanding limitation of using quality adjustments for outcome measures is the publication delay as many of these measures can take a long time to be produced. For example, the rate of re-offending, which is used as a quality adjustment for the criminal justice services, is based on re-convictions by a court or other legal process. As it can take up to two years before a case is labelled as a re-conviction, the statistics becomes less useful for policy

2 More information on this aspect of the Review can be found in ONS (2023) and ONS (2024d).

purposes.[2]

Environmental and Local Services

Before one can begin to produce estimates for different activities, one must define their scope to ensure completeness and prevent duplication. However, such definitions also need to reflect public and technical understanding of the scope of different activities, and to be internationally agreed, so comparisons can be undertaken.

The definitions used in the United Kingdom align to the UN Classification of Functions of Government (COFOG), which were last updated in the 2010s (The last revision was 2019). It is recognized internationally that these need to be refreshed. For example, while decommissioning nuclear reactors—currently a major component—is undoubtedly important as an environmental protection service, there is significant scope to broaden the definition to better reflect the full range of government activities relevant to the environment.

In this instance, the Review identified the challenge that around 50 per cent of those activities (by cost-weight) which today an informed citizen may expect to be considered as being concerned with the environment are classified under different sections of the classification system. For example, forestry is listed under 'economic affairs' because its primary function was previously perceived to be the production of timber, an economic asset. Today carbon sequestration and cultural services from forestry are, at least in the United Kingdom, given greater weight, so one could argue it should be moved to environmental services.

Redefining environmental and local services would have two benefits. First, it would better show what share of public services are actually targeted at protecting the environment; conceivably 5 per cent of public services on a broad reading could be classified this way. Second, it would compel us to address whether various locally managed services should be within this envelope, for example local planning functions or waste collection and management.

As such the Review submitted a recommendation to the current global consultation that a wider Environmental Services section should be created, split into three parts: Environmental Protection (broadly equivalent to the current 'Environmental Protection' COFOG category), Natural Resource Management to cover areas such as forestry, planning and waste management, and Climate Change and Net Zero to cover adaptation and other similar activities.

Tax Administration

Tax administration is a service area which traditionally had not benefited from a direct measure of output. As such, this service was calculated on an 'inputs = outputs' basis. The first course of action therefore is to ask: what is the output that is being delivered, accompanied by a subtler question of what output do citizens benefit from? In areas like education where the citizen is directly receiving the service this is relatively simple to map. Similarly, if the health service delivers more health interventions this is broadly understood as additional output which is of value (one imagines no one will agree to an unnecessary

surgical intervention – so there is a natural limit on volume of activity).

In an area like tax this is less clear. If the tax collecting agency collects more tax than is mandated by law, or collects it from the wrong people, this is clearly not a socially positive output: citizens would not place a positive value on this output. Akin to the 'broken bricks' argument, there would not be a positive price amongst citizens for over-taxation.

One could therefore simply count the number of taxpayers validly caught under each tax regime (Income Tax, National Insurance, etc.), and cost-weight these together. However, this again misses the point of where the value is created. The value is created for citizens by the tax agency (HM Revenue and Customs or HMRC) collecting the quantity of tax specified in law so that the government can spend funds on the delivery of public services. Collecting too little tax or collecting more than legally specified are both of less value to citizens.

This means that not all tax schemes are of equal value. Some taxes are relatively expensive to administer and raise smaller levels of tax, while others are relatively cheap to administer and raise large quantities of tax (e.g. PAYE Income Tax). Just eleven specific taxes schemes in the United Kingdom have harvested around 88-89 per cent of all tax revenues in recent years It feels appropriate to assume the public and ministers are more concerned about the efficiency and productivity of those schemes rather than the smaller ones.

The Review resolved this by 'revenue-adjusting' the various tax schemes, so their value in aggregation better reflects the tax-revenue collected rather than the costs of delivery. This adjustment reflects an intermediate step whilst the ONS and HMRC explore methods to adjust for fraud and error as a quality adjustment. The ONS is also exploring how to take account of taxes raised outside HMRC, primarily via local government, and customs and excise duties raised by HMRC.

Public Order and Safety

Another problem caused by the outdated COFOG structure is the grouping together of services which are now perceived to perform a diverse range of functions. Public order and safety is a prime example, being an amalgam of policing, immigration, fire services, civil and criminal courts, probation, prisons and other criminal justice activities. This covers a mixture of civil (some immigration activities, civil and county courts, some police activity, fire) alongside criminal detection, prevention and punishment services. There is therefore a variety of inputs and funding models. For example, some services use fee regimes, such as some courts and tribunals, whereas others are funded from taxes.

For this reason in 2018, the ONS split criminal justice services and fire services from policing and immigration. These services all benefit from input and output metrics alongside outcome measures. Whilst most of these have been subject to extensive revision and updating under the Review, primarily focused on ensuring up to date data sources are used, the more extensive changes relate to policing and immigration.

Historically, policing and immigration

services relied on an 'inputs = outputs' approach. However, these two areas differ significantly in terms of input growth, particularly in recent years as immigration has gained greater political salience. They also differ on data availability to develop comprehensive direct output measures, although both appear now to be feasible.

A key issue that remains is determining the appropriate weighting of different outputs within individual services, particularly policing where the Review has identified a number of discrete data sources across different types of activity. While a successful investigation leading to a conviction is clearly a positive output, more ambiguous cases raise important questions. For instance, if an investigation identifies a suspect but fails to proceed due to an unwilling witness, should this outcome be considered equivalent to a conviction? Or should it be down-weighted, perhaps through a quality adjustment on the grounds that, while it may not yield immediate results, it could still contribute positively in the future by generating intelligence or evidence that supports later cases? In addition, sourcing appropriate data to weight together criminal and 'non-criminal' outputs is difficult when inputs are not always clearly delineated between the two. Overall, the ONS identified sufficient data to progress developing metrics which are likely to be delivered in 2026 and 2027.

