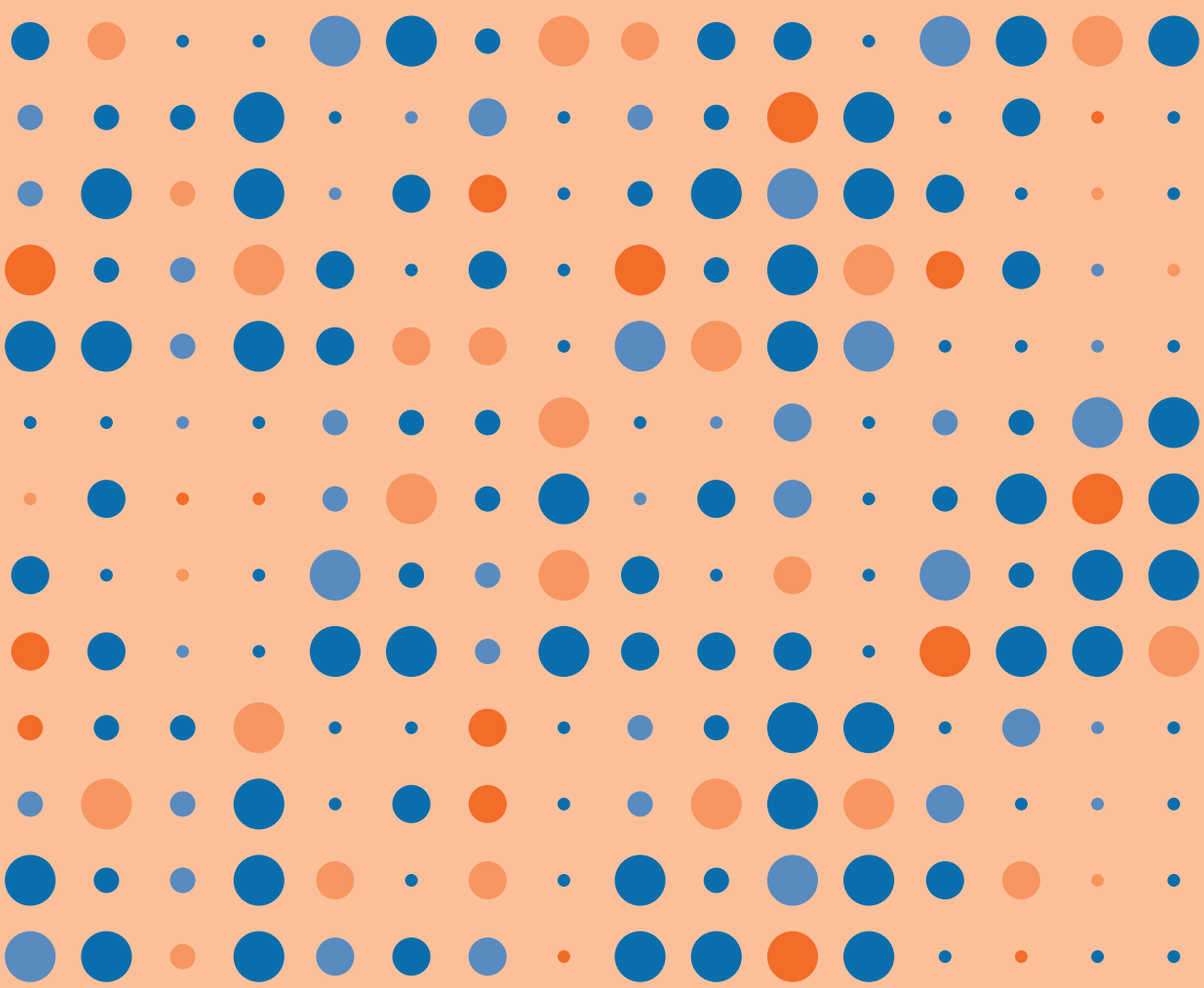
Office for National Statistics

National Statistician's independent review of the Measurement of Public Services Productivity

March 2025



Contents

National Statistician's foreword

I am delighted to publish this report today, summarising the significant work the Office for National Statistics (ONS) has undertaken over the last 18 months to review and improve the measurement of public service productivity in the UK.

In 2023 the then Chancellor of the Exchequer commissioned me to undertake this review. Since then, the productivity of our public services continues to be under scrutiny, particularly in the wake of the coronavirus pandemic. Having good quality data is more important than ever to provide insight and support government decision making in this area.

Measuring public service productivity has historically been recognised as a challenging area. In the UK, around 20% of Gross Domestic Product (GDP) is accounted for by the outputs of public services, comparable to most other western economies. The last significant review of how to measure this substantial section of the economy was carried out in 2005 Atkinson Review Final report [see link 1, Annex I]. The ONS responded to this with implementation of recommended methods for the largest services, particularly Healthcare and Education and led in this field of statistics internationally. Following the Bean Review [see link 2, Annex I] in 2016, work in this area was rejuvenated and further new methods were developed for Children's Social Care, Criminal Justice and Adult Social Care, but that still left numerous other services with methods which have benefitted from this review.

During this review we have revisited the 'Atkinson Principles', considered new methods and sought previously untapped data sources to enable introduction of productivity estimates for public services not measured up until now, as well as improving our existing ones. Today's report outlines improvements we have already made over the last 18 months and those being incorporated in our next statistical publication this Spring. We have also improved the timeliness of provision of data and made recommendations for improving the coherence of our productivity estimates.

To complement the development of our measures of public service productivity, we have conducted groundbreaking survey work on how public sector workers spend their time, and management practices in the public sector, enabling us to make comparisons between the private and public sectors for the first time and understand key drivers of the trends our statistics reveal. Furthermore, we conducted new qualitative research to better understand the lived experiences and views of public sector workers around the factors impacting their productivity and where innovation could improve it.

These analyses have provided new insight on the drivers of productivity in the public sector, and opportunities and barriers to increasing automation, for example through the use of Artificial Intelligence (AI). Future adoption of these data collections as repeat surveys would add to the evidence base on which policy and spending decisions are made.

We have also looked ahead strategically to how our improvements should be incorporated into the compilation of GDP and the UK National Accounts, in the context of the anticipated publication of the new System of National Accounts 2025 [see link 3, Annex I] classification and its subsequent implementation.

Beyond this review we plan continuous improvement of our measurement of public service productivity where, despite the significant progress we have made, there remains more that can be done. The report makes a number of recommendations for the future.

I want to give my thanks to colleagues across the UK Government, in the devolved governments, and partners in various organisations and wider academia who have worked with us, and without whom we could not have delivered this review.

Professor Sir Ian Diamond
National Statistician
March 2025

Executive summary

In the UK, around 20% of Gross Domestic Product (GDP) is accounted for by the outputs of public services, comparable to most other western economies. The recent coronavirus pandemic, and desire to improve public services without increasing spend or taxes has brought the focus back onto public services.

The coronavirus pandemic highlighted that public services can be subject to significant changes which measurement systems designed for more standard times can struggle to accommodate. In 2023 the then Chancellor of the Exchequer asked the National Statistician, Professor Sir Ian Diamond to review the measurement of public service productivity, noting the changing data and policy landscape that may arise in coming years. Automation, the use of Artificial Intelligence (AI) and other forms of innovation also offer opportunities to transform public service delivery.

Measurement of the productivity of public services is historically and internationally acknowledged as being challenging. This is because of the absence of prices to demonstrate the value to citizens of what is produced. This report builds on previous work undertaken by the Office for National Statistics (ONS) to make improvements to published statistics, building on the Atkinson Review: Final report Measurement of Government Output and Productivity for the National Accounts (2005) [see link 1, Annex I] and the Bean Review (2016) [see link 2, Annex I].

This report summarises the work of the review over the last 18 months, presenting the key challenges identified; the cross-cutting methodological improvements the Review has developed; the service specific improvements implemented; and recommendations for further work.

The Review has collaborated extensively with government departments, academics and other bodies across the UK to identify new data sources and innovative methods to improve existing ONS published estimates of public services productivity. The quality, granularity and timeliness of estimates have been improved.

The review has:

- delivered a once in a generation update to the Atkinson Principles (2005)
- improved the quality and granularity of UK public services productivity estimates. Some changes were implemented in published estimates from March 2024; the majority will be incorporated in the Spring 2025 ONS release of data

- significantly improved the measurement of Healthcare, Education, Social security administration, Tax administration, Public order and safety, as well as made research progress on measuring long standing conceptual challenges in Policing and Defence
- improved the timeliness of estimates through publication of experimental annual 'nowcasts' since November 2023 to address the two-year lag in data availability and the consequent lack of quality adjustment
- introduced publication of quarterly experimental estimates of Healthcare productivity from February 2025, with a plan to extend this increased frequency to other services
- developed a plan to improve the coherence of the quarterly and annual estimates to improve understanding for users, which is already underway
- influenced the UK recommendations to the United Nations review of the Classification of the Functions of Government (COFOG) [see link 4, Annex I]. to improve the categorisation of government expenditure data essential for measurement of public services productivity
- delivered groundbreaking research and insight into public sector managers' and workers' views on the opportunities and barriers associated with improving productivity, use of automation and Artificial Intelligence
- produced a strategic 'roadmap' for incorporation of these public services productivity improvements into the UK National Accounts and Gross Domestic Product, aligned with the wider ONS planning for implementation of the new System of National Accounts 2025 [see link 3, Annex I]

With continued focus, further improvements should be achievable in 2026 as well as beyond. It will be important for the ONS to continue collaboration with the growing network of service-specific bodies exploring public service productivity measurement at more granular levels, to ensure consistency and transparency of data in the public domain. The coherence of devolved governments' data underpinning the UK-wide public service productivity estimates could be improved if sufficient user need exists and investment into development is made available.

A significant number of farther reaching recommendations for further work have been identified. Some impact or depend upon other government departmental initiatives, and may need additional funding to bring to fruition.

The review offers a substantive step-change in UK public services measurement, which over the next few years can be applied into the UK National Accounts as well as productivity measures, to support the coherence and accuracy of the UK economic measurement system.

Recommendations

Measurement principles

Recommendation 1: The principles outlined by the Atkinson Review should remain the underpinning intellectual methodology behind the measurement of public services, both in UK National Accounts and public service productivity statistics.

Recommendation 2: There are instances where cost provides a weaker weighting metric than alternative approaches. These alternatives can and should be applied, but only where a clear case can be made against the principle that the alternative metric better reflects the value of the service than the relevant cost data.

Recommendation 3: The value of transfers can be used to weight together individual components within a service or in a quality adjustment but should not be a direct measure of output.

Recommendation 4: Where different weights are used within a service, the cost weighted activity index will still be used to weight in any elements which cannot be addressed with the new weighting approaches. When service level statistics are estimated, these should be aggregated in a cost weighted activity index to produce the national aggregates.

Recommendation 5: Adjustments should be considered to equivalise cost weights between services of equivalent value but different cost.

Recommendation 9: The weights used to bring together quality adjustment components need to, as closely as possible, reflect societal preferences in as objective a fashion as possible.

Recommendation 14: The ONS should continue to work with the devolved governments to understand the devolved service-delivery landscape and improve data coverage, quality and consistency in the UK measure of public service productivity.

Recommendation 26: The ONS should continue to work with other organisations undertaking development activity in the measurement of public service productivity for mutual knowledge sharing, adoption of coherence between estimates where possible, and transparency around differences in measurement.

Recommendations implemented for the spring 2025 ONS release of data

Healthcare

Recommendation 31: The ONS should implement abdominal aortic aneurysm, bowel cancer, breast and cervical screening services within healthcare outputs in Spring 2025.

Recommendation 35: The ONS should use the Department of Health and Social Care Index of Services data to estimate dental activity growth from 1996 until 2006 and improve the method of linking this activity data to the current 'Units of Dental Activity' data source to avoid a discontinuity between data sources.

Recommendation 36: The ONS should ensure the overall weight of ophthalmic services and dental services is consistent with expenditure as reported in the Department of Health and Social Care annual accounts, and uprate historical NHS telephone and website services unit costs to account for NHS cost inflation.

Recommendation 37: The ONS should implement equivalised weights in Spring 2025. Where service weights are equivalised to account in productivity for cost-savings from moving services to lower-cost modes of provision, these should be incorporated in the non-quality adjustment measures. Quality adjustment should account for changes in value of services delivered that goes beyond cost-saving, such as improvements to the estimated health improvement from treatment.

Recommendation 38: The ONS should implement equivalisation of weights for ambulance services in Spring 2025.

Recommendation 40: The ONS should adopt the improvement to remove excess bed days in Spring 2025.

Recommendation 43: The ONS should extend the application of the health gain quality adjustment, excluding the waiting times component, from elective to non-elective procedures in Spring 2025.

Education

Recommendation 50: The ONS should implement improvements identified for Education inputs in Spring 2025 related to salaries.

Recommendation 52: The ONS should implement improvements to the quality adjustment of Education to better account for the impact of the coronavirus pandemic and to account for student well-being and Further Education attainment.

Defence

Recommendation 58: A new method implementing a more granular direct measure of labour inputs into Defence should be applied.

Recommendation 59: A UK National Accounts consistent intermediate consumption deflator should be derived and applied to Defence intermediate consumption spending, using the His Majesty's Treasury Online System for Central Accounting and Reporting (OSCAR) data to weight components.

Recommendation 60: The review recommends that the national accounts capital deflator should be applied to the Defence capital expenditure data, but the combined implied deflator of intermediate consumption and capital should be compared to the Defence contracts price index to determine whether subsequent manual intervention is then required.

Policing

Recommendation 64: The ONS should apply bespoke Police and Immigration deflators to Central Government expenditure data used to indirectly measure intermediate consumption in Spring 2025.

Recommendation 65: The ONS should implement more granular salary data into its direct labour measure for Police, incorporating salary information for individual police ranks (including Uplift officers) in Spring 2025.

Recommendation 66: The ONS should incorporate Northern Ireland workforce data into its direct labour measure for Police in Spring 2025.

Public order and safety

Recommendation 79: The ONS should apply bespoke intermediate consumption deflators to expenditure data in all Public Order and Safety service areas, with the exception of fire.

Recommendation 80: A new method implementing a direct measure of fire service labour inputs into Public Order and Safety should be introduced, expanding in future years to cover devolved governments.

Recommendation 88: Timeliness data by case type should be used to produce a more granular quality adjustment that aligns with the improved Crown Court output measure.

Recommendation 89: A straight linear interpolation should be applied to the re-offending quality adjustment applied for 2018 to 2021 where the coronavirus pandemic interrupted the normal provision of data.

Recommendation 90: Court quality adjustment weights should be revised to be 40% recidivism and 60% timeliness.

Taxation

Recommendation 93: The ONS should 'revenue-adjust' direct measures of Tax Administration output whilst developing, with HMRC, appropriate quality adjustments, including using factors such as the 'tax gap' to reflect the difference between the desired and achieved levels of tax collected, customer satisfaction and call waiting times.

Recommendation 94: The ONS should publish provisional estimates of Tax Administration productivity in Spring 2025 and adjust the "Other" grouping accordingly.

Social security administration

Recommendation 102: The ONS should measure the output of the administration of Universal Credit and its predecessor legacy benefits using a benefit-weighted index to account for the allocative efficiency of the transition. This measure should be seen as provisional whilst undergoing further testing.

Recommendation 104: The ONS should adjust the output of Universal Credit administration for changes in the number of entitlements included in payments, to account for the heterogeneity of cases.

Recommendation 105: The ONS should apply a quality adjustment in Spring 2025 based on existing Department of Work and Pension fraud and error rates.

Recommendations for future work

Concepts and methods

Recommendation 6: Further consideration should be made of which services are considered as preventative services. An imputed valuation for the preventative service A should then be used where this is the product of the probability in reduction in use of service B and the actual cost of service B.

Recommendation 7: For pre-selected preventative services where high quality data on impact of downstream services can be found, the probability weighted cost of these downstream services can be used as a proxy valuation of the preventative services in the cost weighting methodology.

Recommendation 8: For latent capability, further research is required to identify instances where this method could be piloted using high quality data.

Recommendation 10: Further research should be undertaken to consider the potential to use alternative weighting regimes proposed for quality adjustment in replacing cost weights, as per recommendations 3 and 9.

Recommendation 11: The ONS should exploit methods developments around technology and other deflators to improve the measurement of volume input and output in public services, and continue to seek out methods improvement.

Recommendation 12: Further research should be undertaken to consider the potential and value or necessity to apply an adjustment to account for labour inputs undergoing training.

Recommendation 13: The ONS should consider how best to reflect private sector productivity growth within the measurement of public service productivity, where this captures private sector delivery of services.

Recommendation 15: The ONS should further investigate the feasibility and user need for devolved metrics on public services productivity, particularly the education sector in collaboration with the devolved governments.

Recommendation 16: The ONS should continue to improve annual and quarterly public service productivity estimates to take account of available quality adjustment data and, where this is not possible, keep nowcasting models under annual review to provide the most accurate and timely data possible.

Recommendation 17: The ONS should continue the roll-out of publication of service estimates on a quarterly basis having started with the largest sector (Healthcare) in February 2025.

Recommendation 18: The ONS should replace the current 'contribution to growth' compilation method with 'chain volume measures', and then implement reconciliation of the quarterly estimates with the annual estimates each year, in order to align with the UK National Accounts protocols and improve coherence and understanding for users.

Recommendation 19: The Quarterly cumulative Average Growth Rates (QAGR) method should be applied to provide more timely nowcast estimates for annual estimates as further research is undertaken to evaluate the efficacy of alternative methods in the light of the coronavirus pandemic. The performance of the QAGR model should be evaluated on an annual basis.

Recommendation 20: The ONS should proceed with best practice improvements to align quarterly and annual production statistics.

Recommendation 21: The ONS should keep under review whether there is convergence of the HM Revenue and Customs expenditure data and the ONS Government Expenditure on Research and Development Survey estimates to allow future consolidation of the two data sources.

Recommendation 22: The ONS should incorporate the methods and data developments from this review into the UK National Accounts as part of its implementation of the 2025 System of National Accounts revision, in line with decisions made by the National Statistician.

Recommendation 23: The high level 'roadmap' developed by this review for incorporation of quality adjustments, and improvements to public sector output in the UK National Accounts should be reviewed and maintained by the ONS as part of wider strategic planning of the implementation of the System of National Accounts 2025 develops.

Recommendation 24: The ONS should continue to influence the United Nations Classification of the Functions of Government Review to maximise the opportunity for improved future categorisation of departmental expenditure (underpinning inputs measurement) via the HM Treasury Online System for Central Accounting and Reporting system.

Recommendation 25: The ONS should work with HM Treasury to plan for future upgrading of the Online System for Central Accounting and Reporting system to enable implementation of the revised Classification of the Functions of Government.

Recommendation 27: Although not a high priority, the ONS should explore the feasibility of improvements to the measure of the ONS' productivity, to provide a measure that is adequate to feed into the estimates of total public service productivity.

Healthcare

Recommendation 28: The ONS should evaluate the benefits and costs of switching from Finished Consultant Episodes to the person-level data provided by the Hospital Episode Statistics for measuring hospital output in England.

Recommendation 29: The ONS should review, together with the Department of Health and Social Care, how the data needed for the quality adjustment for healthcare services should be produced or commissioned in the future.

Recommendation 30: The ONS should continue to work with the NHS to improve data on NHS-funded services contracted from the independent sector.

Recommendation 32: The ONS should monitor the quality of National Cost Collection activity data for England for abdominal aortic aneurysm, bowel cancer, breast, and cervical screening services, and transition to using these data when they are of adequate quality to bring it in-line with expenditure data used.

Recommendation 33: The ONS and Department of Health and Social Care should explore sourcing data for other preventative activities, such as glaucoma screening for inclusion when data permits.

Recommendation 34: The ONS should continue to engage with Wales, Scotland and Northern Ireland to assess the feasibility of including abdominal aortic aneurysm, bowel cancer, breast, and cervical screening services in the healthcare output measures for the devolved governments in the future.

Recommendation 39: The ONS should continue to explore the feasibility of applying the improvements on handling equivalent treatment across different modes of provision made for England to Scotland, Wales and Northern Ireland.

Recommendation 41: The ONS should keep under review whether incorporation of an adjustment for unnecessary A&E admissions would be feasible, material and proportionate.

Recommendation 42: The ONS should continue to monitor the development of patient satisfaction surveys conducted by the ONS, the Care Quality Commission and others, with a view to further expanding the quality adjustment for patient experience in the future.

Recommendation 44: The ONS should continue to work with the Department of Health and Social Care and the NHS to identify additional cases of lower-cost services being substituted for higher-cost services.

Recommendation 45: Further research should be conducted to continue to explore the feasibility of improving and expanding the existing Healthcare quality adjustment for waiting times.

Recommendation 46: The ONS should explore how to improve the measurement of preventative services in Healthcare output, including commissioning a literature review of data sources to ensure consistent application across the range of preventative treatments.

Recommendation 47: The ONS should continue to investigate whether aspects of the incentives in NHS Payment Scheme can be incorporated in the relative weighting of different services in Healthcare productivity.

Recommendation 48: The ONS should explore approaches to disaggregate the Healthcare service, to allow the relative performance of different components of this large service to be better understood.

Recommendation 49: The ONS should actively explore working with NHS Wales to access its health data and develop stronger productivity metrics for Wales.

Education

Recommendation 51: The ONS should continue to review if Annual Survey of Hours and Earnings remains the best data source for labour data for Education, or whether alternative sources may be preferable.

Recommendation 53: The ONS should continuously engage with the Department for Education over indicators of student well-being.

Recommendation 54: The ONS should work with the devolved governments to improve Education data sources as far as possible.

Recommendation 55: The ONS should, as part of its research agenda, continue to explore the links between Education, the current Education quality measures, health expenditure and human capital acquisition, with specific attention to labour market returns, to better understand the output of these.

Recommendation 56: The ONS should engage with academic researchers and stakeholders to understand if the impact of grading policy on Education productivity is substantial enough to warrant further research.

Defence

Recommendation 57: As input to the United Nations Review of the Classification of the Functions of Government (COFOG) the ONS should recommend that COFOG 2.1 Military defence should be split as follows:

- Division: 02 – Defence
  - Group: 02.1 – Military defence
    - Class: 02.1.1 – military defence of which contains operations of land forces
    - Class: 02.1.2 – military defence of which contains operations of air forces
    - Class: 02.1.3 – military defence of which contains operations of sea forces

Recommendation 61: The ONS should continue to explore the direct measurement of Defence output with Ministry of Defence and other experts following the publication of this report.

Recommendation 62: The ONS should continue to work with Ministry of Defence to explore the potential of using their new readiness measure within the calculation of Defence outputs, and to continue to develop appropriate methods.

Policing

Recommendation 63: The ONS should endeavour to source more granular police workforce data for Scotland.

Recommendation 67: The ONS should continue to explore with Home Office the potential to run an England and Wales Police Activity Survey as part of ONS's survey portfolio.

Recommendation 68: The ONS should explore the potential to source police activity data for Scotland and Northern Ireland.

Recommendation 69: The ONS should continue to collaborate with key police partners on appropriate methods to develop and weight different types of output activity.

Recommendation 70: The ONS should draw together existing data sources for Police 'public safety and welfare' activities and develop a plan to fill data-gaps.

Recommendation 71: The ONS should commission a literature review of the impact of crime prevention activities which fall under the Police 'criminal prevention' activity category, with a view to future development.

Recommendation 72: The ONS should continue to explore the feasibility of introducing a combined Police crime and Police public safety and welfare output measure by critically assessing alternative activity and weighting sources.

Recommendation 73: The ONS should continue to collaborate with policing partners to develop appropriate quality adjustments for Policing services.

Immigration and citizenship services

Recommendation 74: The ONS should seek to convince the United Nations Review of the Classification of the Functions of Government (COFOG) to separate Immigration and citizenship activities from COFOG 3.1 Police services.

Recommendation 75: The review recommends commencing the split of Policing from Immigration and citizenship services from 2004, and retaining the combined series for earlier years.

Recommendation 76: The ONS should review the composition of the Immigration component of the combined Police and Immigration intermediate consumption deflator annually to ensure accurate representation.

Recommendation 77: The ONS should review whether full cost recovery fees, for example for passports, should be treated as market-equivalent prices, which do not require quality adjustment, as prices should internalise quality change.

Recommendation 78: The ONS should continue discussions with the Home Office to identify the most appropriate service activities and suitable weights for these to construct a direct volume output measure for Immigration and citizenship services.

Public order and safety

Recommendation 81: The ONS should explore possible new prison data sources for Scotland to allow for differential cost weights for different categories of prison, and data for Northern Ireland in the ambition of creating a UK-wide output measure.

Recommendation 82: The ONS should continue to work with the Ministry of Justice to obtain expenditure data that enable the relative weights of different groups of probationers to vary with time in future years.

Recommendation 83: The ONS should explore the data available on probation services for Scotland and Northern Ireland with a view to creating a complete UK measure for probation output.

Recommendation 84: The ONS should continue to work with the Ministry of Justice to obtain expenditure data to derive improved weights for individual law courts service areas.

Recommendation 85: The ONS should explore the data available for Scotland and Northern Ireland with a view to creating a complete UK measure for criminal court output.

Recommendation 86: The ONS should explore the data available for the Scotland and Northern Ireland with a view to creating a complete UK measure for legal aid output.

Recommendation 87: The ONS should undertake further research into whether legal aid should be considered an intermediate stage in the provision of justice as opposed to an output in itself.

Recommendation 91: Work should continue on identifying appropriate data to develop a direct measure of labour input for the prison service component of Public Order and Safety.

Recommendation 92: Work should continue on developing a direct measure of labour input for the courts service component of Public Order and Safety, ensuring no double-counting with intermediate consumption input in this area.

Taxation

Recommendation 95: The ONS should work with HM Revenue and Customs to continue to ensure data on different taxes is utilised in as granular a fashion as benefits productivity estimates.

Recommendation 96: The ONS should explore how to complement existing HM Revenue and Customs data to complete coverage of Tax Administration activities.

Recommendation 97: The ONS should investigate further developments that could be made to account for changes in quality such as fraud and error, customer satisfaction and waiting times, but recommends that treatment of the 'tax gap' should be the priority area for focus.

Recommendation 98: The ONS should review data on taxes administered locally and in devolved governments in relation to inputs and outputs, to explore the feasibility of further improving coverage of Taxation Administration.

Recommendation 99: The ONS should investigate the feasibility of extending the time period for Tax Administration productivity measure to pre-2018.

Social security administration

Recommendation 100: The ONS should undertake further work to identify the best treatment of the Coronavirus Job Retention Scheme.

Recommendation 101: The ONS should review whether data exists to better understand, for Social Security Administration, whether non-administrative costs for concessionary fares and Housing Benefit are correctly recorded, and whether they are sufficiently impactful to merit further work.

Recommendation 103: The ONS should seek to acquire data to incorporate Tax Credits, Housing Benefit and Child Benefit in both output and inputs measures of Social Security Administration, while making appropriate adjustments to the "Other" grouping of inputs.

Recommendation 106: The ONS should work with the Department of Work and Pensions to replace stock measures of fraud and error with more appropriate flow measures to better reflect in-period performance, in Social Security Administration quality adjustment.

Recommendation 107: The ONS should investigate data availability accounting for benefit administration in Northern Ireland, as the current measure does not include that activity.

Recommendation 108: The ONS should further investigate whether administrative costs associated with benefits delivery are being captured in Social Security Administration.

Recommendation 109: The ONS should continue to work to identify the most appropriate route to capture user satisfaction in Social Security Administration across the UK.

Recommendation 110: The ONS should continue to collaborate with the Department of Work and Pensions to develop data on timeliness for Social Security Administration.

Recommendation 111: Weighting mechanisms for baskets of quality adjustment factors, such as in Social Security Administration, should be considered and if necessary, data collected to ensure quality adjustment estimates reflect societal preferences and that these are aligned with how other relevant areas are treated, such as health waiting lists.

Local Government, including adult and children's social care

Recommendation 112: The ONS should continue to work with the Ministry of Housing, Communities and Local Government to determine a strategy to capture the remaining local services, with a particular focus on ensuring alignment of treatment with Environmental Services, noting the potential impact of the United Nations Classification of the Functions of Government Review.

Recommendation 113: Local government data across the devolved governments should be sought to enable UK-wide local government estimates to be prepared.

Recommendation 114: The ONS should seek to separate elements of Children's Social Care where output measures cannot be identified, with particular reference to Early Years Childcare, and in the absence of direct output measures estimate output in this area through the 'inputs = outputs' methodology until further measures can be developed.

Recommendation 115: The ONS should periodically investigate the methodology into Children Social Care in relation to non-government sectors in order to make further improvements to the measure.

Recommendation 116: The ONS should continue to keep the Adult Social Care system under review with the aim, utilising ongoing data developments, to address areas of this service which continue to use the 'inputs = outputs' methodology.

Recommendation 117: The ONS should consider the user need for a separate Adult Social Care, England publication in the suite of public service productivity publications.

Environment

Recommendation 118: The ONS should work with relevant departments to recommend a definition of 'Environmental Services' to the upcoming consultation on the international Classification of the Functions of Government (COFOG) definitions to encourage the creation of a wider COFOG sector covering 'Environmental Services', for use in developing measures, using the work of this review to inform developments in this space.

Recommendation 119: An improved Classification of the Functions of Government structure for 'Environmental Services' should differentiate between 'Environmental Protection', 'Natural Resource Management' and 'Climate Change and Net Zero', with further detail embedded below this structure.

Recommendation 120: The ONS should undertake research to identify appropriate deflators for inputs into, and outputs from, Environmental services.

Report summary

In the UK, around 20% of Gross Domestic Product (GDP) is accounted for by the outputs of public services, comparable to most other western economies. The recent coronavirus (COVID-19) pandemic, and the desire to find innovative ways to improve public services without increasing spend or taxes has brought the focus back onto public services, both in terms of measurement and making improvements to delivery itself.

The coronavirus pandemic highlighted that public services can be subject to significant changes which measurement systems designed for more standard times can struggle to accommodate. In 2023 the then Chancellor of the Exchequer asked the National Statistician, Professor Sir Ian Diamond to review the measurement of public service productivity, noting the changing data and policy landscape that may arise in coming years. Automation, the use of Artificial Intelligence (AI) and other forms of innovation also offer opportunities to transform public service delivery.

The measurement of the productivity of public services is historically and internationally acknowledged as being challenging. This is because of the absence of prices to demonstrate the value to citizens of what is produced. This report builds on previous work undertaken by the Office for National Statistics (ONS) to make improvements to published statistics, building on the Atkinson Review: Final report Measurement of Government Output and Productivity for the National Accounts (2005) [see link 1, Annex I] and the Bean Review (2016) [see link 2, Annex I].

The following principles, which are referenced in recommendation 1 of this review, are reproduced from recommendation 4.1 of the Atkinson Review (2005) and continue to apply:

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

This report summarises the work of the review over the last 18 months, presenting the key challenges identified; the cross-cutting methodological improvements the Review has developed; the service specific improvements implemented; and recommendations for further work. The review offers a substantive step-change in UK public services measurement, which over the next few years can be applied into the UK National Accounts as well as productivity measures, to support the coherence and accuracy of the UK economic measurement system.

Status of published ONS estimates pre and post review

Tables 1 and 2 illustrate the status of the ONS published estimates, by public service, prior to the review, and the position achieved by Spring 2025 when improvements will be incorporated into the Annual Public Service Productivity (PSP) statistical publication [see link 5, Annex I] on an ongoing basis.

In these Tables a yes indicates robust estimates; a no indicates the use of simpler methods. The use of 'Research' indicates the review was able to undertake research and develop recommendations for further work, but not implement improvements.

Table 1 Published estimates status prior to this review, June 2023

Service	Robust inputs	Activity based outputs	Quality adjusted outputs
Social security administration*	No	No	No
Taxation administration**	No	No	No
Policing and immigration*	No	No	No
Defence*	No	No	No
Education	Yes	Yes	Yes
Healthcare	Yes	Yes	Yes
Public order and safety	Yes	Yes	Yes
Environmental services***	No	No	No
Adult social care	Yes	Yes	Yes
Children's social care	Yes	Yes	Yes
Other local government services***	No	No	No

Notes:

* Input = outputs (and hence by definition the growth of productivity is zero)

** Pre-review, taxation was part of "other" grouping

*** Environment and other local government Services are within "other" grouping

Table 2 Published estimates status post this review, Spring 2025

Service	Robust inputs	Activity based outputs	Quality adjusted outputs	Incorporated into PSP published estimates
Social security administration* **	Yes	Yes	Yes	Yes
Taxation administration* **	Yes	Yes	Research	Yes
Policing and immigration*	Yes	Research	Research	No
Defence*	Yes	Research	No	No
Education	Yes	Yes	Yes	Yes
Healthcare	Yes	Yes	Yes	Yes
Public order and safety	Yes	Yes	Yes	Yes
Environmental services **	Research	Research	Research	No
Adult social care	Yes	Yes	Yes	Yes
Children's social care	Yes	Yes	Yes	Yes
Other local government services**	No	No	No	No

Note:
*Input = outputs (and hence by definition the growth of productivity is zero)
**Environment and other local government services remain within "other" grouping
*** Statistics under development (experimental statistics)

Further work has been identified to build upon this progress. See the list of recommendations in the previous chapter.

Main findings

Key challenges

Even before the global coronavirus pandemic, the delivery models of some public services were changing substantially (for example, the transition of the benefits system to Universal Credit) necessitating new methods to better account for these. The review makes key proposals to address this change and make the system more robust to any future changes.

As well as changes to services, over time there has been, and will be, changes in data availability. This may be because data are no longer available or sufficiently robust, or new data becomes available which allows further disaggregation of measures or improved robustness in measurement.

Some services have multiple outputs or outcomes which makes it difficult to create robust metrics. A number of considerations need to be taken into account such as mapping inputs to each activity, establishing relative weights and attributing the output and outcomes to participating bodies.

The coronavirus pandemic, as a significant shock to the UK economy, fundamentally impacted models of service delivery and has necessitated intervention in some public service productivity models. In some cases the need for this will continue as the pandemic's impact continues to affect public service delivery outcomes.

The coronavirus pandemic is impactful not just on service delivery, but also directly on the statistical models themselves. This is particularly the case with regards to 'seasonal adjustment' (the statistical methodology employed to smooth out the effect of seasonal impacts from the quarterly time series) where new structural breaks have required the ONS to refresh it's understanding of when different activities happen and how to best reflect this. This review is very much in step with wider domestic and international methods development across the public sector.

Using cost to value services, as Atkinson (2005) noted, leads to a failure to recognise the value added by public services. This is most extreme for preventative services, which are often low cost but deliver significant later savings, for example a lower-cost tobacco cessation programme which would reduce a future higher-cost expense, in this example – cancer operations. The review has carefully considered how to address this challenge.

The UK is not alone in recognising the importance of this topic. In 2024 the United Nations (UN) launched a review of the Classification of Functions of Government (COFOG) [see link 4, Annex I]. This presents the opportunity for the UK to feed the findings of this review into international negotiations on recommendations to update the current classification system used for public service measurement to better reflect current (and expected future) patterns of government expenditure.

Areas the review has proposed which merit particular attention by the international process include gaining a better line of sight of Environmental spending, Tax Administration, and Defence.

Adding to the evidence base, improving methods and data

The review has undertaken a line-by-line audit of existing methods and proposed multiple improvements based on new data, new methods and the experience accrued in delivering estimates since the Atkinson Review (2005).

The review has resulted in implementation of improved methods and data into the accredited annual and experimental quarterly Public Service Productivity statistics [see link 6, Annex I] published by the ONS, improving transparency for users.

Investment in new pilot surveys of how public sector workers spend their time, and the management practices employed in the public sector, has yielded valuable insight into the drivers of productivity in the public sector.

Qualitative research on the lived experience of public sector workers has provided insights into understanding the impact of administrative tasks on productivity and highlighted opportunities and barriers to innovation, including the use of automation and Artificial Intelligence (AI) to improve productivity.

The review has worked with ongoing work on the measurement of Government Research and Development funding to help ensure the understanding of this key driver of productivity is well understood and evidenced.

The review has worked to improve deflators used to derive public service volume estimates of inputs and outputs, with particular focus on improving the relevance and accuracy of intermediate consumption (IC) inputs deflators across all service areas (apart from Healthcare and Adult Social Care). There have also been small improvements to labour inputs deflators, and exploratory work on the capital inputs deflator for Defence.

The review has focused methodological development work on addressing outstanding gaps relating to the measurement of preventative services and latent capability, to better reflect the value of such services where current cost weighting models may under-represent their importance in modern public service delivery.

The review has challenged the accepted concept of ‘collective services’, which has been used to brand some areas of public services as ‘too hard to measure’. Whilst the ONS has not been able, in the time available, to fully develop methods in every area, it is increasingly clear that, given modern data availability, it should move away from dividing services into ‘individual’ and ‘collective’, and address opportunities to tailor solutions given the nature of the service.

It remains the case that some services deliver multiple outcomes and how to reflect these in quality adjustment, and particularly how to weight these different outcomes together remains to be addressed. Similar issues apply over time – is a GCSE point in 1998 worth the same as a GCSE point in 2023 – and the review has developed recommendations for further work in this regard.

The review has led to a number of countries to directly approach the ONS to discuss methods and concepts, and in some cases to launch their own parallel domestic processes.

Improving timeliness of public services productivity estimates

Timeliness of published National Statistics accredited estimates of annual public services productivity has been improved through development of experimental 'nowcasting' modelling [see link 7, Annex I], to address the two-year time lag in the necessary administrative data sources becoming available. These will continue to be under review given the challenges of modelling associated with the coronavirus pandemic.

Publication of experimental quarterly estimates of productivity by service sector has begun, with the release of estimates for the Healthcare service [see link 8, Annex I] from 10 February 2025 (the largest service by spending) with plans for release of Quarterly estimates of Education productivity, followed by a roll-out to other services.

Coherence of public services productivity estimates

The principles of the Atkinson Review (2005) remain relevant, but this review has developed innovative methods to implement these, building on the knowledge the ONS has gained in producing these statistics over the past twenty years.

Coherence between annual accredited National Statistics and quarterly experimental statistics on public services productivity can be improved, and a programme of methodological work to achieve this has been developed.

Coherence of estimates across the devolved governments of the UK could be improved. The review has found numerous examples where different measures have evolved in different systems and where consistency across data sources would improve coherence in the UK statistics produced.

The development or proliferation of productivity measurement in local areas of the UK, and the establishment of organisations focused on increasing productivity (and efficiency) in specific public services, means that coherence between these measures and the national measures produced by the ONS becomes increasingly important.

Summary of sector specific findings

Recognising work which was recently undertaken on Children's and Adult Social Care following the Bean Review (2016), this review has identified and implemented further improvements in the measurement of public services productivity in the Healthcare, Education, Social Security Administration, and, Public Order and Safety services, exploiting new datasets.

In Healthcare, Education, and Public order and safety the review has also identified key improvements which react to the impact of the pandemic on these services and better reflect the true pattern of productivity growth during and following the pandemic.

Material improvements in the measurement of the productivity of the Police service are possible: first by separating policing from Immigration and citizenship services, and secondly by categorising police outputs into three themes: crime and criminal investigation, public safety and welfare outputs, and crime prevention, to provide a structure to understand the diversity of services delivered. The review prioritised 'crime and crime investigation' and improvements should be implemented in 2026 following final refinement. Quality adjustments have also been considered but require data: the review has explored the Home Office ad hoc Police Activity Survey and recommends that this survey is made regular, with the possibility of inclusion in the ONS survey portfolio.

Defence productivity, a long-standing conceptually challenging area to measure, is an international measurement issue rather than a UK-specific one. It has been possible to make some improvements on measuring UK Defence inputs, which will be implemented in the Spring 2025 forthcoming statistical release. Conceptual improvements to output measures have also been researched but appear dependent on work at the Ministry of Defence, who has developed a measure of output "readiness", which is a defined and measured state of military preparedness.

Given the increased awareness of, and public spending on climate change, net zero and environmental protection it is clear that the classification of environmental services needs to be better brought together to aid visibility. Currently these are dispersed across multiple public services. The review has submitted its considerations on this issue to the UN COFOG review. The review recommends improvements need to be in the light of the changes expected from the UN process, and anticipates this will not report until 2027, but notes this area, when fully mapped should be seen as a large service area, at least comparable in size to policing.

High quality data sources, which exist for several smaller local government services, including waste management, environmental services, highways and planning, would allow improvements to be made at pace, although whether at least some of these will be classed within environmental services by the COFOG UN review needs to be resolved.

Historically the ONS has grouped smaller areas of government into an "Other" category, which has grown through time and is now the second largest reported unit in ONS statistics. This is analytically unhelpful, so the review has explored how to disaggregate this "Other" category and has prioritised improving estimates of the productivity of Tax Administration working with HM Revenue and Customs, with implementation from Spring 2025. As a result, the "Other" grouping will have a lower expenditure share and in some years will now be the largest in terms of expenditure after Healthcare and Education. Further work is recommended, particularly on environmental services, which should reduce this further.

Social Security Administration improvements can be made by applying similar innovative methods as those for Tax Administration. That is, the ONS should measure the productivity of services which award benefits to (or receive taxes from) citizens in a comparable way. In addition, the review identified and implemented a solution to the transition of the legacy benefits system to Universal Credit, but continued work is required to identify the transition of recipients of Tax Credits to Universal Credit.

For an 'at a glance' summary of improvements and further work for each service examined by the review see Annex A of the full report [see link 9, Annex I].

Impact of this review on published ONS estimates of public services productivity

In developing improved estimates of public services productivity the review has worked extensively with relevant UK government departments, arm's length bodies and other organisations in the quality assurance of these data. The impact on published estimates of public services productivity has been extensively analysed and considered. The ONS productivity statistics publications incorporated improvements in Healthcare and Education [see link 10, Annex I] from March 2024, with further improvements being implemented in Spring 2025 for other services as shown in Table 2. Subsequent continuous improvement will be possible if recommended research is successful.

New and improved estimates with associated revised back series are being implemented in the annual statistical publication of public services productivity estimates by the ONS in Spring 2025. A separate, accompanying article detailing the impact on previously published estimates and any revisions is planned for release simultaneously.

Impact on the UK National Accounts

The National Statistician's acceptance of the recommendation of the National Statistician's Committee for Advice on Standards for Economic Statistics' (NSCASE) for the UK to move from the European System of Accounts 2010 ('ESA10') to the revised System of National Accounts in March 2025 ('SNA25') necessitates the incorporation of public services 'quality adjustments', such as those used in public service productivity estimates into the UK National Accounts and compilation of Gross Domestic Product (GDP) for the first time since 2010. The work of the review to move towards a comprehensive set of quality adjustments is a substantive contribution towards this goal. A high level 'roadmap' to continue this work has been compiled as part of the current ONS planning for implementation of SNA25 more widely. This roadmap will need to be maintained and updated in alignment with the final SNA25 implementation plans. The roadmap can be seen at Annex B within the full report [see link 9, Annex I].

The incorporation of quality adjustments into the UK National Accounts is one of a number of changes that will be incorporated as a result of implementing SNA25. Incorporation of quality adjustments is one of the more significant changes in measurement along with the recognition of data as an asset. The impact of these changes on UK GDP and National Accounts estimates will be evaluated by the ONS in future as part of the wider plans for implementing SNA25.

The ONS will also update outputs measures used in the national accounts to mirror those implemented in productivity estimates as a result of this review.

Conclusions and further work

The review has made significant progress delivering a step-change improvement in the published ONS estimates of public services productivity, which will be made available from the Spring 2025 release of data, improving the accuracy, granularity and timeliness of these statistics.

With continued focus, a number of further improvements should be achievable in 2026 and beyond. In addition a work programme to improve the coherence of the annual and quarterly public services productivity estimates in 2026, has been produced.

Valuable new insight into public sector workers' use of time and managers' views on the opportunities and barriers to use of automation and Artificial Intelligence can be obtained through pilot surveys. Future investment would be needed to add these to the ONS' survey portfolio.

The coherence of devolved governments' data underpinning the UK-wide public service productivity estimates could be improved if sufficient user need exists and investment into development is made available.

There is a growing network of service-specific bodies exploring public service productivity measurement at more granular levels than the ONS can deliver under current budgets, such as the Centre for Police Productivity. These propositions should make local data more transparent. Consistency of methods will require continued engagement.

A significant number of farther reaching recommendations for further work have been identified utilising new data and methods which could deliver further improvements. Some impact or depend upon other government departmental programmes or initiatives, and may need additional funding to bring to fruition.

Further step change improvements in productivity measurement are possible, particularly in Policing, Immigration, Defence and Environmental services, to continue to improve the evidence base for policy decisions relating to the productivity of public services.














CHAPTER 1

Background to the Review

1.1 The Atkinson Review (2005)

In the UK, around 20% of Gross Domestic Product (GDP) is accounted for by the outputs of public services, comparable to most other western economies. How best to measure this substantial section of the economy has been a key question in the UK National Accounts for a long time, but the recent experience of the coronavirus (COVID-19) pandemic, and the desire to find innovative ways to improve public services without increasing spend has brought the focus back to public services, both in terms of measurement and improving delivery itself.

In 2023 the then Chancellor of the Exchequer asked the National Statistician, Sir Ian Diamond, to review the measurement of public service productivity.

Previous in-depth investigation took place in 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 UK National Accounts, with a Final Report [see link 1, Annex I] produced in 2005. This informed the development of the System of National Accounts (SNA) 2008 [see link 11, Annex I] in how to conceptualise and empirically measure output of the public services in GDP. Importantly, it established a set of key principles which informed future work. The UK, alongside several other countries [see link 12, Annex I], pressed ahead with implementing these methods, particularly in relation to healthcare and education. This work addressed the largest parts of public services, but gaps remained.

The Bean Review [see link 2, Annex I] (2016) commended the Office for National Statistics (ONS) for this work but identified that renewed efforts were needed to update methods where quality adjustments were in place, and to create new adjustments where these were not. The ONS then extended the landscape by developing measures related to Adult Social Care, Children's Social Care and Public Order and Safety (excluding Police and Immigration). This work led to 55% of public services by volume being quality adjusted.

The coronavirus pandemic highlighted that public services can be subject to significant changes which measurement systems designed for more standard times can struggle to accommodate. In particular, ONS and the Organisation for Economic Co-operation and Development joint research [see Link 13, Annex I] evidenced that when different countries apply different methodologies in their national accounts international comparisons can be hard to interpret.

The potential for Artificial Intelligence and other innovations to transform public service in future years was an added driver to ensure the data presented by the ONS is as robust and useful as possible, and measures are future-proofed and able to account for innovation and technological changes.

This report presents the key challenges identified, key cross-cutting methodological improvements developed and consequent sector specific improvements.

1.2 The Atkinson approach

This Review is very much in the tradition of the Atkinson Review (2005) and should be seen as a periodic update of that approach and methodology. However, the first recommendation reflects the longevity and intellectual power behind Professor Atkinson's proposals.

> **Recommendation 1:** The principles outlined by the Atkinson Review should remain the underpinning intellectual methodology behind the measurement of public services, both in National Accounts and public service productivity statistics.

Central to these principles is that the objective is to deliver a dataset which is useful for analytical purposes. The Review experienced the desire from policy-makers to understand these data, often at granular levels. Whilst the Review has attempted to develop metrics which meet this need, it has also been aware of data burdens where, as with Atkinson, the ONS has looked to use only the data departments should already be using to manage their services.

1.3 The national accounts roots of the Atkinson approach

The Atkinson Review derived its principles from core measurement principles contained in the national accounts, particularly those articulated in the SNA2008. It took valuation methodologies used in the market sector and considered their application to the non-market sector, specifically public services. Importantly, in some services national accounts concepts may differ from how policymakers would normally view their service: for example in Tax Administration, whilst the ONS measures the number of taxpayers being processed, the HMRC notes the importance of the total tax-take gathered. This issue is discussed further in Chapter 13.

Underpinning the Atkinson Principles is the fact 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 balance two alternative perspectives on value:

- the costs of production, including an appropriate margin
- the discounted sum of benefits the consumer believes at the time of purchase they will accrue from acquisition of the product

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, in that 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, in a competitive framework, it must be the case that:

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

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

- a sum of costs methodology (which the Review will refer to as 'inputs = outputs') where the value of the output is set equal to the value of the inputs which go into its production
- a methodology which looks to proxy the discounted sum of benefits approach by splitting the analysis into two computable components: a direct measure of the volume of the relevant output (e.g. number of hospital operations) and a direct measure of the quality of the final output (so an operation where on average more patients see a greater improvement in the quality of life is valued more highly than one where the improvement is less strong)

In both cases, subtracting the volume of inputs from the volume of outputs would deliver an estimate consistent with the concept of Gross Value Added (GVA), as applied in the private sector. Similarly, dividing the volume of output by the volume of inputs would present a productivity measure. However, under the 'inputs = outputs' approach to measuring output, dividing output by input always gives a result of one. That is, productivity is implicitly assumed to be constant, unless an assumption is made to apply a 'margin' or 'profit' to the costs the state incurs in producing the service.

This is clearly circular: productivity would equal whatever assumption was used to define this 'margin'. Hence this method is considered inferior as without being able to apply an objective method for the margin discussed, it is a weak proxy for the exchange equivalent price, and hence is only used where stronger alternatives are not available.

The stronger method is to directly estimate the volume of output, and an objective quality adjustment, using evidence and data, to reflect how the value of this activity changes over time, as received by citizens. This model rests on two assumptions:

Assumption one

That the quality adjustments act in the same way as 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. This has two implications:

- a quality change is implicitly equivalent to an output change in volume terms (that is, if someone was to buy car A, which is twice the price of Car B, this is equivalent to purchasing two Car Bs, so replacing a Car B with a Car A in this instance doubles volume output). As such, a 1% increase in quality is equivalent to a 1% increase in output.
- this is clearly equivalent to market transactions. Imagine a factory makes one good brick in one time period, and then two broken bricks in the next. The broken bricks cannot be sold for a positive price and can only be given away (price = £0.) Simply because there are two bricks does not mean quantity has doubled, instead it has fallen one hundred percent. In the public sector, there can be activities (say hospital operations) where the outcome delivered worsens; so comparing one successful hospital operation with two unsuccessful operations, the Review would not wish to suggest the two unsuccessful operations were worth double the one successful operation. In this instance the quality adjustment would fall to ensure the final value reflected the value of the operations being delivered.

It is this recognition that the ONS need to adjust measures of output to deliver closer analogy of the methods used in the market sector, that marks Atkinson's Review as the landmark which it is.

Assumption two:

The Government prioritises social welfare in decision making.

Government would, within this, make decisions relating to the quantities of each public service to deliver, such that costs are spent up until the quantity where citizens would no longer desire additional units of output. 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 may not be feasible to take all optimal decisions and hence:

$$\text{Sum of Costs} \leq \text{Discount sum of Expected Benefits}$$

On the assumption that the budget available to government is insufficient to deliver all valid opportunities, it could be assumed that under this constraint interventions would be chosen where downstream benefits on average exceed the sum of costs.

Sum of costs would then inherently be expected to under-estimate the value produced, even if accurate assumptions could be made about the margin which the public sector would apply (it would be expected for governments to draw from the top of the distribution). For this reason, applying a quality adjustment is essential to accurately reflect the value created by the public sector in a form comparable to those observed within the market.

CHAPTER 2

The measurement challenges in 2024

Whilst the Atkinson Review (2005) [see link 1, Annex I] remains a comprehensive and valid study, with principles which have demonstrated longevity, the landscape of public services has changed materially in this time. This Review identified key challenges identified through the implementation of the Atkinson Review and more recently as a result of coronavirus (COVID-19). Six key challenges are described under each of the following headings. These are considered on a sector-by-sector basis as each issue may differentially impact on different public services, or in some areas the Office for National Statistics (ONS) may already have adequately addressed this issue since 2005.

2.1 The passage of time

The simple passage of time delivers two distinct challenges:

- changes experienced within services – a service may have been re-designed in a fundamental fashion such that the measures no longer reflect the landscape. For example:
  - during the coronavirus pandemic, National Healthcare Service created new 'Test and Trace' capability, and other new services, which were outside the existing measurement framework
  - from 2013, Universal Credit began to replace a number of benefits, which formed the core of the existing measurement model. In 2018, the ONS suspended the existing measure of Social Security Administration as the model struggled with this change.
- changes within the measurement of services – the service may have remained consistent through time but the measurement system may have deteriorated. This may have been for a number of reasons:
  - data sources may have ceased to be published by various agencies
  - data quality may have deteriorated because of falling sample sizes or other statistical reasons
  - the coronavirus pandemic directly impacted on statistical models. This is particularly the case with regards to 'seasonal adjustment' (the statistical method employed to smooth out the effect of seasonal impacts from the quarterly time series), where new structural breaks have required the Review to refresh its understanding of when different activities happen and how to best reflect this.
  - data are nowcast to cover more recent time periods, but the model may need updating and bringing up-to-date, particularly if the model was adversely affected by the discontinuity of the coronavirus pandemic

2.2 The opportunities presented by new data

Significantly more data are available across government today than in 2005. The Review has considered where services can now be reliably measured where this was not previously possible, or where data can be disaggregated to give more refined measures, better marrying 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 to derive the headline volume series.

2.3 The challenge of collective services

Atkinson recognised a pre-existing national accounts norm that services are divided into 'individual' and 'collective' services. That is, those where the service affects 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 – no one can 'opt-out' of the UK-wide nuclear deterrence operated by the Ministry of Defence's submarine fleet.

How to value this deterrence, and how the valuation would change ("Would citizens feel more defended if the UK purchased an additional nuclear submarine?") remains a significant question. There is not any fundamental rule requiring the Review to treat collective services differently, but they present complexity against many of the key principles Atkinson established. A service which can be received by one individual is inherently easier to measure because it generally delivers outputs which can be more objectively measured.

There is also a question about where to draw the boundary: only a small fraction of citizens actively engage with the police or the criminal justice system: should the Review measure value against the services received by victims and (unwillingly) by perpetrators, or should it instead value the protection the system grants everyone? Is policing and criminal justice an individual or collective service? Are vaccination services individual or collective, given the whole of society benefits even if every single individual is not vaccinated?

This Review concludes the debate around the labels 'collective' and 'individual' is a futile one; what matters are the individual methodological challenges presented by each service and how the Review unravel these to give a meaningful measure of value and the volume of output being produced. This challenge remains as fundamentally difficult today as faced by Atkinson and those who have worked on this topic in the interim.

2.4 The challenge of services with multiple outcomes

Measuring the output and outcomes delivered by a service can be complex when they are focussed on one or a small number of outcomes (for example Healthcare Services primarily deliver health interventions). Others are characterised by delivering multiple, varied, outcomes. Policing is the clear example, with the police working to 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 resolve anti-social behaviour, alongside counter-terrorism and addressing organised crime. This raises a number of distinct challenges:

- mapping inputs to each activity
- accessing good quality and consistent data for each activity, with no double-counting between activities
- being able to adjust the relative weights of these activities in the aggregation process based on accurate and timely data
- being able to attribute 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 these two agencies?

2.5 The direct impact of the coronavirus pandemic

Both Healthcare and Education Services, alongside others, saw the operating model for their existing services fundamentally transformed by the coronavirus (COVID-19) pandemic, either in terms of the quantity or nature of what they were asked to deliver. In some areas this impacted the measures used in productivity estimates.

In the area of Education, for example, where output is measured via exit qualifications (e.g. GCSEs in England), it is not just this year's teaching and learning which shape this year's results: the previous ten years of formal education, and pre-primary early year's provision, need to be taken into account.

As such the existing methodology pro-rated qualification results back through the cohort's education. So, for example, only 30% of a student's success in Year 11 in 2020 was attributed to the year of education they received in 2020: the other 70% was attributed to their previous educational career. The coronavirus pandemic fundamentally disrupted this model as weaker performance in the pandemic years, caused by disruption to education at that time, could not be taken to imply that education delivered pre-pandemic was similarly affected.

It would be inappropriate to model that a student sitting their exams in 2021, and whose results suffered because of disruption in 2020 should see their 2017 performance downgraded. The Review looked to address this.

2.6 Preventative services and latent capability

How to consider preventative services is at the heart of the conceptual challenges raised by the coronavirus pandemic, but which extend far wider. Preventative services present the most extreme example of needing to quality adjust output data to reflect the true value created by a service. They are generally designed to cost significantly less than the down-stream benefits they may unlock.

For example, a low-cost tobacco cessation programme aims to reduce future demand for expensive cancer operations. The sum of cost approach, even if quality adjusted, may be insufficient to reflect the valuation produced, especially if a cost weighted activity index is used to aggregate these services alongside other services: a far greater weight would apply to the expensive operations. How to resolve this is addressed in Chapter 3.

CHAPTER 3

General methods improvements

Some of the methods challenges raise wider strategic questions which apply across multiple services and require a consistent treatment. These are considered here and are also reflected in the following chapters specific to service areas.

The topics considered in this chapter are:

- alternatives to cost weighting
- application of these principles
- the treatment of preventative services
- the treatment of latent capability
- the treatment of complex baskets of quality metrics
- deflation
- the treatment of labour inputs undergoing induction training
- the treatment of purchased services
- the impact of devolution

3.1 Alternatives to cost-weighting

The absence of prices has long presented the main difficulty in valuing different non-market services, including for the purposes of weighting them into a single output measure. In line with System of National Accounts guidelines, weighting is conventionally done using cost weights. Cost weighting offers a feasible, consistent alternative in the absence of prices, given cost weights are available and feasible, and internally consistent. In addition they weight output against a consistent 'floor' for the value of public services as determined by the provider, if not the consumer, of these services, given that a provider would not provide a service at a higher cost than the value they assign it.

However, as explored in depth in paras 6.17 to 6.22 of the Atkinson Review [see link 1, Annex I], it is the feasibility of cost weights rather than their representation of value that result in their choice:

'...we may not be able at present to apply estimates of (value) ... So, for the present, the only feasible approach appears to be to continue to use cost weights.' (para 6.19)

The disadvantage is that costs may be biased in terms of representing the value of the service to citizens. The Review considered this issue as a priority: in terms of the key challenges described in Chapter 2, biased weights are a significant driver of all of them save, arguably, the challenge of collective services.

Cost weights are not a strong metric, but in many cases there are no practical alternatives. The Review audited alternatives from the perspective that any changes proposed had to deliver superior benefits, at least in terms of being more accurate and being able to interact with cost weights. As a minimum any alternative would continue to need to be in money weights in market equivalent prices.

> **Recommendation 2:** There are instances where cost provides a weaker weighting metric than alternative approaches. These alternatives can and should be applied, but only where a clear case can be made against the principle that the alternative metric better reflects the value of the service than the relevant cost data.

The principles guiding the application of this recommendation are:

1. Introduction of alternative weighting methodologies should not permit the subjective or arbitrary selection of which method to apply.
2. Services should be selected based on a clear assessment that cost weights are either inappropriate or inferior to an alternative method.
3. Selection of an alternative method should be dependent on existing strong evidence on which to base any necessary assumptions or factors. Cost weighting has the advantage of consistency: if alternative methods are not more robust it does not make sense to remove that consistency without gain elsewhere.
4. Methods should be piloted in experimental series for consultation with users before being introduced into core statistics.

3.2 Application of these principles

The two services where feasible alternatives to cost weights exist are Tax and Social Security Administration. Payments between government and citizens are viewed by the national accounts as 'transfers' rather than additional output. That is, they move money from one agent to another (from taxpayers to the government, or from the government to benefit recipients) without changing the overall value of goods and services produced.

Transfers are, therefore, excluded by convention from measurement of output of private and public industries, otherwise Gross Domestic Product (GDP) could double simply by transferring all earnings to a different agent, even if all agents end up with the same funds.

The challenge is, in productivity terms, ideally the minimal inputs would be used to deliver each pound of tax collected or each pound of benefits distributed. Weighting different casefiles by cost fails to take into account the amount of tax raised, or the amount of benefits dispersed by different taxes or benefits.

The key example the Review examined was in the social security system: the replacement of six benefits with Universal Credit meant that multiple case files were replaced by a single claimant file. A new benefit, which delivers an equivalent value of benefits to several older ones, may have a lower value in a cost weighted output measure than the older benefits it was replacing if the related costs of processing one file rather than six had fallen. If the Office for National Statistics (ONS) used the number of case files as the measure of output, and they all cost the same, assuming inputs remain constant, this would appear to deliver up to an 86% reduction in productivity, dependent on how inputs varied, which would be counter-intuitive given the purpose of rationalising of the benefits system is to improve productivity by removing duplicative activities.

For this reason, in 2018, the ONS deactivated the Social Security Administration output metric it used at that time and reverted to 'inputs = outputs' so as not to present a misleading picture of productivity. The coronavirus (COVID-19) pandemic then delayed plans to revisit this; the Review therefore treated this as a priority.

The Review identified that use of case file measures of output inadequately reflects the fact that not all casefiles are equal in value terms, and thus some method of taking into account the amount of taxes collected or benefits dispersed, other than cost, needs to be considered to address this inequity.

In the case of Universal Credit the Review recommends the use of 'benefit-weights' rather than cost weights to reflect that the relative value of Universal Credit is higher if it consolidates various benefits into a single case file. In the case of Taxation, the Review recommends an intermediate 'revenue-adjustment' should be applied in lieu of a fully developed quality adjustment, which reflects that the tax-take should be bounded – the fiscal authority determines the amount of tax it wishes to be collected as a social optimum and Tax Administration should not aim to gather more tax than this, even though normally more output would be considered beneficial to society.

Whilst the Review continues to adhere to the principle to exclude transfers in the direct measure of output, given the function of the output is to collect or disperse transfers, these data can be used as a weight to reflect the value of different case-file types in the measure of output, or within a quality adjustment. For further detail see Chapter 3.

> **Recommendation 3:** The value of transfers can be used to weight together individual components within a service or in a quality adjustment but should not be a direct measure of output.

Using benefit-weights for those parts of Social Security Administration affected by the transition to Universal Credit does not mean every aspect of this service receives this treatment so cost weights need to be retained. Chapter 14 provides more detail.

> **Recommendation 4:** Where different weights are used within a service, the cost weighted activity index will still be used to weight in any elements which cannot be addressed with the new weighting approaches. When service level statistics are aggregated, a cost weighted activity index should be used to produce the national aggregates.

In some cases, specific adjustments should be applied to cost weighting to avoid treating similar outputs delivered via different channels at different costs in a way which moving from expensive to lower-cost service providers suggests lower output. For example, the Healthcare output measure sets the weights equal in the case of the same hospital procedure being carried out with or without an overnight stay, when patient characteristics and needs are constant.

> **Recommendation 5:** Adjustments should be considered to equivalise cost weights between services of equivalent value but different cost.

It is important to note this is a different issue to the substitution of higher-cost services for lower-cost services as an alternative if this does not represent a reduction in inputs. The classic example is the replacement of branded drugs and pharmaceuticals with 'generic' versions, which are generally lower cost but identical in impact. This question is addressed in detail in the chapter on Healthcare in Chapter 7.

3.3 The treatment of preventative services

Preventative services are those that, by definition, efficiently reduce demand for future, higher cost services, and hence overall cost. The challenge around preventative services is that they are only commissioned if the cost of delivery is lower than the cost avoided. This is made more complex if citizens receive additional benefits from avoiding the high-cost service. There are numerous examples:

Healthcare interventions which reduce demand for higher cost interventions, which citizens may find invasive, painful, or have a wider impact on their life. They may also improve patients' quality of life prior to and after when they would otherwise have received treatment. Avoiding treatment may also deliver wider benefits, such as not losing time in work, or suffering pain arising from treatment, or subsequent complications.

Children's Social Care services, such as family-support and similar policies, which reduce demand for high-cost services, such as placements in residential children's homes.

Enforcement activities around Tax Administration, which aims to deter non-payment, and therefore reduce the cost of future enforcement activity by incentivising payment. Chapter 13 discusses this within the issue of fraud and error.

Defence services act as a deterrence which reduce the risk of future military actions.

The Review commissioned a report [see link 14, Annex I] from Professor Martin Weale, Kings College London, which used diabetes programmes as an exemplar in this area. Prof. Weale identified two problems: how to value services which are in operation, and how to value new preventative activities (akin to the classic 'new good' pricing challenge in national accounting).

Professor Weale argues that the volume effects of introducing preventative services should be derived from the implicit price consequences. In the latter case, in the period before introduction a 'reservation price' for the service is derived. This is the amount at which the demand falls to zero because the benefits of the service are exceeded by its costs. In the case of deriving such a reservation price, the impact on final outcomes, such as quality adjusted life years (QALYs) would be taken into account.

In the first instance, the Review focussed its attention on existing services, and methods to implement Prof. Weale's proposals, which imply the identification of an imputed valuation for the preventative service which better represents its true value. Once a decision is reached on which services are in scope, the Review considers that an 'imputed avoided cost' should be used as a substitute for the actual cost where:

Imputed avoided cost of preventative service A =
(probability in reduction in use of service B) x (actual cost of service B)

Recommendation 6: Further consideration should be made of which services are considered as preventative services. An imputed valuation for the preventative service A should then be used where this is the product of the probability in reduction in use of service B and the actual discounted cost of service B.

Note that this is not the only feasible method Prof. Weale argued the aggregate of avoided costs and increases in the quality of life, as measured in QALYs, could be used as the measure of value. The Review acknowledges this proposal but recognises health improvements of this nature have generally not been intensively used in quality adjustment of healthcare services. This is because not all health outcomes are a result of public health interventions: for example, people's health improves if they give up smoking, reduce their alcohol consumption, take up exercise, or eat more healthily.

Weale, however, acknowledges that while this proposed approach provides a conceptual basis for accounting for preventative programmes, there are practical challenges in its implementation. In particular, the need to define a total amount for the gain in QALYs attributable to Healthcare Services overall. For example, the approach in Cutler et al (2022) [see link 15, Annex I] would require estimates of expected QALYs of the population.

Atkinson's key principle of attribution is also relevant: the output of public services cannot be increased for value not created by them. Prof. Weale (2024) makes the point that if properly established evaluation data are used which have addressed causality, the issue of attribution is no longer a binding constraint.

The Review agreed, but concluded further research is needed on how many such studies exist and are applicable before proceeding. If the ONS applied such an adjustment for diabetes, but not other health issues, would that incentivise investment into diabetes which may not be reflective of true relative value? Methods should be tailored to each service so they are considered fairly, even if that means applying different methods to different questions to arrive at a suitable and consistent level of accuracy.

The Review therefore proposes the methods applied in the short-term focus on existing cost data for avoided activities in the first instance. This should start by agreeing a defined list of preventative services so other items cannot be treated in this way inappropriately.

They would require a number of data items:

- each preventative service needs twinning to a particular service which is being prevented (or a basket of such services). This should be defined before calculation to prevent any appearance of arbitrary decisions.
- the cost of the prevented service needs to be available, including relevant changes through time and deflators as appropriate. As these data should already be available within the system, this should be the easiest data to access.
- the availability of robust probability data around the causal relationship between the delivery of the preventative service and the service avoided. The new Cabinet Office Evaluation Registry, which is accessible to civil servants, provides a single venue for all government evaluations which can form the basis for this dataset. The key issue will be whether these probabilities are constant through time. Most UK National Accounts 'fixed factors' of this type would be subject to five yearly revision, so this will need to be carefully considered, including deflation.

Importantly, this methodology does not deviate from the principle that the cost of services is relevant and comparable. By utilising cost data for avoided services the principle of cost weighting is adapted rather than wholly replaced: the preventative services however priced are to be cost weighted. As potentially one of the most substantial potential areas for methods revision, this continuity is valuable, and is intuitively in line with the way public servants think about this type of activity.

> **Recommendation 7:** For pre-selected preventative services where high quality data on impact of downstream services can be found, the probability weighted cost of these downstream services can be used as a proxy valuation of the preventative services in the cost weighting methodology.

3.4 The treatment of latent capability

The section on preventative services relates most to the delivery of public services, but there is a distinct sub-category which relates to capital investment, and hence a type of input, rather than output. This is when additional capital investment is undertaken not for immediate use but to cope with periods of peak demand in the future. This capital expenditure therefore counts toward the inputs denominator of the productivity equation, but without impacting on the outputs numerator, hence depressing productivity.

The treatment of such 'latent capability' is particularly relevant given the coronavirus pandemic in 2020. The construction of ultimately unused 'Nightingale hospitals' demonstrated that latent capability could bias productivity calculations.

If such institutions had been maintained through 2021 and required for use in 2022, how would the ONS reflect this expenditure in 2020 and 2021? The normal treatment of capital as an input is to take account of the consumption of capital (the rate at which it is worn out), but unused facilities neither depreciate at the normal rate nor contribute to the productivity of the Health Services until they come into active use.

Whilst this example did not come to pass, more routine examples exist in Defence where stores of ammunition and weapons etc are held in contingency of future demand. Whilst defence output measurement is difficult because of its collective nature (see Chapter 9) this is compounded by this aspect of latent capability, on top of being a preventative service. The Review noted the argument that to activate latent capital, staffing would often be necessary.

In the case of latent capability, extending the logic of the proposed methodology relating to preventative services would suggest the ONS needs to consider whether capacity which is not intended for final use in the delivery of the service at that time, should be regarded as a separate final output as investment is made, on an 'inputs = outputs' basis. This would smooth output through demand peaks, although this would appear to be a similar approach to those taken by many other comparable countries in their approach to events such as the coronavirus pandemic (see ONS 2022 [see link 13, Annex I]). This is therefore a compromise that attempts to reflect the value of the asset as created and 'used' as a latent capability.

The other alternative would be to consider the latent capability as an insurance service, where again the ONS would need to define what the resultant output was. The pros and cons of all options would require further consideration.

The principles the Review has applied for preventative services appear to also apply here. The proposed model requires, in particular, robust data on the probability of peak demand events occurring to justify moving away from the traditional approach. The number and value of instances of latent capability are few, and relatively small outside of defence, so this is not a priority, but provides a sensible model for consideration as public service measures become more sophisticated.

> **Recommendation 8:** For latent capability, further research is required to identify instances where this method could be piloted using high quality data.

3.5 The treatment of complex baskets of quality metrics

One of the most common challenges identified by the Review was that some services deliver more than one type of outcome. The classic example is Policing, as discussed in Chapter 2, where crime prevention, crime solution, crowd-management, finding missing persons, and fighting organised crime, alongside counter-terrorism work, are outcomes which are balanced in terms of delivery. Another example is schools, which primarily deliver educational qualifications, but also develops citizens as functioning members of society.

Weighting together such baskets of outcomes into a single quality adjustment necessitates identifying a mechanism to expose the relative value that people as a collective place on different outcomes or outputs in an unbiased fashion.

The Review considered that an objective rather than subjective method to aggregate different outcomes is preferred.The Review considered that an objective rather than subjective method to aggregate different outcomes is preferred.

There are two ways to consider this: identify a metric that reveals people's view through their actions (revealed preference) or directly reported through some selection methodology (stated preference), and which can act as a weight within the calculations. The Review recognised the five possible mechanisms identified in Heys (2024) [see link 16, Annex I] across these two approaches:

- the use of prices as a weight, noting that for public services the challenge is that prices are not available
- the use of time as a weight, on the basis that people allocate their time on the basis that time has an equal value – so an activity people devote on average two hours to a week is valued twice as much as an activity people devote on average one hour to a week
- the use of voting or surveys to directly report preferences
- the use of legislative and regulatory decisions as a proxy for social preferences
- the use of a value of a wellbeing year (WELLBY) (as per Layard and De Neve (2023) [see link 17, Annex I] and HM Treasury's wellbeing supplementary guidance to the Green Book (2022) [see link 18, Annex I] using the change in WELLBYs multiplied by the value as the weights)

When the ONS developed new metrics for Public Order and Safety in 2018, the relative weight of topics within the inspectorate regime was used to allocate weights between the different components of quality (reducing re-offending, safety and decency in prison, keeping the public safe etc). The logic was that the inspectorate regime set out by Ministers in Parliament served as the best approximation of social preferences across this set of outcomes that the ONS could feasibly access, as Ministers are elected by society.

This demonstrates that feasible weighting approaches derived from existing structures can be identified and implemented from legislative and regulatory models. Another example is the legal 'purposes of sentencing' give potential weights for the court service within the criminal justice system. Services will face different challenges in this regard, but the principle is clear.

> **Recommendation 9:** The weights used to bring together quality adjustment components need to, as closely as possible, reflect societal preferences in as objective a fashion as possible.

This recommendation applies to the weighting together of quality adjustments, but there could also be relevance to the weighting together of inputs or outputs.

> **Recommendation 10:** Further research should be undertaken to consider the potential to use alternative weighting regimes proposed for quality adjustment in replacing cost weights, as per recommendations 3 and 9.

3.6 Deflation

Inputs are a challenging measurement area, specifically when the costs of different inputs change over time, in part because the quality of the inputs are changing. Deflators are used to calculate the price change for like-for-like products over time so there can be control for quality change of inputs. This can be difficult if the quality of the input is changing as this may lead to volume change being misinterpreted as price change, and hence will directly affect productivity calculations.

The derivation of appropriate, accurate deflators is an important part of the compilation process across national accounts. In line with recent developments the Review has focussed on where more granular data allows deflation to occur at lower levels of disaggregation. This is particularly important for the improvements proposed in Defence and Policing. More detail is provided in the relevant chapters.

Again in common with the wider national accounts, the degree to which deflators fortechnology products keep pace with rapid moving technology change is important for public services, and will increasingly be so as technology such as artificial intelligence and cloud computing are adopted. This is a wider area of methods and data improvement which the ONS is undertaking alongside the Review.

> **Recommendation 11:** The ONS should exploit methods developments around technology and other deflators to improve the measurement of volume input and output in public services, and continue to seek out methods improvements.

How to deflate output volumes is also an important issue, going back to at least Fisher and Shell (1972) [see link 19, Annex I]. Whilst there is a case for using the final consumption deflator, the GDP deflator is generally applied where necessary and local deflators are not available.

The Review also considered the valuation of quality adjustments. This has been debated since the Atkinson Review: does the volume of health services output increase if the value of a QALY goes up? The Review has considered a price change of this nature to be a price effect which increases the value of output, but not the volume; in the same way that in the private sector the volume of gas consumed does not go up if the price changes, but is open to academic feedback on this point.

3.7 The treatment of labour inputs undergoing induction training

An emerging issue across a number of services is where new entrants are brought in as trainees, with salaries which reflect long-term potential, but may over-estimate the value of their contribution whilst they are still undergoing their induction training. Where the volume of trainees is relatively constant this should not affect growth rates observed in the measure of total inputs, but in a period where a substantial additional recruitment regime for trainees is implemented, this may over-estimate the value of the inputs being used in the service, and hence under-estimate productivity.

How to resolve this is dependent on a number of factors: should an adjustment be made to scale-back the inputs delivered by trainees relative to more experienced staff? If so, how big should this be; for how long should it apply; and should the adjustment decrease over time reflecting skills acquisition? Alternatively, an approach similar to how capital inputs are treated might make sense: the labour inputs of new entrants who are undergoing intensive training programmes could be accrued over later accounting periods.

The Atkinson Review (2005) recommends weighting labour by skill level, hence the Review looked to use salary bands in the compilation of labour inputs. Where sliding pay scales are available this weighting will give a reflection of staff experience. The detail and the availability of these pay scales dictates whether these are sufficient to account for the level of staff experience, including new entrants undergoing a training phase.

This approach allows the most consistency across services however, where these data are not sufficient, or there are more extreme changes to recruitment or training practices, then further adjustment could be considered. It will be more difficult to capture the quality effects on inputs from training for fully qualified staff. In this case you have examples of dedicated time given by employers for training, and on-the-job training conducted by staff during working hours that may not result in a change in pay. Where this remains constant between periods there will be minimal impact on growth rates, but the ONS should consider the impacts of larger changes in training practices. Atkinson referenced that a better way to measure staff skill levels would be implementing an estimate of human capital into the input measure, however, there is insufficient information to do this at present.

The important principles to be applied are:

- if such an adjustment is made to one service for staff induction, it should be applied similarly across all areas
- treatment needs to be symmetric – if an intervention is made when the share of trainees is above average and likely to bias productivity down, it should intervene in the opposite direction if the share of trainees is below average
- any such adjustment would need to be evidenced-based in terms of determining its scale

> **Recommendation 12:** Further research should be undertaken to consider the potential and value or necessity to apply an adjustment to account for labour inputs undergoing training.

3.8 The treatment of purchased services

The Review refers to 'public service' rather than 'public sector' because the UK uses a mixed-market in delivery. Public services can be delivered by private as well as third sector and public entities. The public service productivity data takes account of this by recognising private and third sector provision as inputs and often utilising 'inputs = outputs' to derive their output.

This may impose similar problems to those identified elsewhere, that is, changes in delivery design may result in improvements in allocative efficiency that only capture technical efficiency – changes in the delivery of an existing service.

It would be anticipated that part of the reason for using an alternative provider is to secure productivity gains, or the potential for future productivity gains.

Given the interest in the relative performance of the productivity of the public and private sectors, how to reflect this to ensure the productivity of public service is correctly estimated is clearly an important question. Given the productivity of private sector delivery of services can be derived from other ONS statistics, there is merit in further research to apply appropriate Gross Value Added growth factors to publicly procured private sector delivery for inclusion in estimates of public service productivity. But this would not be the case with public sector output, as clearly this would double-count the private sector's contribution to GDP.

> **Recommendation 13:** The ONS should consider how best to reflect private sector productivity growth within the measurement of public service productivity, where this captures private sector delivery of services.

3.9 The impact of devolution

The published ONS public service productivity statistics are UK-wide. Across the public sector some aspects are reserved (decisions are taken by the UK Parliament) whilst others are devolved (decisions are taken by the devolved governments) and in some cases services are partially devolved. In devolved services differences in policy and subsequent legislation can lead to differences in data availability and its definition.

In compiling the UK-wide estimate of public service productivity, the ONS currently estimates some components for the devolved nations. The Review has worked with the devolved governments to map data availability for Wales, Scotland and Northern Ireland, as well as England. This has enabled identification of data gaps and investigation of potential data sources which would lead to improved coherence across the UK measure.

Methodological and data improvements for each service are outlined in Chapters 7 to 16. Information is also provided on whether the improvements are being made or proposed across the UK, highlighting which sectors are devolved or reserved. Moving forward, the ONS will continue to build relationships with the devolved governments to share updates and understand differences across countries.

> **Recommendation 14:** The ONS should continue to work with the devolved governments to understand the devolved service-delivery landscape and improve data coverage, quality and consistency in the UK measure of public service productivity.

The Review looked at the feasibility of creating metrics at a devolved level, using education as a test case. This work is still in very early development, and to move forward further investigation is needed on whether data are suitably comparable, the nature of the user need for devolved public service productivity statistics, and handling of potential misinterpretations.

Regardless of whether the ONS publishes the estimates at a devolved level, it is a useful exercise to complete to better understand the data and differences across the four countries of the UK.

> **Recommendation 15:** The ONS should further investigate the feasibility and user need for devolved metrics on public services productivity, particularly the education sector, working with the devolved governments.

CHAPTER 4

Coherence of public services productivity published data and the demand for faster statistics

The Office for National Statistics (ONS) publishes a range of statistics on public service productivity (PSP). This has been developed incrementally to meet user needs in the years following the Atkinson Review [see link 1, Annex I].