Social Security Administration

Social Security Administration historically benefited from a direct measure of output, but only for benefits administered by the UK Department for Works and Pensions (DWP). Tax credits and child benefit are both administered by HMRC, which is the tax and customs authority, and were omitted alongside housing benefits administered by local government.

Immediately pre-Covid, the implementation of a new benefit system, Universal Credit (UC), replaced seven benefits with a single consolidated benefit. Since, as with Tax Administration, the activity measure used was the number of case-files, this change had three implications. First, it consolidated seven case files into one case file, which would show up as an 85 per cent drop in output even though this likely came with significant input savings. Secondly, as implementation was phased, with easier cases being ported into UC first, the remaining activity within the legacy benefits appeared to increase in average costs, as only the more complex cases remained, whilst UC appeared artificially cheap on the same basis. Finally, Tax Credits were one of the schemes being replaced but were not included in the existing output measure. Their inclusion would appear to replace no output with a positive output biasing productivity upwards.

To prevent these impacts, ONS reverted to 'inputs = outputs' in 2018 with the aim of rapidly developing a new model to cope with these effects. The new method, which has now been introduced, adjusts UC outputs for complexity and for the number of component benefits received. This gives a cleaner output metric, which could be combined with the legacy benefits to better reflect overall output.

However, this measurement change left a final challenge. Consider a simplified exam-

ple: two separate benefits, each providing a citizen with £50 in year 1, are replaced by a single benefit of £100 in year 2. All three benefits (the original two benefits in year 1 and the combined benefit in year 2) cost £1 each to administer. From the citizen's perspective, the value received remains unchanged—£100 in total. Under the traditional model, which weights outputs by the number of case files and their associated costs, productivity also remains unchanged. In year 1, two case files (each with a cost-weight of £1) yield an output of £2, divided by £2 of input, resulting in a productivity score of £1. In year 2, one case file (cost-weighted at £1) divided by £1 of input also gives a productivity score of £1, despite the fact that the same outcome was achieved with half the administrative effort.[3]

The Review found this approach increasingly difficult to defend, because cost-weights do not necessarily have a strong alignment with the concept of value.[4] To correct for this distortion, a benefits-weighting approach was used. The single case file in year 2 is assigned double the weight of each case file in year 1, reflecting the consolidation of two benefits into one. This adjustment results in an output of £2 divided by £1 of input, yielding a productivity score of £2. While this is a highly stylized example, it illustrates the core principle which builds on the foundational work of Atkinson (2005), which flagged that cost-weighting, whilst it had the attractive qualities of being readily available and in consistent market prices, crucially can deviate from value significantly enough to be the worst of all viable alternatives. There is a strong argument that cost weights, whilst probably unavoidable, are the weakest component of the core methodology and effort should be taken to find alternatives.

Healthcare

Healthcare is a well-established sector, which has benefited from substantial efforts to improve measurement of inputs and outputs since 2005. As part of the Review numerous smaller remaining issues were addressed. However, two issues were noted as key at the start of the process, with one further issue arising as the work developed.

Preventative Activities

Better reflecting preventative services involved firstly ensuring those services are captured as distinct units of output, and then exploiting new methods to measure their impact on outcomes. Preventative activity of NHS providers is already accounted for elsewhere in healthcare output,

3 Schreyer (2010:p 11) makes a particular effort to defend cost-weighted activity indices: "For non-market producers, unit costs can replace prices to value different kinds of services. However, unlike market prices that combine consumer and producer valuations of products, unit cost weights reflect in the first instance the producer or supply side (or government's willingness to pay). This implies that it is the production value and not necessarily the societal value that is attributed to education or health care. However, the purpose of output measurement is not to provide estimates of the societal value, so the use of cost weights does not constitute a major drawback in the context of the national accounts."

4 Indeed, if it did, there would be no need to move beyond 'inputs = outputs' methods as one can read 'costs' for inputs.

taking into account the following considerations:

- Pharmacological treatments were already captured meaning only behavioural support were added
- Drug and alcohol misuse output were estimated using the total number of psychosocial interventions; although this did not distinguish between the intervention setting, as there are not enough data to disaggregate unit costs by setting.
- For smoking cessation, the number of quit attempts was used as the activity measure.

In 2023, ONS reviewed the coverage of preventive healthcare leading to the introduction of new activity measures to capture the growth in the volume of activities provided by local authorities:

- Local authority commissioned treatments for drug or alcohol misuse excluding NHS providers
- Local authority commissioned smoking cessation services

Equivalization of Services Delivered in Different Providers

The granular data available for healthcare activity and costs enables a high degree of differentiation in weights between different services. However, where process improvements lead to lower-cost service delivery methods, which are recorded as separate activity types, they are assigned a lower weight in output, meaning efficiency gains from moving to lower cost treatment are not represented in the productivity measure. This is particularly notable in the case of elective surgery, where procedures may be carried out either as an inpatient procedure, a day case procedure or as an outpatient procedure. Historically, separate unit costs have been used for each procedure type. Therefore, if procedures transition from overnight hospital stays to same-day treatments and costs fall, this results in the measurement system in more lower-weighted activity and so appears as a reduction in output, even though in reality the same care is being provided more efficiently

The ONS has developed equivalized unit costs for equivalent treatments across different modes of provision. These are applied by combining activity and expenditure across different services categories within each Healthcare Resource Group (HRG). HRGs are clinically meaningful groupings of patient activity derived from NHS patient records, primarily using procedure and diagnosis codes. They provide a means of determining fair and equitable reimbursement for healthcare services by providing consistent 'units of currency', based on expected resource use.

This approach generates a new unit cost, calculated as a weighted average of the previously separate unit costs, reflecting both higher and lower-cost modes of care. For inpatient and day case procedures, there is no restriction on inclusion in the equivalization. If a HRG exists in more than one of those components, an equivalized weight will be applied. For outpatient procedures, unit costs are only equivalized where the HRGs tariff (the price paid by commissioners under the NHS Payment Scheme) is equal to that of elective inpatient and day cases.

Disaggregation of Healthcare as a Domain

Since 1997, healthcare has accounted for an increasing share of public services, reaching just under 40 per cent in the latest year of data (2022). It is delivered by the NHS in each of the four UK nations, alongside the UK Health Security Agency (UKHSA), and various Public Health bodies which often form a collaboration or partnership between the NHS and local government. The NHS itself can be divided into primary (GPs, opticians, dental, etc.), community (services such as district nursing, wheelchair provision, palliative care, alongside some Covid-facing services such as Test and Trace and Vaccination) and secondary (hospitals) services. As each of these healthcare services has expanded, they have individually grown larger than any other public service category reported in the Public Service Productivity statistics. The publication of quarterly healthcare productivity data during the Review led to increased public and policy interest, particularly when compared to NHS England data which only covered the hospital sector. As such, the need to disaggregate overall 'healthcare' into components which better enable policy analysis is a clear priority for future work.

Education

Education, like healthcare, benefited at the start of the Review from a mature set of productivity measurements, taking account of examination results at age 16, as a general proxy for educational attainment, as well as a measure of bullying to proxy for student well-being outcomes. During the Review, alongside the common work of improving datasources, work focused on ensuring COVID-19 did not impact estimates of previous years. Improved output metrics covering primary and further education were also incorporated, alongside GCSEs which reflect secondary achievement. Finally, the bullying measure, which acted as a proxy for a range of well-being issues was replaced with a wider 'well-being' measure drawn from the Understanding Society survey funded by the ESRC.[5]

The most substantive intervention related to the distribution of achievements in GCSEs back through previous years of schooling. This used a cubic splining approach to pro-rate fractions of overall attainment back through the student's career, with the greatest weights put on the most recent years. However, as explained above, this model of 'casting back' had the flaw that a one-off event such as Covid-19 in 2019-20 and 2020-21 acted to reduce schools outputs for earlier years.

To address this, the average attainment of earlier cohorts who had recently sat the same school year was utilized as a proxy. The rationale was that, for example, the attainment of those who sat Year 7 in 2015 was better measured by the outcomes of students who had sat Year 7 immediately prior to this cohort and who completed their final qualifications in the years preceding Covid-19. This was considered more robust than trying to retain the ex-

5 https://www.understandingsociety.ac.uk/.

isting methodology and strip out a common 'Covid' effect'.

The common 'Covid effect' for 2019/20 was calculated as a residual: once the attainment of the previous year's schooling was fixed for the cohort who sat GCSEs in that year, the difference between actual attainment and the fixed contribution from previous years was attributed to the pandemic. This residual effect was applied as a common factor to the attainment of every cohort in school during the pandemic. As these cohorts reach their final examinations, their in-year contributions are calculated using the traditional method, given the impact of the fixed elements during and pre the Covid years. This method will be utilized until all students who were in school during Covid-19 have worked their way through the system and the full 'original' method can be applied without adjustment.

Before this change could be implemented, the Review first had to address the underlying raw qualifications data. The COVID-19 pandemic caused widespread disruptions, including school closures and the cancellation of examinations. In response, teacher-assessed grades (TAGs) were provided for GCSE results in place of typical attainment grades. This approach resulted in grades becoming inflated, making the data unsuitable for use in the quality- adjusted output measure, as this could overestimate both output and productivity.

Initially, ONS used a "learning loss" metric based on Renaissance Learning data (2022), commissioned by the Department for Education, to estimate the impact on reading and maths. However, this metric only covered 2019–2020 and risked double-counting output. Experts recommended the National Reference Test (NRT) as a better alternative. Introduced in 2017, the NRT assesses Year 11 students in English and maths without the pressures of formal exams. Results are benchmarked against 2017 GCSE outcomes to track trends. Though not a final attainment measure, the NRT was adopted by ONS to provide more consistent qualifications data during the pandemic years.

Finally, the Review applied changes to reflect the changing governance models in England's schools, predominantly the shift from local authority control to academy providers to better account for variations in the funding (inputs) received by different providers. By 2022/23, over 40 per cent of primary, 80 per cent of secondary, and nearly 45 per cent of special schools were academies. To better capture this change, the number of categories of institutions increased from five to ten, distinguishing between academy and local authority schools across phases, including alternative provision. Updated expenditure weights now allow for more accurate cost-weighting and assessment of each phase's contribution to productivity. This will support evaluation of whether academization has improved educational efficiency.

Defence

Defence is one of the largest activities, reflecting around 10 per cent of total public services, and is currently measured on the 'inputs = outputs' basis due to the conceptual challenges in deriving a measure of output. These challenge have been

long-standing, and were neither resolved in the Atkinson Review (2005) nor in subsequent research, including Prtak (2019) and RAND Europe, (2021). Whilst the Ministry of Defence has considered the issues of measuring defence output in the intervening years, it has also been unable to definitively deliver an appropriate approach. Defence, therefore, remains the largest single service which is treated in this fashion.

In essence the challenge is that a direct measurement of defence output would not be appropriate because defence has a primary function of deterrence, in which active deployment of the armed forces is to be avoided if possible. For example, while defence capital assets, such as aircraft and ship are designed to perform their function in rare combat situations, they primarily act as a deterrence to prevent conflicts from arising and/or escalating.

However, estimating outcomes such as wars being avoided is highly speculative, as is measuring threats such as terror attacks. Deriving a suitable 'unit' of deterrence, or the capability of the military to deliver against its five priority outcomes, is challenging, as the capabilities of the armed forces vary over time, depending on the type of threat, technological advances, and military strategy.[6] Additionally, there is a considerable degree of confidentiality with regard to the activities of the military for reasons of national security thus limiting data availability.