The ONS publishes two estimates for PSP. Methodological differences between the annual and quarterly statistics are explained in Section 9 of Sources and methods for public service productivity estimates [see link 20, Annex I]:

- annual estimates which are badged as accredited National Statistics with data on a two-year time lag and are quality adjusted [see link 21, Annex I]
- quarterly estimates which are badged as Official Statistics in development; these are timelier but experimental, and only published for the most recent time periods. Unlike the annual series, these hold quality adjustment constant from the last quality adjusted annual period available, accounting for published quality levels, but not recent growth.

One of the key issues identified by users is the time lags involved in sourcing some of the data required for producing the annual estimates, particularly the quality measures.

The second is the complexity in interpreting the published quarterly and annual estimates given methodological differences, and the degree to which these present a unified and coherent dataset.

4.1 Coherence of ONS public service productivity publications

Coherence is a cross-cutting theme of the Code of Practice for Statistics [see link 22, Annex I], which states that:

> "producers must demonstrate that they do not simply publish a set of numbers, but that they explain how they relate to other data on the topic, and how they combine with other statistics to better explain the part of the world they describe."

The Review has considered how coherence can be improved across the ONS suite of publications on PSP, and with other ONS productivity publications, given user interest in comparison with whole economy and market sector productivity.

Using the Office for Statistics Regulation's framework for achieving coherence [see link 23, Annex I], a review of web analytics has provided insight into how the publications are currently used. A user engagement survey launched in January 2025 will identify changes needed to optimise the purpose of each publication and the relationship between them, to improve the clarity of the overall narrative.

4.2 Work programme to improve coherence and timeliness of ONS estimates

The Review found that the post-review steps needed to improve the coherence and timeliness of the ONS PSP estimates were:

- development of 'nowcasting' methods aiming to bring changes in quality into the most recent quarterly time periods, and to address the two-year time lag in availability of annual estimates. This was a key priority for the Review, and nowcasts were initially incorporated from November 2023 [see link 24, Annex I], then methods were reviewed in December 2024 [see link 7, Annex I] and updated from February 2025. This is described in more detail in the section Nowcasting methodologies for quality adjustment.
- taking faster account of quality adjustment data with shorter time lags (for example, education data are available after one year, but is implemented alongside other quality data with a two-year lag)

> **Recommendation 16:** The ONS should continue to improve annual and quarterly public service productivity estimates to take account of available quality adjustment data and, where this is not possible, keep nowcasting models under annual review to provide the most accurate and timely data possible.

- improvement of the quarterly experimental PSP publication to include disaggregation for individual services. The ONS introduced publication of quarterly experimental service estimates starting with the largest sector (Healthcare) from February 2025 [see link 8, Annex I] with the aim of rolling out to other sectors.

Recommendation 17: The ONS should continue the roll-out of publication of service estimates on a quarterly basis having started with the largest sector (Healthcare) in February 2025.

- Replacement of the current 'contribution to growth' compilation method with that of 'chain volume measures' to better align quarterly and annual compilation methodology in line with UK National Accounts best practice. The differences in methods in the historic series make interpreting the published quarterly and annual estimates relatively complex. Addressing this will allow the ONS to present a more unified and coherent dataset. This work is underway.
- Reconciliation of the quarterly estimates with the annual accredited National Statistics PSP estimates. For the years covered by the annual publication the sum of the quarterly series will equal the annual figure (known as 'benchmarking the quarterly series to the annual series'). Quarterly data after this point follows this new annual level. This is in line with the ONS' UK National Accounts protocols and will simplify understanding for users in interpreting the growth shown in the quarterly and annual estimates.

Recommendation 18: The ONS should replace the current 'contribution to growth' compilation method with 'chain volume measures', and then implement reconciliation of the quarterly estimates with the annual estimates each year, in order to align with the UK National Accounts protocols and improve coherence and understanding for users.

4.3 Nowcasting methodologies for quality adjustments

The timeliness of outcome data, which are frequently related to longer-term changes for citizens means that annual estimates are currently produced with a two-year time lag. For example, the measure of re-offending used counts the number of offenders who commit further offences in the year following the end of their sentence. A further year is then given for the courts to complete the process of re-conviction. These are explained in Section 9 of Sources and methods for public service productivity estimates [see link 20, Annex I]. The Review therefore prioritised the development of 'nowcasts' (estimates for the years annual data are not yet available).

Nevertheless, prior to the Review, the ONS already published quarterly estimates, holding quality constant at the last published date, alongside 'annualised' estimates (the sum of the relevant four quarters) at the end of each calendar year as a 'nowcast' to provide timelier annual estimates for users.

However, this approach has limitations. Although both the annual estimates and annualised quarterly estimates track changes in the quantity of services delivered, only the annual series includes an up-to-date picture of quality adjustment. Moreover, the quarterly indices (which are in themselves experimental) have lower coverage and are less granular than the accredited National Statistics annual estimates.

The inference from using an annualised approach of the quarterly time series is that the annual, quality adjusted productivity series will follow the same path as the non-quality adjusted series (derived from the quarterly estimates), implying that quality increases or falls over time with a mean of zero, whereas the Review knows it has a generally upward path.

With the aim of improving upon this the Review developed and implemented a 'nowcast' method to produce new experimental estimates of annual PSP, incorporating the ability to estimate the likely change in quality adjustment, and started publishing this from November 2023 [see link 24, Annex I].

The Review has iteratively improved these techniques. The ONS initially produced an experimental modelled estimate using Dynamic Regression (DR). This is an extension of autoregressive integrated moving average (ARIMA) modelling which allows inclusion of predictor variables. Unlike forecasting, which relies heavily on projections and assumptions about the future economic situation, the DR approach uses data on annual quality adjusted productivity trends for previous years, and information on more recently published quarterly PSP, to provide timelier estimates of annual PSP.

To address the two-year time lag, these nowcasts are a product of the observed annualised quarterly series for the latest two years, and the relationship between the observed annual series (including quality adjustments) and annualised quarterly series preceding those years (excluding quality adjustment).

The publication of the 2021 accredited official annual PSP statistics allowed assessment of the accuracy of the initial DR model. It was found that the nowcast estimate was an over-estimate, particularly in some service areas. Given nowcasts are reliant on recent data, the pandemic and its impact on the UK economy made this nowcasting exercise particularly challenging. The DR model under-estimated the effect of the pandemic. Consequently, the Review analysed options to further improve this approach, testing different treatment of 2020 (the year most affected by the pandemic).

In December 2024 the Review published an article [see link 7, Annex I] recommending a model utilising Quarterly cumulative Average Growth Rates (QAGR), which in general performed better in terms of accuracy than dynamic regression. Productivity nowcasts for total productivity, and the Healthcare and Public Order and Safety services saw improvements using the QAGR method.

However, the new Education nowcast continued to over-estimate growth relative to published accredited annual statistics.

From February 2025 [see link 8, Annex I], in quarterly publications, the ONS has published experimental annual nowcast estimates using the QAGR approach. The performance of the two most optimal methodologies in terms of accuracy (QAGR and DR) should be reviewed on an annual basis to ensure this conclusion remains valid as the analysis moves further away from 2020 (the biggest impact from the pandemic). Future developmental work could also include exploring other nowcast methodologies such as a hybrid approach or state-space models. For further information please see Developing nowcast methodologies for public service productivity, UK [see link 7, Annex I].

> **Recommendation 19:** The Quarterly cumulative Average Growth Rates (QAGR) method should be applied to provide more timely nowcast estimates for annual estimates as further research is undertaken to evaluate the efficacy of alternative methods in the light of the coronavirus pandemic. The performance of the QAGR model should be evaluated on an annual basis.

4.4 Future work

The work programme developed by the Review is a substantial programme of methodological work that has been started and partially implemented by the Review as described in section 4.2. An associated plan to deliver the work programme has been developed for financial year ending 2026. Delivery will require resources to conduct this development work in parallel with production and publication of estimates using the existing methodologies until they can be replaced.

> **Recommendation 20:** The ONS should proceed with best practice improvements to align quarterly and annual production statistics.

CHAPTER 5

Exploring drivers of public service productivity

The level of public sector output produced reflects not only the amount of inputs but the efficiency in which labour and capital are combined to produce that output. This highlights not only the importance in understanding the type of activities undertaken by those delivering public services, but also the effectiveness with which these are managed.

It is therefore important to be able to measure a) what activities are undertaken; b) the amount of time spent on work-related activities by those delivering them; and c) how well organisations are managed. The Review commissioned two new 'pilot' surveys of the public sector to provide such insights. Piloting these surveys would then inform decision-making on future user need and cost associated with adding them to the Office for National Statistics' (ONS) survey portfolio.

The potential of new drivers of public service productivity are also important. Given possibilities through use of Artificial Intelligence (AI), this chapter also explores where data can be sourced to aid understanding of how public services are making use of these new tools, as well as other examples of Research and Development within the government sector.

5.1 Public Sector Time Use Survey (PSTUS) and associated qualitative research

The ONS has a long experience of running Time Use Surveys across the household sector. This Review provided an opportunity to develop a consistent measure of the amount of time spent on different activities by public sector workers through the Public Sector Time Use Survey (PSTUS). The PSTUS gathered data (in February 2024) from public sector workers living in Great Britain through an online diary tool to record their activities (both detailed work and high-level non-work activities) during a 24-hour period.

The ONS also commissioned the National Centre for Social Research (NatCen) to conduct qualitative research with a subset of the PSTUS participants to further explore views on self-perceived productivity, impact on administrative tasks on their productivity and on possible changes to improve their productivity, for example through delegation or automation.

Particular interest was placed on collecting data on administrative tasks, to understand whether these divert resources from more 'productive' tasks. These were based on the following principles:

- **subsidiarity:** the activity or process could be legally, accurately and exclusively undertaken by an alternative staff member or by an automated process without detriment to the operation of the service
- **pertinence:** the activity or process is not a key part of the professional role of the frontline worker (those who tend to work with the public directly)
- **proportionality:** the activity or process is a key part of the professional role of the frontline worker, but the frequency, length or depth of the model of delivery is disproportionate to the requirements of efficient and effective delivery of the service

For further information about the survey and detailed findings from the survey data and qualitative research, please see the Time use in the public sector, Great Britain publications [see link 25, Annex I]. The ONS will review the frequency that may be justified and beneficial for any future survey, for example by exploring changing use of AI.

5.2 Public Sector Management Practices Survey (PSMPS) and associated qualitative research

The importance of management for businesses' performance is recognised in published literature [see link 26, Annex I]. Evidence from the private sector shows that the lower the average 'management score' is, the lower the productivity level, where management is important in organisational efficiency.

The ONS measures private sector management practices through the Management and Expectations Survey (MES) [see link 27, Annex I]. A comparable measure did not exist for the UK public sector and is limited in scope even in the World Management Survey [see link 28, Annex I]. Smaller scale studies have found that better structured management practices are associated with delivery of better public service outcomes.

The ONS adapted the MES to the public sector, measuring management practices in a consistent way across the UK for the first time. A pilot survey was conducted from April to July 2024. This also established a baseline performance across public sector organisations, which could be followed up in future surveys. It would also enable the ONS to compare management practices in the public and private sector.

The survey was designed to measure management quality across organisations and 'score' them against:

- **continuous improvement:** how well organisations monitor and adapt to unexpected situations
- **key performance indicators:** how many, and how frequently they are reviewed
- **targets:** how targets are set, tracked, and reviewed
- **employment practices:** processes of promotion, management, and training of employees

Leaders of public sector organisations from central government, local government, police and fire services, education, and health and social care sectors were sampled to complete the survey on behalf of their organisation.

The ONS commissioned NatCen to conduct accompanying qualitative research to develop understanding of public sector managers' views on the types of administrative tasks carried out and their impact on productivity, and explore opportunities and barriers to innovation, including the use of automation and AI.

For further information about the survey and accompanying qualitative findings, please see the Public Sector Management Practices Survey pilot, UK: 2023 [see link 29, Annex I] publication and the Public sector managers' views on management practices, Great Britain: August to September 2024 [see link 30, Annex I] publication.

The ONS will be making the microdata available for researchers to use via the Secure Research Service and Integrated Data Service to improve potential research in understanding how management practices might impact upon the provision of public services. The ONS plans to release further analysis of the PSMPS pilot survey in Spring 2025, which would show how management practices compare in the public and private sectors. The ONS will also review the merits of future waves of this survey.

5.3 Artificial Intelligence: research by the Alan Turing Institute

AI presents opportunities to transform public services. The UK Government has identified the potential for large-scale productivity gains from the adoption of AI across the public sector, as outlined in the 2024 National Audit Office (NAO) report [see link 31, Annex I].

However, to date a unified view on the feasibility or cost of delivering these improvements has not been reached. While developing such assessment is outside the scope of this Review, both PSTUS and PSMPS data collection

offer the opportunity to generate new valuable insights on the applicability of these technologies to public sector activities, and the barriers and facilitators currently in place towards its adoption.

In partnership with the Department for Transport (DfT), the Review commissioned research working with the Alan Turing Institute (ATI). The aim was to develop an approach for identifying, prioritising and evaluating generative AI applications within DfT, while also measuring the resulting productivity gains. This would offer a valuable blueprint for adoption that can be used by public sector organisations.

The ATI conducted qualitative research with senior civil servants from across DfT to understand and map business processes that are both high-frequency and repetitive in nature. This research allowed DfT and ATI to focus on two business processes where generative AI tools could bring productivity gains in the short term:

1. Accurately answering internal HR recruitment queries.
2. Answering policy questions for policy briefs, ministerial correspondence, urgent questions from parliament or questions from journalists and the general public, and information retrieval for policy briefs.

For the first process, DfT and ATI will develop and evaluate, respectively, a Human Resources (HR) chatbot tool to automatically answer questions related to recruitment processes, based on current HR policy documents. For the second, ATI will develop a proof-of-concept of an information triage tool that can sort through large amounts of heterogenous policy documents, signposting the most relevant information to inform responses to policy questions.

ATI also carried out a mapping exercise using the PSTUS data [see link 32, Annex I] to assess the extent to which work activities are potentially amenable to generative AI adoption. ATI built on established methodologies from the field of labour economics by adapting an 'automation exposure' rubric developed by Eloundou et al. (2023) [see link 33, Annex I] to assess the potential for applying generative AI to work activities in the public sector based on their exposure to technology.

The rubric assessment was based on criteria including the activity's contents, methods and tools of work, and the capabilities of technology. Each activity was scored against 'automation exposure' to indicate the potential for generative AI application to reduce the amount of time spent on it.

5.4 Key messages from the ONS surveys, accompanying qualitative research and ATI research

The research conducted by the ONS on new pilot surveys and accompanying qualitative research, and through working with DfT and ATI has generated new and complementary insights. These findings are summarised under over-arching themes.

Barriers to improvement

Findings from the PSMPS showed that the most common perceived barriers to improving the way public sector organisations are managed were cost (58% of organisations) or there was too little time to think about or implement them (41% of organisations).

Managers from the accompanying qualitative research said that lack of capacity and financial resources were barriers to innovation, particularly in sectors with more public-facing duties such as education and healthcare. Resistance to change and unwillingness of staff to learn something new was mentioned by some managers. While organisations were open to change, innovation was challenging when there was pressure to be productive in day-to-day tasks.

Impact of administration tasks on productivity

The PSTUS and qualitative research showed that while administration tasks can be viewed as time-consuming, they are perceived to be important and often essential to be completed.

Managers (from the qualitative research) also found administration tasks to be time-consuming, whereby those in more public-facing sectors (such as health, education, fire and police) found it difficult to reserve time specifically for administration and sometimes needed to work out of hours.

The PSMPS respondents reported lack of resource (66% of organisations) an ad hoc work requests (56% of organisations) as common reasons for not being able to complete administration work on time.

Use of automation and streamlining processes to save time and make administration work more efficient was a common finding across the PSMPS survey, accompanying qualitative research and PSTUS research.

Use of AI

Findings from the PSMPS showed that 42% of public sector organisations had tested, used or planned to use AI; with common reasons for AI adoption being to automate tasks performed by labour (45% of organisations) and improve quality of processes (37% of organisations).

Accompanying qualitative research on managers and public sector workers showed that they tended to respond favourably to introducing AI in their workplace and saw its potential to free up staff time, reduce time spent on administration tasks, and improve the accuracy and speed of reporting. However, participants did caveat the need for careful implementation and oversight, particularly in more public-facing sectors (such as healthcare, education, fire and police).

The qualitative research conducted by ATI identified five key areas that could benefit from the introduction of AI to support delivery:

- **drafting correspondence:** such as emails, letters, briefings for ministerial, public or internal departmental purposes, ensuring they accurately reflect governmental stances, policies, and legislation, and are consistent with previous correspondence
- **data and information management:** gathering information and accessing data (that is often spread across different locations and in different formats) quickly, and with assurance of its accuracy
- **project and budgets assessment:** processes related to management and oversight of projects and budgetary allocations to monitor project progress
- **public relations and engagement:** drafting consultation questions, transcribing interviews or conversations and recording queries from the public or press
- **process tracking:** whilst overlapping with correspondence and data information management, additional examples included maintaining logbooks to register press questions and responses, tracking where a policy brief is along the completion pipeline and tracking recruitment process (such as application status)

ATI will publish its report in Spring 2025.

5.5 Analysing Government research and development

Key to innovation is investment in the creation of new ideas and technologies. AI is just one example, but new processes, new methodologies, and new working practices are equally relevant.

The ONS conducts an annual survey of Government Expenditure on Research and Development (GOVERD) which collects information on how much the UK government spends on Research and Development (R&D). Not all UK government R&D expenditure directly relates to public service improvement as the UK government funds projects undertaken by the private sector. Equally, other types of capital investment (for example more advanced scanners in hospitals) will be equally important.

The Review compared two sources of UK Government expenditure on R&D: HM Treasury Online System for Central Accounting and Reporting data on expenditure, and the ONS GOVERD survey. This revealed that while both data sources are valuable, they serve distinct purposes and user needs. The first focuses on budgetary data, making it essential for fiscal policy and expenditure monitoring, whereas the ONS survey provides detailed survey data on R&D activities, meeting international reporting requirements.

The initial comparison highlighted that these sources are unlikely to consolidate or substitute each other because of their different definitions and purposes. The ONS will, however, continue to monitor whether the guidance, definitions and purposes of these data converge in future, which may enable the consolidation of the datasets.

> **Recommendation 21:** The ONS should keep under review whether there is convergence of the HM Treasury expenditure data and the ONS Government Expenditure on Research and Development Survey estimates to allow future consolidation of the two data sources.

CHAPTER 6

The relationship between public service productivity and other statistical systems and outputs

This Review has been conducted within the wider context of national and international developments influencing the measurement of public services productivity (PSP). This chapter outlines how the Office for National Statistics (ONS) can integrate the findings of this Review, in terms of methods improvements, with two parallel streams of activity:

- firstly, the implementation of the System of National Accounts (SNA) 2025 [see link 3, Annex I] will provide the conceptual framework to take forward these recommendations
- secondly, the Review has undertaken its work in full awareness of the forthcoming United Nations (UN) led review of the Classification of Functions of Government (COFOG) [see link 4, Annex I], which defines the service areas which countries use to produce these statistics. Changes to the COFOG may impact the scope of some of the methods proposals contained in this Review

The Review has also become aware of considerable local activity within various services being undertaken by different branches of government. The latter often analyse at a finer degree of disaggregation, with focus on the performance of individual units within their system, rather than, as the ONS does, comparing between public services. Coherence with these local metrics is key to delivering a robust and useable set of data.

6.1 System of National Accounts and Gross Domestic Product

The UK National Accounts is the pre-eminent source of macroeconomic data on the UK economy, including headline indicators such as Gross Domestic Product (GDP), household consumption and business investment. As public services make up around 20% of total GDP, their measurement is a vital contributor.

There are two key distinctions between the methods used in national accounts compilation, and parallel measurement of PSP:

- Some measures of input and direct measures of output are currently out of alignment. This is because PSP is a self-contained statistic able to innovate at a faster pace; the national accounts has a time lag in absorbing some of these data because of its more comprehensive and structured framework. Also, national accounts reports primarily at a higher level of aggregation,

therefore prioritising data which are required at this level, whereas PSP estimates are generally more disaggregated.

- The UK National Accounts currently excludes, by convention, quality adjustments because of the use of the existing national accounts frameworks in the UK. The concept of quality adjustments are included in the SNA 2008 and 2025. However, in July 2023 the National Statistician [see link 34, Annex I] agreed the National Statistician's Committee for Advice on Standards for Economic Statistics (NSCASE) recommendation that quality adjustments should be incorporated into the UK National Accounts, which presents the opportunity to close this gap.

National accounts data are used to understand what is changing domestically over time within an economy, but these data are also used for international comparisons. Both these purposes require consistency of methods to allow valid comparisons to be drawn. To support this, since 1948 the international statistical community has via the UN, agreed a statement of principles and methods which all countries align to (the SNA). The SNA is revised periodically and the latest version was published in 2008 [see link 11, Annex I]; the next is scheduled for sign-off by the UN in March 2025. SNA 2008 (paragraph 15.122) explicitly permits quality adjustment of the volume metrics (not the current price values) of public services.

However, through its membership of the European Union (EU), the UK was required to align to a separate standard: the European System of Accounts (ESA). This was revised in 2010 and in most areas aligns to the SNA 2008. The biggest difference in the context of this Review is in the treatment of public service quality adjustments: where SNA 2008 permits these, ESA 2010 forbids them.

This is because of the legislative and financial requirements of ESA 2010 and the ways contributions by Member States to the EU budget are set as a fraction of Gross National Income (GNI) in current prices. The methods to calculate GNI are defined within ESA 2010, and included as a regulation within European law, to ensure all countries pay an equitable share of the costs of the EU.

The ESA 2010 position, that quality adjustment should be excluded to prevent inconsistency amongst EU Member States given that standard methods do not exist, is one which the UK has long disputed. Whilst the UK has formally exited the EU there are two reasons the UK has not migrated away from applying ESA 2010. The first is under the terms of the various exit and trade agreements between the UK and the EU there is a continued requirement for coherent data for an agreed period of time post-exit. The second is, knowing that SNA is planned for revision in 2025, it was sensible to make one change to the latest framework.

Following a period of testing and quality assurance through application into PSP estimates, the methods outlined in this Report will be implemented through the usual ONS UK National Accounts 'Blue Book' challenge processes to ensure appropriate application into the national accounts in line with other changes

to conform to the new 2025 SNA. Pending funding decisions, the Review anticipates this will occur in alignment with other SNA 2025 related changes, although implementation may be phased. Necessary changes to better align public sector stocks and flows data in the national accounts should be prioritised.

The Review is aware there are a number of recommendations in the SNA update, which will interact with elements of the Review when the process to take these improvements into the UK National Accounts is completed. The most important of these applies a rate of return to public sector capital. This will change input values included in the productivity calculations described in this Review. This should be viewed separately from the work described throughout this Review to augment output measures with quality adjustments, or the recommendations included in this Report which consider improvements to input measures on the current (SNA 2008 and ESA 2010) basis. The ONS should review how these effects interact as part of implementing the package of changes to deliver SNA 2025, once further international guidance on the exact methods to implement is made available.

> **Recommendation 22:** The ONS should incorporate the methods and data developments from this Review into the UK National Accounts as part of its implementation of the 2025 System of National Accounts revision, in line with decisions made by the National Statistician and the National Account's standard procedures.

In preparation, the Review has developed a high level 'roadmap' to incorporate quality adjustments into the UK National Accounts. This has been compiled as part of the ONS planning for implementation of SNA25 more widely. This roadmap will need to be maintained and updated in alignment with the final SNA25 implementation plans. The roadmap can be seen at Annex B.

> **Recommendation 23:** The high level 'roadmap' developed by this Review for incorporation of quality adjustments, and improvements to public sector output in the UK National Accounts should be reviewed and maintained by the ONS as part of wider strategic planning of the implementation of the System of National Accounts 2025 develops.

6.2 The Classification of Functions of Government (COFOG) update

Alongside the SNA, the other major instrument the international community uses to achieve consistency of data is to agree a classification structure for public service activity. This is called the Classification of Functions and Government (COFOG) [see link 4, Annex I], and is maintained by the United Nations.

The international community, led by the UN, has commenced a review of COFOG to ensure it remains fit for purpose in measuring government expenditure in the most useful way possible and to a lesser extent, assure alignment to the new SNA.

The review of the COFOG structure is outside the remit and life of this Review. However, the opportunity presented to influence future improvement of these classifications is significant. This Review has therefore made a substantive submission to the COFOG Review, included as Annex C of this report. The Review's observations:

- the existing structure is insufficiently detailed in some areas making measurement challenging. For example, defence as a large proportion of spending could benefit from being sub-divided into land, sea and air capabilities to provide a similar level of granularity to other services.
- the structure is outdated now that the environment, climate change, and Net Zero are of far more policy relevance. Services targeted at addressing these are located in numerous COFOG classifications, with several, such as forestry, being recognised for their economic contribution (selling timber) rather than their environmental contribution (sequestrating carbon). This Review has undertaken research into how to best reflect which functions should be brought together into a single clearly defined environmental service which would allow spending on this area to be better understood.
- some large services are grouped together in ways which simplify the system but may hinder analysis. For example, police and immigration services are merged into a single COFOG with the criminal justice system. In some countries this may be appropriate, but in the UK context, the Review notes the substantial benefits from separately identifying and analysing each of these.

Countries' perspectives may differ on how much detail is desirable or feasible. The UK is in the lead internationally in the measurement challenge associated with PSP: the Review cannot identify another G7 country which has reviewed this area of statistics equivalent to this Review, or the Atkinson Review. Therefore, it is clear the ONS is seeking more data than its peers. However, a number of countries have contacted the ONS to discuss their parallel work in this field, and government efficiency and productivity remains hotly debated. The Review encourages other countries to proactively engage with the COFOG Review.

Changes caused by amendments to the COFOG structure would need to be implemented not just by ONS but also HM Treasury and the relevant spending departments in how they collect and submit data to HM Treasury's Online System for Central Accounting and Reporting (OSCAR) system. Changes may create discontinuities with historical data, and resources to refresh the relevant IT systems to meet their new requirements, and ensure comparability of data through time would need to be prioritised.

Data published directly by departments may in some instances allow for more granular breakdowns than those in the COFOG structure, which may allow for more detailed analysis. As long as it is possible to aggregate data to the agreed COFOG categories there is nothing to prevent the ONS operating a more detailed system.

Recommendation 24: The ONS should continue to influence the United Nations Classification of Functions of Government Review to maximise the opportunity for improved future categorisation of departmental expenditure (underpinning inputs measurement) via the HM Treasury Online System for Central Accounting and Reporting system.

Recommendation 25: The ONS should work with HM Treasury to plan for future upgrading of the Online System for Central Accounting and Reporting system to enable implementation of the revised Classification of Functions of Government.

6.3 Coherence of local and national estimates

The Review has identified where some parts of public services have been developing frameworks to compare the productivity of their individual institutions. This has yielded opportunities to work with and share knowledge on methods development as well as, importantly, creating coherence in estimates in future. Where differences are necessary between the ONS official estimates of PSP and those of other organisations, this work enables transparent narrative for users.

Recommendation 26: The ONS should continue to work with other organisations undertaking development activity in the measurement of public service productivity for mutual knowledge sharing, adoption of coherence between estimates where possible, and transparency around differences in measurement.

National Health Service England

The National Health Service (NHS) England is undertaking a programme developing new measures of productivity for the acute and non-acute care sectors. These will be produced more frequently and with less of a time lag than existing measures and be produced at an organisational level (such as for NHS trusts and Integrated Care Systems), enabling localised comparisons of productivity performance.

The ONS is one of several bodies engaged with this project as part of a methodological advice and oversight group; the mutual knowledge sharing has informed development work the ONS has undertaken as part of this Review. NHS England is also developing supporting metrics associated with productivity. The objective of the new productivity measures and associated metrics is to enable the NHS to identify, share, implement and adapt the best policies for improving productivity more quickly.

There are differences in the various measures produced by different bodies, for example, NHS England produce data on an England basis whilst the ONS produces UK statistics, and there can be differences in time periods being compared. Being able to explain these differences is vital to ensuring the useability of different estimates of similar concepts.

University of York

The Centre for Health Economics (CHE) at the University of York measures healthcare productivity in their long-running series, 'The Productivity of the English NHS' [see link 35, Annex I]. This measure was developed in partnership with the Atkinson Review [see link 1, Annex I]. CHE use volumetric measures of inputs and output, with output adjusted for quality. As a result of CHE's engagement with the Atkinson Review, there is a degree of commonality between CHE and the ONS measures for England. In particular, the most substantial part of the quality adjustment used for healthcare output is common to both the CHE and the ONS measures, as it relies on Hospital Episode Statistics data processed by CHE.

However, there are a number of key differences. While the CHE measure focusses specifically on NHS productivity, the ONS measure is defined according to all general government final consumption expenditure defined as healthcare under the COFOG definitions. As a result, the ONS measure includes additional services, including public health services and NHS-funded privately-provided care (although the latter of these is measured on an 'inputs = outputs' basis because of the paucity of data on private providers). This difference resulted in a substantial difference during the coronavirus pandemic, where COVID testing, tracing and vaccination services were included in the ONS measure but not the CHE measure.

Other differences between the two measures include:

- different sources and methods used in the calculation of output, particularly Hospital and Community Health Services
- different data sources used in labour, intermediate consumption and capital inputs
- differences in the quality adjustment, particularly for primary care and out-of-hospital services

Centre for Police Productivity

Measuring productivity in policing is seen as vital to both restoring public confidence in the service and unlocking investment. The independent Policing Productivity Review (November 2023) [see link 36, Annex I], identified 26 cross-cutting recommendations with considerable scope to improve productivity in policing. It suggested that up to 15 million hours of police time could be saved through technologies such as redaction, robotic process automation and video response, and a further 3.4 million by overhauling policing's outdated and siloed approach to sharing and analysing data.

In response the Home Office established a new Centre for Police Productivity (CPP), based in the College of Policing, to take forward these recommendations. It was also announced that, as part of the Centre, a National Policing Data Hub would be established allowing policing and government to access the data required to maximise improvements to police productivity and the further use of technology, including Artificial Intelligence.

The Centre's aim is to support forces to harness innovations that will free up significant officer and staff hours, that can be redirected to the high impact activities that drive performance. The ONS should continue to work with the centre and other relevant policing bodies on development of metrics around police activity.

Local Government measures

The departments of state which have responsibility for local government have, for a prolonged period, worked to understand and estimate the productivity of the various local government functions.

The majority of these, by volume, are already accounted for in the services that the ONS currently measure, specifically Education, Adult Social Care and Children's Social Care. These are discussed in detail in the relevant chapters.

In addition, there are a number of smaller services, specifically waste management and environmental services, highways and planning and local democracy that are currently part of the "other" group. The Review, because of resource constraints, has been unable to prioritise these. The opportunity for further work and associated recommendations can be seen in Chapter 16.

The productivity of ONS

A simple measure of the productivity of the ONS was developed during the coronavirus pandemic to assess how productivity changed in response to the pandemic. It defines input as the number of full-time equivalent staff on the payroll, and output as the number of statistical bulletins and articles published, each quarter. Estimates have been produced several times since its development, to feed into financial reporting and planning.

Significant increases in expenditure related to the Census in 2020 meant that inclusion of expenditure on goods and services and the consumption of fixed capital, to provide a more comprehensive measure of inputs, was not fit for purpose. Therefore, the current measure of the productivity of the ONS considers labour productivity only.

There are limitations to the existing estimates: only a subset of the ONS outputs is included (additional published outputs, such as datasets, and any non-published outputs are excluded), and the quality of outputs is not measured. Having considered the feasibility and impact of possible improvements to the methodology, the potential improvements most worth pursuing are:

- expanding the measure of published outputs to include other types of publication beyond statistical bulletins and articles
- developing a quality adjustment based on the number of major statistical errors

Given that the ONS is relatively small in public sector terms, the Review has not prioritised assessing feasibility in depth. However, this could be explored further. Although limitations of the methodology would remain, it is considered that it would provide a measure that is adequate to feed into overall estimates of total public service productivity.

> **Recommendation 27:** Although not a high priority, the ONS should explore the feasibility of improvements to the measure of the productivity of the ONS, to provide a measure that is adequate to feed into the estimates of total public service productivity.

CHAPTER 7

Healthcare

Healthcare is the largest public service by spend, representing around 41.5% of total public service expenditure in 2021. It has consistently been prioritised for measurement development since the Atkinson Review [see link 1, Annex I] in 2005.

By the time of the Bean Review (2016) [see link 2, Annex I], methods were sufficiently well advanced for standalone National Statistics to be published, on the usual calendar year basis. This was further enhanced in 2018 with the addition of England specific financial year statistics to meet the needs of the Department of Health and Social Care (DHSC) and National Health Service (NHS) England. These statistics have always made significant use of measures produced by the University of York. Detailed information on the methodology can be found in Public service productivity estimates: healthcare quality and methodology information (QMI) [see link 37, Annex I].

7.1 Core methodological and data challenges identified

The coronavirus pandemic (COVID-19) introduced significant challenges for measurement of healthcare output and hence productivity, not least the emergence of new activities established to manage and mitigate its impact. Furthermore, since the start of the pandemic, notable changes have been made to the delivery of many services, such as 'virtual wards' where patients receive care at home that traditionally would have been provided in hospital. Such changes present substantial challenges to the measurement of productivity as little data are available to assess the relative value of services provided by traditional and new modes of delivery. From 2020, the Office for National Statistics (ONS) captured the volume of coronavirus related testing, tracing and vaccination services provided, both in the UK National Accounts and the public service productivity estimates. These represented a sizeable contribution to public service healthcare output between 2020 and 2022.

Despite already being at a high level of quality and coverage, and improvements made during the coronavirus pandemic (COVID-19), there remain measurement challenges for healthcare output around three main issues:

1. data quality and coverage,
2. the suitability of cost weights for assigning relative values to different services and
3. the treatment of preventative services

7.2 Data quality and coverage

While the data used for measuring much of healthcare output draws on the most detailed data used for any public service, the availability and detail varies by sub-sector. There are currently large gaps in activity data for NHS-funded services contracted from the independent sector, public health services and substantial portions of healthcare output in the devolved governments.

In addition, even more detailed data are available. NHS England's Hospital Episode Statistics (HES) provide person-level data, which are used by the University of York in their healthcare output measure. These data enable output to be measured using Continuous Inpatient Spells (CIPS). This is a broader activity measure than the Finished Consultant Episodes (FCEs) used by the ONS. The different data used mean differences are expected between ONS's output and quality adjustment estimates and those produced by the University of York.

For FCEs, additional activity is recorded when patients move between consultants during a hospital stay. In contrast, with CIPS, one hospital stay accounts for one activity regardless of any moves between consultants, thus preventing increases in the transfers of patients between consultants from affecting output growth. While the challenges of acquiring and processing HES data would require more time than enabled by the Review, the ONS will consider the use of these data in future.

> **Recommendation 28**: The ONS should evaluate the benefits and costs of switching from Finished Consultant Episodes to the person-level data provided by the Hospital Episode Statistics for measuring hospital output in England.

The ONS relies on data produced by the University of York, funded by the National Institute of Health and Care Research, to produce the quality adjustment for healthcare services. If quality adjustment is to be included in the UK National Accounts to comply with the System of National Accounts 2025 (SNA25) [see link 3, Annex I] (see Chapter 6), the ONS needs to ensure these priority statistics can be published with minimal risks beyond its control.

> **Recommendation 29:** The ONS should review, together with the Department of Health and Social Care, how the data needed for the quality adjustment for healthcare services should be produced or commissioned in the future.

The suitability of cost-weights

Cost weights are standardly used in public service output measures to assign relative values to different activities. While costs can serve as a proxy for value, using cost weights within a productivity measure can result in that measure missing a key source of productivity growth – the replacement of higher-cost service designs with lower-cost service designs which deliver the same outcome but at lower cost.

A hospital moving to treating non-complex knee replacements as day surgery, whilst its neighbour retains an in-patient delivery model for cases of the same complexity would, if cost weighted at the lower price, appear to have reduced its output, even if it delivered the same volume of knee replacements. This is obviously erroneous. Within healthcare, there are a range of services where the NHS has sought to deliver equivalent outcomes for patients by employing lower-cost forms of healthcare provision. Such service changes are seen as an important contributor to improvements in productivity.

Treatment of preventative services

Preventative services merit special consideration as the value of the output they generate is not in the number or cost of patients treated, but in the number or cost of patients who avoid the need for treatment in future. A naive application of cost weights to preventative services, which typically have relatively low costs and hence low weights, therefore fails to reflect the relative value being generated by such services and changes in this over time, as discussed in Chapter 3.

7.3 Improvements to inputs estimates

A review of healthcare inputs, specifically the value of goods and services used in the provision of healthcare activities (intermediate consumption), was prioritised in Year One of the Review. This found that the ONS was largely using the best available existing data, and therefore only small improvements were required, such as the inclusion of legal and audit services. These were incorporated into published statistics from Spring 2024. The weights used to calculate labour inputs were also updated to better align with our measure of full-time equivalent (FTE) staff. The overall impact of these changes was minimal and there appears little scope to deliver further advances, at least in England, at this time.

7.4 Improvements to outputs estimates

The ONS has, following the advice of the Review, made improvements to the coverage of preventative healthcare activities by reviewing the extent to which local authority commissioned public health activities are captured in its output measure. Many local authority funded services were already included, as a wide

range of services are commissioned to NHS providers. However, some services are also commissioned to a wide range of alternative providers. Treatments for alcohol misuse, drug misuse, and smoking cessation services commissioned to non-NHS providers were added to the ONS output measure from Spring 2024. However, the ONS has not yet been able to apply the alternative weighting methods proposed in the Treatment of preventive services section of Chapter 3. Given the growing importance of preventative services the Review considers this should remain a priority for future work.

The ONS remains in conversation with the devolved governments around improving coverage of their data (as per Recommendation 14). There also remains a lack of data on NHS-funded services contracted from the independent sector and this limitation is acknowledged within the healthcare sector.

> **Recommendation 30:** The ONS should continue to work with the NHS to improve data on NHS-funded services contracted from the independent sector.

From the Spring 2025 annual release the ONS will enhance the healthcare output measure to incorporate certain screening services, improved measures for primary care, equivalised unit costs for equivalent treatments across different modes of provision, and removal of excess bed days activity.

Screening services

Historically, abdominal aortic aneurysm, bowel cancer, breast, and cervical screening services have been excluded because of data limitations. These can now be included because of improved expenditure data availability from the financial year ending 2023 National Cost Collection (NCC) [see link 38, Annex I] and long-term activity data from NHS England and DHSC.

The ONS will use expenditure proportions from the financial year ending 2023 NCC as a baseline and activity growth data from NHS England and DHSC to establish a reliable back series from financial year ending 2015. This method accounts for both the relative cost differences between screening programmes and the annual shifts in activity. This provides a more accurate reflection of change over time than other methods considered (applying the financial year ending 2023 unit costs to each year in the back series, or assuming the cost of one screening programme relative to the others in financial year ending 2023 has remained consistent over time).

There are currently quality issues with the NCC activity data, therefore the ONS is only using this source for expenditure and using other sources for activity. When the quality of the NCC activity data improves, the ONS intends to make this the sole source to measure screening services, streamlining the process, and reducing reliance on other datasets.

These changes are being applied to England only initially, as regular activity and expenditure data for these services for Wales, Scotland and Northern Ireland are not currently available to the ONS. The ONS will continue to work with the devolved health administrations to assess the feasibility of incorporating these adjustments for the other UK nations in the future. However, the inclusion of these services for England increases coverage of healthcare services, ensuring a more comprehensive and accurate reflection of healthcare activity and expenditure in the UK productivity measures.