Finally, there is a pertinent question of ethics and political sensitivity when considering the frameworks for defence output. For example, it would not be appropriate to have defence output fall in the absence of active deployment, nor would it be appropriate to measure the number of adversaries neutralized in operations.

The Review commissioned Smith (2024) to consider the measurement of defence productivity further, who concluded:

> "For defence, there is a problem of comparability because the nature of the activities, capabilities and objectives of defence change over time, and for good reasons, as threats, technology and strategy evolve... When these activities or capabilities are discontinued to reorient to the new context, measured output will have fallen while the military are fully occupied doing different activities. For defence, performance measures include elements such as success in operations, maintaining readiness, and stopping equipment being delivered late, over budget and not meeting technical requirements. These are difficult to convert into indicators that would match national income accounting criteria."

Although progress in measuring output remained limited, the Review made significant improvements in the treatment of inputs by replacing indirect estimates, which are typically based on deflated finan-

6 https://www.gov.uk/government/publications/ministry-of-defence-outcome-delivery-plan/ministry-of-defence-outcome-delivery-plan-2021-to-2022

cial data, with more direct volume measures. Where indirect methods continued to be used, the Review incorporated higher-quality source data and deflators. Additionally, it proposed new methods for developing a direct output measure that distinguishes between active deployment and deterrence as separate components.[7]

Results of the Review

Whilst only 25 of the 120 recommendations of the Review have been implemented at the time of drafting this article (June 2025), comparing the 2025 vintage of data to the 2022 vintage, published prior to the commencement of the Review, growth was slower than originally estimated in the pre-2010 period, and faster in the following period. As shown in Chart 1 the total public services productivity compound annual growth rate (CAGR) between 2010 and 2019 was estimated to be 0.9 per cent, up from 0.8 per cent.

As Chart 1 also shows, while recovery from the Covid-19 pandemic has occurred, productivity levels across all public services together have not recovered to the peak seen in 2019. The key question is: to what extent is the 2019 peak a fair comparator? This depends on how far citizen behaviour and needs have adjusted through Covid-19. For example, the cancellation of cancer and other diagnosis activity during the pandemic has resulted in more instances of complex cases being overrepresented in the post-pandemic period (see, for example, Barclay *et al.* 2024). While this effect may erode over time, the data suggest it continues to drag which may imply that previous achievements should be perceived as an aspirational goal.

Chart 2 also showcases the large and growing importance of healthcare in understanding public services in the United Kingdom. Table 1 presents the change in the relative shares of public sector expenditure by sector between 1997 and 2025.

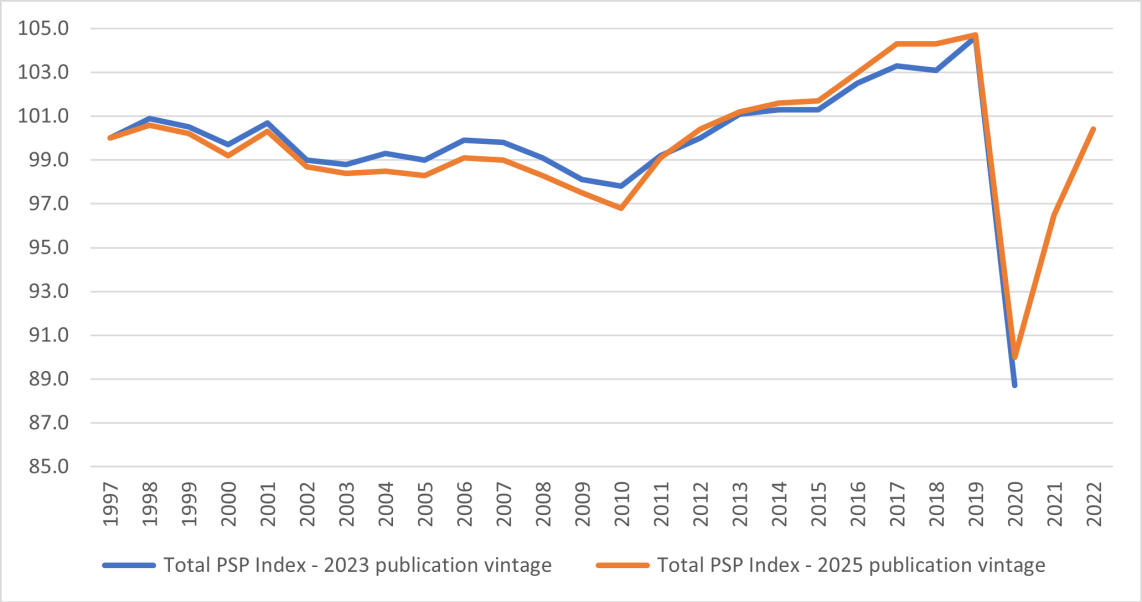
Health spending, as a fraction of all public spending has grown by 11.8 percentage points, whilst education has fallen by 2 percentage points and, despite the aging demographic adult social care has grown by only 0.1 percentage points. Defence has shrunk by 5.3 percentage points, while public order and safety, combined with policing and immigration has fallen 1.5 percentage points. Health's relative strong productivity performance appears correlated with sustained funding and investment, whereas other sectors appear to have faced greater struggles without that investment, as shown in Chart 3. While productivity in education has increased it shows a decline in total expenditure on public services.

One of the key challenges in assessing the share of total public services that are quality-adjusted following the Review, is that this proportion has shifted for two main reasons:

1. Following its rapid growth between 2020 and 2021, the expenditure share for

7 The Review considered this issue further and developed some research methodologies which will be written up in greater depth as a standalone article to follow. These are summarised in Annex E of the Review. https://uksa.statisticsauthority.gov.uk/publication/national-statisticians-independent-review-of-the-measurement-of-public-services-productivity/pages/26/.

Chart 1: Updated Whole Public Service Productivity – 2025 Vintage Compared to 2022 Estimates (1997=100)



Source: Authors calculations based on data contained in ONS (2023) and ONS (2025)

Table 1: Expenditure Weights in Per cent and Percentage-point Change – 1997 and 2025

	Health	Education	Adult Social Care	Public Order & Safety	Children's Social Care	Defence	Police & Immi gration	Other
1997	28.0%	18.0%	5.3%	4.2%	1.9%	14.4%	5.5%	22.7%
2025	39.8	16.0	5.4	3.1	2.7	9.1	5.1	18.8
Change (pp)	11.8	-2.0	0.1	-1.1	0.8	-5.3	-0.4	-3.9

Source: Table 8 ONS (2025)

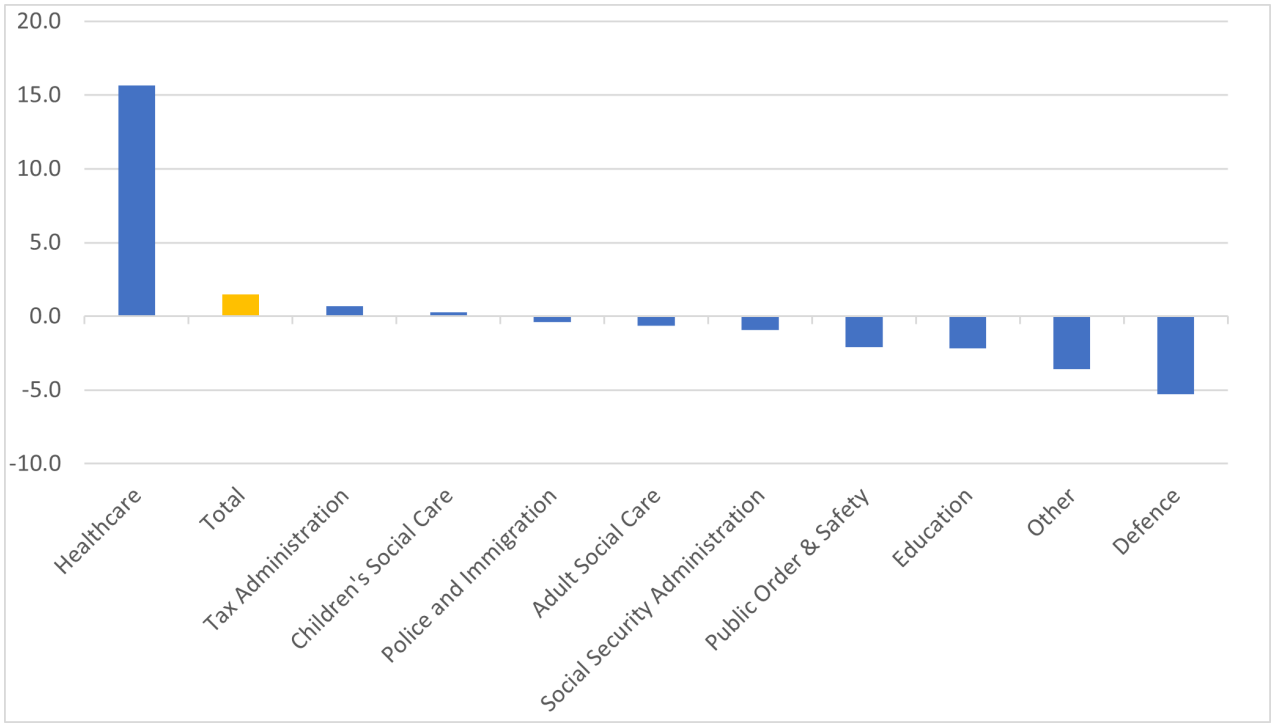
test, trace and vaccinations activities decreased between 2021 and 2022. Because test, trace and vaccinations are not adjusted by quality and the contribution to growth is calculated based on the previous year's expenditure share (which is 2021 for growth in 2022), this results in a lower share for quality-adjusted output in 2022.

2. Service areas where expenditure is not quality-adjusted, such as "Other" government services and Police and Immigration also increased their shares of expenditure in 2022. Therefore Table 2 presents the resulting data in two ways: the full up-to-date data and a version of the data, holding the 2021 weights constant into 2022, to show a like-for-like counter-factual reflecting the impact of the methods changes. In the latter case the share of quality-adjusted estimates increased by about 3 percentage points.

Conclusions

This article presents an update on nearly 20 years of methodological development by the ONS for the measurement of public services productivity, following finalization of a Review commissioned by the UK government. Improving the headline statistics is not sufficient though. Understanding the differences between sectors is vital. Work

Chart 2: Contribution to Whole Public Service Productivity Growth by Type of Service – 1997-2022 (percentage points)



Note: Weighted contribution reflects the change in productivity between 1997 and 2022 times the change in the relative weights amongst total public services. For example, whilst Defence remains 'inputs = outputs', the change reflects its diminishing share of total public services.
Source: Authors calculations based on data contained in https://www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/datasets/publicserviceproductivityestimatestotalpublicservice(ONS2025)

also needs to be done to deliver explanatory supporting information, including disaggregations. Hence the ONS has also invested in surveying new management practices and use of time for the public sector and focused effort on producing more detailed data, particularly relating to healthcare.