Recommendation 31: The ONS should implement abdominal aortic aneurysm , bowel cancer, breast and cervical screening services within healthcare outputs in Spring 2025.

Recommendation 32: The ONS should monitor the quality of National Cost Collection activity data for England for abdominal aortic aneurysm, bowel cancer, breast, and cervical screening services, and transition to using these data when they are of adequate quality to bring it in-line with expenditure data used.

Recommendation 33: The ONS and the Department of Health and Social Care should explore sourcing data for other preventative activities, such as glaucoma screening for inclusion when data permits.

Recommendation 34: The ONS should continue to engage with Wales, Scotland and Northern Ireland to assess the feasibility of including abdominal aortic aneurysm, bowel cancer, breast, and cervical screening services in the healthcare output measures for the devolved governments in the future.

Primary care

The Review considers that the consistency of the dental activity time series can be enhanced by using an alternative source of data to estimate activity growth from 1996 until 2006, and an improved method of linking different activity data sources. The Review also found that the cost weighting for ophthalmic services and NHS telephone and website services can be improved, which means these services are weighted more appropriately relative to other Healthcare outputs.

For ophthalmic services this can be achieved by ensuring the overall weight of the component is consistent with expenditure as reported in the DHSC annual accounts. For NHS telephone and website services this can be achieved by uprating the historical unit costs to account for general inflation in NHS costs.

> **Recommendation 35:** The ONS should use the Department of Health and Social Care Index of Services data to estimate dental activity growth from 1996 until 2006 and improve the method of linking this activity data to the current 'Units of Dental Activity' data source to avoid a discontinuity between data sources.

To better represent activity undertaken over the pandemic period, for NHS dental services, the ONS will reduce the size of the expenditure weight used to calculate the contribution of dental output in financial year ending 2022. Rather than the standard approach of taking net dental expenditure as the dentistry weight, the ONS will uprate the financial year ending 2020 unit costs by a growth factor equivalent to growth in other healthcare services between financial year ending 2020 and financial year ending 2021. This approach ensures that the resulting implied expenditure weight for dentistry is more reflective of the cost of the activities performed by dentists over the year, rather than the full contract value. This approach is more consistent with the cost weighting applied to other healthcare services and ensures the weight given to the fall in output from dentistry in financial year ending 2021 and recovery in financial year ending 2022 do not result in an unrepresentative contribution to growth from dentistry.

> **Recommendation 36:** The ONS should ensure the overall weight of ophthalmic services and dental services is consistent with expenditure as reported in the Department of Health and Social Care annual accounts, and uprate historical NHS telephone and website services unit costs to account for NHS cost inflation.

Equivalisation of weights for (i) acute care

As discussed in Chapter 3, in the absence of prices, cost weights are typically used as a next best alternative to approximate the value of different services. The granular data available for healthcare activity and costs enables a high degree of differentiation in value 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 an outpatient procedure. Historically, separate unit costs have been used for each procedure type. Therefore, where more procedures transition from overnight hospital stays to same-day treatments and costs fall, this results in more lower-weighted activity and so appears as a reduction in output, even though the same care is being provided more efficiently.

The ONS has developed equivalised 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 equivalisation. If a HRG exists in more than one of those components, an equivalised weight will be applied. For outpatient procedures, unit costs are only equivalised where the HRGs tariff (the price paid by commissioners under the NHS Payment Scheme) is equal to that of elective inpatient and day cases. This follows the approach used by NHS England in their new productivity measure.

> **Recommendation 37:** The ONS should implement equivalised weights in Spring 2025 where service weights are equivalised to account in productivity for cost-savings from moving services to lower-cost modes of provision, these should be incorporated in the non-quality adjustment measures. Quality adjustment should account for changes in value of services delivered that goes beyond cost-saving, such as improvements to the estimated health improvement from treatment.

Equivalisation of weights for (ii) ambulance services

There is a similar issue with the reduction of avoidable ambulance service activity through increasing provision of medical advice by telephone and care in situ by ambulance crews without conveying patients to hospital. This approach improves patient outcomes and reduces unnecessary hospital admissions, easing pressure on emergency care services. Over time, this should lead to a decline in the more expensive 'See, treat and convey' service and an increase in the lower cost modes of ambulance service delivery. Further details on reducing avoidable

conveyance can be found in Planning to Safely Reduce Avoidable Conveyance [see link 39, Annex I]. This aligns with the NHS Long Term Plan [see link 40, Annex I] which introduces new standards to ensure sickest patients receive the fastest possible response, while also ensuring that all patients receive appropriate care the first time, within a clinically suitable timeframe.

The ONS has addressed this issue for ambulance services by calculating an average unit cost across the following components of treatment:

- hear and treat or refer – when a person does not require an ambulance, but a clinician is able to provide treatment and advice over the phone
- see and treat or refer – when a person does not require hospital care but instead a paramedic or another clinician provides treatment at the scene
- see, treat and convey – when a clinician or paramedic assesses and provides treatment at the scene and determines that the patient requires further care, resulting in the patient being transported to a hospital or other healthcare facility for additional treatment

As a result, a shift in the composition of ambulance activity towards more calls being responded to with advice or care in-place without conveyance will not result in a reduction in output. A shift is likely to lead to a reduction in inputs, hence productivity would improve. Emergency 999 calls and other ambulance services, such as those referred from NHS 111, are not included in the equivalisation because the ONS intends to retain the difference in the weight of these services as they are not necessarily alternative services to ambulance conveyances.

> **Recommendation 38:** The ONS should implement equivalisation of weights for ambulance services in Spring 2025.

Equivalisation of weights for (iii) the Devolved Governments

These changes have currently only been applied to England's healthcare output. In Wales and Northern Ireland, inpatient and day case activity are already aggregated and equivalised in an equivalent manner to methodological developments applied to England in the source data providedby these administrations. For Scotland, inpatient and day case equivalisation has not yet been implemented, but the feasibility of implementation can be evaluated in future.

The equivalisation of outpatient procedures and ambulance services are not currently applicable to the devolved governments because of differences in data reporting and availability. The ONS will continue to investigate opportunities to develop methods for Wales, Scotland, and Northern Ireland.

Recommendation 39: The ONS should continue to explore the feasibility of applying the improvements on handling equivalent treatment across different modes of provision made for England to Scotland, Wales and Northern Ireland.

Excess bed days

A further improvement developed is the removal of excess bed day activity from historic volume calculations. Excess bed days are additional hospital stays extending beyond the expected treatment period for clinical reasons e.g. complications or other medical needs. Prior to financial year ending 2019, excess bed days were treated as additional activity categories within NHS England's National Cost Collection. From financial year ending 2019 onwards, the excess bed day activity data were removed and the costs were captured within the elective and non-elective care components.

This improvement integrates expenditure from excess bed days into the respective elective and non-elective care categories and removes excess bed days from the activity measure entirely. This ensures that growth in the number of excess bed days no longer contributes to output growth during the financial year ending 2015 to financial year ending 2019, aligning with the NHS' financial year ending 2019 update.

Excess bed days are not identified as a distinct component in the data from the devolved governments. Therefore, the change the ONS is implementing for England does not need to be applied for the devolved governments.

Both the removal of excess bed days and the equalisation of weights across admitted patient care and ambulance services have been applied to the growth rate from financial year 2015 onwards, as changes in how activity was recorded in financial year ending 2014 present challenges for reprocessing earlier years. The impact is expected to be largest on earlier years, particularly financial year ending 2015, when significant shifts in care delivery methods began. The ONS has also applied minor consistency adjustments to data from financial year 2015 onwards.

Recommendation 40: The ONS should adopt the improvement to remove excess bed days in Spring 2025.

Unnecessary accident and emergency (A&E) admissions

The Review also considered a further improvement to output to account for reducing unnecessary A&E admissions by treating more patients in primary care settings being another potential source of productivity gains. A&E attendances carry a higher unit cost, giving them a higher weight in output calculations,

relative to General Practice (GP) consultations. The Review developed a method to adjust the output weight for unnecessary A&E admissions to represent that of a typical GP consultation. This approach highlights that equivalent care could be delivered through a lower cost GP consultation, while acknowledging that A&E inputs remain higher, thereby indicating lower productivity when A&E resources are used for these cases.

However, early estimation of the effect of this new method suggests the impact on output is very small, although this should be monitored. Given the additional processing burden for a minimal impact on overall output, the ONS has not implemented the new method at this time, although the ONS should be ready to return to this if it becomes more significant.

> **Recommendation 41:** The ONS should keep under review whether incorporation of an adjustment for unnecessary A&E admissions would be feasible, material and proportionate.

7.5 Improvements to quality adjustments

Indicators of quality are conceptually easier to identify for healthcare than for most other public services and a number of quality adjustments that account for patient experience and effects on health-related quality of life are already applied to the productivity estimates. However, the Review identified opportunities to improve some of the existing quality adjustments. For GP patient outcomes, the ONS measures the change in the proportion of patients managing different health conditions through a selection of quality indicators from the Quality and Outcomes Framework (QOF), a tool that helps assess the performance of general practice.

Through consultation with primary care experts at Imperial College London additional outcome metrics were identified, concentrating on indicators with a sufficient time series, or without substantial changes in definition over time. The Review also assessed the breadth of conditions covered, identifying two additional health conditions with suitable coverage to include in our measure – asthma and diabetes mellitus.

The means of weighting the significance of the indicators used to measure outcome improvements were then reviewed. Previously the change in each indicator was weighted equally; since Spring 2024 the ONS has used QOF points to weight indicators relative to one another. This represents the value given by the NHS to each indicator, meaning that indicators for activities perceived as more valuable are given a higher relative weight than lower value indicators.

The ONS' existing quality adjustment for patient experience prior to the Review included patient satisfaction measures for admitted patient care, emergency services and mental health services taken from surveys conducted by the Care Quality Commission. Patient satisfaction in primary care had been a gap in coverage since financial year ending 2008, when the National Patient Survey Programme was discontinued. The Review identified that the GP Patient Survey, which has been in operation since financial year ending 2009, can be used as an alternative data source for measuring patient experience in general practice and dentistry services.

A range of other data sources investigated to further expand patient satisfaction measures include the 'Family and Friends Test', which is used within a wide range of NHS services. However, NHS England advised against its use because of the voluntary nature of the test, likely providing biased results. The ONS will continue to monitor progress of the Care Quality Commission in rolling out patient satisfaction surveys for other healthcare services.

In 2024 the ONS was commissioned by NHS England to introduce the Health Insight Survey, with results on satisfaction and access with various NHS services published quarterly in Experiences of NHS healthcare services in England [see link 41, Annex I]. While this data source does not yet have the length of time series needed to be included in the patient satisfaction quality adjustment, the ONS will evaluate its suitability for this purpose in future.

> **Recommendation 42:** The ONS should continue to monitor the development of patient satisfaction surveys conducted by the ONS, the Care Quality Commission and others, with a view to further expanding the quality adjustment for patient experience in the future.

A quality adjustment is also applied that uses the Quality-Adjusted Life Years conceptual framework to estimate the health gain from treatment over remaining life expectancy. This has 2 dimensions: 1) health gain, with an adjustment accounting for survival rates and average health gain from the procedure, and 2) time, where the health gain is applied over discounted remaining life expectancy after accounting for waiting times to treatment. Until now, the ONS has applied this adjustment in a way which means no adjustment for the time component is applied to non-elective treatments because waiting times are not relevant to these treatments. The Review has identified that the ONS can improve the quality adjustment by assigning a time quality adjustment factor for non-elective treatments based on just the life expectancy. This is consistent with the approach taken by the University of York and allows for any changes in life expectancy in non-elective treatments to be accounted for in the quality adjustment measure. This improvement has been implemented back to financial year ending 2015 for consistency with other developments implemented as part of the Review.

Recommendation 43: The ONS should extend the application of the health gain quality adjustment, excluding the waiting times component, from elective to non-elective procedures in Spring 2025.

7.6 Recommendations for further work

Substitutions of higher-cost services for lower-cost services

As already described, the Review has identified a method for ensuring that efficiency gains from process improvements that lead to lower-cost service delivery methods are represented in the productivity measure. While the Review has implemented methods to account for these productivity gains in acute care and ambulance services, new ways of delivering equivalent outcomes for patients by employing lower-cost forms of healthcare provision continue to emerge. In particular, the planned moves from hospital to community services and greater use of digital services could have large implications for productivity. Identifying these substitutions in the future and addressing them in the productivity measure would help to ensure that important contributors to improvements in productivity are captured.

Recommendation 44: The ONS should continue to work with the Department of Health and Social Care and the NHS to identify additional cases of lower-cost services being substituted for higher-cost services.

Improving quality adjustment of waiting times

The Review started to explore whether the ONS' existing quality adjustment for waiting times could be a) expanded, so that an adjustment for waiting times is applied to a wider range of healthcare services, and b) improved, by taking account of longer waiting times leading to worse health outcomes.

The current waiting times adjustment is applied to elective inpatients and day cases only and concentrates on the delay in health gain. Any change in waiting times is spread over remaining life expectancy, meaning the effect is trivial. Accounting for changes in waiting times for a wider range of services and adding a further adjustment to account for longer waiting times leading to worse health outcomes for patients would mean that a change in waiting times would carry more weight in the productivity estimates.

A literature review identified a link between delays and increased risk of death for some treatments, but a lack of evidence exists to suggest this approach is generalisable.

Therefore, further consideration of this approach is needed before any changes can be implemented. Some data sources have been identified that could potentially be used to create a waiting times quality adjustment where the ONS does not currently have one, however, further research is needed to consider how the adjustment would be weighted and applied for different service types.

> **Recommendation 45:** Further research should be conducted to continue to explore the feasibility of improving and expanding the existing Healthcare quality adjustment for waiting times.

Preventative services – a case study

As discussed in Chapter 3, preventative services, such as public health interventions aimed at reducing the incidence of disease, merit special consideration in the measurement of public service output as the public value they generate is typically disconnected from the cost weighted bundle of activities delivered. Preventative services instead reduce the volume of cost weighted activities demanded by the public in future, such as reducing the need for surgery, while also leading to improved public welfare.

As part of the Review the ONS commissioned Weale (2024) to consider potential methodologies for measuring preventative services in public service output and productivity, using the Diabetes Prevention Programme as a case study.

Further practical challenges to implementation include the lack of time-variant data on the outcomes of preventative programmes and the wide variety of such schemes, with each having only a marginal impact relative to healthcare overall. The ONS intends to further investigate the feasibility of the approach proposed by Weale (2024) to scope other preventive programmes where the methodology could be applied, and determine if the challenges can be sufficiently addressed to merit the implementation of such a bespoke approach to the output of preventative services.

> **Recommendation 46:** The ONS should explore how to improve the measurement of preventative services in Healthcare output, including commissioning a literature review of data sources to ensure consistent application across the range of preventative treatments.

NHS Payment Scheme

NHS trusts are funded through payment for activity based on the cost of delivering different activities. However, for some activity, NHS England modifies the costs used in the payment scheme to incentivise trusts to adopt improvements in clinical practice. This may be through improving outcomes or adopting new practices or technology to deliver similar services at lower cost.

The ONS has started to explore the feasibility of adjusting the cost weighting of certain HRGs within the output measure to incorporate these incentive-based modifications for acute care. The ONS found that while it may be possible to use the NHS Payment Scheme weights for acute care, there remain significant difficulties with implementing this approach into healthcare output.

In particular, the matching of the NHS Payment Scheme data with the NCC data is challenging, particularly prior to financial year ending 2018, because of many HRGs in the NCC data not having equivalent categories in the NHS Payment Scheme data. Another issue is that acute trusts do not only receive funding for acute care through the NHS Payment Scheme, which means they have additional costs which would still need to be accounted for when weighting their care against other services which are cost weighted.

> **Recommendation 47:** The ONS should continue to investigate whether aspects of the incentives in the NHS Payment Scheme can be incorporated in the relative weighting of different services in Healthcare productivity.

Coherence of the ONS and NHS England healthcare productivity estimates

In developing this report, the Review has had the opportunity to observe the reaction of stakeholders to the data which has been published during the life of the Review, especially where new data have been made available. In February 2025 the ONS published new quarterly estimates of healthcare productivity, which differed to those published concurrently by NHS England. These differences are explained by:

a. the ONS measuring the UK whilst NHS England measures England,

b. the ONS measuring acute, GP, and community services, including COVID-19 vaccination services and NHS England only measuring acute services,

c. the ONS comparing quarter with the same quarter the previous calendar year whilst NHS England compares the first seven months of the financial year,

d. the ONS applying seasonal adjustment while NHS England do not, and

e. minor methodological and data differences

Nevertheless, the Review noted that this debate allowed the relative productivity performance of the hospital sector versus other parts of the healthcare system to be exposed as a potentially important area of investigation. It suggests that where healthcare accounts for 41.5% of public services, to report this as one service whilst services such as tax and social security administration (which account for less than 2% each), are given equal treatment, may hinder user understanding of the key drivers of productivity growth. Taking this example, where NHS England are able to produce statistics coherent with the ONS which allow greater granularity to be shared with users, the ONS should look to develop a more granular breakdown of the healthcare service to be included in its statistics.

> **Recommendation 48:** The ONS should explore approaches to disaggregate the Healthcare service, to allow the relative performance of different components of this large service to be better understood.

In addition, during the completion of the Review, the potential to explore working with NHS Wales to explore its datasets to augment the measures of Wales for Healthcare services presented itself. This occurred too late in the process to be taken forward as part of the Review itself, but is an opportunity the ONS should take forward.

> **Recommendation 49:** The ONS should actively explore working with NHS Wales to access its health data and develop stronger productivity metrics for Wales.

CHAPTER 8

Education

Education is the second largest individual area in Government spending terms and was therefore a priority for Atkinson (2005) [see link 1, Annex I] and subsequent Office for National Statistics (ONS) research. Along with Healthcare, these were the only two services with quality adjustments by 2016. This, however, masked several known deficiencies in the measure, leading to its lack of use in policy discussion and subsequent suspension of the publication of a standalone education National Statistics bulletin.

Post the Bean Review (2016) [see link 2, Annex I], improvement work was prioritised ultimately addressing one of the largest gaps in the application of Atkinson's methods, namely the apportionment of attainment at GCSEs across the full educational experience of the student. This was considered a significant improvement until the advent of the coronavirus pandemic, when the model apportioned some of the learning loss exhibited during the pandemic to previous years. This, and discontinuities in the sitting and marking of public examinations, forced the suspension of parts of the method from 2019.

8.1 Scope and higher education

The Education productivity measure includes education offered from the ages of 0 to 18, through Early Years, Primary, Secondary, and Further Education. All Special Needs services are included within this, but importantly adult learning and Higher Education (HE), which covers university and those parts of colleges offering degree qualifications, is excluded.

The exclusion of HE is because there is a 'market price': there is a tuition fee charged to students to attend these courses. In the context of national accounts, alternative methods are applied to public services solely because no such market price is available (see Chapter 3). Given there is such a price, there is no need to consider alternative methods. Existing methods of ensuring prices are on a like-for-like basis should suffice.

However, whilst the Review excludes HE, it has observed there are two prices for HE: a regulated price set by government for domestic students and a competitively set price charged to international students. Given there is no assumed difference in the teaching output experienced by these two groups, how this is tackled is a question which can be considered by the ONS elsewhere, but the Review would encourage any consideration of this question to reflect the principles underpinning Atkinson and its work.

8.2 Core methodological and data challenges identified

The challenges in measuring education productivity are the change in the coverage of the service over time, the definition of the quality of the service, and the differences in the education system across the UK nations. In particular: capturing the expansion of compulsory education to age 18 in some nations, the increase in academisation of schools in England, and updating the treatment of healthcare and teacher training to reflect changes in provision.

The ONS uses several sources to create inputs, outputs and quality adjustments for each nation in the UK and coverage is good. However, some data sources became unavailable during the coronavirus pandemic, particularly in relation to outputs and quality adjustments. The Review has considered proxies and methodological adjustments to avoid breaks in the series.

The Review looked at quality adjustment, which is a composite of attainment and data on the prevalence of bullying, recognising this is a 'service' which delivers multiple outcomes. Currently, attainment is the main indicator of quality and assumes that, while if students achieve higher grades, this is an increase in the volume of output just as much as a school educating more students (and achieving the same grades).

The Review explored whether attainment (and the other quality adjusted measures currently used) should be complemented with a longer view of the impact of education towards lifetime outcomes, in particular human capital acquisition. This is discussed in the further work section and in greater depth in Annex D.

The Review also explored improvements to the 'cohort split model' used to allocate attainment at age 16 across the different years that cohort were in education. This is to address the point that (usually) an improvement in attainment cannot just be attributed to better teaching in Year 11 (i.e. the final year in which GCSEs are sat), but rather are a factor of the full educational career.

As previously noted, this model was unable to cope with the impact of the coronavirus pandemic, when the impact on attainment for the cohort sitting their exams in 2020 and subsequent years should be allocated only to that year. The quality adjustment section details how the impact of the coronavirus pandemic was prevented from impacting previous years.

8.3 Areas prioritised by the Review

Within the first year of this Review the ONS implemented some improvements to output:

- better reflecting the conversion of local authority (LA)-maintained schools to academies
- capturing the increase in activity from the increase in funding entitlement for pre-primary
- updating the treatment of healthcare and teacher training to reflect changes in provision

These improvements were included in the first wave of improvements [see link 42, Annex I] carried out as part of the Review and published in the Total Public Services productivity [see link 10, Annex I] bulletin in March 2024.

In collaboration with the Department for Education (DfE), the Review identified further areas for improvement relating to quality adjustment:

- review of indicators of attainment during the coronavirus pandemic using the Office of Qualifications and Examinations Regulation (Ofqual) National Reference Test data [see link 43, Annex I]
- improvement of the adjustments applied to attainment to account for the cumulative nature of education following the coronavirus pandemic (Cohort-split model)
- inclusion of attainment as a quality measure for Further Education (FE)
- exploring if student wellbeing is a better indicator than the prevalence of bullying as a quality adjustment

8.4 Improvements to inputs estimates

The current approach of using deflated expenditure with a combination of indirect and direct labour for Education remains the preferred methodology for obtaining a comprehensive and accurate measure of education inputs. Whilst no major methodological changes have been required in this area, some minor improvements to improve data quality and increase harmonisation with other service areas are being made in the ONS Spring 2025 annual release.

Currently, the labour inputs are created based on the full-time equivalent (FTE) teacher and support staff numbers (split by school and academies for England only) and weighted together using data on salaries from the Annual Survey of Hours and Earnings (ASHE). In the past, the ONS used salaries of all workers, while the Review recommends moving to the salary of full-time workers. This change aligns with the methodology used in other services and shields the measure from any artificial growth in salary due to a change in the relative mix of full-time and part-time workers. Movement in the full-time salary reflects only actual changes in the salary bands themselves.

Further minor remedial amendments have been made within the ASHE salary indices to increase the robustness of data by applying a compound annual growth rate to impute missing values. In using the ASHE data, the Review notes the definition of some categories of staff in education has changed since 2020. This could raise some issues for comparability over time. The Review has worked to ensure the best match of these categories and the ONS will continue to review them in the following years.

> **Recommendation 50:** The ONS should implement improvements identified for Education inputs in Spring 2025 related to salaries.

> **Recommendation 51:** The ONS should continue to review if Annual Survey of Hours and Earnings remains the best data source for labour data for Education, or whether alternative sources may be preferable.

8.5 Improvements to outputs estimates

As part of this Review, developments on Education have been adopted and implemented into PSP. These developments were implemented in Public service productivity: total, UK, 2021. These developments include:

- the refinement of categories used to assign relative weights to different school types in output to reflect changes to the landscape of Education
- improvements to pre-primary inclusion, better capturing pre-primary activity
- removal of teacher and healthcare training to avoid double-counting from Further Education

Refinement of education categories for weighting

Academies, state-funded schools that are financed by central, rather than local, government and are granted a greater degree of operational independence from local authority control, were introduced in England in 2002. Initially, the academisation of schools was concentrated almost exclusively on secondary schools. However, from 2010, the process of academisation spread to primary schools and in 2011 to special schools.

In the academic year ending 2023, just over 40% of primary, 80% of secondary and just under 45% of special schools in England were academies. With this large change in the education landscape, it is important to capture the contribution of each phase of education (primary or secondary etc) to productivity as accurately as possible, such that any further changes in the landscape will give a more accurate picture of overall education productivity.

Previously, compulsory education in England was captured through three education categories: primary, secondary and special. These are now split into a further five categories, splitting out primary, secondary, and special academies from their LA-maintained counterparts, as well as splitting out alternative provision (education for students who cannot go to mainstream schools) into two separate categories (academies and LA-maintained). Updates to the expenditure weights have allowed cost weighting of each category, allowing the contribution of each to overall education productivity to be more accurately captured.

Pre-primary education

Pre-primary education was previously captured through pre-primary (all LA-maintained pre-primary schools plus pre-primary classes in primary schools) and private, voluntary and independent (PVI) pre-primary schools.

However, expenditure data for PVI pre-primaries are not available separately from LA-maintained pre-primary expenditure. Similarly, for Scotland and Wales, there are no PVI enrolment data. The scarcity of PVI data inhibited a consistent LA-maintained and PVI split in pre-primary education across the nations, so LA-maintained pre-primary schools and PVI pre-primary schools have been included as a single pre-primary figure. This allows a more accurate cost weighting for pre-primary and therefore gives a clearer contribution to productivity.

Previously, for England, those within primary schools but who are of pre-primary age were not being captured as pre-primary but rather primary pupils. All pupils under 4 years old as of 1 September on the year prior to the school census (which is taken in the January) have been considered as pre-primary, and contribute to the LA-maintained pre-primary figure. This allows a more accurate estimate of primary-age pupils.

In the academic year ending 2014, funding became available for some 2-year-olds in pre-primary education. While included in the LA-maintained pre-primary enrolment figure, these have not previously been included in the PVI pre-primary enrolment. These enrolments are now included from the academic year ending 2014 onwards.

Similarly, in 2017, the extended entitlement of 30 hours was introduced for 3 to 4 year olds in families meeting certain eligibility criteria. The ONS implemented these improvements to capture the extended childcare entitlement in output. Note that the pre-primary inclusion for Wales, Scotland and Northern Ireland will not be affected by these changes.

For enrolment, full-time equivalence (FTE) is used to calculate activity. FTE is calculated as 50% for part-time pupils and 100% for full-time pupils. This has a small impact on primary, secondary or special schools, where there are very few part-time pupils; however, it is more of a consideration for pre-primary pupils.

Where part-time pupils are not identifiable (England PVI pre-primaries, Northern Ireland PVI pre-primaries, all Scotland and Wales pre-primary), pre-primary enrolment (headcount) figures were previously given a factor of 0.5 to proxy FTE (that is, pre-primary pupils were assumed to attend school 50% of full time).

Where part-time pupils are identifiable (predominantly, the England school census and some of the Northern Ireland school census), the FTE factor is higher and growing. Therefore, where part-time pupils are not identifiable, the pre-primary figure is now given a factor equal to that of the calculated FTE factor for each year, instead of the current constant factor of 0.5. As the FTE factor is not identifiable for Wales and Scotland, the weighted average of the FTE factor for England and Northern Ireland is applied.

Teacher and healthcare training

Historically teacher and healthcare training were included within Education outputs. However, by the end of the 1990s, most teacher training colleges were absorbed into universities and have therefore been increasingly reflected within higher education measures. Similarly, when degrees became compulsory for nurses joining the NHS in 2009, the number of nurses training via non-degree routes declined. Nursing degrees, delivered through universities, are also included within the UK National Accounts through existing methods. To avoid double counting, teacher and healthcare training have been removed from Education outputs.

For more details on the effects that these changes have on output, see Improved methods for total public service productivity: total, UK, 2021 [see link 42, Annex I].

8.6 Improvements to quality adjustments

Using the National Reference Test (NRT) to inform attainment during the coronavirus pandemic

As a result of the coronavirus pandemic, exams were cancelled in schools and typical attainment data were not available. Data were either unpublished (such as the case with England Primary schools) or informed by teacher assessed grades (such as the case with England Secondary schools) which were determined not to be appropriate to inform PSP estimates due to concerns on grade inflation.

Prior to the Review, the ONS used learning loss measures published by the DfE to inform attainment from academic year ending 2020, and the first year of the Review sought to explore alternative indicators of attainment that were consistent with historical data and provided a more robust overview of academic performance during the pandemic.

The Review identified the National Reference Test (NRT) [see link 43, Annex I] as a more appropriate indicator of attainment during the coronavirus pandemic.

This was primarily because the NRT is a strong indicator of GCSE-level performance and is independent of teacher assessed grades.

Please refer to Annex D for more information.

Attainment

In the second year of the Review, the treatment of attainment was further explored. The key issues regarding improving the quality adjustment of the Education estimates relate to adapting the 'Cohort-split' model in the years of the coronavirus pandemic. The Review highlighted that the ONS needed to develop adjustments to the existing model, as it apportioned some of the learning loss exhibited during the pandemic to previous years. The key elements are:

- how to address the years prior to the coronavirus pandemic for students in school during the pandemic
- how to align cohorts in subsequent years to actual data, given the impact of the intervention described in the previous bullet
- whether in the light of the Review the weights used in the cohort-split model needed to be updated
- how to handle missing data in this new model

The adjustments made to the cohort split model were reviewed extensively in collaboration with DfE, and are described in detail in Annex D. In effect the updated model confines the effects of the coronavirus pandemic on attainment to the years directly affected by the coronavirus pandemic without apportioning any learning losses to previous years. Furthermore, the weights assigned to year groups in Secondary schools were revised so that each year group has equal weighting and only considers year groups in secondary school.

Concerning the treatment of missing data, there was a distinct lack of attainment data for primary schools across the UK for some academic years. In the absence of alternative data, the NRT was used to inform data gaps for primary schools. This was because the NRT is an indicator of academic performance, which is relevant to primary schools, and there is a concern that leaving the data gaps issue untreated could produce estimates that do not account for the broader effects of the pandemic on attainment.

> **Recommendation 52:** The ONS should implement improvements to the quality adjustment of Education to better account for the impact of the coronavirus pandemic and to account for student well-being and Further Education attainment.

Further education attainment

The Review also explored quality adjusting Further Education (FE) according to attainment at this stage of schooling. In collaboration with DfE, the Review identified the percentage of students meeting the minimum requirements for Level 2 and Level 3 qualifications by age 19 as the most appropriate and robust indicator of attainment at FE level. It must be noted that these relate to FE institutes in England only.

Level 2 and Level 3 attainment were individually processed according to the updated cohort-split model and were weighted and aggregated into an FE attainment index according to enrolment figures for each qualification.

FE marking practices were also influenced by the coronavirus pandemic, and no alternative data are available to indicate FE performance during this period, therefore the attainment indices for L2 and L3 will need to be held constant from academic year ending 2020 to academic year ending 2024 for Level 3, and academic year ending 2026 for Level 2. The NRT was not an appropriate indicator to inform data gaps for FE due to its link to academic performance whereas FE does not have a pure academic focus and has a shift to technical and vocational qualifications.

Please refer to Annex D for more information.

Student well-being

The ONS has historically used bullying rates as a component to inform quality adjustments for Education, however the Review identified that this does not capture the wider effects of school on students outside of attainment.

The Review identified that a measure of general student well-being would be more appropriate to inform quality adjustment, given the links between well-being and attainment and engagement at school. Areas such as bullying would also be accounted for under a well-being measure.

A student well-being measure was derived from the Understanding society's harmonized UK Household Longitudinal survey (UKHLS) [see link 44, Annex I], which covers households across the UK. A well-being index was prepared based on responses to 'how do you feel about your school?' and 'how do you feel about your schoolwork?'.

In terms of the weighting of well-being so that it may be aggregated alongside other components such as attainment, planned annual expenditure share (%) allocated to funding pupil deprivation as declared in the DfE's National Funding Formula (NFF), was used to assign weights to well-being.

Please refer to Annex D for more information.

The DfE are currently exploring other methods to assess student well-being in schools, and these will be kept under review to determine if more robust estimates of student well-being can be obtained.

Recommendation 53: The ONS should continuously engage with the Department for Education over indicators of student well-being.

8.7 Devolved governments

The coronavirus pandemic had a strong impact on the availability and quality of data for all of the UK nations. The ONS is working with England, Scotland, Wales and Northern Ireland to overcome these limitations and improve the sources of data in the upcoming years.

Recommendation 54: The ONS should work with the devolved governments to improve Education data sources as far as possible.

8.8 Recommendations for further work

Lifetime outcomes as alternative to attainment

The Review recognises the broader conceptual issue of whether attainment (and the other quality adjusted measures currently used) should be complemented with a longer view of the impact of education towards lifetime outcomes, lifetime earnings and human capital acquisition. However, there are related issues around attribution. This is discussed in greater depth in Annex D.

Recommendation 55: The ONS should, as part of its research agenda, continue to explore the links between Education, the current Education quality measures, health expenditure and human capital acquisition, with specific attention to labour market returns, to better understand the output of these.

Grading policy

There is a second significant topic which the Review has not considered, and this emerges from one of the risks which were highlighted in Foxton, Grice, Heys et al (2019) [see link 43, Annex I], namely arising from decisions made within a service which affect the measures used in these estimates. In this instance, the issue is grading policy.

Grading policy has varied under the different administrations in operation over the time period of these data (1997 to the present day). Essentially, GCSE grades can change for one of two reasons: a quantity effect or a price effect. Either students have achieved greater knowledge and better skills in their years of learning and are achieving an increase in the volume of output (more attainment), or there is a phenomenon of 'grade inflation' – the perception that

students are achieving the same level of skill and knowledge from year to year so improvements in grades represents a drift in the grades awarded for any particular level of skill.

Broadly, if one believes improving grades reflect improved attainment, then this allows a larger share of students over time to be awarded the highest grades. Alternatively, if one believes improving grades is predominantly a result of 'grade inflation', a fixed proportion of the student population can be set who can receive the highest grade, and another fixed proportion receive the next highest grade and so on across the whole distribution.

The Education Act 2011 put in place a new regime, where the The Office of Qualifications and Examinations Regulation (Ofqual) was given the powers and duties to set grading policies and standards to deliver consistent and comparable standards. Ofqual's approach to maintaining standards uses statistical evidence to allow grade boundaries to reflect changes in attainment over time (if they can be evidenced). Taking account of attainment at the primary stage, this establishes a broad expectation that grade distributions are unlikely to change significantly year-on-year.

The impact of this legislation (first introduced by the previous Labour administration) is that from 1997 to 2010 the number of students attaining the highest grades increased, whilst from 2012 to 2024, growth was much more muted. The result is the statistics published prior to the Review suggested faster productivity growth in education under the 1997 to 2010 Labour administrations than under the 2010 to 2024 Conservative administrations, although the Review considers this an unhelpful simplification which does not recognise the complexity of the policy landscape and the impact of the Education Act 2011.

It is, nevertheless, a concern to the Review that the change in regime and the creation of a more robust grading policy framework under Ofqual may be misinterpreted as a change in productivity trends. As it is a long-standing feature of the statistics and there has been no concern raised externally on this topic, the Review has not proceeded to tackle this question, but it may be an issue academic researchers may wish to interrogate.

> **Recommendation 56:** The ONS should engage with academic researchers and stakeholders to understand if the impact of grading policy on Education productivity is substantial enough to warrant further research.

CHAPTER 9

Defence

Reflecting around 10% of total public services, Defence has consistently been one of the top five largest categories since the public services productivity (PSP) statistics commenced from 1997. Nevertheless, the challenge of measuring productivity in Defence has been a long-standing conceptual issue, both in the UK and elsewhere [see link 46, Annex I], which was not resolved in the Atkinson Review (2005) [see link 1, Annex I], or subsequent research, including Prtak , 2019 [see link 47, Annex I], RAND Europe, 2021 [see link 48, Annex I].

Post-Atkinson, whilst the Office for National Statistics (ONS) published a number of conceptual papers, no substantive progress had been achieved to move beyond an 'inputs = outputs' model. Defence, therefore, remains the largest single service which is treated in this fashion. Whilst the Ministry of Defence (MOD) has considered the issues of measuring defence output in the intervening years, it has also been unable to definitively deliver an appropriate approach.

In Defence, as with other areas without strong measures, the ONS measures of productivity may not adequately reflect departmental activities to improve productivity. However, MOD is establishing a Productivity Portfolio to drive a co-ordinated and systematic approach to capturing both cashable and non-cashable benefits, as announced in the Public Spending Audit [see link 49, Annex I] for financial year ending 2025. As part of this, a consistent methodology to track benefits will provide a view on the efficiency of intermediate consumption in MOD. Furthermore, MOD has developed a measure of output termed "readiness", which is a defined and measured state of military preparedness.

The Review commissioned a report from Professor Ron Smith [see link 50, Annex I] (Emeritus Professor of Applied Economics, Birkbeck, University of London) via the Economic Statistics Centre of Excellence [see link 51, Annex I], to consider measurement of defence productivity. Smith, 2024 [see link 52, Annex I] summarised the historical evidence and 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."**

The Review took a two-stage approach to this work:

- developing improved inputs measures. This is a necessary condition of developing stronger methods, whether 'inputs = outputs', or utilising direct output measures
- developing methods to introduce a direct measure of output

The work on input measures reached a state of maturity sufficient to enable recommended improvements which will be implemented into the ONS statistics from Spring 2025. The output work is still at a conceptual level and whilst some data are presented to inform discussion, the Review recommends further work to develop this further, not least because new readiness data could become available in the near future from MOD, which needs to be considered before final decisions can be made.

9.1 Core methodological and data challenges identified

Measurement of defence inputs and outputs

Input measurement in this area has relied on indirect approaches, essentially deflated expenditure at the aggregate level. As with many services, data improvements since Atkinson (2005) allow a move to direct measures, as described in this chapter.

Defence output is difficult to directly measure for several reasons:

- the service is not provided to the individual and can be hard to define: is deterrence or active deployment, or both, or something else,

the delivered output? This leads to three problems:

  - how to value deterrence?
  - what to include in the measure of active deployment?
  - how to weight active deployment and deterrent outputs together? This matters as assumptions made could result in productivity automatically jumping or falling when forces move into or out of an active deployment. Whether this seems appropriate is a valid question.

- defence output could be conceived as either one output or multiple outputs.
- defence outputs are closely related to the UK's allies and their spending. Periodic proposals to develop jointly new capabilities, such as the AUKUS (a trilateral security partnership between Australia, the United Kingdom, and the United States) submarine programme; sharing intelligence, such as through the Five Eyes alliance (an Anglosphere intelligence alliance comprising Australia, Canada, New Zealand, the United Kingdom, and the United States); multinational peacekeeping missions, such as through the United Nations (UN); and collective security agreements, such as the North Atlantic Treaty Organization (NATO), mean there is a need to consider how to reflect expenditure by other governments which, by principle, Atkinson argued should be excluded.

Recognising these challenges, the Review explored several conceptual models and derived some methods for determining a value for deterrence discussed in Annex E. The remainder of the chapter addresses these, including the strengths and limitations of each approach, as well as making some improvements to the ONS input measure.

Granularity of data collected on military defence

The Classification of the Functions of Government (COFOG) [see link 4, Annex I] defines the different services this Review is considering and as described in Chapter 6, is currently under review itself. The ONS continues to work closely with MOD and the UN Task Team on the revision to COFOG, with defence being high on the UK's list of priorities for potential improvements.