Nevertheless, this article has focused on a strategic set of headline conclusions:

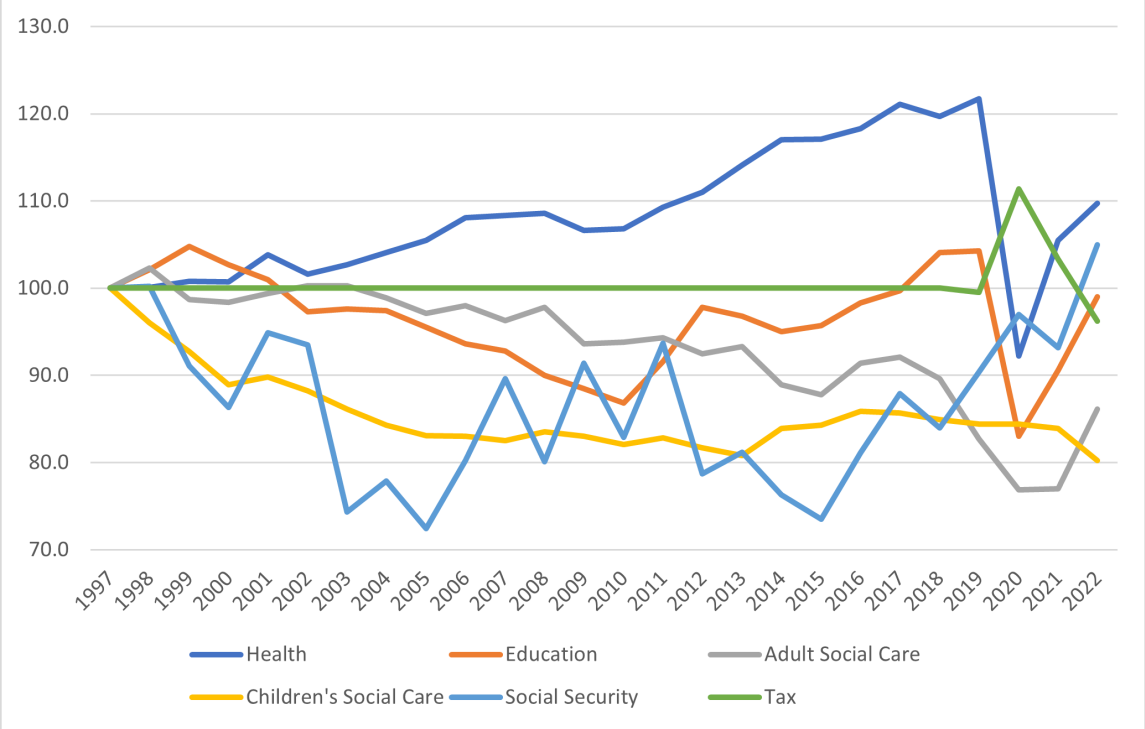
First, the principles underpinning the Atkinson Review appear to still hold true and implementation is becoming increasingly feasible as different areas of government have invested in their data. Assuming the same is true in other countries, the potential to apply these methods elsewhere to achieve a generalized improvement in

Table 2: Shifts in the Shares of Quality Adjusted and Non-quality Adjusted Estimates

Publication Vintage	Quality adjusted direct output	Quality adjusted indirect output	Non-quality adjusted direct output	Non-quality adjusted indirect	Total quality adjusted
2021	48.68%	3.75%	11.79%	35.78%	52.43%
2022	48.63	3.57	11.49	36.32	52.20
2022, holding 2021 weights constant	51.62	3.75	9.40	35.24	55.37

Source: Authors calculations from ONS (2025)

Chart 3: Public Sector Productivity Growth by Service, 1997-2022 (1997=100)



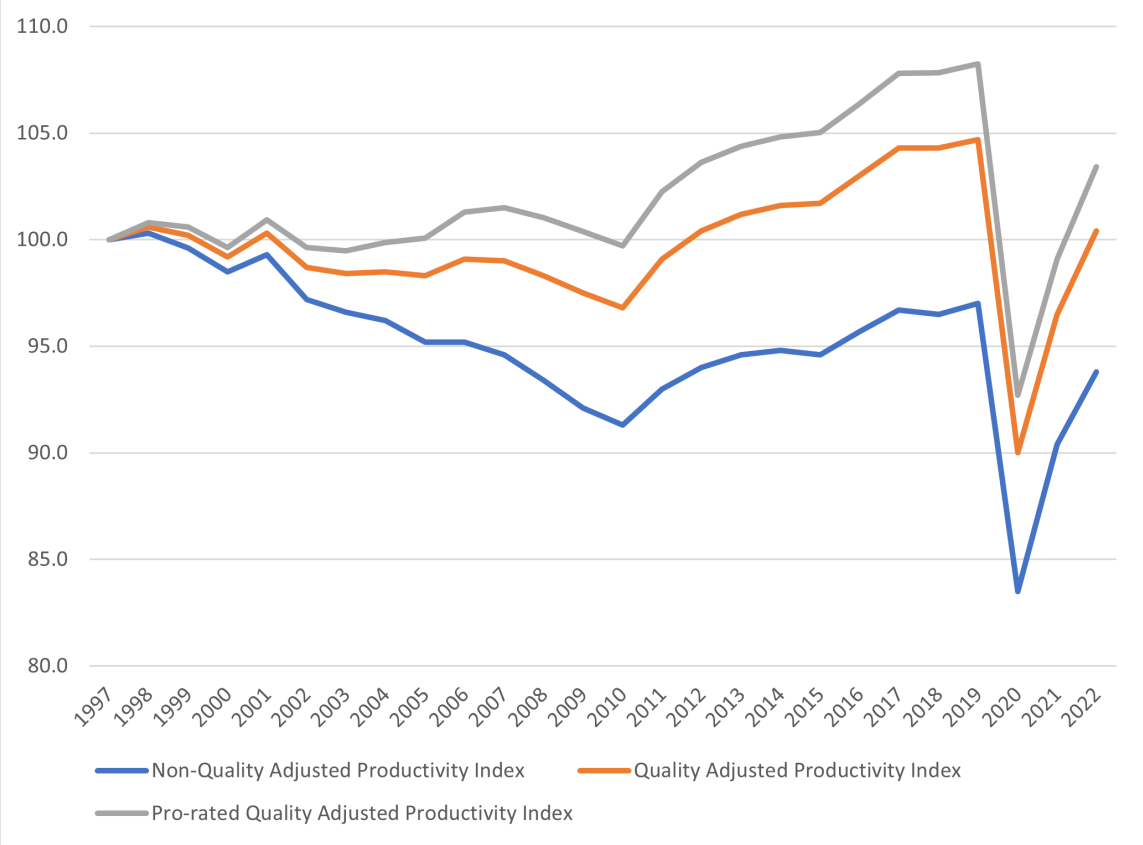
Authors calculations based on data contained in ONS (2025)

volume output and productivity is within reach.