Irrespective of the methods discussed in this Chapter, the Review has identified potential improvements regarding how defence spending is currently classified and reported to HM Treasury. The COFOG structure for Defence does not have the granularity of other services of the same relative scale. It categorises expenditure under Division 2 – Defence, which includes:

- 2.1 Military Defence
- 2.2 Civil Defence
- 2.3 Foreign military aid
- 2.4 R&D Defence
- 2.5 Defence not elsewhere classified

Much of the MOD's expenditure is therefore reported under the broad category of 2.1: Military Defence. Spending on NATO-specific operations, such as peace-keeping overseas and similar expenditure is captured under 2.3. This lack of granularity hinders detailed analysis and productivity measurement. In particular, it prevents comparison of a country's army, navy and air force through grouping them together. Following discussions with experts, it is the recommendation of the Review that it would be advantageous to split COFOG 2.1 further, to differentiate at least the operation of land, air, and sea forces (and possibly also cyber and space).

Recommendation 57: As input to the UN Review of the COFOG the ONS should recommend that COFOG 2.1 Military defence should be split as follows:

Division: 02 - Defence

- group: 02.1 – Military Defence
  - class: 02.1.1 – military defence of which contains operations of land forces
  - class: 02.1.2 – military defence of which contains operations of air forces
  - class: 02.1.3 – military defence of which contains operations of sea forces

9.2 Improvements to inputs estimates

Labour measurement

Labour can be measured directly in terms of staff numbers, or indirectly via the expenditure on staffing divided by a relevant deflator. Accounting for the skill mix of the workforce by including rank or grade and salaries allows for a more accurate time series to be produced, showing how the workforce changes throughout time and with policy. Annex F provides greater depth as this topic is addressed by a number of chapters.

Following Atkinson's (2005) recommendation, the ONS and MOD worked together on evaluating how the measurement of defence could be improved in the UK National Accounts. Anagboso and Spence (2009) [see link 53, Annex I] recommended using a direct measurement of defence labour, considering the skill of the work force by disaggregating civilian and military staff based on their rank or grade and salary.

Further ONS research, Kamarudeen (2010) [see link 54, Annex I], examined the feasibility of incorporating the improvements to labour suggested in the 2009 research paper. This stated that a chained Laspeyres volume index that considers the quantity and quality of the defence labour force was being included

in the ONS' Public Sector Output and PSP publications, and hence proposed its inclusion in the UK National Accounts. If a direct labour measurement were to be introduced in defence, the statistics would hold more value as the public would have a better understanding of how the flow of personnel numbers differs across time, and with changes in policy.

However, these recommendations were not implemented because of resource constraints. Defence labour inputs are currently captured indirectly by deflating current price expenditure on defence by a derived deflator based on the COFOG.

The Review revisited this debate and is satisfied this methodological development should be implemented based on existing available data. A precedent for this approach can be seen in policing where similar measures are used to break down labour inputs by rank or grade, average annual pay, and expenditure, and then subsequently deflated using a generic Central Government pay deflator.

For civilian personnel employed in defence, which comprise 'non industrial' (e.g. civil servants) and 'industrial' (e.g. auxiliary staff, apprentices) personnel, the Review obtained a complete time series of grade data, but pay data only from 2007 onwards because of historic pay groupings. For consistency, the Review used data which groups the pay of civil servant grades for senior civil service (SCS), Grade 6 with Grade 7, senior executive officer (SEO) with higher executive officer (HEO), and administration officers (AO) with administrative assistant (AA); executive officers (EO) were their own category. Pay was imputed prior to 2007 by negative compound average growth rates.

In years prior to 2004, 'Non-industrial Civilian' Full Time Equivalents (FTEs) for grades 6 and 7, SEO and HEO, and, AO and AA, respectively, were banded together. In the absence of disaggregated data, the ONS could forecast individual grade FTEs by their average proportions in the three most-recent years with observed data. For example, in 2003 the total FTE of Grades 6 and 7 combined was 2,470. To get an approximation of grade 6s, the total, 2,470, is multiplied by the ratio of Grade 6s relative to the combined sum of Grade 6s and 7s from 2004 to 2007. This is repeated for all grades that are constructed this way.

In years prior to 2009, 'Industrial Civilian' FTEs were banded together. The same methodology as previously described is applied to get approximate disaggregation of industrial grades.

The ONS follows the direct process separately for military and civilian personnel and combine each total growth contribution by the implied expenditure share of each staff division split. For example, if the total civilian labour volume grows by 5% in year t, and total civilian labour comprise 60% of total defence labour expenditure in year t-1, the contribution of civilian staff to total defence labour inputs growth is 3% (i.e. 5% of 60% which is 3% of total civilian labour). A similar calculation is completed for military personnel and then these are summed to create the final contributions.

The derived direct labour volume index was compared with the index presented in Kamarudeen (2010), who similarly adopted a direct labour estimation approach. Figures 1 and 2 show an encouraging similarity of trend. The proposed index falls at a steeper rate, particularly in the early period (1997 to 2002), but the general growth patterns are similar. Both depict a sharp fall in 1998, followed by transient growth to 2004, before another sharp fall to 2007. To note, methodological and data intricacies mean they are not expected to match perfectly, however it is encouraging to verify the broader growth pattern.

Figure 1: Direct labour volume Kamarudeen (2010)

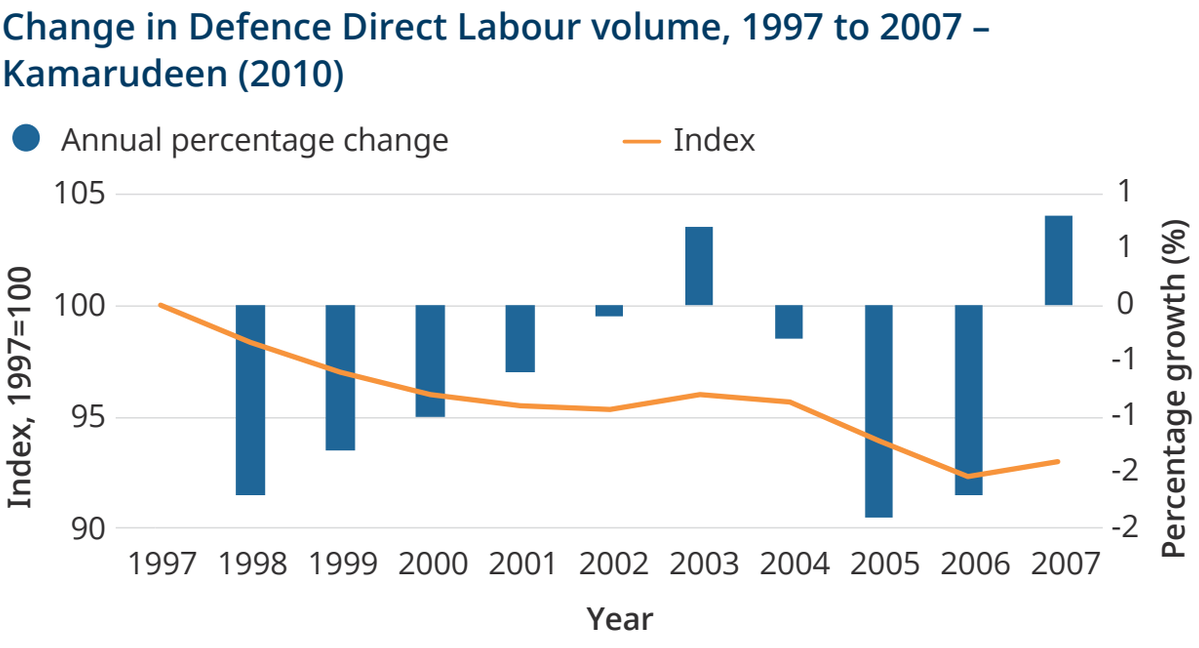
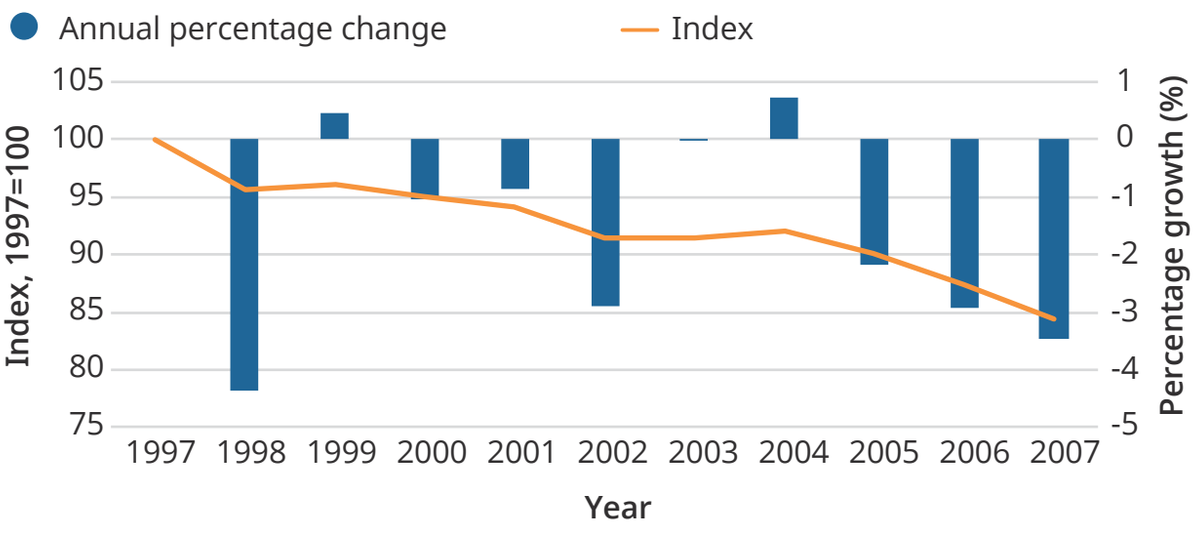


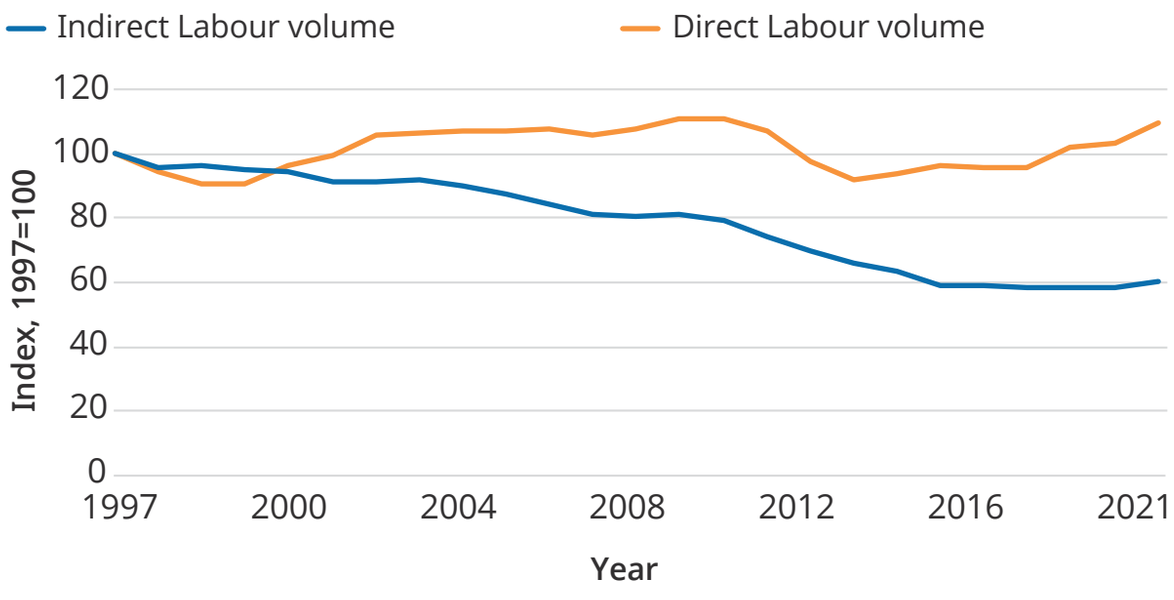
Figure 2: Direct labour volume growth, 1997-2007 (2024 version)



Recommendation 58: A new method implementing a more granular direct measure of labour inputs into Defence should be applied.

Comparing the proposed series results with the previously published series shows a significant difference which begins in the early 2000s. The Review has worked with experts to understand this, and the key factor appears to be changes in the pension regime for military and civil service personnel via the employer contribution. This increased current price expenditure, without being visible in the deflator used to convert staff expenditure into volume. This led to price change being mis-labelled as volume change. The new methodology more accurately portrays the change in the volume of labour inputs into the Defence service through this time period.

Figure 3: Indirect and Direct Labour Volume Index, Comparison, 1997-2021



Intermediate consumption

Given the proposed move to a direct labour estimation approach, intermediate consumption (IC) current price expenditure will need a bespoke deflator applied to it to create volume metrics. The Review recommends deflators be applied to Defence IC expenditure items with observed expenditure used to weight items. In the absence of weights prior to 2013, an average weight of expenditure is used as weights are fairly consistent across observed series. In the absence of deflators in the early part of the time series, growth rates from comparable deflators should be applied to the earliest data point available to impute the historical series.

Recommendation 59: A UK National Accounts consistent intermediate consumption deflator should be derived and applied to Defence intermediate consumption spending, using the His Majesty's Treasury Online System for Central Accounting and Reporting (OSCAR) data to weight components.

Capital

The Review analysed two proposals to provide new capital deflators, neither of which is entirely satisfactory. Both use data on capital expenditure supplied from the Online System for Central Accounting and Reporting (OSCAR) system:

- Proposal one uses the implied deflator from the national accounts for defence capital items produced in the UK. The weakness of this approach is that simply because an item is manufactured in the UK, this does not mean it was purchased by MOD, nor that MOD's pattern of consumption mirrors that of defence purchases from overseas.
- Proposal two is to use a deflator derived from the defence contract prices index published in MOD's 'Defence Inflation Report'. However, the Review has been advised that the defence contract prices include elements of labour and intermediate consumption and hence is not strictly a capital deflator.

In light of this, the Review recommends incorporating both methods to quality assure the results, given neither approach can be considered superior, using the national accounts capital deflator as the lead metric.

Recommendation 60: The Review recommends that the national accounts capital deflator should be applied to the Defence capital expenditure data, but the combined implied deflator of intermediate consumption and capital should be compared to the Defence contracts price index to determine whether subsequent manual intervention is then required.

9.3 Improvements to outputs estimates

Working with experts, the Review has attempted to develop theoretical solutions to the question of how to measure defence output. This remains work in progress and Annex E summarises the research. These methods are not proposed for implementation at the present time but do act as a major step forward in how to think about this issue.

Whilst the Review considers these as being the most feasible methods currently available, this does not mean that data are always available, or do not lead to non-volatile results. Feedback and comments on the content of Annex E are welcome to contact psp.review@ons.gov.uk.

Recommendation 61: The ONS should continue to explore the direct measurement of Defence output with Ministry of Defence and other experts following the publication of this report.

9.4 Quality adjustments

Until a suitable output metric has been agreed, the Review concludes it is too early to consider quality adjustments. The treatment of output, either reflecting active deployment, deterrence, or both, and the exact methods and measures deployed will so fundamentally alter the perspective that considering quality at this point may be superfluous.

9.5 Devolved governments

Given the UK government has sole jurisdiction in this service, there are no devolved government considerations.

9.6 Recommendations for further work

It is clear that the methodological complexities of measuring defence are exacerbated by challenges around data. MOD's programme to create a new readiness measure is a key step to enable further methodological work. It is the understanding of the Review that long time series of readiness data is not planned to be created. Nevertheless, the ONS should work to seize the opportunity these data may present.

Recommendation 62: The ONS should continue to work with Ministry of Defence to explore the potential of using their new readiness measure within the calculation of Defence outputs, and to continue to develop appropriate methods.

CHAPTER 10

Policing

Police services account for 4.5% of total Government expenditure, and yet, like Defence, measures have been unable to mature past the 'inputs = outputs' stage, with the result that productivity growth is automatically zero. This is due to a number of inter-locking challenges: data availability, the complexity of the range of outputs which policing delivers, the attributability of crime outcomes to solely police activities, the preventative nature of many policing activities, and the way the Classification of Functions of Government (COFOG) [see link 4, Annex I] groups policing with both Immigration services and the Criminal Justice services.

In 2017, the Office for National Statistics (ONS) published new methods for measurement of the Criminal Justice System, Quality adjustment of public service criminal justice system output: experimental method (2017) [see link 55, Annex I] as a first step to developing this area, separately publishing these from the rest of the Public Order and Safety COFOG. Policing and Immigration are therefore published in the ONS statistics within the 'Other' grouping.

This Review made policing one of its highest priorities, particularly in the positive light of the National Police Chiefs' Council led Policing Productivity Review published in 2023 [see link 56, Annex I]. This identified new data sources and encouraged the Home Office to create a Centre for Police Productivity (see Chapter 6), whom the Review has collaborated with in developing new methods.

10.1 Core methodological and data challenges identified

A complicating factor in identifying an appropriate methodology for measuring police service productivity is that the relevant COFOG category also encompasses a range of Immigration and Citizenship services, as discussed in Chapter 11. Consequently, substantial changes in expenditure in these services make spending in the overall category appear volatile and hard to relate to policing experience in practice. The Review therefore prioritised separating spending and output related to Immigration and Citizenship services to create a separate productivity measure (see Chapter 11). This enabled a clear line of sight on policing inputs.

Following the decision to separate Policing from Immigration and Criminal Justice, and in the light of new data becoming available, the remaining major challenges are:

- identifying and weighting the wide range of policing activities in what is an operationally conceptually complex service to deliver an output metric
- the attributability of crime outcomes to police activity in terms of quality adjusting that activity

- the preventative nature of police activity, and how best to capture and weight this within the measurement of output

The Review has, drawing on wider practice, divided police activity into three broad areas: 'crime and criminal investigation', 'crime prevention' and 'public safety and welfare'. The Review has prioritised 'crime and criminal investigation' as this area has the strongest data sources. However, it does not underestimate the conceptual challenges and data limitations in bringing together all the aspects of police activity into a single output measure. In other services, cost weighted indices are standardly applied, but it may be necessary to consider alternative methods. The ONS continues to work with Home Office and the Centre for Police Productivity to consider options for weighting.

The Review explored improvements to inputs estimates but focused on producing direct measures of output. The Review has considered quality adjustment, but implementation is dependent on the exact design of output measures. This is clearly an area for further work which may benefit from considering analogous services elsewhere: comparing police custody suites with prisons, potentially in respect to measures of safety and decency such as deaths in custody.

10.2 Improvements to inputs estimates

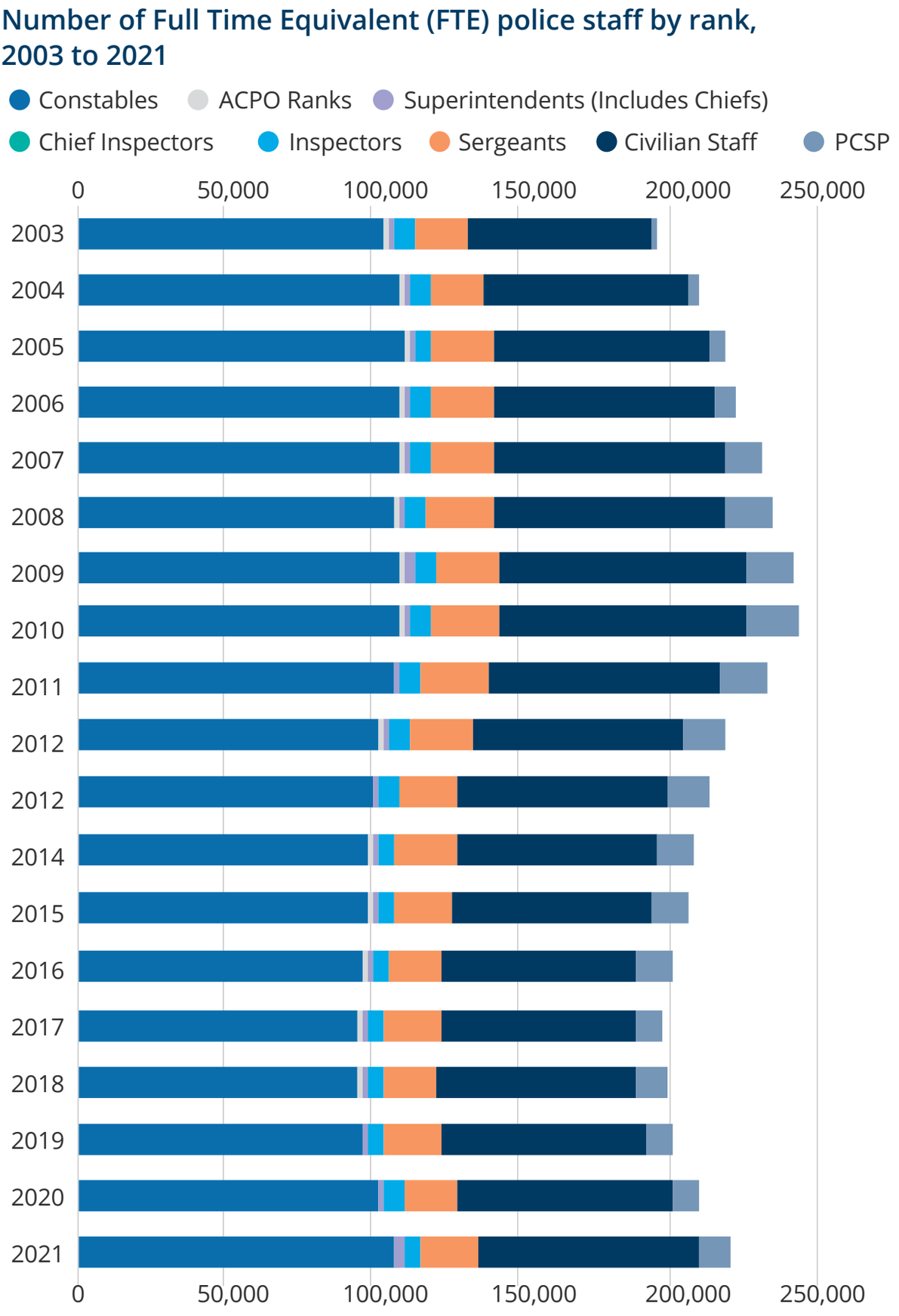
Unlike Defence, where the Review explored moving labour measures from indirect to direct (see Annex F), direct measures for Policing were already available. The degree of detail achievable has been examined.

As in almost all services, labour accounts for the greatest contribution to inputs, and most of this is directly measured using workforce data and associated average salary information. However, the Review recognised that the granularity and coverage of these data may be improved. The Review investigated a number of labour measurement improvements:

- granularity of police salary data: Improving granularity of direct police labour data by disaggregating officer salaries by their respective ranks
- accounting for the Police Uplift Programme (2019 to 2023): Accounting for different skill-level of police labour arising from the Uplift programme
- accounting for reduced capacity: Accounting for different quality of police labour arising from the increasing proportion of officers on limited duties
- improving police labour data coverage for the devolved governments: Improving the comprehensiveness of devolved police labour data. This area is covered under the 'Devolved Governments' subsection.

Full Time Equivalent (FTE) police staff numbers (police and civilians) by rank in England and Wales is shown in Figure 4. Overall numbers have fluctuated over time, rising to a peak in 2010 of 244,497 from a low of 198,375 in 2003, with material variation in the share in specific categories.

Figure 4. Number of FTE police staff by rank, England and Wales, 2003 to 2021



Granularity of police salary data

The Review has used newly acquired average salary data from the Home Office to split the consolidated groups of salaries to the same level of granularity as the full-time equivalent data used in the ONS direct labour measure from 2014 onwards. Prior to this, the average salary information used was grouped for (i) Constables and Sergeants, and (ii) Inspectors and Above.

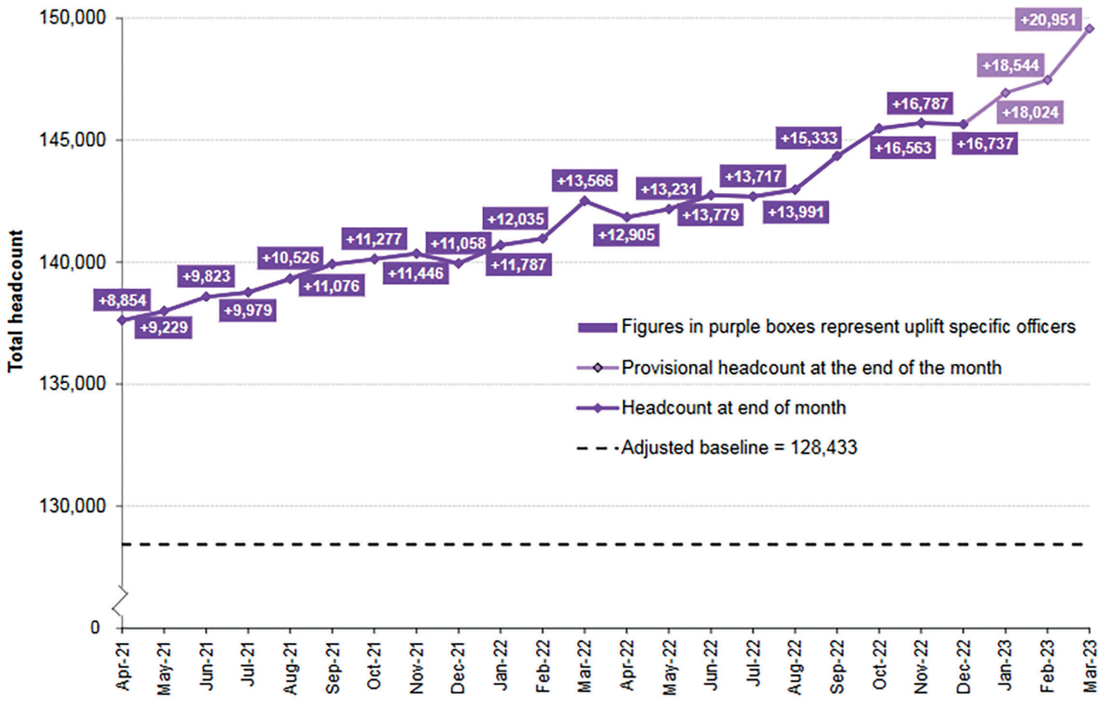
Average salary data for Police Community Support Officers (PCSOs) are also available from ONS Annual Survey of Hours and Earnings (ASHE) data and the Review proposes using these to disaggregate PCSOs in the measure from 2011 onwards. This will result in the most accurately measured police direct labour index to date.

While the indicative impact on annual input growth rates may be small, the improved granularity ensures the measure is more reflective of changes in the level of labour input at each rank, as even a small impact on the inputs measure can reflect a considerable amount of underlying expenditure. This impact is also made more visible when policing is presented as a service distinct from Immigration and Citizenship services, which will also enhance the useability of these data for practical purposes.

Accounting for Police Uplift Programme

The Police Uplift Programme [see link 57, Annex I] was a police recruitment programme running from October 2019 to March 2023. Its goal was to recruit 20,000 new officers within this period, which was achieved, with approximately 21,000 officers recruited. The implication of such a recruitment regime for trainees on productivity measurement is that it has the potential to cause an over-estimation of the value of the inputs being used in a service, and hence under-estimate productivity.

Figure 5: Number of police officers recruited under the Police Uplift Programme, to end March 2023



Source:
Home Office: Accredited official statistics: Police officer uplift, final position as at March 2023

Given the availability of distinct salary rates for staff on uplift, the Review concluded, as discussed in Chapter 3 that further intervention would 'double-count' this effect: the lower salary rate of staff when on the Uplift scheme is assumed, as with all other services, to reflect the lower contribution made by these staff at this time. These data were used to derive separate weights to apply to FTE growth rates to account for the impact of the uplift on labour input.

Accounting for reduced capacity through officers on limited duties

Limited duties include officers on recuperative and adjusted duties, as well as those on restricted duties. Recuperative duties typically occur when an officer has been injured whilst on duty and requires a period of recuperation with less or no deployment to either regain their full duties or move onto adjusted duties. Typically, this lasts for up to six months, however, there are exceptions where this can be increased to 12 months. During this time the officer is paid their full wage. Restricted duties are for officers who have not met the standards required for policing, typically by poor behaviour.

The Review considered whether an adjustment should be applied to account for reduced capacity arising from an increased proportion of staff on limited duties.

The Review attempted to identify an appropriate discount factor to apply to this group of staff.

The Winsor Review [see link 58, Annex I] (2012) calculated the risks that are encountered by fully operational police relative to other public sector workers: this "x-factor" equates to 8% of pay. The Police Remuneration Review Body 2022 report [see link 59, Annex I] presents a potential replacement for the "x-factor": the "p-factor", valued at 13%, which is defined as "element of police basic pay that reflected the responsibilities, obligations, constraints, expectations and terms and conditions that are unique to a police officer's work". Therefore, the Review investigated the impact of incorporating an FTE reduction factor of both 8% and 13% in the direct police labour measure to account for the impact of limited duties. This had a negligible impact on the measure, with the 13% discount factor only reducing the 2021 annual inputs growth rate by 0.01 percentage points, as the share of staff on limited or restricted duties changes.

However, the lack of impact was not the reason for the Review deciding not to recommend incorporation to the ONS measure. The Review concluded that, while effectiveness at delivering frontline service has been impacted, the system is designed to accommodate such staff within the wider range of services and processes delivered by policing, and hence is 'priced-into' the operating model of the service. Unless there is a material change in the number of such staff through time causing the operating model to change, the 'average input' of a police officer has not changed and as such adjustment is unnecessary.

Input deflators

The Review conducted a full audit of the input deflators currently used in the compilation of estimates (labour, intermediate consumption and capital consumption). The central government intermediate consumption deflator (GDP implied deflator) will be replaced with a bespoke police and immigration deflator, based on detailed expenditure from the Online System for Central Accounting and Reporting (OSCAR) from Spring 2025. The Review also developed separate bespoke intermediate consumption deflators for Police and Immigration and Citizenship services, in preparation for the intended disaggregation of estimates for these two areas.

Devolved governments

Current input data sources include labour data from England, Wales, and Scotland. The Review was able to source labour data from Northern Ireland for the first time, adding a time series from 2000 onwards. There is a need for more disaggregated labour data (FTE by rank) for Scotland for earlier in the time series to fill in current gaps.

> **Recommendation 63:** The ONS should endeavour to source more granular police workforce data for Scotland.

Recommendation 64: The ONS should apply bespoke Police and Immigration deflators to Central Government expenditure data used to indirectly measure intermediate consumption in Spring 2025.

Recommendation 65: The ONS should implement more granular salary data into its direct labour measure for Police, incorporating salaryinformation for individual police ranks (including Uplift officers) in Spring 2025.

Recommendation 66: The ONS should incorporate Northern Ireland workforce data into its direct labour measure for Police in Spring 2025.

10.3 Development of outputs

On the output side, the Review has divided police outputs into three categories to simplify working through the diversity of services delivered:

- 'crime and crime investigation' relating to services addressing crimes which have already occurred. Crime refers to a deliberate act that causes physical or psychological harm, damage to or loss of property, and is against the law (police.uk) [see link 60, Annex I]
- 'crime prevention' relating to services to prevent the occurrence of crimes which are yet to occur. Crime prevention is activity taken by the police that actively prevents a crime incident from occurring (such as visible patrols)
- 'public safety and welfare' refers to police activity other than that related to crimes, such as that related to road traffic or anti-social behaviour incidents

For 'crime and criminal investigation', the Review developed a proposed method using police recorded crime and outcomes data, weighting the activity associated with these by the police resource used in investigating different crimes and achieving different outcomes. For 'public safety and welfare', the Review investigated a range of potential metrics. 'Crime prevention' is particularly challenging, for the same reasons faced by multiple services and discussed in depth in Chapter 3.

A direct volume output measure is required to construct a measure for police productivity. This entails:

- defining the activities to be included as output (across the three different categories of crime, public safety and welfare and crime prevention)
- collecting data on these activities completed

- developing weights for each category of activity, and if necessary within each category

In February 2023 the Home Office undertook a Police Activity Survey (PAS), a currently unpublished time-use study covering England and Wales. In collaboration with Home Office, the Review assessed the utility of this dataset in developing output metrics for policing activity and found this to be favourable. If such a data source were to be available on an ongoing basis this would facilitate the development of policing output estimates. Moreover, there could be potential in extending the survey to Scotland and Northern Ireland in the context of developing UK wide estimates.

The PAS collects police activities by crime and criminal investigation, public safety and welfare and 'other' activities and can therefore be used for the cost weighting of much of police outputs. 'Other' activities are activities that are not related to a specific crime or public safety and welfare incident. However, as these activities relate to those which support the delivery of crime and public safety and welfare services (and would not normally be considered as service output), only crime and public safety and welfare were considered for police output.

Using the PAS, the Home Office has estimated police 'resource unit costs' by considering time spent on activities to be representative of the population, and weighting by the rank pay of the officer carrying them out. The time weighting is calculated by comparing the survey activity percentages with the annual total using Police Recorded Crime (PRC) for the crime incident groups available. Note that these weights therefore only consider the relative labour costs of different activities, although usage of capital and intermediate consumption inputs can be assumed to be correlated with labour inputs.

Whilst the PAS provided a 'snapshot' in time, the Review understands that Home Office intend to repeat the survey in financial year ending 2026 subject to available funding. Initial exploration has begun of ONS adding the survey to its existing survey portfolio.

Recommendation 67: The ONS should continue to explore with Home Office the potential to run an England and Wales Police Activity Survey as part of the ONS survey portfolio.

Recommendation 68: The ONS should explore the potential to source police activity data for Scotland and Northern Ireland.

Crime and criminal investigation

Police Recorded Crime and Outcomes data [see link 61, Annex I] provides a time series of police criminal investigation data from all England and Wales police forces on the number of recorded offences committed and reported. These are produced quarterly by the Home Office and are disaggregated by force, type of offence, and the outcome for each case. Full outcomes data has been published since financial year ending 2016. As each crime recorded in the data is associated with police activity, this can serve as a time series for police crime and criminal investigation activity.

The PRC classifies crimes into 10 categories, which can be broken down into a further 23 sub-categories. PAS uses the same categories for reporting the breakdown of staff resource costs for crime activity. The Review, therefore, used these two data sources to produce a 'proof of concept' cost weighted activity index for crime and criminal investigation activity, albeit one where the costs do not include non-staff inputs and constant weights are applied, where these should change through time. This would require further waves of the PAS.

Whilst PAS may not cover crimes which are not reported to the police, retaining a focus on the crimes police are made aware of feels sensible in terms of measuring police activity. One potentially important factor affecting how much resource is deployed is whether the crimes included in PRC are uncovered through proactive policing or reported by victims or witnesses. To date, no data has been uncovered that would facilitate this, but this omission means changes in the extent to which crimes are uncovered through proactive policing is not captured in the output measure.

As Action Fraud [see link 62, Annex I] is not included in the PAS, the proof-of-concept work produced excluded the fraud crime category meaning that only nine categories were used: Criminal damage and arson; Theft offences; Violence against the person; Miscellaneous crimes against society; Sexual offences; Drug offences; Possession of weapons offences; Public order offences; and Robbery.

Police crime output can be further refined by disaggregating the weights allocated to each crime category by the outcome achieved, using the 12 outcomes categories captured in PRC. There is clearly an overlap here with the concept of quality adjustment: identifying an offender and bringing them to justice is clearly preferable to closing a case without identifying a suspect and would normally be considered within a quality adjustment factor. However, in this case, the input variation is also suggestive that these outcomes represent a different quantity of output. The Review has therefore considered how to reflect differences in the police resource used to achieve an outcome in the quantity of output but further development reflecting differences in the social value of different outcomes should be investigated for the quality adjustment.

Data on average time taken to reach a particular outcome by crime category are also available. By considering what outcome is associated with each crime, as well as time taken to reach that outcome, it could be possible to get a more accurate idea of how much activity went into a particular crime event.

The Review developed a method for sub-dividing the weights for each crime category to take into account both outcomes and time taken to reach the outcome. The outcomes can be split into high, medium and low-value outcomes, each given their own weight according to the number of days taken to reach the outcome, with high-value outcomes taking more time to achieve than low-value outcomes and thus reflecting a greater quantity of police activity. The identification of which outcome falls into each group has been completed following expert consultation.

The outcomes are weighted in three levels as shown in Table 3:

Table 3: Crime outcome by proposed weighting level used to subdivide weights for each crime category

Outcome	Weighting level
Charged or summonsed	Full weight
Out-of-court (formal)	Full weight
Out-of-court (informal)	Full weight
Taken into consideration	Full weight
Diversionary, educational or intervention activity, resulting from the crime report, has been undertaken and it is not in the public interest to take any further action.	Full weight
Evidential difficulties (suspect identified; victim supports action)	Medium weight
Evidential difficulties (victim does not support action)	Medium weight
Further investigation to support formal action not in the public interest – police decision	Medium weight
Prosecution prevented or not in the public interest	Medium weight
Responsibility for further investigation transferred to another body	Medium weight
Investigation complete – no suspect identified	Low weight

Recommendation 69: The ONS should continue to collaborate with key police partners on appropriate methods to develop and weight different types of output activity.

Public safety and welfare activity

In contrast to crime and criminal investigation activity there is no single source for all public safety and welfare activities, and the individual sources capturing these activities are not always comprehensive.

The PAS splits public safety and welfare activities into 37 categories. However, as many of these take up very little police time and some do not lend themselves to measurable activity ("police generated resource activity", "unknown", "messages", "other public safety and welfare" and "other incident"), our investigation has focussed on a refined list of seven categories, comprised of 12 of the 37 public safety and welfare categories in PAS:

- missing persons
- concern for safety
- anti-Social Behaviour (ASB), comprising three categories: 'Nuisance', 'Personal' and 'Environmental'
- domestic incident
- transport comprising four categories: 'Road related offence', 'Road traffic collision – death or injury', 'Road traffic collision – damage only', and 'Other transport')
- sudden death
- pre-planned events

Cost weighting public safety and welfare activity is more challenging than for crime and criminal investigation. While the PAS provides resource unit costs for all crime activities, there are only limited resource unit costs available for public safety and welfare activities. These are: anti-social behaviour, transportation, administration, and public safety and welfare. Of these, only anti-social behaviour and transport can easily be mapped to activity, with the other two categories being too generic, but further work is encouraged here, particularly if future waves of the PAS can be refined to better map onto the desired categories.

One source that contains data on multiple categories is the Police Efficiency, Effectiveness and Legitimacy (PEEL) [see link 63, Annex I] data collected by HM Inspectorate of Constabulary and Fire and Rescue Services (HMICFRS). HMICFRS use this to assess the performance of all 43 police forces in England and Wales. Data collected includes information on missing persons, anti-social behaviour and domestic incidents, and therefore provides potential activity data for three of the seven public safety and welfare activities.

However, the annual statistical report [see link 64, Annex I] from the National Crime Agency Missing Person Unit contains more comprehensive data on missing persons and could be used in place of or in conjunction with the PEEL data for that activity.

The broad category of transport comprises of four sub-categories: Road related offence, road traffic collisions – death or injury, road traffic collision – damage only and other transport. Data taken from the Department for Transport on collisions provides potential data for road traffic collisions – death or injury. Useable data have not currently been identified for any of the other categories within transport.

While activity data could potentially exist for the remaining categories, this has yet to be identified. This means that concern for safety, sudden death and pre-planned events could not currently be included in police public safety and welfare output.

> **Recommendation 70:** The ONS should draw together existing data sources for Police 'public safety and welfare' activities and develop a plan to fill data-gaps.

Crime prevention activity

As addressed in Chapter 2 and Chapter 3, accurately measuring prevention has its challenges. This is particularly the case with crime where it is not always clear what crimes would be prevented from particular activities. For example, having visible patrols may prevent certain crimes but it would be difficult to allocate this prevention measure to an existing group of activities for which the ONS has measures. Similarly, estimation requires evaluation evidence on the effectiveness of prevention activities and the probabilities of prevention which would beneficially be of equivalent academic quality as those applied to other services, such as Healthcare.