Second, Charts 2 and 3 demonstrate that whilst it may appear that the quantitative impact of these revisions is modest, this disguises large variations by service area. This shows that even at high levels of aggregation the use of 'inputs = outputs' or 'sum-of-costs' methods are deeply flawed approaches to estimating public sector output. While the new 2025 System of National Accounts supports the inclusion of quality adjustments in measuring public services, it also stresses the practical and methodological challenges of such adjustments thereby permitting the continued use of the 'inputs = outputs' convention in national accounts.

Third, even when applied to just 60 per cent of public services, quality adjustment has a significant effect by transforming flat or negative productivity growth into positive trends, as illustrated in Chart 4. Using an index where 1997=100, productivity without quality adjustment declined to 97.0 by the pre-COVID peak, while quality-adjusted productivity rose to 104.7. This shift changes the narrative from one of stagnation to one of modest improvement before the pandemic, followed by a recovery right after the collapse in 2020 due to the pandemic. Similarly, as of the latest data, the quality-adjusted index has rebounded to above 100, whereas the non-quality-adjusted index remains at 93.8 (ONS 2024c). A simple simulation further illustrates the potential impact: if all currently unadjusted services experienced quality improvements at the average rate of those already adjusted, the resulting trajectory—shown as the highest line in Chart 4—would imply a substantial uplift. If this

Chart 4: The Impact (Actual and Imputed) of Quality Adjustment (1997=100)



Note: Pro-rated quality adjustment is calculated by subtracting the non-quality index from the quality-adjusted index, dividing by the share in each year of expenditure which is quality adjusted, and multiplying by 100.
Source: Authors calculations based on data contained in https://www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/datasets/publicserviceproductivityestimatestotalpublicservice(ONS2025)

estimate reflects the true value of quality improvements, incorporating them into national accounts could add an average of 0.1 percentage points to annual GDP growth over the period 1997–2025.

The increased availability of data for quality adjustments, when accurately measured, demonstrate the clear capability of public services to demonstrate productivity growth. This requires us to reconsider the validity of Baumol's Cost Disease as a useful way of conceptualizing services of this type, as the new estimates show reasonable public service productivity growth in the United Kingdom.

Understanding public service productivity should not be a niche activity. Anyone who wants to understand, or set policy relating to, public services should review what these data communicate. Quality of outcomes matter and if universally applied, could increase have a positive impact on measured average per annum GDP growth. Finally, with government debt approaching 100 per cent of GDP in the United Kingdom, the need to ensure government services are being delivered in such a way that it maximizes the delivery of outcomes and outputs per input is as important as it has ever been.

All of this is of particular importance as the world's National Statistical Institutes

prepare to implement the new revision of the System of National Accounts, which expects countries to apply quality adjustments to the volume measure of the public services within the national accounts. The methods discussed above to improve public service productivity should therefore not be viewed in isolation from the wider economic statistics system. The Review includes a roadmap for implementation into UK national accounts and this will likely form one of the most substantive changes in the SNA 2025 revision, in terms of impact on the level of GDP. To ensure continued international comparability, and meaningfulness of GDP data, other countries should look to prioritize similar improvements also.

References

Ab Iorwerth, Aled (2006) "How to Measure Government Productivity: A Review Article on 'Measurement of Government Output and Productivity for the National Accounts' (The Atkinson Report)" *International Productivity Monitor*, Number 13, https://www.csls.ca/ipm/13/IPM-13-Iorwerth-e.pdf.

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

Barclay, N. L., M. Pineda Moncusi, A. M. Jodicke, D. Prieto-Alhambra, B. Raventos, D. Newby, A. Delmesti, W. Y. Man, X. Chen, and M. Catala, (2024) "The Impact of the UK Covid-19 Lockdown on the Screening, Diagnostics and Incidence of Breast, Colorectal, Lung and Prostate Cancer in the UK: A Population-Based Cohort Study," *Frontiers in Oncology*, March 27.

Baumol, W., (1967) "Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis," *American Economic Review*, Vol. 57, pp. 415–426.

Bean, C., (2016) "Independent Review of UK Economic Statistics." https://assets.publishing.service.gov.uk/media/5a7f603440f0b62305b86c45/2904936_Bean_Review_Web_Accessible.pdf

Bojke, C., A. Castelli, K. Grasic, A. Mason, and A. Street, (2018) "Accounting for the Quality of NHS Output," CHE Research Paper 153, Centre for Health Economics, University of York.

Cutler, D. M., K. Ghosh, K. L. Messer, T. Raghunathan, A. B. Rosen, and S. T. Stewart, (2022) "A Satellite Account for Health in the United States," *American Economic Review*, Vol. 112, No. 2, pp. 494–533. https://www.aeaweb.org/articles?id=10.1257/aer.20201480

Davies, J. (2020) "International Comparisons of Healthcare Quality," Nuffield Trust, Indicator Update, 25 June 2020. https://www.nuffieldtrust.org.uk/news-item/international-comparisons-of-healthcare-quality-1.