In addition, to ensure different aspects of crime prevention are addressed broadly consistently, data on different interventions are required. To support this, the Review recommends a literature review to identify data.

> **Recommendation 71:** The ONS should commission a literature review of the impact of crime prevention activities which fall under the Police 'criminal prevention' activity category, with a view to future development.

Weighting outputs to derive an aggregate output index

The core issue identified by the Review is the lack of a time-series for weighting both within crime and criminal investigation or public safety and welfare and, importantly, between crime and public safety and welfare over time. In order to combine these two activity types into a single police output, an appropriate time-varying weight would need to be sourced.

The Review has explored the Police Objective Analysis data from the Chartered Institute of Public Finance and Accountancy (CIPFA). These split funding by different categories which it may be possible to align to the crime and public safety and welfare categories currently being explored. More importantly, this is a time-series from financial year ending 2015 and could therefore allow for a time-varying cost weight. Alternatively, future runs of the PAS would be beneficial to enabling a Laspeyres index to be constructed.

> **Recommendation 72:** The ONS should continue to explore the feasibility of introducing a combined Police crime and Police public safety and welfare output measure by critically assessing alternative activity and weighting sources.

10.4 Development of quality adjustment estimates

The Review has considered potential quality adjustments (i.e. outcomes) of crime investigation outputs (explored for England and Wales in the first instance) to better reflect the value citizens receive from these functions, and how this istranslated through to citizens in terms of the quality of the activity undertaken. The following were considered:

- police legitimacy
- victim (of crime) satisfaction
- casefile quality

Police legitimacy

Police legitimacy or 'policing by consent' is key to the nine principles [see link 65, Annex I] developed by the founder of the Metropolitan Police, Robert Peel. This consent is dependent upon carrying out the 'law fairly, impartially and by using minimal force' (Policing in the UK, House of Commons Library) [see link 66, Annex I]. It is important to reflect this as it determines to what extent crimes and criminality are reported to the police: a police force which is not made aware of crimes cannot be fulfilling the function it is designed to deliver to citizens, both at the national and local levels.

The views of the general population are likely a useful gauge of public sentiment around police legitimacy (for example, as collected in the Crime Survey of England and Wales (CSEW) or other surveys such as YouGov Monthly Tracker, Ipsos Veracity Index, HMICFRS Public Perceptions of Policing Surveys, IPSOS Global Trustworthiness Index, IOPC Public Perceptions Tracker, or the ONS' Opinions and Lifestyle Survey). Although there is an argument it may prove a more informed quality adjustment to use responses from members of the public who have had direct contact with the police through being victims of crime, this misses the universal aspect of the police offering which should ensure all members of society feel safe and protected.

The CSEW is a large sample survey (approx. 31,000 in 2023) that asks household residents about their experience of a range of crimes in the 12 months prior to being surveyed. The survey provides an estimate of the level of crime committed, covering both reported and unreported crimes. It is considered the most reliable indicator for long term trends, as it is unaffected by changes in police recording, crime reporting rates or police activity and the Review considers it the best data source for this purpose. While not exhaustive, potential indicators are shown in Potential Police legitimacy indicators from the CSEW.

Potential Police legitimacy indicators from the CSEW

Local

- they (the police in this area) would treat you with respect if you had contact with them for any reason
- the police in this area treat everyone fairly regardless of who they are

National

- in general how much do you trust the police as an organisation?
- experiences that have affected your trust in the police as an organisation? (options given)

Victim satisfaction

One of the most straightforward ways of adjusting crime investigation outputs would be to utilise data on victims of crime and their experience of the police at different stages of the police investigation. Whilst the CSEW is the most likely data source, the Review also investigated the potential of other victim related surveys (e.g. Victims' Commissioner's Victim Survey Reviews and reports – Victims Commissioner [see link 121, Annex I], Home Office Annual Data Requirement [see link 67, Annex I] on forces to interview victims of a range of crimes).

Further research is required to demonstrate feasibility of CSEW data, including around different crime types or different aspects of the victim experience, as shown in Victim satisfaction indicators from the CSEW, specifically in relation to whether samples yielded enough victims to allow valid comparisons through time.

Victim satisfaction indicators from the CSEW

Initial contact

- Were you satisfied or dissatisfied with the way in which you were able to report the matter?
- result of the initial contact

Victim experience

- Do you think the police treated you fairly?
- Did the police treat you with respect?
- Overall, were you satisfied or dissatisfied with the way the police handled this matter?

Victim support

- Do you think the police treated you fairly?
- Did the police treat you with respect?
- What types of information, advice or support, if any, did you receive?

Case file quality

Using administrative data recorded in the criminal justice system, the timeliness and quality of investigation as evidenced through the case file, could act as a quality adjustment, through for example, the share of case files that are returned to police by the Crown Prosecution Service (CPS) due to problems with case notes. This can result in delays in prosecution. The Police Productivity Review (2023) [see link 68, Annex I] noted that 47% of case files were not accepted at first triage by the CPS and recommended that case file timeliness and quality be added to its proposed Model Process tool.

Wider quality adjustments

In relation to 'public safety and welfare' and 'crime prevention' how one may quality adjust depends on the final output measures which are agreed on, but this is clearly an area for future work.

> **Recommendation 73:** The ONS should continue to collaborate with policing partners to develop appropriate quality adjustments for Policing services.

CHAPTER 11:

Immigration and Citizenship Services

Measurement of Public Order and Safety (POS) services, as described in Chapter 12, has always been made more complex as the Classification of Functions of Government (COFOG) category (3) combines a range of very different services. Including Immigration and Citizenship services alongside criminal justice systems in the same group as police services makes this category hard to interpret. In the absence of a better method, given this constraint, Atkinson [see link 1, Annex I] and subsequent work adopted the 'inputs = outputs' approach.

In 2017 the Office for National Statistics (ONS) introduced a new model for the criminal justice system component of the POS COFOG category, separately presenting it in the statistics. Nevertheless, the residual POS services – essentially policing and immigration – continued to be volatile and hard to interpret. The Review has identified that splitting Immigration services from Policing is necessary to provide data which would have policy relevance. The ONS has already recommended that the COFOG review recently commissioned by the United Nations should consider implementing such changes to improve data internationally.

COFOG 3.1 'Police Services' is described in Figure 6. The Review considers the activities highlighted in green to be those of Immigration and Citizenship services.

Figure 6: Classifications of Functions of Government definition for police services

03.1 POLICE SERVICES

03.10 POLICE SERVICES (CS)

Administration of police affairs and services including alien registration, issuing work and travel documents to immigrants, maintenance of arrest records and statistics related to police work, road traffic regulation and control, prevention of smuggling and control of offshore and ocean fishing; operation of regular and auxiliary police forces, of port, border and coast guards, and of other special police forces maintained by public authorities; operation of police laboratories; operation or support of police training programs.

Includes: traffic wardens.

Excludes: police colleges offering general education in addition to police training (09.1), (09.2), (09.3) or (09.4).

Source:
Manual on sources and methods for the compilation of COFOG statistics (PDF) [see link 4, Annex I], Eurostat

Recommendation 74: The ONS should seek to convince the United Nations Review of the Classification of the Functions of Government (COFOG) [see link 4, Annex I] to separate Immigration and Citizenship activities from COFOG 3.1 Police services.

The Review made preliminary investigations into the data that could be used, as well as the methodology involved in creating estimates for this service. Examining the historical time series for Immigration and Citizenship inputs, the Review concluded that future separation of Policing from Immigration and Citizenship services will be more viable from 2004. The earlier data face two challenges: the level of spending on labour and intermediate consumption in Immigration and Citizenship services has materially increased since 1997, partly due to a broad increase in spending, but also because of spending having previously been allocated to other areas of Police services, which makes it difficult to disentangle.

11.1 Core methodological and data challenges identified

Being a new service, the Review had to consider questions of definition, data and methods from scratch, applying the Atkinson Principles alongside ensuring methods improved the ability to interpret both Policing and Immigration and Citizenship services. This development work is described in the next section and whilst these methods need further refinement to include into the statistics, the Review considers implementation to be highly feasible in the short-medium term.

11.2 Inputs development.

Immigration inputs are calculated by deflating expenditure. The Review explored existing expenditure data to determine whether these permit development of indirect labour, intermediate consumption and capital expenditure input data for Immigration and Citizenship services separately. The deflators used will need to be reviewed alongside reviewing the expenditure data.

At present, Immigration and Citizenship is included in the Police and Immigration service, and inputs are measured using existing expenditure data and deflated using the implied Gross Domestic Product (GDP) and central government pay deflators. These data are available for five areas of Home Office: Asylum and Protection; Border Force; Immigration Enforcement; Passport Office; and Visas and Immigration.

In 2022, Asylum and Protection had the largest expenditure on intermediate consumption, followed by Visas and Immigration. These data have sub-divisions for categories of spend.

The Review examined the spending of each of the five areas on IT, travel, etc. This informed the development of improved deflators for Policing and Immigration, which will be implemented in the upcoming ONS annual release in Spring 2025.

An example of the improved deflators is that the Review identified the amount typically spent on posting passports, and using price indices for postal services, the deflator is able to reflect price changes more accurately on spending in this area, as opposed to using the implied GDP deflator, or another less relevant price index. The Review used these data to compile a 'proof of concept' inputs index solely for Immigration and Citizenship. This used deflated data on aggregated final consumption expenditure but did not include capital expenditure data. The Review has identified a suitable approach for apportioning expenditure on capital consumption within the COFOG category, alongside methods to split consumption and labour expenditure, although the implementation of this separation requires more time and resource and is held as a future objective.

This inputs index demonstrates clear increases in 2004, then again in 2012 and 2013. The Review has inferred this is due to reorganisation of bodies including Border Force, and the increasing political salience of immigration. The size of the increase in this index in 2004 is sufficiently large and without further research into the reason behind this, the Review does not think it is appropriate to split the Police and Immigration and Citizenship prior to this point. The Review does recommend that Police and Immigration and Citizenship should be split from 2004 as these are two significant policy-areas and distinct enough to each merit their own focus.

> **Recommendation 75:** The Review recommends commencing the split of Policing from Immigration and Citizenship services from 2004, and retaining the combined series for earlier years.

Associated deflators required

The Review recommends reviewing spending patterns that would inform Immigration and Citizenship deflators annually. Categories in the historical time series change frequently over time, which means that the improved deflators to be implemented in the next ONS annual release (Spring 2025) may become less accurate.

Asylum and Protection, as a broad area of Home Office, only appears in the Online System for Central Accounting and Reporting (OSCAR) data for the first time in 2022, with the activities in this area previously forming part of Visas and Immigration. Examining the broad areas in previous years, these change frequently.

For example, a deflator developed in 2020 is unlikely to have been an accurate representation of spending in 2023, owing to sharp rises in Asylum and Protection spending in 2022 and 2023. Therefore, the broad areas of Home Office, and the categories of spend within, should be reviewed on an annual basis.

> **Recommendation 76:** The ONS should review the composition of the Immigration component of the combined Police and Immigration intermediate consumption deflator annually to ensure accurate representation.

11.3 Developing measures of output for Immigration and Citizenship

Currently, the output of Immigration and Citizenship services is measured using the 'inputs = outputs' approach as part of the same service as policing. This means that output for Immigration and Citizenship services is not separately identified, and because its volume growth matches the inputs, this implies no productivity change. To measure change in the productivity of Immigration and Citizenship services, the ONS would require data on the number of activities carried out by type of activity and the expenditure on each activity type.

The Review considered identification and assessment of output measures which align with the five relevant services. An initial exercise identified a range of activity data and, in some cases, associated unit cost data, as well as potential quality adjustment metrics. The ONS continues to work with colleagues in the Home Office to identify the most suitable data.

As with other services the incorporation of preventative work is a methodological challenge (for example work to prevent illegal immigration). However, to make progress in this area, data on service costs alongside data on the causal relationship between the preventative service and the service avoided would be required.

Within each of the five services, the Review has evaluated the data available on the main service activities to assess their suitability in producing a volume output measure.

Asylum and protection

This considers requests for asylum and processes applications. The key services identified are the processing of asylum applications and asylum appeals, and the housing of asylum seekers. The Home Office publishes data on asylum applications and appeals in Immigration system statistics [see link 69, Annex I] , and maintains internal datasets on the housing of asylum seekers.

In developing an output measure, a key conceptual question is whether asylum applications, appeals and accommodation should be considered separate aspects of valuable output or whether a single aspect of the service, for instance applications could be used to represent the growth of output for asylum and protection overall. Future discussions with Home Office and other experts would be required to determine the most appropriate activity measure and identify any further cost data required to utilise this in the output measure.

Border force

Border Force is responsible for carrying out immigration and customs control for people and goods entering the country. Data relevant to measuring the activities undertaken by the Border Force include the number of passengers processed at the border, tax revenue protected and seizures.

Activity data are published by the Home Office in the quarterly Border Force Transparency Data publication [see link 70, Annex I] . These data include:

- Numbers of passengers processed.
- Tax revenue that is protected through detecting goods where excise duty has not been declared.
- Firearms, knives, and other offensive weapons seizures.
- Drug seizures.
- Convention of International Trade of Endangered Species seizures.

In common with other services, a significant challenge in measuring output is establishing suitable weights for combining different activity measures into a single output index. Where there are limited records on how much staff time and other resources are directed at different activities, it may not be possible to aggregate different activity measures into a single index. In such a case, the ONS would seek to determine if a single activity measure can provide an overarching view of output growth. In the case of the Border Force, it may be possible to use the total number of passengers processed as that activity measure.

Immigration enforcement

Immigration Enforcement is responsible for enforcing immigration law. Published Home Office [see link 71, Annex I] data relevant to these activities are the number of immigrants returned from the UK and people in detention. The measurement of the output of immigration enforcement presents a conceptual question in terms of which activities are relevant to include in output.

For instance, the number of people in detention by the immigration authorities may not be considered an output, but an interim stage within the process of removing irregular migrants. However, as a costly government deliverable, it may be best compared to the prison population, which is treated as a standalone output despite often being part of the offender journey augmented by probation supervision. The prison population is measured as an output, and quality adjusted to reflect their impact on re-offending, for example. In that context, treating the share of those in detention who are removed may be considered a similar quality adjustment.

Passport office, and Visas and Immigration

The Passport Office processes applications for, and produces, passports. The key activity identified is the number of passport applications processed. Two Home Office published datasets [see link 70, Annex I]: passport application intake volumes, and domestic and overseas applications for passports, could be utilised.

Visas and Immigration process applications to stay in the UK, both short and long term, and citizenship applications. The key services identified are visa processing, citizenship applications, and short and long-term extensions of stay within the UK, and the Home Office publish activity data on all these services.

A further consideration in the case of the Passport Office and Visas and Immigration is the fees recovered. Currently, within public service productivity statistics cost weights net of fees are used, with activity data adjusted accordingly. However, if the fees deliver full cost recovery of a service, this fee should be considered a 'market equivalent price', which suggests the ONS can use traditional national accounts methods to identify the value of this service within the output produced in this service.

As the fees charged for passport and visa and immigration services are highly significant relative to the cost of providing these services, the ONS would need to consider whether changes ought to be made to this approach.

Recommendation 77: The ONS should review whether full cost recovery fees, for example for passports, should be treated as market-equivalent prices, which do not require quality adjustment, as prices should internalise quality change.

Recommendation 78: The ONS should continue discussions with the Home Office to identify the most appropriate service activities and suitable weights for these to construct a direct volume output measure for Immigration and Citizenship services.

CHAPTER 12

Public Order and Safety, including the Criminal Justice System

As discussed in Chapter 10, the Public Order and Safety (POS) Classification of the Functions of Government (COFOG) category is a complex mix of different services currently grouped under a single COFOG category. This report therefore discusses productivity measurement for the policing and immigration and citizenship services separately (see Chapter 10 and Chapter 11 respectively) given the distinctive nature of these services. The remaining services included in the POS measure are law courts (which itself has five further sub-components: Magistrates' Courts, County Courts, The Crown Court, Crown Prosecution Service, and legal aid), prisons, probation, and Fire and Rescue.

In 2017, the Office for National Statistics (ONS) introduced a new model for measurement of the Criminal Justice System (CJS) components of the POS COFOG category – primarily courts, legal aid, prisons and probation, and supporting activities such as electronic monitoring and prisoner transportation.

This new model implemented methodological innovations around weighting quality adjustments and heralded consideration of the criminal justice system as a model of integrated service delivery. Quality adjustments currently included in the POS measure are:

- re-offending
- courts' timeliness
- prison safety
- custody escapes

In the context of this Review the Ministry of Justice (MoJ) proposed a number of potential further improvements, which this chapter discusses. In particular, the impact of the coronavirus (COVID-19) pandemic and the consequent staff furlough scheme resulted in reduced court services being offered that impacted on the measurement of productivity estimates for the courts and the quality adjustment based on re-offending rates.

12.1 Core methodological and data challenges identified

Even after separately dealing with Policing and Immigration and Citizenship services, together with the various components of the CJS, there remains a final service, Fire and Rescue, which is accommodated in this service grouping. Fire and Rescue services have been historically captured in this area and the Review considered this continued to be appropriate, pending any update of the international COFOG structure.

In 2020, due to the coronavirus pandemic, court services were significantly impacted and reduced due to the furlough scheme. Consequently, fewer convictions were passed down. Given the re-offending metric used to quality adjust the CJS is, in reality, a 're-conviction' metric, this meant that the ONS was forced to suspend the re-offending section of the quality adjustment of this service during the pandemic. The Review explored options to re-apply this.

The inputs of the CJS productivity measure are currently calculated indirectly (Labour, Intermediate Consumption (Goods and Services) and Capital Consumption), using annual UK National Accounts data and a range of deflators. As in other services, the Review explored moving to direct measurement by using appropriate full-time equivalent (FTE) workforce data and associated average salary information.

Direct measures of quantity output have already been included in the total public service productivity (PSP) series for, as this chapter will henceforth refer to them, the four primary CJS components (courts, legal aid, prison, and probation). However, the Review has identified improvements given the availability of updated data sources to bring measures of the court service up to date, alongside methodological improvements to reflect changes in the composition of prisons by category. The MoJ also queried the pertinence of some of the current quality adjustments to the court system. The rationale for the current model and the reflection of the Review on this are explained in the Improvements to Quality Adjustments section.

Except for Fire and Rescue services, all output is quality adjusted using a range of metrics. There are several metrics used to derive quality adjusted CJS outputs and the Review identified within these that updating data sources for the output of criminal courts and updating the quality adjustments with the best data on case timeliness in courts, were priorities.

Identifying and sourcing unit cost data to facilitate output methods improvement across the POS services is particularly challenging.

12.2 Improvements to inputs estimates

The inputs element of the POS productivity measure uses an indirect approach of deflated expenditure to derive the volume of Labour, Intermediate Consumption (Goods and Services) and Capital Consumption inputs. An aggregate index is then compiled for total POS inputs and weighted using expenditure levels for the four primary CJS components.

The Review focused on assessing the latest availability of workforce data to explore the feasibility of moving to a direct measure of labour inputs for each of the four primary CJS components, as per Annex F. Where data permitted, the Review compared the direct approach against the existing indirect series, as recommended in the Atkinson Review [see link 1, Annex I].

Criminal Justice System

Data were sourced from a variety of areas including the ONS Annual Survey of Hours and Earnings (ASHE) and published management information. The Review also investigated options to obtain unpublished data directly from relevant organisations. Data availability varied, with only a limited timeseries available in some cases. Certain services have also implemented changes to grades or job roles over the years, which results in issues of comparability over time.

Work undertaken to derive direct measures of labour for the four primary CJS components was advanced as far as data availability and time constraints allowed. Following engagement with MoJ officials the Review feels that progress in these areas is possible (see 12.5 Recommendations for further work).

In addition to work on improving the measurement of labour input, the ONS has improved the central and local government deflators used to determine volume growth in the Intermediate Consumption (IC) measure. Due to the different types of expenditure across services, a composite bespoke IC deflator has been generated for each services, with the exception of the Fire services, where IC expenditure will be deflated using the headline CPI-based deflator, both because of relative scale and absence of data.

> **Recommendation 79:** The ONS should apply bespoke intermediate consumption deflators to expenditure data in all Public Order and Safety service areas, with the exception of fire.

Fire and Rescue

The Review developed a direct labour input measure for Fire and Rescue services. This relates to local government labour, which accounts for more than four-fifths of Fire and Rescue service labour expenditure.

To account for the different FTE volumes by rank or grade the ONS matched the FTE in each rank with relevant average salary data from ASHE. ASHE was deemed the most suitable data source with a long time series, as this currently uses aggregate ranks and grades. The Review considers that given the comparably small size of this service, ASHE provides sufficient detail and further work to find more granular data is feasible, but not necessary or recommended at this time.

From ASHE, the ONS extracted the data series for the following Standard Occupational Classification codes with availability back to 1997, consistent with the treatment of other areas in taking account of generic administrative functions:

- 3313 – Fire service officers (watch manager and below)
- 1163 – Senior officers in fire, ambulance, prison and related services
- 4113 – Local government administrative occupations

A comparison of the new direct measure and pre-existing indirect approach showed they follow a pattern of negative growth between 2011 and 2016, but the rates are more pronounced in the indirect measure. Using the direct approach, negative growth extended to 2018. The fall in direct labour captured in this analysis was in line with previous ONS published analysis [see link 72, Annex I] indicating a fall in workforce across Fire and Rescue services.

Based on the data available, the Review recommends implementing this change from 2011 onwards into published ONS estimates. Home Office has developed a comprehensive Fire salary model, but at present there is an insufficient time series for incorporation as an alternative to ASHE. The available Home Office data on FTE relates to England only. The ONS would seek to expand coverage of the direct labour measure to the devolved governments should data be available.

> **Recommendation 80:** A new method implementing a direct measure of fire service labour inputs into Public Order and Safety should be introduced, expanding in future years to cover devolved governments.

12.3 Improvements to outputs estimates

Prisons

The activity measure used for English and Welsh prisons output is MoJ published monthly estimates [see link 73, Annex I] of the prison population. This does not contain data on differences in the cost of different categories of prison and therefore the effect of changes in the composition of prisons.

The Review has identified data to improve the estimation of prisons output to enable this heterogeneity to be reflected.

The new prisons output measure will weight activity by their security categories, with higher weights applied to higher cost prisons. For example, prisoners in 'Dispersal' prisons (those prisons which hold Category A prisoners, alongside other prisoners) will have a greater weight than those in 'Male open' (establishments where prisoners are lowest risk).

The current prisons output measure uses total monthly prison population figures for England, Wales and Scotland. Northern Ireland is not included in the current output measure for prisons. The proposed output measure will continue to cover England and Wales, although these will now be split by security category. Scotland will also be captured as a single category due to the different design of Scottish prisons compared to the English and Welsh dispersal, local, trainer or open prison model. Appropriate new data sources may exist for Scotland.

> **Recommendation 81:** The ONS should explore possible new prison data sources for Scotland to allow for differential cost weights for different categories of prison, and data for Northern Ireland in the ambition of creating a UK-wide output measure.

The cost weights come from prison performance data which provides the expenditure of individual prisons annually for England and Wales. These are published by security category which are used to split the activity data. For England and Wales, the ONS uses the categories provided in the annual cost publication to allocate cost weights for each of nine categories:

- male dispersal (category A)
- male reception (category B)
- male Trainer (category B)
- male category C and Young Offenders Institute (YOI), Trainer and Resettlement
- male open (Category D)
- female closed
- female local
- female open
- male YOI young people (ages 15 to 17)

As the ONS currently lacks data disaggregating the prison population in Scotland by category, the average cost for English and Welsh prisons is used to weight the Scottish prison population figures. These matching activity and expenditure figures will then be used to create the Cost Weighted Activity Index (CWAI) for prisons output.

Probation

The current probation output measure uses number of offenders being supervised to capture activity. This is taken from data published by the MoJ [see link 74, Annex I]. Data are currently only available for England and Wales. Currently only a single activity measure is used with all probationers given the same weight.

However, as this measure does not reflect differences in the cost of different groups of probationers, the Review developed a new probation output method to improve measurement of productivity by splitting the activity into two categories of probationers:

- community order and suspended sentence order
- on licence

Community order and suspended sentence order probationers will be summed into a single activity. Relative weights will be assigned to these two activity categories using the marginal unit cost of each group of probationers provided by MoJ, allowing the ONS to create a CWAI for Probation. However, as these data are available only for financial year ending 2024 the ratio of unit costs for the two groups will be fixed over time. The cost weight for probation output relative to other aspects of POS will continue to come from HM Treasury data, meaning the average unit cost of probationers will still vary over time.

> **Recommendation 82:** The ONS should continue to work with the Ministry of Justice to obtain expenditure data that enable the relative weights of different groups of probationers to vary with time in future years.

> **Recommendation 83:** The ONS should explore the data available on probation services for Scotland and Northern Ireland with a view to creating a complete UK measure for probation output.

Criminal Courts

The output measure used by the ONS for criminal courts is a CWAI covering the activity of the courts of first instance for criminal offences, that is, the Magistrates' Courts and Crown Court. Magistrates' courts deal with cases sanctionable with up to 12 months in prison and motoring offences. The Crown Court deals with more serious or complex cases, such as serious assault or large fraud. The activity of appeals courts is not included in the output measure as firstly the number is small, but secondly, if the desired outcome is successfully completed trials, including the costs of appeals but not an additional output metric allows the ONS to see the system cost of delivering that outcome.

The Crown Court

The current output measure for the Crown Court uses three activity categories: net committals for trial, net cases for sentence and net appeals. These activity series have been forecast since 2018 and the relative weights for the three activity categories uses a set of unit costs from 2000. The Review has developed an updated cost weighted activity measure for the output of the Crown Court. The new output measure uses the criminal court statistics quarterly [see link 75, Annex I] published by MoJ, who have recommended using disposals as the activity measure.

The analytical tool 'Pivot Table Analytical Tool for England and Wales' [see link 76, Annex I]. published alongside the activity data includes data on the average number of hours taken for each type of court activity (hearing type, offence type and plea). By splitting by the hearing type and plea the ONS is able to weight the different types of activity with their own cost weights and is therefore able to better capture changes to the landscape of the criminal courts system.

The data are available from 2014 and, where unavailable, the existing activity data series will be used to generate the output index for earlier years.

Note that splitting by offence type as well as hearing type and plea has also been explored. The resulting index varied little from that split by hearing type and plea alone, yet the resource demand of splitting by offence type was significantly larger. As such, the Review has discounted this option.

There are four hearing types (triable-either-way, indictable only, committal for sentence and appeals) with their four respective plea options. Both 'Triable-either-way trials' and 'Indictable only trials' hearing types can have a plea of Guilty, Not guilty, No plea entered and dropped case. This is not applicable to 'committed for sentence' and 'appeals' hearing types. Where a plea is not applicable (non-trial hearings), the hearing category will be taken as a whole leading to a total of 10 activity categories.

To weight these ten categories, the Review has used the analytical tool 'Pivot Table Analytical Tool for England and Wales' discussed to estimate overall hourly cost for each of the hearings – plea type using overall average hearing time divided by overall expenditure, using HM Treasury data, which includes both labour and non-labour costs. Other MoJ expenditure data provided was considered, but only contain labour costs. With hourly cost calculated, the differential in the number of hours per hearing type or plea can be used to get a unit cost for each category.

Magistrates' courts

For magistrates' courts only a single activity measure is used, defined as caseload. This activity series has been forecast since 2011. As with the Crown Court, the ONS used disposals as the activity measure, with this taken from the criminal court statistics quarterly [see link 75, Annex I] published by MoJ.

Due to data limitations, magistrates' courts activity will be taken as an aggregate measure.

The activity data used to improve these criminal court output measures covers England and Wales only. Further work is required to identify suitable data for Northern Ireland and Scotland to produce a complete UK measure.

Cost weighting

Weights to aggregate the various components are derived from the Online Systems for Central Accounting and Reporting (OSCAR) dataset, HM Treasury's expenditure dataset used for Central Government spending in the UK National Accounts and much of PSP inputs. COFOG 03.3 refers to 'public order and safety: law courts', with a further breakdown derived by the ONS for the Crown and magistrates' courts expenditure respectively. As this breakdown is coded using OSCAR segment codes according to a methodology set up when OSCAR started, it is believed these breakdowns are out of date and likely result in both the Crown Court and magistrates' courts being underweighted relative to the other services in POS.

Work is ongoing with MoJ to obtain more suitable expenditure data to address this issue.

> **Recommendation 84:** The ONS should continue to work with the Ministry of Justice to obtain expenditure data to derive improved weights for individual law courts services.

> **Recommendation 85:** The ONS should explore the data available for Scotland and Northern Ireland with a view to creating a complete UK measure for criminal court output.

Legal Aid

Legal Aid is the provision of legal services to those who require them in and around court hearings. In this area the ONS only measures public provision of Legal Aid, rather than privately provided legal advice. This is because the consumer of these privately provided services has 'opted out' of using the public model (either voluntarily or involuntary, such as eligibility criteria). This is the same as not including private schools within the measure of education services: the opting out has removed the function from public services.

The Review has developed a new measure of Legal Aid output. This greatly increases the coverage from four lower crime legal aid activity categories to:

- all lower crime categories which represents Legal Aid work carried out in police stations and magistrates' courts

- all higher crime categories which represent Legal Aid work carried out in the Crown Court and above
- all of civil legal aid such as Special Children Act Proceedings

The new output measure continues to use the MoJ's Legal Aid statistics tables quarterly publication [see link 77, Annex I]. This source publishes both activity and expenditure using matching categories, which enables the construction of a CWAI. The ONS is expanding the scope of use of this source to encompass almost all legal aid output activity. The new method will take the most disaggregated data available, increasing the number of categories from 4 to 72, as this provides the greatest level of granularity and means the CWAI is more sensitive to changes in higher-cost activities and less sensitive to others.

> **Recommendation 86:** The ONS should explore the data available for the Scotland and Northern Ireland with a view to creating a complete UK measure for Legal Aid output.

Alongside this, the Review is assessing the concept of Legal Aid being included as intermediate consumption in the wider criminal justice system. Whilst the ONS treats Legal Aid as a standalone function with its own outputs, increasingly the Review has accepted the argument that Legal Aid cannot operate outside the wider court system: if the Legal Aid solicitor does not arrive, the case is adjourned, in the same way that if the judge is unavailable the court session will not commence.

The Review therefore recommends the ONS considers the possibility of linking legal aid to criminal courts data and using expenditure on this within the weighting of the delivery of justice services more widely, removing the distinct Legal Aid output measures. Feedback on this proposal would be welcomed by the Review.

> **Recommendation 87:** The ONS should undertake further research into whether Legal Aid should be considered an intermediate stage in the provision of justice as opposed to an output in itself.

12.4 Improvements to quality adjustments

The Review focussed on the existing quality adjustments for criminal court (Crown and Magistrates' Courts) output within POS. The quality adjustments, which are given equal weight, applied at present are:

- **timeliness:** The timeliness adjustment relates to the average time taken for criminal cases to reach completion, on the basis that the delivery

of a sentence in a timely manner is favourable. This is measured through the reciprocal of mean time until completion.

- **recidivism:** The recidivism (or re-offending) measure is based on the number of proven re-offences, adjusted for the 'offender mix' of the cohort (i.e. their likelihood to re-offend), weighted to account for the severity of the proven crime types, and expressed as a rate per person. This is calculated as the observed reoffending rate for a particular cohort plus any difference between the Offender Group Reconviction Score (OGRS) 4/G score in that cohort and a baseline cohort in 2011.

Areas considered for improvement were:

- Improvement of the granularity of the timeliness quality adjustment in line with the new criminal court output measures being developed through the Review (see Improvements to outputs estimates).
- Introduction of an additional quality adjustment to account for change in the case backlog in the criminal courts service.
- Re-introduction of a recidivism quality adjustment to account for the period when the adjustment was held constant due to the impact of the coronavirus pandemic on the source data.
- Review of weighting levels for both adjustments.

Improving granularity of the timeliness quality adjustment

Data for the existing timeliness adjustment are available from 2011; the current data for Magistrates' timeliness have been used since inception, while Crown courts changed to an improved method in 2014. The adjustments for both court types are:

- magistrates' courts: Average number of days between charge and completion
- crown courts: Average number of days from case being sent by Magistrates to the Crown courts, to completion in the Crown courts

These adjustments are then indexed and applied to the respective courts output data.

Criminal cases do not always take a linear path through the courts system. Depending on their severity, Magistrates can send cases to Crown Courts either before trial, dependent on the nature of the crime and the plea, or after a trial, if they believe the sentence should exceed their sentencing powers, so the contribution of both courts to the timeliness of cases can be intertwined and difficult to fully reflect in this adjustment. To ensure the ONS factor this complexity into its adjustment, the Review identified relevant data from the MoJ 'timeliness tool' [see link 78, Annex I] which puts substantial data into the public domain. The ONS and MoJ have collaborated to ensure that the most appropriate source for the variables are included into the quality adjustment.

The MoJ 'timeliness tool' also allowed the opportunity to improve the granularity of the timeliness quality adjustment in line with the new case disposal data being used to derive the criminal court output measure. As a result, the ONS will now be able to ensure the timeliness adjustment is reflective of the underlying mix of case disposal activity at Crown Court (i.e. aligning with the relative proportion of both case type and plea), and incorporate Triable-either-way, Indictable, Committed for Sentence and Appeals case types.

Both indictable and triable cases are further disaggregated by plea types (not-guilty, guilty plea, no plea entered, dropped case). Average timeliness was used for appeals and those cases where no plea is entered (as these are excluded from the MoJ timeliness calculations). This composite timeliness adjustment better accounts for the differing case complexity and more serious cases by considering the respective timeliness and relative expenditure as opposed to using a simple average based on the number of cases alone.

The Review also analysed whether a similar, granular approach to timeliness quality adjustment is possible for Magistrates' Court output (see Recommendations for further work).

> **Recommendation 88:** Timeliness data by case type should be used to produce a more granular quality adjustment that aligns with the improved Crown Court output measure.

Consideration of case backlog

While the timeliness adjustment is used to account for changes in the speed at which cases progress through the courts system, the Review has also considered whether an additional adjustment should be made to account for change in the volume of outstanding cases (backlog). It was felt that the term 'backlog' implies that all cases are somehow past their due date, whereas the published data refers to the 'open caseload', and as such reflects both 'work-in-progress' as well as backlog.

The Review has concluded that there is a clear relationship between backlog and timeliness (i.e. where backlogs rise, timeliness generally gets worse) and that including both risks compounding the impact of the quality adjustments. The Review does not recommend including an additional backlog quality adjustment.

Re-introduction of recidivism adjustment

The proven re-offending data used within the PSP measure were materially impacted because of the disruption to court proceedings during the coronavirus pandemic as re-offending is measured through re-convictions, which requires courts holding cases. With courts furloughed, the ONS made the decision

to hold the recidivism quality adjustment constant at the level applying to the last cohort unaffected by court closures over the lockdown periods. The rate has now returned to a level within historical ranges, and the Review considered how best to address the reintroduction of the adjustment, over the coronavirus pandemic impacted period.

Several options were explored including (1) retaining the current approach of keeping the adjustment constant, (2) removing the adjustment completely during the impacted period and using the published re-offending data, and (3) re-introducing the adjustment based on published data for the 2022 cohort and using linear interpolation to estimate the impacted period (2018 to 2021).

The Review concluded that option (3) was most appropriate. As the resulting trend is one of gradual decline in the severity-adjusted rate of re-offending the ONS uses in its measure, this will result in a positive impact on the quality adjustment through the impacted period. This option was seen as preferable to the volatility which would have been introduced to the series from removing the adjustment completely during the impacted period.

> **Recommendation 89:** A straight linear interpolation should be applied to the re-offending quality adjustment applied for 2018 to 2021 where the coronavirus pandemic interrupted the normal provision of data.

Review of weighting levels

In light of the re-introduction of a recidivism quality adjustment, the Review considered the relative weights attributed to each of the components of the quality adjustments applied to criminal courts output. When considering a baseline framework for capturing a quality adjustment measure for Courts productivity, the requirement was a measure of intrinsic core values which hold stable across time for a fair measure, resistant to the changes in policy of the government of the day.

This framework is provided by the five principles of sentencing, introduced in the 2003 Criminal Justice Act, which lists the legal purposes of sentencing by the court as the following:

a. the punishment of offenders.
b. the reduction of crime (including its reduction by deterrence).
c. the reform and rehabilitation of offenders.
d. the protection of the public.
e. the making of reparation by offenders to persons affected by their offences.

Two of these principles (b and c) relate to recidivism while the remaining principles can be considered to have a greater association with the timeliness component of the measure. Noting this, there is no reason in the 2003 Criminal

Justice Act to assume any of these five purposes has greater importance than others. The Review, therefore, recommends a revision to the equal weighting of the quality adjustments, which treats the five purposes as equivalent to each other and will be incorporated into the Spring 2025 annual ONS statistical release as follows:

- 40% weighting to Recidivism.
- 60% weighting to Timeliness.

> **Recommendation 90:** Court quality adjustment weights should be revised to be 40% recidivism and 60% timeliness.

12.5 Recommendations for further work

Development of direct labour measures for Criminal Justice services

Prison and Probation: Initial investigations and discussions with MoJ highlighted published HM Prisons and Probations quarterly workforce statistics [see link 79, Annex I]. Based on this, the ONS extracted FTE data for HM Prison and Probation employees dating back to 2010 for England and Wales.

Accurate historic salary data prior to 2018 were not available due to a change in payroll systems, and sufficient historical data are not available to produce average costs. It was therefore not possible to obtain bands that aligned appropriately with published workforce statistics. Ongoing engagement with MoJ indicates that an extended back-series of FTE data and more granular management information on the prison workforce may be available in the future, if so this could be used to develop an enhanced direct measure.

In terms of developing a direct labour measure for the Probation service, a lack of disaggregated salary data, combined with the part-privatisation and renationalisation of services in the recent past resulted in the Review concluding that probation labour input should continue to be measured indirectly at present.

> **Recommendation 91:** Work should continue on identifying appropriate data to develop a direct measure of labour input for the prison service component of Public Order and Safety.

Courts: This is another area where the Review considers a direct labour measure will be possible in the near future. MoJ publish management information detailing FTE for services, which covers England and Wales. MoJ has advised that data prior to 2011 are not readily available, may not be robust and would require significant resource to obtain.

While the Review derived a measure of direct labour using FTE and available salary band data, other labour components within HM Courts and Tribunals Service were not captured, and there were expenditure items which may already be captured within the intermediate consumption inputs measure.

A comparison of the direct measure the Review derived with the existing indirect measure showed that while the long-term trends for both indices are broadly similar, the direct labour Index had more pronounced negative annual growth between 2012 and 2017. The Review recommends retaining an indirect labour approach while engaging with MoJ to ensure no double-counting in both labour and intermediate consumption by moving to a direct labour approach.

> **Recommendation 92:** Work should continue on developing a direct measure of labour input for the courts service component of Public Order and Safety, ensuring no double-counting with intermediate consumption input in this area.

CHAPTER 13

Tax administration

Tax Administration accounts for approximately 0.6% of total public expenditure on services captured within the scope of the Review (therefore excluding transfers). It reflects government activity in the administration and collection of taxes (such as Income Tax, National Insurance, Value Added Tax (VAT) etc.

By far the largest contributor to this service is HM Revenue and Customs (HMRC), with other activity provided by local authorities in terms of collecting Council Tax. However, not all HMRC activity is included, for example Tax Credits and Child Benefit relate to Social Security, rather than Tax Administration, although there are challenges in identifying these in the available spend data.

Tax Administration is currently consolidated within the "Other" grouping in the Office for National Statistics (ONS) public services productivity (PSP) statistics. This "Other" grouping contains only services calculated on the "inputs = outputs" basis and as a result, Tax Administration productivity is currently constant by assumption.