Dawson, D., H. Gravelle, M. O'Mahony, A. Street, M. Weale, A. Castelli, R. Jacobs, P. Kind, P. Loveridge, S. Martin, P. Stevens, and L. Stokes, (2005) "Developing New Approaches to Measuring NHS Outputs and Activity," CHE Research Paper 6, University of York.

Diewert, W. E., (2018) "Productivity Measurement in the Public Sector," in E. Grifell-Tatjé, C. A. K. Lovell, and R. C. Sickles, (Eds.), *The Oxford Handbook of Productivity Analysis*, Oxford Academic. https://doi.org/10.1093/oxfordhb/9780190226718.013.7

Dunleavy, P., (n.d.) "Public Sector Productivity: Measurement Challenges, Performance Information and Prospects for Improvement." https://eprints.lse.ac.uk/87207/1/Dunleavy_Public%20sector%20productivity1.pdf

Fisk, D. and D., Forte (1997) "The Federal Productivity Measurement Program: Final Results," Monthly Labor Review, Bureau of Labor Statistics, May, pp. 19- 28. http://www.bls.gov/opub/mlr/1997/05/art4full.pdf.

Foxton, F., J. Grice, R. Heys, and J. Lewis, (2019) "Measuring Government Output: Twenty Years of Lessons Following the Atkinson Review." https://ec.europa.eu/eurostat/cros/content/measurement-public-goods-lessons-10-years-atkinson-united-kingdom-fred-foxton-joe-grice-richard-heys-james-lewis_en

Gu, W., and A. Wong, (2015) "Productivity and Economic Output of the Education Sector," *Journal of Productivity Analysis*, Vol. 43, pp. 165–182. https://doi.org/10.1007/s11123-014-0414-y

Hall A.E. (2017) "Adjusting the Measurement of the Output of the Medical Sector for Quality: A Review of the Literature," *Med Care Res Rev.* 2017 Dec, Vol. 74, No. 6, pp.639-667.

HM Treasury, (2022) "The Green Book: Appraisal and Evaluation in Central Government." https://www.gov.uk/government/publications/the-green-book-appraisal-and-evaluation-in-central-government/the-green-book-2020

Kapetanios, G., and F. Papailias, (2023) "Nowcasting Unpaid Production Activity and Quality Adjustments in Public Service Productivity," ESCoE DP 2023-20. https://www.escoe.ac.uk/publications/nowcasting-unpaid-production-activity-and-quality-adjustments-in-public-service-productivity/

Martin, J., and R. Riley, (2024) "Productivity Measurement: Reassessing the Production Function from Micro to Macro," *Journal of Economic Surveys*. https://doi.org/10.1111/joes.12615

Mitchell, J., J. Lewis, J. Zwijnenburg, R. Dkhissi, and T. Prendergast, (2022) "International Comparisons of the Measurement of Non-Market Output During the COVID-19 Pandemic." https://www.oecd.org/publications/international-comparisons-of-the-measurement-of-non-market-output-during-the-covid-19-pandemic

ONS (2023) "Public Service Productivity, UK: 1997 to 2022" Available at www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/articles/publicserviceproductivityuk/latest.

ONS(2024a) "Improved Methods for Total Public Service Productivity: Total, UK, 2021," Available at https://www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/methodologies/improvedmethodsfortotalpublicserviceproductivitytotaluk2021.

ONS(2024b) "Public Service Productivity, Quarterly, UK: January to March 2024" Available at: https://www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/bulletins/publicserviceproductivityquarterlyuk/latest

ONS (2024c) "Public Service Productivity: Total, UK, 2021" Available at: https://www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/articles/publicservicesproductivityestimatestotalpublicservices/totaluk2021

ONS (2024d) "Developing Nowcast Methodologies for Public Service Productivity, UK" Available at: https://www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/articles/developingnowcastmethodologiesforpublicserviceproductivityuk/2024-12-11.

ONS (2025) "Public Service Productivity: Total, UK, 2022" Available at https://www.ons.gov.uk/releases/publicserviceproductivitytotaluk2022.

Prtak, D., (2019) "Measuring Defence Productivity – A First Step." https://quarterly.blog.gov.uk/2019/10/30/measuring-defence-productivity-a-first-step/

RAND Europe, (2021) "Understanding the Contribution of Defence to UK Prosperity." https://www.rand.org/randeurope/research/projects/2021/defence-contribution-to-uk-prosperity.html

Renaissance Learning, (2022) "Understanding Progress in the 2020/21 Academic Year." https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1063488/Understanding_Progress_in_the_2020_to_2021_Academic_Year_Extension_report_covering_the_first_half_of_the_autumn_term_2021.pdf

Schreyer, Paul (2012) "Output, Outcome, and Quality Adjustment in Measuring Health and Education Services", *Review of Income and Wealth,* Series 58, Number 2, June 2012.

Schreyer, Paul (2010) "Towards Measuring the Volume Output of Education and Health Services: A Handbook' OECD Statistics Working Papers 2010/02" Available at https://www.oecd.org/en/publications/towards-measuring-the-volume-output-of-education-and-health-services_5kmd34g1zk9x-en.html.

Smith, R., (2024) "Challenges of Measuring Defence Output." https://www.escoe.ac.uk/challenges-of-measuring-defence-output/

UK Statistics Authority, (2025) "National Statistician's Independent Review of the Measurement of Public Services Productivity." https://uksa.statisticsauthority.gov.uk/publication/national-statisticians-independent-review-of-the-measurement-of-public-services-productivity/

Weale, M., (2024) "New Medical Treatments, Quality Improvements and Public Health Measures in the National Accounts: The Effects of the Diabetes Prevention Programme." https://www.escoe.ac.uk/publications/new-medical-treatments-quality-improvements-and-public-health-measures-in-the-national-accounts-the-effects-of-the-diabetes-prevention-programme/