The Review has developed a provisional method focused on HMRC taxes in the first instance to directly measure tax administration output, and hence productivity for the first time. Although relatively small by expenditure weight, exploring this area alongside social security administration has provided the Review valuable insights which can be applied more widely in terms of how to best address administrative functions of this type.

As the Tax Administration method is new, the ONS determines this measure to be 'Official statistics in development' [see link 80, Annex I]. Once this has undergone further testing, it will become an 'Accredited official statistic' in line with the rest of the estimates.

The Coronavirus Job Retention Scheme was launched in April 2020 and aimed to protect jobs affected by the coronavirus (COVID-19) pandemic. This is not related to taxation as a service and expenditure is therefore recorded within "Other" groupings. Further consideration of this matter is advised in relation to whether this scheme should be considered social security administration.

13.1 Core methodological and data challenges identified

The Review has collaborated with academics and HMRC experts to develop innovative approaches to measuring Tax Administration productivity, whilst ensuring comparability with other services and adhering to UK National Accounts standards.

Inputs

Tax Administration activity, as with other services discussed in this Review, suffers from sharing a Classification of the Functions of Government (COFOG) code 1.1.2 (Financial and Fiscal Affairs) with other non-tax activity such as the management of public funds and public debt. In terms of accessing data on inputs, it is therefore not independently identifiable or measurable in the national accounts, despite being an essential service undertaken in some form by almost all governments, often by a distinct body with responsibility for tax collection alone (rather than wider public finance management).

The Review has considered viable alternative data sources to measure inputs, and the ONS has separately proposed that tax administration be considered as a potential separate category within the COFOG structure in its response to the United Nations Statistical Commission's consultation [see link 81, Annex I] on updating the COFOG. As with other services, implementation of any such change would need to be coordinated and funded across government.

Importantly, once a data source is identified, the scope of activity for which equivalent output metrics needs to be considered and input split accordingly.

Outputs and outcomes

Alongside gaining clarity on inputs, the Review has reflected on how to best assess measures of output, which enable productivity to be measured over time. To achieve this the Review has identified the need to balance two perspectives. Firstly, the need to identify a suitable measure for changes in the amount of service activities delivered and weights for these. Secondly, a reflection of the outcome of relevance from Tax Administration, which is to raise the right amount of revenue to fund government public services whilst spending as little to achieve this as possible.

The key challenge in this area therefore relates to traditional metrics of activity, such as the number of taxpayers' casefiles processed poorly reflecting the outcome desired, which is the correct collection of tax. Some tax-files may result in no tax being paid, and the relationship between cost and taxes collected fundamentally relies on the design of the tax regime. Real Time Information Pay as You Earn (PAYE) is a relatively low cost form of collecting income tax and national insurance, whereas Self-Assessment is materially more expensive. Conversations with HMRC revealed their operational preference and experience in using a 'cost of collection' model, focused on the expenditure per pound of tax collected. The Developing outputs estimates section discusses this in greater detail.

One of the key issues arising however is the bounded nature of the desired outcome. Taking Healthcare as a counterfactual, if more operations are delivered, this would be perceived as an increase in output; each additional operation may have diminishing marginal utility, but as citizens are unlikely to volunteer

for operations they do not need, one can assume there is no upper bound on demand. This is not the case for Tax Administration. The tax regime is designed to deliver a certain value of tax which is considered necessary by the fiscal authority to achieve macroeconomic aims. If HMRC were to collect too little tax compared to the theoretical liabilities (the 'tax-gap [see link 82, Annex I]') or too much (effectively 'over-taxing' citizens), either of these outcomes would be viewed as having social detriment.

Therefore it cannot be the case that if HMRC was to continuously increase the tax collected as a fraction of Gross Domestic Product that this would be seen as a positive outcome, if this was not the mandate set for it by the fiscal authority. In other words, unlike private activity where, as long as there is a positive price extra output is always 'a good thing', HMRC could over-supply tax revenue and the fiscal authority would, if it was in the position of paying for this service, not be willing to pay for this – the market equivalent price for this activity would be zero and hence the additional output would have no value. As such it cannot be the case that the raw tax take can be used to quality adjust direct measures of output.

This Review therefore proposes an additional model to 'revenue-adjust' direct measures of output in lieu of a more complete quality adjustment which the Review has begun to develop with HMRC, but did not reach a state of suitable maturity for implementation at this time.

> **Recommendation 93:** The ONS should 'revenue-adjust' direct measures of Tax Administration output whilst developing, with HMRC, appropriate quality adjustments, including using factors such as the 'tax gap' to reflect the difference between the desired and achieved levels of tax collected, customer satisfaction and call waiting times.

13.2 Developing inputs estimates

The following data sources were considered but the Review deemed them not suitable for PSP purposes due to a lack of specificity in their coverage.

- the ONS' UK National Accounts' data [see link 83, Annex I]: It is not possible to measure tax administration inputs via this traditional method because it does not have a unique COFOG code. Therefore, alternative sources were explored in developing a suitable approach to measuring Tax Administration inputs for PSP.
- OSCAR (Online System for Central Accounting and Reporting): this cross-government management information system provides HM Treasury with key spending data from each department. For the Review's purposes, it was deemed unsuitable because of its inability to differentiate HMRC expenditure on Tax Administration activity from other financial and fiscal affairs within COFOG 1.1.2.

- HMRC Annual Report and Accounts [see link 84, Annex I]: these accounts did not allow the Review to specifically identify HMRC's Tax Administration activity, and detailed input breakdowns were also not fully or appropriately defined.

HMRC Cost of Collection as a data source

Expenditure data used in HMRC's 'Cost of Collection' statistics was the only identified source that provided suitable coverage on Tax Administration activity. 'Cost of Collection' statistics are published annually in the HMRC annual reports and accounts [see link 84, Annex I] and are presented as the ratio of Tax Administration expenditure per pound collected in tax revenue.

The expenditure component is calculated by allocating expenditure to Tax Administration functions based on reported staff time spent on those activities. Only the ratio is published (available via the 'Historical data series' [see link 85, Annex I] file), with the expenditure component being provided directly to the ONS.

Cost of collection expenditure figures do not provide a breakdown by input components, so OSCAR data are used to estimate the proportion allocated to labour, intermediate consumption and consumption of fixed capital within this expenditure. Input breakdowns in OSCAR are preferred over national accounts data because it can be assumed that Tax Administration is more homogenous to wider HMRC activity than it is to wider activity under COFOG 1.1.2.

Due to the availability of these data the new Tax Administration productivity measure begins in 2018, and the ONS plans to publish these from Spring 2025 as part of the Public service productivity: total, UK [see link 86, Annex I]. An equivalent amount of expenditure will be removed from the "Other" grouping series from 2018 to avoid double counting.

> **Recommendation 94:** The ONS should publish provisional estimates of Tax Administration productivity in Spring 2025 and adjust the "Other" grouping accordingly.

13.3 Developing outputs estimates

HMRC has worked closely with the Review, and shared information on the range of performance indicators published in their Annual Report and Accounts, including cost of collection which is their preferred measure of the productivity of taxation. The inverse of the cost of collection, defined as the ratio of revenue collected over Tax Administration expenditure is a measure of the productivity of Tax Administration.

The Review decided not to use this measure because it relates the outcome directly to the inputs (the spend) without utilising as with other services a direct measure of output. Whilst there is a precedent for this in adult social care, this

was not viewed as an optimal approach, in part because tax revenue is viewed by the national accounts as a 'transfer' rather than value added and using the value of a transfer as the measure of output of public (and private) industries is not permitted.

In consultation with experts in HMRC, the measure developed by the Review uses an estimate of the number of taxpayers for each regime as a direct measure of output, a similar approach to that already used for social security administration, where the number of benefit claims and caseload processed is used as the activity measure. However, it should be noted this approach prices in 'false positives' into the output measure. That is, there are a number of casefiles processed where no tax is collected due to the circumstances of the tax-payer. For example, PAYE will automatically report on all staff in a payroll to HMRC, not just those eligible to pay tax in the relevant time period. Equally whilst all self-employed workers will complete a Self-Assessment, whether they have any additional tax to pay under that regime will depend on their activity in the year and their eligibility for tax reliefs etc.

This 'over-head' of 'false positives' is obviously dependent on tax policy (e.g entitlements and thresholds) and clearly in this context changes in such policies would result in productivity falling or rising dependent on whether the number of 'non-tax-generating' casefiles increases or falls. Whilst the Review accepts this 'overhead' is a necessary cost of ensuring tax is properly collected, its impact on productivity estimates is an issue which would merit further research once statistical estimates are being created.

Importantly, therefore, this direct measure of output requires a formal treatment to relate it to the desired outcome, which is the collection of tax.

The Review has been able to source direct output coverage of eleven taxes accounting for a broadly consistent 88% to 89% of total HMRC Tax Administration expenditure for the period 2018 to 2023 using the HMRC's 'Cost of Collection' measure in the calculation of inputs. This covers Income Tax, VAT, Corporation Tax and National Insurance Contributions and several other taxes.

Table 4: Taxes included and basis of measurement

Tax	Activity measure
Income Tax	Number of taxpayers
National Insurance Class 1 (employees)	Number of taxpayers
National Insurance Class 4 (self-employed)	Number of taxpayers
Corporation Tax	Number of taxpayers
Capital Gains Tax	Number of taxpayers
Value Added Tax	Number of registered traders
Insurance Premium Tax	Number of registered traders
Air Passenger Duty	Number of operators
Landfill Tax	Number of operators
Climate Change Levy	Number of operators
Aggregates Levy	Number of operators
Inheritance Tax	Liabilities upon death (Estates eligible for Inheritance Tax) plus Lifetime Transfers

Source:
HMRC publication [see link 87, Annex I] "Number of taxpayers and registered traders", with the National Insurance component being provided directly to ONS.

Unit costs can be calculated by dividing the expenditure on administering each tax by the number of taxpayers of that tax. Detailed cost of collection data has been supplied by HMRC to the ONS, which will enable the calculation of unit costs for each of the eleven taxes in Table 4 (treating Class 1 and Class 4 National Insurance as two instances of a single tax). This may be a topic worthy of further investigation as Income Tax is collected via PAYE and Self-Assessment and National Insurance is collected via the same two channels, which may be better understood in a more disaggregated fashion.

> **Recommendation 95:** The ONS should work with HM Revenue and Customs to continue to ensure data on different taxes is utilised in as granular a fashion as benefits productivity estimates.

There are a number of taxes without activity data, which account for around 11% of HMRC administrative expenditure in recent years. These are treated on an "inputs = outputs" basis at this time due to data availability. The inclusion of this "inputs = outputs" component enables the output measure to account for changes in the administration of taxes that lack activity data. However, as the productivity of the administration of these taxes cannot be measured, the inclusion of the "inputs = outputs" component has a slight 'dampening' effect on productivity, slightly reducing the size of productivity changes.

While this measure covers nearly 90% of the output of tax collection, it does not cover all tax-related administrative activity undertaken by HMRC. In particular, it does not cover customs, borders and trade activity and tackling non-compliance (i.e. tax evasion and avoidance). As explained in Chapter 14 the primary non-tax elements of the HMRC portfolio, such as Tax Credits and Child Benefit are accounted for in social security administration. The developed metric currently also omits local authority functions in relation to council tax and business rates. Initial investigations identified data on customs duties and import taxes and the ONS intends to further investigate the possibility of expanding to full coverage in these services.

> **Recommendation 96:** The ONS should explore how to complement existing HM Revenue and Customs data to complete coverage of Tax Administration activities.

These metrics are combined using a cost weighted activity index. Importantly, as with other areas, and as discussed in Chapter 3, cost weighting is widely used and is generally the only feasible method, but that does not make it conceptually strong.

The Review has devoted significant resource to understanding how best to address a problem also observed in other services and discussed in Chapter 3 of this report. Namely that the relative costs of undertaking various public services are only a proxy for the value of that activity. In short, this approach captures well the changes to the rate of technical efficiency (productivity gains from improving existing processes) but struggles more when trying to capture allocative efficiencies (gains from changing to an entirely different process). As such this chapter discusses how to take account of tax revenue.

13.4 Quality adjustment

The outcome which matters in relation to Tax Administration is the collection of the right amount of tax, as defined by the decisions taken by the fiscal authority. However, tax revenue is viewed by the UK National Accounts as a 'transfer' rather than value added and transfers are excluded from the measurement of output of public (and private) industries. Therefore, the Review has had to consider the most appropriate way to balance the inclusion of tax revenue whilst not breaching national accounting rules if the ONS wishes these metrics to flow into the national accounts data.

This issue has traditionally been avoided by focussing on the administration of taxation. This administration function has been included as part of general government final consumption expenditure and contributes to overall economic output in the UK National Accounts and therefore the PSP measures, and the purpose of this Tax Administration is to collect revenue for government. However, this approach fails to account for quality change in the service.

The Review has reflected on the gap between a measure of administrative activity, in terms of the number of taxpayers and the concept of revenue collected against the tax plans of the fiscal authority.

Pending this process, the Review has developed a temporary adjustment to direct output. This applies a 'revenue adjustment' to the output index that weights tax administration activity by the revenue raised from different taxes, divided by the spend to collect this revenue. Whilst this uses a simple measure of tax revenue it enables a better reflection of the relationship between taxpayers, spend per taxpayer and revenue per taxpayer. This enables the measure to account for allocative efficiency, whereby a shift towards lower cost-per-£-raised taxes will increase productivity.

Data for revenue adjustments can be sourced from the publication HMRC tax receipts and National Insurance contributions [see link 88, Annex I] for the UK which provides the receipts by financial year for the taxes covered in this analysis.

Although applying a revenue adjustment is in keeping with the spirit of HMRC's preferred measure of value for money, which is the ratio of tax revenue relative to the cost of collecting it (referred to as the "cost of collection"), other aspects of quality in tax collection are not currently addressed by the measure. The ONS intends to investigate further developments that could be made to account for changes in quality. In particular, the ONS is interested in exploring the possibility of using the "tax gap" [see link 82, Annex I] as a quality adjustment. This is the difference between the theoretical amount of revenue that should be collected by HMRC and the amount that is actually collected.

Recommendation 97: The ONS should investigate further developments that could be made to account for changes in quality such as fraud and error, customer satisfaction and waiting times, but recommends that treatment of the 'tax gap' should be the priority area for focus.

13.5 Devolved and Local Government

The developments in this Review reflect tax collection by HMRC only, excluding Customs. Locally administered taxes and taxes collected by devolved governments have not been prioritised.

13.6 Recommendations for further work

While substantial progress has been made in the Review, the measures developed only cover taxes administered by HMRC, excluding Customs Duties. Locally administered taxes, such as council taxes and business rates, are not currently included. Further work to expand the scope to all taxation activities and to consider quality adjustment of these areas will depend on the exact form of the direct measures of output which are chosen for these.

Recommendation 98: The ONS should review data on taxes administered locally and in devolved governments in relation to inputs and outputs, to explore the feasibility of further improving coverage of Taxation Administration.

In addition to further work to develop coverage and quality adjustment, the time period covered by Tax Administration productivity measure should be considered. The ONS' new initial Tax Administration productivity measure will begin at 2018. This is due to limited input data being available prior to this, however it may be possible to extend this further to 2010 if Customs were to be included and aspects such as weights estimated using more recent data. The ONS should continue to work with HMRC to investigate data sources and potential solutions to this.

Recommendation 99: The ONS should investigate the feasibility of extending the time period for Tax Administration productivity measure to pre-2018.

CHAPTER 14:

Social Security Administration

Social Security Administration (SSA) is one of the smaller services, accounting for approximately 1.7% of total public service productivity expenditure. It covers the administration of benefits to individuals and households including Universal Credit, Housing Benefits and more.

In the UK in financial year ending 2023 around 87% of expenditure on SSA was spent by the Department of Work and Pensions (DWP), with the residual being the responsibility of either a) HM Revenue and Customs (HMRC) in relation to the administration of Tax Credits and Child Benefit, noting that the process of migration of Tax Credit recipients to Universal Credit has been continuing for some time, or b) top tier local authorities in relation to the administration of Housing Benefit.

Historically, Atkinson (2005) [see link 1, Annex I] and subsequent Office for National Statistics (ONS) research made progress in delivering a direct measure of SSA output, driven by the number of case-files being processed, or maintained on the caseload, but without any quality adjustment. As described in Chapter 3, the value of the benefits themselves is considered a transfer and is not a direct measure of output. This measure, however, was a partial and incomplete solution.

Further to this, the historical methodology was suspended in 2018 following the introduction of Universal Credit, which caused significant methodological issues to arise given its rationalisation of various benefits into one. The existing measure, whilst strong in terms of technical efficiency (e.g. reducing the costs of processing each type of existing case-file), was insufficient to address allocative efficiency (e.g. the merger of various types of benefit into a single case file).

14.1 Prioritisation of issues

The Review prioritised resolving immediate issues relating to the introduction of Universal Credit and the coronavirus (COVID-19) pandemic, rather than longer-term challenges relating to the measurement of Tax Credits and Child Benefit. As a result, the measures developed focus on DWP activity, covering output from 2013 (upon development of the output model to account for Universal Credit) and quality adjusted output from 2008. The Review also investigated DWP and HMRC inputs pre-2013 to gain an understanding behind historic volatility, however no methodological changes were implemented from this.

The in-depth analysis of input and output data of the Review (back to 1997 and 2013, respectively), however, indicates that expenditure on Tax Credits and Child Benefit administration by HMRC is not separately identifiable within data returned via the HM Treasury Online System for Central Accounting and Reporting (OSCAR) system, potentially recorded as costs of General Public Services (COFOG 1), which are not included in the SSA classification.

As a result, administrative costs for Tax Credits and Child Benefit are likely not being accounted for in the SSA inputs model. Therefore, the ONS do not include the output measures of Tax Credit in the calculation of SSA. Including this information in the outputs, but not the inputs could lead to misleading changes in productivity estimates. The Review believes that the DWP share of SSA has been historically over-stated as the large amount of HMRC Tax Credits have not been included, which suggests the historical method applied may have been subject to bias.

As a priority, the Review has developed a new method, working closely with DWP and HMRC, that better captures the administrative activity in the transition from legacy benefits to Universal Credit, particularly from 2018.

The new measure addresses the challenge of addressing allocative efficiency, but this means that comparisons between the existing measure and the new one are relatively challenging to make and interpret. Previously compositional changes in the caseload demographics may appear erroneously as productivity movements, although this remains the case for those elements which remain cost weighted.

However, this improved methodology currently only captures DWP activity whilst inputs and outputs data on Tax Credits and Child Benefit cannot be sourced. Equally, the Review has not had the resources to explore local authority data on Housing Benefit inputs and outputs.

The data which the ONS will release in the Spring 2025 statistical bulletin will therefore focus on DWP activity but are expected to be liable to substantive revision, particularly as Tax Credit information is brought into the calculation (importantly, in terms of both inputs and outputs), which is likely to change the relative weights of DWP and HMRC activity within this series, particularly historically. These data should therefore be considered as 'Official statistics

in development' [see link 80, Annex I]. Once this has undergone further testing, it will become an 'Accredited official statistic' in line with the rest of the estimates.

The Coronavirus Job Retention Scheme was launched in April 2020 and aimed to protect jobs affected by the coronavirus pandemic. This is not currently considered as part of the SSA services and expenditure is therefore recorded within the "Other" grouping. Further consideration of this matter is advised

> **Recommendation 100:** The ONS should undertake further work to identify the best treatment of the Coronavirus Job Retention Scheme.

This development of the output measure of SSA is supplemented by a new quality adjustment reflecting DWP Fraud and Error. Again, HMRC expenditure and performance data on Fraud and Error related to Tax Credits should also feed into this metric, as well as wider measures of quality across the SSA service.

The partial coverage of the quality adjustment means this should be considered 'experimental' until further notice.

Finally, the Review has introduced an improved intermediate consumption deflator in line with the UK National Accounts development, impacting slightly on inputs. These improvements will be incorporated into the upcoming annual publication in Spring 2025.

14.2 Methodological and data challenges addressed by this Review

Setting aside the longer-term issues related to accessing data on Tax Credits, Child Benefit and Housing Benefit addressed at the start of this chapter, the major methodological issue tackled by the Review was the introduction of Universal Credit to rationalise numerous legacy benefits into one, which was exacerbated by the impact of the coronavirus pandemic.

The three main challenges considered by the Review can be summarised as:

- Inputs volatility: SSA inputs show considerable volatility over time, more so than other services. The ONS explored the data sources and systems to assess the main drivers of these movements to determine their legitimacy, and if any methodological changes need to be considered.
- Output: Developing a method to appropriately treat the transition from multiple legacy benefits into one single benefit (Universal Credit) while considering conceptual implications around the capture of technical versus allocative efficiencies, and updating the methods used for those benefits which have not been replaced by Universal Credit. This has an obvious implication that there is a need to also review those benefits not impacted by the change to ensure wider consistency.

- Quality Adjustment: Consideration of outcomes such as fraud and error, customer satisfaction and timeliness to better capture the quality in the delivery of benefits to recipients. The rationalisation of numerous benefits into Universal Credit presented an opportunity to improve the productivity of benefit delivery. However, under the traditional cost weighted method the relative values assigned to Universal Credit and its legacy benefits would have been based on their cost, so in the absence of commensurate cost savings, this scheme would have weakened, instead of strengthened, productivity estimates. As described in previous chapters, this is an example where existing methods failed to suitably address allocative efficiency improvements. Methods revisions in a number of services have, under the Review, been undertaken with a view of 'future-proofing' methods against this risk re-occurring.

The existing methodology was therefore judged to be severely limiting in the light of the development of Universal Credit, and prior to the coronavirus pandemic, the ONS suspended publication of SSA's direct measure of output. Though the intention was to rapidly develop an alternative improved methodology to deal with the transition, the emergence of the coronavirus pandemic shifted resource away from doing so, until the launch of this Review. In the meantime, the "inputs = outputs" convention was applied from 2018 onwards and, as a result, SSA productivity has been constant since.

14.3 Improvements to inputs estimates

The methods applied by the ONS to estimates SSA inputs have not changed since the Atkinson Review. This method is based on deflated national accounts data (including compensation of employees, intermediate consumption, and consumption of capital) for activity in the relevant service (which in this case is COFOG 10 – Social Protection).

The inputs data were reviewed in detail to understand the key drivers of volatility in the back-series. The Review investigated the compilation of the national accounts' COFOG 10: Social Protection data back to 1997 (when the series began) and corroborated that with HM Treasury's OSCAR database back to its starting point in 2012 to 2013.

Summary of SSA inputs volatility

The SSA inputs index begins in 1997 and year-on-year volatility is most evident between 2003 and 2013. The index's annual average growth rate in absolute terms before this period is 6%, during this period it is 13%, and after this period it is 4%. Further interrogation of the data showed that trends in intermediate consumption were driving the volatile movements in the inputs index, as opposed to labour or capital consumption.

Although a large increase was observed for local government intermediate consumption expenditure in 2008, it was clear that most of the volatility stemmed

from central government intermediate consumption expenditure. It particularly drives movements in the following periods:

- 2003: coinciding with the introduction of pension credit, the Income Tax (Earnings and Pension) Act, and establishment of Jobcentre Plus
- 2007 to 2013: coinciding with changes to the demand in claims for benefits such as Job Seekers' Allowance
- 2013 to 2015: coinciding with increased expenditure on 'Clinical and Medical' spend in OSCAR, which rose in accordance with an increase in Disability Living Allowance (DLA) claims, and the need for additional resources to meet the demand of assessing claimants' health conditions and thus eligibility for DLA

Local government intermediate consumption also adds to some of the volatility, rising sharply in 2008. This reflects the average rise of £9 a week for Housing Benefit claims, which led to increased administration costs (+£0.2billion). There was also a £0.7 billion increase associated with concessionary transport fares, coinciding with the extension of the English National Concessionary Travel Scheme to free off-peak national travel from 1 April 2008.

However, it is uncertain whether these costs are solely attributable to administrative burden (see Recommendations for future work) or may contain some element of the value of the benefit or subsidy for concessionary schemes. Whilst recognising these are small, given these components are not costs of the administration of these benefits, but rather the benefit themselves, this may merit further investigation.

> **Recommendation 101:** The ONS should review whether data exists to better understand, for Social Security Administration, whether non-administrative costs for concessionary fares and Housing Benefit are correctly recorded, and whether they are sufficiently impactful to merit further work.

Verifying data sources

Investigations were conducted into the compilation of UK National Accounts data to ascertain whether inputs movements from 1997 onwards are attributable to actual DWP spending data, or adjustments that are applied to the data by the ONS. More than two-thirds of the data feeding into SSA intermediate consumption, of which most of the volatility stems from, is actual DWP expenditure from the OSCAR dataset. The remaining third is made up of other data, including adjustments.

These adjustments are applied for reasons such as error correction, redistribution of quarterly expenditure across a financial year, or re-allocation of expenditure (e.g. allocating some defence expenditure to healthcare).

In all cases, adjustments are agreed and implemented as part of a collaborative exercise with HM Treasury on an annual basis.

Nevertheless, the adjustments account for a similar amount of volatility as the DWP data, despite having a much smaller share of contribution to intermediate consumption levels. The Review has investigated this and overall, is confident in the legitimacy of SSA inputs. The movements reflect changes in social security over time, they are predominantly driven by DWP-reported accounting, and the adjustments resolve clear oddities, such as in 2005 when the raw data appears unfeasibly low.

COFOG structure

The Review also investigated the COFOG structure within SSA, noting that it should cover benefits administered by DWP alongside Housing Benefit which is administered by local authorities and Tax Credits and Child Benefit, both of which are administered by HMRC. These non-DWP components, the Review has concluded, may be excluded from either or both inputs and outputs and therefore this appears to be an area for further improvement, as noted in Recommendation 95. Given this, the time series appears a valid estimate of those DWP services currently in-scope of the data collection, but in time should change from being DWP-focussed to SSA-focussed as data improves.

14.4 Improvements to outputs estimates

The ONS SSA output measure used data from DWP for benefits administration in Great Britain until 2018 when the previous measure ceased to be produced and the "inputs = outputs" convention was used due to the complexities in relation to the implementation of Universal Credit, specifically:

- The replacement of multiple "legacy" benefit schemes by Universal Credit.
- The changing composition of Universal Credit claims over time, as the transfer from legacy benefits to Universal Credit started with the simplest cases, with progressively more complex benefit claims transferred over time.

The first of these issues prevented the measurement of any productivity change resulting from the transfer to Universal Credit. If Universal Credit could deliver equivalent benefits at lower cost, then a conventional index, using cost weights, would place a lower value on the new Universal Credit activity, and any resultant fall in inputs would be matched by a fall in output, and productivity would not change. This is clearly incorrect: delivering the same benefits or support to citizens at a lower administrative cost should be seen as a productivity enhancement.

For the second issue, the application of a conventional index, using cost weights, may result in productivity appearing to fall as a result of more complex claims having a higher cost of processing, but no greater value in output.

Introducing use of benefit weighting as an alternative to cost weighting.

The new output measure developed by the Review therefore involves:

- Producing a benefit-weighted index for the administration output of Universal Credit and legacy benefits, where the changes in the number of claimants of each benefit are weighted by their average benefit payments.
- Producing a cost weighted index for the remaining benefits, which uses an update to the existing claims and load activity data and associated unit costs.
- Producing an overall index for SSA output by weighting together the two indices described in the bullets discussed using the total administrative expenditure on all benefits within each index.

Using a stylised example of benefit weighting, a benefit payment bundle of equivalent value from Universal Credit and legacy benefits would be given equivalent weight in output and so output would not change where a benefit claimant was given equivalent support under legacy and Universal Credit payment schemes.

This means that where there is a difference in the cost of delivering equivalent benefits under the legacy and Universal Credit schemes, this will manifest as a change in inputs but not output. Therefore, the productivity measure will account for changes in the cost of delivery resulting from the introduction of Universal Credit. This is an over-simplification, used to illustrate the theoretical approach. In reality, the legacy and Universal Credit systems are not always directly equivalent. The formula for the Benefit Weighted Activity Index is as follows:

Figure 7

$$L_t = L_{t-1} \times \frac{\sum_i b_{t-1,i} \times q_{t,i}}{\sum_i b_{t-1,i} \times q_{t-1,i}}$$

Where:

L = index

b = unit benefits (analogous to a unit cost, but weighting for benefits)

q = quantity of activity

t = time period

i = activity type

> **Recommendation 102:** The ONS should measure the output of the administration of Universal Credit and its predecessor legacy benefits using a benefit-weighted index to account for the allocative efficiency of the transition. This measure should be seen as provisional whilst undergoing further testing.

The legacy benefits included in the benefit weighted activity index include Employment and Support Allowance, Income Support, Jobseeker's Allowance and Housing Benefit. Tax Credits, which were administered by HMRC, are not included as the inputs used to process these are not accounted for in the SSA inputs measure due to challenges in isolating these inputs from other HMRC activity. The Review therefore excluded Tax Credits from the SSA output measure and the ONS intends to continue investigating the appropriate data sources for Tax Credits in future.

> **Recommendation 103:** The ONS should seek to acquire data to incorporate Tax Credits, Housing Benefit and Child Benefit in both output and inputs measures of Social Security Administration, while making appropriate adjustments to the "Other" grouping of inputs.

In terms of accounting for changes in the composition of claims in Universal Credit, the increase in the number of entitlements per household over time represents an increase in the administrative output of Universal Credit. The challenge is that the calculation needs to account for the changing composition of Universal Credit claimants in output.

This means it is inappropriate to treat the administration of the average Universal Credit claim as having an equal value over time. Activity for Universal Credit should therefore be the number of Universal Credit claimants (households) adjusted for the change in composition and therefore assumed administrative burden of these cases. In the absence of evidence on the relationship between claimant composition and administrative cost the method assumes a relationship between these using the benefit award amounts to approximate their value.

> **Recommendation 104:** The ONS should adjust the output of Universal Credit administration for changes in the number of entitlements included in payments, to account for the heterogeneity of cases.

These changes are planned for incorporation into the ONS Spring 2025 annual statistical release. The ONS will account for the changes in composition of Universal Credit claims by applying a composition adjustment to the activity for Universal Credit. This estimates changes in the composition of Universal Credit claims and therefore processing of Universal Credit payments, while controlling for uprating of Universal Credit payments over time to account for inflation. Detail is provided in Annex G.

Although Universal Credit was introduced in the financial year ending 2014, the existing SSA output index for the period 2013 to 2016 did not include Universal Credit. It should be noted the significant increases in compositional change to Universal Credit claims came into place post-2016, which along with the large-scale roll-out of Universal Credit coming after 2016, means the absence of a composition adjustment factor before 2016 should have a limited effect on output over this period.

However, activity and unit cost data provided by DWP has enabled the ONS to calculate a conventional cost weighted activity index including Universal Credit for the period financial years ending 2014 to 2017. Claims and load for Universal Credit activity is therefore included in the index for this period in the new measure but is not adjusted for compositional changes.

Given feedback to the Review suggests the easiest cases were moved to Universal Credit first, failure to weight-up more complex cases in this period should have less impact. This cost weighted activity index is then combined with the existing SSA output index for 1997 to financial year ending 2014 by retaining the growth rates for this earlier period and applying the growth rates from the new approach for financial year ending 2014 onwards.

For the benefits that are not Universal Credit or its legacy counterparts, a simple cost weighted activity index will continue to be used post-2016. This will retain wider comparability with other services, which continue to use cost weighting.

14.5 Improvements to quality adjustments

There are currently no quality adjustments in SSA productivity estimates. The Review explored several possibilities, namely use of a) DWP fraud and error rates, b) customer satisfaction surveys, and c) timeliness of processing claims. The Review concluded that only the first of these was suitable for implementation at this time.

Fraud and error rates

The methodology has been prepared to incorporate fraud and error indices for benefits administered by DWP, and those administered by HMRC such as Tax Credits and Child Benefit. These are planned for implementation in published statistics from the Spring 2025 ONS annual release, however as HMRC outputs will not be included in the SSA output measure, fraud and error rates for DWP only will be included.

Existing annual National Statistics report DWP's stock of Monetary Value of Fraud and Error (MVFE) [see link 89, Annex I] by benefit and by cause. Much of the quantifiable DWP administration activity has been focussed on enhancing the outflow from the MVFE stock by detecting and correcting erroneous benefit payments. This activity is quantified through targets reported to HM Treasury as delivering Annually Managed Expenditure (AME) savings (financial year ending 2025 target £1.7 billion).

However, a focus of AME savings can create perverse incentives, when the goal should be to prevent the inflow of inaccurate benefit payments at source into the MVFE stock. Those preventative safeguards require investment to verify and block expenditure in which the value for money impacts are difficult to measure. Additionally, a rising propensity for fraud in society (and accepted within Office for Budget Responsibility forecasts), recognise a system not in steady state where additional administration is required just to maintain current levels. See Annex G for more detail.

In addition, these stock measures of fraud and error may not reflect the flow of new fraud and error cases which are a more accurate measure of in-year delivery.

> **Recommendation 105:** The ONS should apply a quality adjustment in Spring 2025 based on existing Department of Work and Pensions fraud and error rates.

> **Recommendation 106:** The ONS should work with the Department of Work and Pensions to replace stock measures of fraud and error with more appropriate flow measures to better reflect in-period performance, in Social Security Administration quality adjustment.

Customer satisfaction

Claimants' views on the effectiveness of the claims process can reveal helpful insights into the quality of a service. The Review explored the potential of using the DWP and HMRC customer surveys. Whilst, in the DWP survey, results cannot be compared before and after 2013, and there were further methods revisions in 2020, this remains an otherwise prime potential data source, although the Review notes the need for comparable HMRC data for completeness. More information is provided in the Recommendations for further work section.

Timeliness

The time taken for claimants to receive benefits can drastically affect their ability to manage their finances effectively. The Review therefore explored the potential of using timeliness metrics related to benefits administered by DWP and HMRC. However, there are complexities in relation to the different time periods permitted for simple and complex cases or various eligibility checks which means the Review could not recommend proceeding with this metric at this time. Further investigation may however be merited.

14.6 Devolved governments

The ONS does not include benefits in Northern Ireland which are administered by the Department for Communities, although the Review recommends further investigation into data sources for including benefits in Northern Ireland in future.

> **Recommendation 107:** The ONS should investigate data availability accounting for benefit administration in Northern Ireland, as the current measure does not include that activity.

14.7 Recommendations for further work

Inputs

Administrative costs associated with benefits delivery are being captured within SSA inputs, however there is a question on whether only administrative costs are being captured for administration of concessionary travel fares, which are governed by local government, or whether the costs for concessionary fares are included, which would be erroneous.

> **Recommendation 108:** The ONS should further investigate whether administrative costs associated with benefits delivery are being captured in Social Security Administration.

Quality adjustment

Customer satisfaction

Two approaches could be considered in the area of customer satisfaction:

- Utilising departmental customer surveys: DWP's Customer Experience Survey appears a strong data source, although the Review would encourage the ONS to work with DWP to consider potential improvements and collaborate on obtaining a consistent long-term time series. For example survey questions could be re-worded to obtain participants' perceptions exclusively on their experiences with Tax Credits or Child Benefit, which would avoid the challenges of merging two surveys, one from HMRC and one from DWP.
- Utilisation of the OPN: The ONS' Opinions and Lifestyle Survey (OPN) has the potential to provide consistent insight into public satisfaction of multiple public services, but obviously with far less depth. This however may act as a helpful means of triangulating and validating movements in the DWP and HMRC data. The OPN, it has to be noted, only covers Great Britain, therefore the Review would recommend augmentation with Northern Ireland's Omnibus Survey

[see link 90, Annex I] and Continuous Household Survey [see link 91, Annex I] for complete UK coverage.

Recommendation 109: The ONS should continue to work to identify the most appropriate route to capture user satisfaction in Social Security Administration across the UK.

Timeliness

There is potential to improve existing timeliness data, including potential new data sources, such as the Operational Performance Research Analysis (OPRA) which aims to amalgamate timeliness measures from all benefits administered by DWP into a single metric for the inclusion in their annual reports and accounts.

Another consideration could be to identify the numbers of claims being processed in relation to actual claims administered to produce a proxy waiting list. The median processing time for benefits may also be explored.

Recommendation 110: The ONS should continue to collaborate with the Department of Work and Pensions to develop data on timeliness for Social Security Administration.

Weighting

Having a single quality adjustment avoids the question of how to weight together disparate measures in the short-term, but clearly remains a potential issue if others are developed. Surveying taxpayers on the relative importance of the three quality adjustments may present the simplest way to determine a weighting mechanism. The existence of the DWP Customer Experience Survey, with subtle methodological improvements, could meet this need. Relative spend on delivering fraud and error, timeliness or customer facing activities would deliver an alternative. A third option could be any inspection regimes which have received Ministerial sign-off.

Recommendation 111: Weighting mechanisms for baskets of quality adjustment factors, such as in Social Security Administration, should be considered and if necessary, data collected to ensure quality adjustment estimates reflect societal preferences and that these are aligned with how other relevant areas are treated, such as health waiting lists.

CHAPTER 15:

Local Government, including Adult and Children's Social Care

Local Government is an area where relatively rich data sources exist, however many of the services delivered by local authorities, particularly Education, are addressed in previous chapters. Two further services: Adult Social Care (ASC) and Children's Social Care (CSC), are primarily although not exclusively delivered via local authorities. Adult Social Care is also significantly delivered by health services and other providers. This results in complexities relating to integrating data sources from these different agencies.

The complexities relating to multiple outputs and outcomes led to both services being areas which Atkinson [see link 1, Annex I] (2005) and subsequently the Office for National Statistics (ONS) were unable to deliver viable output metrics for. At the time of the Bean Review [see link 2, Annex I] (2016) both CSC and ASC were calculated on an 'inputs = outputs' basis.

However, given significant improvements in methods since then, the ONS delivered new methods in 2019 for both CSC [see link 92, Annex I] and ASC [see link 93, Annex I]. This led to improved measures and quality adjustment:

- quality adjusted direct measures of CSC output
- quality adjusted ASC using a mix of direct and indirect measures, according to data availability. This therefore means some measures are quality adjusted "inputs = outputs" series, but in these areas no better direct output measures existed

Estimates for ASC are the timeliest of any individual service, with a dataset for England produced around 11 months after the reference period.

The remaining local government services (for example, planning, environmental and waste services, libraries, democratic services) are either small or potentially eligible for inclusion in Environmental services, in particular planning, environmental and waste services. Environmental services are covered in depth in Chapter 16.

The Review therefore did not prioritise the remaining Local Government services, particularly in the light of the establishment of the Office for Local Government (OfLog), which had a data collection and analysis role in this area which was expected to grow into being a major data supplier and partner for the ONS. This is an area that requires further examination, particularly in light of HM Government's announcement on 16th December 2024 closing OfLog.

Recommendation 112: The ONS should continue to work with the Ministry of Housing, Communities and Local Government to determine a strategy to capture the remaining local services, with a particular focus on ensuring alignment of treatment with Environmental Services, noting the potential impact of the United Nations Classification of the Functions of Government Review.

In addition, further work is required to understand the different nature of local government under the devolved governments and to align strategies to collect and process data on a basis which allows meaningful comparison and analysis.

Recommendation 113: Local government data across the devolved governments should be sought to enable UK-wide local government estimates to be prepared.

15.1 Adult and Children's Social Care

ASC reflects nearly 6% of the total expenditure of public services [see link 94, Annex I] in 2021. Provision of ASC is the responsibility of local authorities in the UK, alongside the NHS. The ONS productivity measure covers publicly-funded ASC services provided by public providers such as local authorities and those contracted from the independent sector. Self-funded provision is excluded, as per other services, such as private schools directly procured by households. CSC accounted for 2.8% of the expenditure share of total public service output in 2021 [see link 94, Annex I].

It covers the provision of social work, personal care, protection or social support services to children in need or at risk.

Despite the developed nature of the ONS measurement of the productivity of ASC and CSC services, the Review identified some areas for improvement.

Children's Social Care

CSC involves a range of services delivered to children and families. These include support to the private and voluntary sector which supply early years childcare via day nurseries, play-groups, pre-schools and childminders. Because these services are delivered outside the government sector, and are reflected in private sector data, it is challenging to find output measures which solely reflect government's contribution here, for example in the form of subsidies or other support. Nevertheless, spend in this part of Children's Social Care has at time been a material proportion of total CSC expenditure.

Without a measure of output related to these inputs, there remains concern that the productivity estimates generated may be downwardly biased if increased spending in these areas is not reflected in traditional output measures of social care provision.

To combat this, the Review considers it would be beneficial to split out the elements of the service related to Early Years Childcare where output data are unavailable and measure this element by the 'inputs = outputs' approach.

> **Recommendation 114:** The ONS should seek to separate elements of Children's Social Care where output measures cannot be identified, with particular reference to Early Years Childcare, and in the absence of direct output measures estimate output in this area through the 'inputs = outputs' methodology until further measures can be developed.

> **Recommendation 115:** The ONS should periodically investigate the methodology into Children Social Care in relation to non-government sectors in order to make further improvements to the measure.

Adult Social Care

The output of ASC is measured using a cost weighted activity index for services where activity data are available, such as the number of weeks of residential and nursing care provided. For services where activity data are not available, output is measured on the 'inputs = outputs' basis.

Where the productivity methodology differs from other services is that in some areas where ASC retains an 'inputs = outputs' methodology, this is augmented by a quality adjustment which was developed as part of previous work. The reason for this is that there remain significant gaps in data provision for some key aspects of this service which prohibits the deployment of superior measures of the number of activities, but conversely, high quality data exists on user satisfaction and objective measures of service outcomes.

Recognising this is a compromise, the Review notes the spirit of the underlying philosophy to this work, particularly 'delivering data which are approximately right is better than delivering data which are completely wrong'. The Review agrees that the alternative of using only 'inputs = outputs' in this instance would be clearly inferior.

For England, a change in data collection between financial year ending 2014, and 2015 resulted in fewer activities being measured (e.g. number of hours of home care and number of 'meals on wheels' delivered). Until financial year ending 2014, activity data was available for residential care, nursing care, assessments of need, day care, home care, provision of meals and provision of equipment. However, from financial year ending 2015 onwards, only residential care and nursing care activity data were included in the output index.

Due to the discontinuation of the activity data source used for ASC output in Northern Ireland in 2020, output for Northern Ireland has been measured on the 'inputs = outputs' basis since this date.

The ONS produces an annual article Public service productivity, adult social care, England: financial year [see link 95, Annex I], and whilst on one hand this is an important release for a growing area of public services, on the other as further improvements have been made to all other services, the ONS must consider whether this product is adding significant value.

> **Recommendation 116:** The ONS should continue to keep the Adult Social Care system under review with the aim, utilising ongoing data developments, to address areas of this service which continue to use the 'inputs = outputs' methodology.

> **Recommendation 117:** The ONS should consider the user need for a separate Adult Social Care, England publication in the suite of public service productivity publications.

CHAPTER 16:

Environmental Services

The measurement of Environmental services faces a number of challenges distinct from those described in previous chapters. These were not addressed in detail by Atkinson [see link 1, Annex I] (2005), in part because the concept of 'Environmental services' has evolved since that date.

Before the Review could commence the normal work of understanding inputs, outputs and outcomes, the review needed to identify the classification which should fall into scope of this category.

Whilst the Classification of Functions of Government (COFOG) [see link 4, Annex I] group 5 relates to Environmental Protection, this is not currently independently published for the UK. The Office for National Statistics (ONS) includes this area within the 'Other' grouping, which includes COFOGs 1, 3.05, 3.06, 4, 5, 6, and 7. All services in this grouping are measured on an 'inputs = outputs' basis. As well as COFOG 5, the Review's analysis suggests that similar activities can be located in numerous COFOG categories. For example, Forestry is currently treated under 'Economic Affairs', because it produces timber, an economic asset, rather than within 'Environmental Protection', because of the environmental impact it has.

A challenge to identifying the correct scope is the changing perspective of the government services which fall into 'Environmental Protection' or 'Environmental services'. A clear and useable definition is important given the UK policy priority on climate change, clean power and net zero, and addressing nature loss and promoting nature recovery. For example, the ONS has recently created a clear and useable definition of scope and indicative estimates for 'green jobs' [see link 96, Annex I].

However, even for functions which are likely to be brought into scope, a second major issue is where departments and local government have not been required to separate these out within their relevant data returns. This means it may not be possible to build realistic historic time series of inputs, even if output data can be identified.

Finally, this is an area that has seen material change since the last update to the COFOG, both in the productivity but also the type of services being delivered. As such, the challenge discussed in numerous chapters, where traditional measures capture technical efficiency (improving an existing process) well but capture allocative efficiencies (swapping to a better process) less well, is likely to be present here. There would be a clear need to consider this in any future work.

16.1 Core methodological and data challenges identified

Potential Improvements to the COFOG categorisation

The current COFOG classification system is more than 20 years old and presently COFOG 5 'Environmental Protection' describes spend across six categories:

- 5.1: Waste management.
- 5.2: Wastewater management.
- 5.3: Pollution abatement.
- 5.4: Protection of biodiversity and landscape.
- 5.5: Research and development on environmental protection.
- 5.6: Environmental Protection Not Elsewhere Classified.

Whether these still describe the range of relevant services, is a pertinent question. As highlighted in Chapter 6, the United Nations has commenced a review of the COFOG structure to ensure that it remains fit for purpose in measuring government expenditure. Although outcomes are not expected until around 2027, it is the Review's understanding that there is broad international consensus that 'Environmental Services' should be added as a service area within the COFOG structure. The UK has made recommendations to better reflect environmental services.

> **Recommendation 118:** The ONS should work with relevant departments to recommend a definition of 'Environmental Services' to the upcoming consultation on the international Classification of Functions of Government (COFOG) definitions to encourage the creation of a wider COFOG sector covering 'Environmental Services', for use in developing measures, using the work of this Review to inform developments in this space.

A more inclusive measure of 'Environmental Services' spending would include a range of activities captured outside of those currently contained in COFOG 5. It is important to consider how best to take account of cross-cutting issues of increasing policy interest such as climate change, climate and health, biodiversity, and the 'circular economy' (whereby waste is recycled and brought back into productive use), alongside disaster risk management and spending that may have multiple impacts.

This chapter presents the Review's research on alternative categorisation for an 'Environmental Services' COFOG, which the Review labels 'Section E' to distinguish it from the existing number-based system. Section E would align more closely to alternative frameworks, such as the proposed international Classification of Environmental Functions described by the European Commission (Eurostat). This provides a broader definition, encompassing all activities, goods and services that have an environmental purpose, i.e. that have as their primary purpose to prevent, reduce and eliminate pollution and other forms of degradation of the environment (e.g. treatment of waste and wastewater, protection of biodiversity), or to make more efficient use of natural resources, and hence safeguarding these from depletion (e.g. recovery and substitution of natural resources, recharges of natural stocks).

To cover all potentially related and inter-related functions that relate to this wider measure of environmental activity, the Review proposes an amalgamation of sections drawn from multiple COFOG categories to recognise and describe three proposed broad categories.

- E.1: Environmental Protection.
- E.2: Natural Resource Management.
- E.3: Climate Change and Net Zero.

> **Recommendation 119:** An improved Classification of Functions of Government structure for 'Environmental Services' should differentiate between 'Environmental Protection', 'Natural Resource Management' and 'Climate Change and Net Zero', with further detail embedded below this structure.

COFOG 5 represents around one seventh of the 'Other' grouping, comprising 2.3 percentage points of the 16.1% of the public services captured by this grouping. Activities identified within categories E.1, E.2 and E.3 from outside COFOG 5 are predominantly also captured by the 'Other' grouping, so this proposal would make the statistics more user-friendly in the aggregate, as well as providing a better picture of Environmental services.

Researching the environment services classification

The Review acknowledges challenges around the conceptual definition of what should fall into scope of 'Environmental services' and how to aggregate these into a single output measure, hence enabling productivity measurement. For example, short-term local-scale activities work includes things such as a community tree-planting projects. Long-term complex landscape-scale outputs work includes things such as trends in national habitat biodiversity. These two areas operate on extremely different timescales and geographies and are delivered via completely different delivery organisations.

Certain environmental policy is also devolved, resulting in differences between UK nations. There is also an increasing role for combined authorities in the environment space (e.g. nature recovery and green space). The Review has learnt that the most challenging services to measure are either those with multiple bodies contributing towards outputs, or those where there are multiple outputs or outcomes being delivered. The measurement of Environmental services appears to face both these challenges.

Privatised and nationalised businesses

There may also be challenges relating to privatised and nationalised industries in this sector. If, for example, a waste-water management company was nationalised resulting in a classification change moving the firm between the private and the public sectors, this could result in significant changes in public service productivity. In such an eventuality, the ONS would need to follow established classifications practice, with considerations including whether market prices would continue to be charged to customers, or whether a nationalised firm received subsidies or similar funding to undertake 'Environmental Protection' activities.

16.2 Inputs

Whilst considering the potential scope of COFOG 5, the Review has investigated both UK central and local government spending, within the current COFOG 5, to understand potential input measures.

Central government spend is spread across all but one of the COFOGs within 'Environmental Protection', there is zero spend currently aligned to 5.2 'wastewater management'. Local Government data currently only exists for inputs into COFOG 5.1 'waste management'. This represents a potentially significant gap for further research into sourcing further detail on spending and alignment of activities.

The Department for Environment, Food and Rural Affairs (DEFRA) is the largest contributor towards currently captured COFOG 5 spending. Additional contributions come from the Department for Energy Security and Net Zero (DESNZ), the Department for Science, Innovation and Technology (DSIT),

the Ministry of Housing, Communities and Local Government (MHCLG), the Department for Transport (DfT), the Department for Culture, Media and Sport (DCMS) and the Department for Business and Trade (DBT).

Potential additional inputs that the Review consider relevant but are not currently included within this COFOG might include, but are not limited to:

- Activity associated with land and habitat management of agricultural land, forestry or urban spaces, including local authority planning functions.
- Activity related to climate change action, including adaptation activities, flood protection and management of water resources.
- Potentially impact management or mitigation activities that affect our natural environment, such as work on greening transport networks.
- Activities related to green spaces in urban communities.
- Expenditure on international activities related to environmental protection.

As an example, the Review compared the activities captured within COFOG 5.4 to the estimated spend captured by Joint Nature Conservation Committee (JNCC) indicator E2 'Expenditure on UK biodiversity', itself part of a broader set of UK biodiversity measures, which addresses one sub-domain of Environmental Protection.

The Review identified that whilst the trends over time follow similar patterns for the two measures, there is a significant difference in the record levels of spending. In financial year ending 2022, JNCC calculated a total public sector spend of £778 million on UK biodiversity (indicator E2). This compares with £317 million calculated from COFOG 5.4, or £655 million from combining COFOG 5.4 and 5.5 (Research and Development (R&D) Environment Protection). Figure 1 of JNCC's annual report [see link 97, Annex I] provides the time series from financial year ending 2001 to financial year ending 2023. There are various differences between the JNCC and COFOG methodology, so care should be taken when comparing the two numbers.

Despite the inclusion of all COFOG 5.5 R&D spend (of which only a fraction can necessarily be expected to fall into this area, as opposed to the other services listed in COFOG 5), the E2 calculation highlights potentially significant gaps in current COFOG methodology. This indicates more work is needed to identify areas of missed spend and improve data coverage, address bias and uncertainty and improve methodologies. The Review is therefore confident that the current COFOG 5 structure does not account for the full range of activity related to 'Environmental Services'.

In summary, data sources for E2 have a wider coverage of producers and are therefore likely to account for more relevant spending. Alongside more robust quality assurance and validation processes, this likely provides a more comprehensive calculation of total spend. However, expert judgement may propose further refinements and should be considered.

The Review undertook a similar exercise for COFOG 5.2 'wastewater management' which could not identify any bodies recording spend against this area. While privatised water companies will be excluded as they are not part of the public sector, the Review's expectation was that some proportion of the Water Services Regulation Authority (Ofwat) and Local Flood Authorities' spending could be expected to be against this item. The Reviews anticipates these inputs have been recorded at a higher level of aggregation via the process of consolidation of accounts.

To identify spend in scope, the Review interrogated data recorded on the Online System for Central Accounting and Reporting (OSCAR), the Government's financial system. Spend was filtered by: 'carbon', 'environment', 'climate change', 'green', or 'sustainable', alongside additional terms such as 'green investment', 'carbon capture', sustainable transport', and 'emissions', rather than 'labour', 'capital', and 'intermediate consumption' to enable identification of activity areas. Some 4,234 lines of possible Environmental Protection spend were identified, aligned to five level 1 COFOG categories, covering a total of 19 sub-categories. For financial year ending 2022 these accounted for around 11% of total government spend (97% of which is Central Government expenditure).

These c. 4,000 lines of spend were mapped to the three proposed sectors (E.1: Environmental Protection, E.2: Natural Resource Management, and E.3: Climate Change and Net Zero). Challenges exist around apportioning spend where activities clearly contribute to multiple sectors, as OSCAR sources do not often provide sufficient detail to robustly and confidently assign them. Given most items are relatively small, the Review would not expect simply mapping them to the lead sector in the first instance would cause significant biases in the resulting statistics.

As an example, table 5 maps the existing COFOG categories which are proposed for inclusion into E.1 (Environmental Protection), which covers mainly COFOG 5 including waste, pollution, protection of biodiversity and landscapes and R&D, with a significant proportion of agricultural and forestry-related activities captured by COFOG 4. Tables for sections E.2 and E.3 are included in Annex H.

Table 5: Classification of Functions of Government (COFOG) expenditure by E.1 activities, UK, 2021

E.1: Environmental Protection	Share of lines of spend within this category to be brought into E.1 (% of category likely aligned to E.1 activities)	Share of Total Spend to be brought into E.1 (% of category likely aligned to E.1 activities)
4.2.2 Other agriculture, food and fisheries policy	20.8%	62.0%
4.2.3 Forestry	49.4%	53.6%
4.5.5 Other transport	1.4%	0.3%
4.8.0 R&D economic affairs	1.3%	0.3%
5.1.0 Waste management	100.0%	100.0%
5.3.0 Pollution abatement	100.0%	100.0%
5.4.0 Protection of biodiversity and landscape	100.0%	100.0%
5.5.0 R&D environment protection	100.0%	100.0%
5.6.0 Environment protection not elsewhere classified	100.0%	100.0%
6.3.0 Water supply	47.2%	99.8%

Section E.1 would account for 8.7% of all Central Government spend in financial year ending 2022. Data breakdowns for local government do not allow the Review to make a direct comparison. However, current spend on COFOG 5.1 (Waste Management) only accounted for 5% of Total Local Government Spending, covering mainly 'waste management'. The other five categories consistently provide a nil return. The Review is confident that activities are carried out at the local government level that would provide services and activities directed towards environmental protection. Initial review of data from MHCLG suggests this activity may be aligned in alternative categories such as Housing, Communities or Transport, which would warrant further investigation.

The Department for Business, Energy and Industrial Strategy (since replaced by Department for Energy Security and Net Zero, Department for Science, Innovation and Technology and Department for Business and Trade) accounted for 96.3% of all E.1 spend in financial year ending 2022, with 96.4% of this coming from one of its seven arms length bodies, the Nuclear Decommissioning Authority.

Defra accounted for 2.1% of E.1 spend. The Northern Irish Executive (1.1%), Scottish Government (0.3%), and Welsh Government (0.3%) were the next largest, with DCMS and DfT responsible for small proportions of the total. Local Government Planning functions were not included in the above analysis but could also be included in E.1.

Deflators

Deflators remove the effect of inflation from inputs and so give clear sight of a consistent measure of volume through time. However, as many of the environmental assets and resources under consideration lie outside the national accounts, whether traditional deflators will suffice would be an issue for consideration. The deflator based on the Consumer Price Index can be used on majority of spending defined as intermediate consumption (P2), but further work would be required to identify appropriate deflators for different types of spend.

> **Recommendation 120:** The ONS should undertake research to identify appropriate deflators for inputs into, and outputs from, Environmental services.

16.3 Output measurement

The Review undertook initial work on outputs, identifying high-level, UK-wide datasets that could be considered as suitable output measures. So far these cover waste management, water waste management and pollution abatement, but more work would be needed on the remaining three categories and activities outside of COFOG 5.

There are conceptual challenges in measuring output relating to the environment as distinct to outcomes which may be better captured in quality adjustments. Equally, the definition of outcomes which fall into scope is more challenging here as one cannot simply focus on human outcomes: environmental policy measures a broader set of objects and systems involving an array of natural and social sciences.

Devolution of certain responsibilities also makes measurement more complex, for example the UK government maintains responsibility for delivering emission reductions nationally within the UN Framework Convention on Climate Change (e.g. Paris Agreement) while, for example, biodiversity conservation is fully devolved.

Potential Environmental outputs or outcomes captures a range of potential output and outcome measures which relate to the current COFOG 5 classification, and for which domestic data exists.

Potential Environmental outputs or outcomes

1. Air, climate and energy

- Reduction and control of air emissions (excluding energy related measures)
- Energy from renewable sources
- Energy savings and management

2. Wastewater and water resources

- Wastewater management
- Water savings and management of natural water resources

3. Waste materials recovery and savings

- Waste management
- Materials recovery and savings

4. Soil, surface and groundwater, biodiversity, and forest

- Protection of soil, surface, and groundwater
- Protection of biodiversity and landscape
- Management of forest resources

5. Noise and radiation

- Protection against noise and vibration
- Protection against radiation

6. Research and Development

- R&D for air, climate and energy
- R&D for wastewater and water resources
- R&D for waste, materials recovery and savings
- R&D for soil, surface and groundwater, biodiversity and forest
- R&D for noise and radiation

7. Cross-cutting and other activities

- Environmental education and training
- General environment administration, management, regulation, dissemination, and consultancy
- Environmental activities not elsewhere classified

However, the environment as an issue extends beyond domestic boundaries, and the Review recognises the importance of considering global targets and indicators. For example, the UK is a signatory to the UN Convention on Biological Diversity, through which it will report on a new Global Biodiversity Framework which is currently being finalised. The framework has 4 goals and 23 Targets, with each target having a set of headline, component and complementary indicators.

CHAPTER 17:

Concluding Remarks and Summary of Recommendations

17.1 Concluding remarks

This Review has made significant progress in delivering a step-change improvement in the published Office for National Statistics (ONS) estimates of public services productivity, relative to the starting line marked by the Atkinson Review [see link 1, Annex I] (2005). Some improvements, for Healthcare and Education, were implemented from March 2024; but many more will be implemented in the next ONS annual release of data in Spring 2025. Further recommendations will be taken forward on a rolling basis thereafter.

The Review has improved the accuracy, granularity and timeliness of public services productivity statistics. As well as building on the already well-developed estimates for the Healthcare and Education services, the Review has implemented innovative solutions to produce experimental estimates for both the Tax Administration service for the first time, and in accounting for the transition to Universal Credit in the measurement of Social Security Administration. The estimates of inputs and output have been improved across a range of services, and sizeable advancements in research have been made into output measurement in the Policing, Immigration, and Defence services, long and internationally recognised as conceptually challenging.

The ONS now regularly publishes experimental 'nowcasted' estimates for the last 2 years, which aim to deliver better data quality compared to previous versions which held quality constant given the lag in data availability for quality metrics. The ONS introduced publication of experimental quarterly estimates of productivity in the healthcare sector from February 2025 [see link 8, Annex I], with plans to extend this increased frequency to other services.

With continued focus, a number of further improvements should be achievable in 2026 and beyond. A work programme has been produced to improve the coherence of the annual and quarterly public services productivity estimates in 2026, subject to resource availability. This would increase the transparency for users of the complex relationship between the accredited annual National Statistics and the experimental (Official Statistics in development) quarterly estimates of public service productivity.

The Review has also found a growing network of service-specific bodies who are exploring public service productivity measurement at more granular levels than the ONS can deliver under current budgets, such as the Centre for Police Productivity. These are exciting propositions which should make local data much more transparent and are to be applauded. Consistency of methods will require continued engagement.

Within available resources the Review has worked collaboratively with the devolved governments to identify data gaps across the four nations, and the ONS is implementing a small number of improvements from Spring 2025. The coherence of devolved governments' data underpinning the UK-wide public services productivity estimates is a priority, but requires consistent investments into coherent datasources across the nations of the UK.

There is also growing international interest in this agenda, and the opportunity to collaborate with national statistics institutes globally. A survey undertaken by the Economic Statistics Centre of Excellence [see link 12, Annex I], funded by this Review has identified a set of interesting lessons the ONS can apply in the coming years.

The Review has also investigated the drivers of public service productivity through designing, running and analysing new 'pilot' surveys of management practices and how time is used in the public services by staff. Both surveys operate at the cutting-edge of research and are unique in the world, giving greater clarity on the delivery models and effectiveness of the public services.

A significant number of farther reaching recommendations for further work have been identified utilising new data and new methods which could deliver further improvements in coming years. Some impact or depend upon other government departmental programmes or initiatives, and may need additional funding to bring to fruition.

The Review has made substantial recommendations to the United Nations Review of the Classifications of Functions of Government (COFOG) [see link 98, Annex I], to update the categorisation and collection of government expenditure data to reflect present day policy, and future proof data collection by HM Treasury. In particular improved classification of Defence, Environmental Services and the disaggregation of Policing and Immigration services would enable future improvement in measurement of these sectors. However, even if the COFOG Review does not fully accept all of these, the UK will still be at liberty to do more than the international minimum and the Review commends this idea to those working on implementation of the new COFOG in years to come. The Atkinson Review pointed out that public service productivity should only require the data government already provides for itself to make informed decisions. In an ever-more complex world, ensuring data detail keeps pace with need is an important function of government in and of itself.

Further step change improvements in productivity measurement of the full landscape of services is clearly possible, as well as in the coherence of the four nations' data underpinning the UK estimates. The Review firmly believes it is possible to continue to improve the evidence base for policy decisions relating to the productivity of public services at relatively low cost. This will only occur, however, with continuing engagement and collaboration across Whitehall departments and other public bodies, as well as investment.

Finally, measurement is not enough: in undertaking this Review the team has observed numerous instances of innovation and improvements, taken forward by civil and public servants as part of their daily work delivering for the UK public. This report and the report team's applause is dedicated to all those making public services deliver and improve.

17.2 Summary of recommendations

Measurement principles

- **Recommendation 1:** The principles outlined by the Atkinson Review should remain the underpinning intellectual methodology behind the measurement of public services, both in UK National Accounts and public service productivity statistics.
- **Recommendation 2:** There are instances where cost provides a weaker weighting metric than alternative approaches. These alternatives can and should be applied, but only where a clear case can be made against the principle that the alternative metric better reflects the value of the service than the relevant cost data.
- **Recommendation 3:** The value of transfers can be used to weight together individual components within a service or in a quality adjustment but should not be a direct measure of output.
- **Recommendation 4:** Where different weights are used within a service, the cost weighted activity index will still be used to weight in any elements which cannot be addressed with the new weighting approaches. When service level statistics are estimated, these should be aggregated in a cost weighted activity index to produce the national aggregates.
- **Recommendation 5:** Adjustments should be considered to equivalise cost weights between services of equivalent value but different cost.
- **Recommendation 9:** The weights used to bring together quality adjustment components need to, as closely as possible, reflect societal preferences in as objective a fashion as possible.
- **Recommendation 14:** The ONS should continue to work with the devolved governments to understand the devolved service-delivery landscape and improve data coverage, quality and consistency in the UK measure of public service productivity.
- **Recommendation 26:** The ONS should continue to work with other organisations undertaking development activity in the measurement of public service productivity for mutual knowledge sharing, adoption of coherence between estimates where possible, and transparency around differences in measurement.

Recommendations implemented for the Spring 2025 ONS release of data

Healthcare

- **Recommendation 31:** The ONS should implement abdominal aortic aneurysm, bowel cancer, breast and cervical screening services within healthcare outputs in Spring 2025.
- **Recommendation 35:** The ONS should use the Department of Health and Social Care Index of Services data to estimate dental activity growth from 1996 until 2006 and improve the method of linking this activity data to the current 'Units of Dental Activity' data source to avoid a discontinuity between data sources.
- **Recommendation 36:** The ONS should ensure the overall weight of ophthalmic services and dental services is consistent with expenditure as reported in the Department of Health and Social Care annual accounts, and uprate historical NHS telephone and website services unit costs to account for NHS cost inflation.
- **Recommendation 37:** The ONS should implement equivalised weights in Spring 2025. Where service weights are equivalised to account in productivity for cost-savings from moving services to lower-cost modes of provision, these should be incorporated in the non-quality adjustment measures. Quality adjustment should account for changes in value of services delivered that goes beyond cost-saving, such as improvements to the estimated health improvement from treatment.
- **Recommendation 38:** The ONS should implement equivalisation of weights for ambulance services in Spring 2025.
- **Recommendation 40:** The ONS should adopt the improvement to remove excess bed days in Spring 2025.
- **Recommendation 43:** The ONS should extend the application of the health gain quality adjustment, excluding the waiting times component, from elective to non-elective procedures in Spring 2025.

Education

- **Recommendation 50:** The ONS should implement improvements identified for Education inputs in Spring 2025 related to salaries.
- **Recommendation 52:** The ONS should implement improvements to the quality adjustment of Education to better account for the impact of the coronavirus pandemic and to account for student well-being and Further Education attainment.

Defence

- **Recommendation 58:** A new method implementing a more granular direct measure of labour inputs into Defence should be applied.
- **Recommendation 59:** A UK National Accounts consistent intermediate consumption deflator should be derived and applied to Defence intermediate consumption spending, using the His Majesty's Treasury Online System for Central Accounting and Reporting (OSCAR) data to weight components.
- **Recommendation 60:** The Review recommends that the national accounts capital deflator should be applied to the Defence capital expenditure data, but the combined implied deflator of intermediate consumption and capital should be compared to the Defence contracts price index to determine whether subsequent manual intervention is then required.

Policing

- **Recommendation 64:** The ONS should apply bespoke Police and Immigration deflators to Central Government expenditure data used to indirectly measure intermediate consumption in Spring 2025.
- **Recommendation 65:** The ONS should implement more granular salary data into its direct labour measure for Police, incorporating salary information for individual police ranks (including Uplift officers) in Spring 2025.
- **Recommendation 66:** The ONS should incorporate Northern Ireland workforce data into its direct labour measure for Police in Spring 2025.

Public Order and Safety

- **Recommendation 79:** The ONS should apply bespoke intermediate consumption deflators to expenditure data in all Public Order and Safety service areas, with the exception of fire.
- **Recommendation 80:** A new method implementing a direct measure of fire service labour inputs into Public Order and Safety should be introduced, expanding in future years to cover devolved governments.
- **Recommendation 88:** Timeliness data by case type should be used to produce a more granular quality adjustment that aligns with the improved Crown Court output measure.
- **Recommendation 89:** A straight linear interpolation should be applied to the re-offending quality adjustment applied for 2018 to 2021 where the coronavirus pandemic interrupted the normal provision of data.
- **Recommendation 90:** Court quality adjustment weights should be revised to be 40% recidivism and 60% timeliness.

Taxation

- **Recommendation 93:**The ONS should 'revenue-adjust' direct measures of Tax Administration output whilst developing, with HMRC, appropriate quality adjustments, including using factors such as the 'tax gap' to reflect the difference between the desired and achieved levels of tax collected, customer satisfaction and call waiting times.
- **Recommendation 94:** The ONS should publish provisional estimates of Tax Administration productivity in Spring 2025 and adjust the "Other" grouping accordingly.

Social Security Administration

- **Recommendation 102:** The ONS should measure the output of the administration of Universal Credit and its predecessor legacy benefits using a benefit-weighted index to account for the allocative efficiency of the transition. This measure should be seen as provisional whilst undergoing further testing.
- **Recommendation 104:** The ONS should adjust the output of Universal Credit administration for changes in the number of entitlements included in payments, to account for the heterogeneity of cases.
- **Recommendation 105:** The ONS should apply a quality adjustment in Spring 2025 based on existing Department of Work and Pension fraud and error rates.

Recommendations for future work

Concepts and methods

- **Recommendation 6:** Further consideration should be made of which services are considered as preventative services. An imputed valuation for the preventative service A should then be used where this is the product of the probability in reduction in use of service B and the actual cost of service B.
- **Recommendation 7:** For pre-selected preventative services where high quality data on impact of downstream services can be found, the probability weighted cost of these downstream services can be used as a proxy valuation of the preventative services in the cost weighting methodology.
- **Recommendation 8:** For latent capability, further research is required to identify instances where this method could be piloted using high quality data.
- **Recommendation 10:** Further research should be undertaken to consider the potential to use alternative weighting regimes proposed for quality adjustment in replacing cost weights, as per recommendations 3 and 9.
- **Recommendation 11:** The ONS should exploit methods developments around technology and other deflators to improve the measurement of volume input and output in public services, and continue to seek out methods improvement.

- **Recommendation 12:** Further research should be undertaken to consider the potential and value or necessity to apply an adjustment to account for labour inputs undergoing training.
- **Recommendation 13:** The ONS should consider how best to reflect private sector productivity growth within the measurement of public service productivity, where this captures private sector delivery of services.
- **Recommendation 15:** The ONS should further investigate the feasibility and user need for devolved metrics on public services productivity, particularly the education sector in collaboration with the devolved governments.
- **Recommendation 16:** The ONS should continue to improve annual and quarterly public service productivity estimates to take account of available quality adjustment data and, where this is not possible, keep nowcasting models under annual review to provide the most accurate and timely data possible.
- **Recommendation 17:** The ONS should continue the roll-out of publication of service estimates on a quarterly basis having started with the largest sector (Healthcare) in February 2025.
- **Recommendation 18:** The ONS should replace the current 'contribution to growth' compilation method with 'chain volume measures', and then implement reconciliation of the quarterly estimates with the annual estimates each year, in order to align with the UK National Accounts protocols and improve coherence and understanding for users.
- **Recommendation 19:** The Quarterly cumulative Average Growth Rates (QAGR) method should be applied to provide more timely nowcast estimates for annual estimates as further research is undertaken to evaluate the efficacy of alternative methods in the light of the coronavirus pandemic. The performance of the QAGR model should be evaluated on an annual basis.
- **Recommendation 20:** The ONS should proceed with best practice improvements to align quarterly and annual production statistics.
- **Recommendation 21:** The ONS should keep under review whether there is convergence of the HM Revenue and Customs expenditure data and the ONS Government Expenditure on Research and Development Survey estimates to allow future consolidation of the two data sources.
- **Recommendation 22:** The ONS should incorporate the methods and data developments from this Review into the UK National Accounts as part of its implementation of the 2025 System of National Accounts revision, in line with decisions made by the National Statistician.
- **Recommendation 23:** The high level 'roadmap' developed by this Review for incorporation of quality adjustments, and improvements to public sector output in the UK National Accounts should be reviewed and maintained by the ONS as part of wider strategic planning of the implementation of the System of National Accounts 2025 develops.

- **Recommendation 24:** The ONS should continue to influence the United Nations Classification of the Functions of Government Review to maximise the opportunity for improved future categorisation of departmental expenditure (underpinning inputs measurement) via the HM Treasury Online System for Central Accounting and Reporting system.
- **Recommendation 25:** The ONS should work with HM Treasury to plan for future upgrading of the Online System for Central Accounting and Reporting system to enable implementation of the revised Classification of the Functions of Government.
- **Recommendation 27:** Although not a high priority, the ONS should explore the feasibility of improvements to the measure of the ONS' productivity, to provide a measure that is adequate to feed into the estimates of total public service productivity.

Healthcare

- **Recommendation 28:** The ONS should evaluate the benefits and costs of switching from Finished Consultant Episodes to the person-level data provided by the Hospital Episode Statistics for measuring hospital output in England.
- **Recommendation 29:** The ONS should review, together with the Department of Health and Social Care, how the data needed for the quality adjustment for healthcare services should be produced or commissioned in the future.
- **Recommendation 30:** The ONS should continue to work with the NHS to improve data on NHS-funded services contracted from the independent sector.
- **Recommendation 32:** The ONS should monitor the quality of National Cost Collection activity data for England for abdominal aortic aneurysm, bowel cancer, breast, and cervical screening services, and transition to using these data when they are of adequate quality to bring it in-line with expenditure data used.
- **Recommendation 33:** The ONS and Department of Health and Social Care should explore sourcing data for other preventative activities, such as glaucoma screening for inclusion when data permits.
- **Recommendation 34:** The ONS should continue to engage with Wales, Scotland and Northern Ireland to assess the feasibility of including abdominal aortic aneurysm, bowel cancer, breast, and cervical screening services in the healthcare output measures for the devolved governments in the future.

- **Recommendation 39:** The ONS should continue to explore the feasibility of applying the improvements on handling equivalent treatment across different modes of provision made for England to Scotland, Wales and Northern Ireland.
- **Recommendation 41:** The ONS should keep under review whether incorporation of an adjustment for unnecessary A&E admissions would be feasible, material and proportionate.
- **Recommendation 42:** The ONS should continue to monitor the development of patient satisfaction surveys conducted by the ONS, the Care Quality Commission and others, with a view to further expanding the quality adjustment for patient experience in the future.
- **Recommendation 44:** The ONS should continue to work with the Department of Health and Social Care and the NHS to identify additional cases of lower-cost services being substituted for higher-cost services.
- **Recommendation 45:** Further research should be conducted to continue to explore the feasibility of improving and expanding the existing Healthcare quality adjustment for waiting times.
- **Recommendation 46:** The ONS should explore how to improve the measurement of preventative services in Healthcare output, including commissioning a literature review of data sources to ensure consistent application across the range of preventative treatments.
- **Recommendation 47:** The ONS should continue to investigate whether aspects of the incentives in NHS Payment Scheme can be incorporated in the relative weighting of different services in Healthcare productivity.
- **Recommendation 48:** The ONS should explore approaches to disaggregate the Healthcare service, to allow the relative performance of different components of this large service to be better understood.
- **Recommendation 49:** The ONS should actively explore working with NHS Wales to access its health data and develop stronger productivity metrics for Wales.

Education

- **Recommendation 51:** The ONS should continue to review if Annual Survey of Hours and Earnings remains the best data source for labour data for Education, or whether alternative sources may be preferable.
- **Recommendation 53:** The ONS should continuously engage with the Department for Education over indicators of student well-being.
- **Recommendation 54:** The ONS should work with the devolved governments to improve Education data sources as far as possible.

- **Recommendation 55:** The ONS should, as part of its research agenda, continue to explore the links between Education, the current Education quality measures, health expenditure and human capital acquisition, with specific attention to labour market returns, to better understand the output of these.
- **Recommendation 56:** The ONS should engage with academic researchers and stakeholders to understand if the impact of grading policy on Education productivity is substantial enough to warrant further research.

Defence

Recommendation 57: As input to the United Nations Review of the Classification of the Functions of Government (COFOG) the ONS should recommend that COFOG 2.1 Military defence should be split as follows:

- Division: 02 – Defence
  - Group: 02.1 – Military defence
    - Class: 02.1.1 – military defence of which contains operations of land forces
    - Class: 02.1.2 – military defence of which contains operations of air forces
    - Class: 02.1.3 – military defence of which contains operations of sea forces
- **Recommendation 61:** The ONS should continue to explore the direct measurement of Defence output with Ministry of Defence and other experts following the publication of this report.
- **Recommendation 62:** The ONS should continue to work with Ministry of Defence to explore the potential of using their new readiness measure within the calculation of Defence outputs, and to continue to develop appropriate methods.

Policing

- **Recommendation 63:** The ONS should endeavour to source more granular police workforce data for Scotland.
- **Recommendation 67:** The ONS should continue to explore with Home Office the potential to run an England and Wales Police Activity Survey as part of ONS's survey portfolio.
- **Recommendation 68:** The ONS should explore the potential to source police activity data for Scotland and Northern Ireland.
- **Recommendation 69:** The ONS should continue to collaborate with key police partners on appropriate methods to develop and weight different types of output activity.
- **Recommendation 70:** The ONS should draw together existing data sources for Police 'public safety and welfare' activities and develop a plan to fill data-gaps.
- **Recommendation 71:** The ONS should commission a literature review of the impact of crime prevention activities which fall under the Police 'criminal prevention' activity category, with a view to future development.

- **Recommendation 72:** The ONS should continue to explore the feasibility of introducing a combined Police crime and Police public safety and welfare output measure by critically assessing alternative activity and weighting sources.
- **Recommendation 73:** The ONS should continue to collaborate with policing partners to develop appropriate quality adjustments for Policing services.

Immigration and Citizenship services

- **Recommendation 74:** The ONS should seek to convince the United Nations Review of the Classification of the Functions of Government (COFOG) to separate Immigration and Citizenship activities from COFOG 3.1 Police services.
- **Recommendation 75:** The Review recommends commencing the split of Policing from Immigration and Citizenship services from 2004, and retaining the combined series for earlier years.
- **Recommendation 76:** The ONS should review the composition of the Immigration component of the combined Police and Immigration intermediate consumption deflator annually to ensure accurate representation.
- **Recommendation 77:** The ONS should review whether full cost recovery fees, for example for passports, should be treated as market-equivalent prices, which do not require quality adjustment, as prices should internalise quality change.
- **Recommendation 78:** The ONS should continue discussions with the Home Office to identify the most appropriate service activities and suitable weights for these to construct a direct volume output measure for Immigration and Citizenship services.

Public Order and Safety

- **Recommendation 81:** The ONS should explore possible new prison data sources for Scotland to allow for differential cost weights for different categories of prison, and data for Northern Ireland in the ambition of creating a UK-wide output measure.
- **Recommendation 82:** The ONS should continue to work with the Ministry of Justice to obtain expenditure data that enable the relative weights of different groups of probationers to vary with time in future years.
- **Recommendation 83:** The ONS should explore the data available on probation services for Scotland and Northern Ireland with a view to creating a complete UK measure for probation output.
- **Recommendation 84:** The ONS should continue to work with the Ministry of Justice to obtain expenditure data to derive improved weights for individual law courts service areas.
- **Recommendation 85:** The ONS should explore the data available for Scotland and Northern Ireland with a view to creating a complete UK measure for criminal court output.

- **Recommendation 86:** The ONS should explore the data available for the Scotland and Northern Ireland with a view to creating a complete UK measure for legal aid output.
- **Recommendation 87:** The ONS should undertake further research into whether legal aid should be considered an intermediate stage in the provision of justice as opposed to an output in itself.
- **Recommendation 91:** Work should continue on identifying appropriate data to develop a direct measure of labour input for the prison service component of Public Order and Safety.
- **Recommendation 92:** Work should continue on developing a direct measure of labour input for the courts service component of Public Order and Safety, ensuring no double-counting with intermediate consumption input in this area.

Taxation

- **Recommendation 95:** The ONS should work with HM Revenue and Customs to continue to ensure data on different taxes is utilised in as granular a fashion as benefits productivity estimates.
- **Recommendation 96:** The ONS should explore how to complement existing HM Revenue and Customs data to complete coverage of Tax Administration activities.
- **Recommendation 97:** The ONS should investigate further developments that could be made to account for changes in quality such as fraud and error, customer satisfaction and waiting times, but recommends that treatment of the 'tax gap' should be the priority area for focus.
- **Recommendation 98:** The ONS should review data on taxes administered locally and in devolved governments in relation to inputs and outputs, to explore the feasibility of further improving coverage of Taxation Administration.
- **Recommendation 99:** The ONS should investigate the feasibility of extending the time period for Tax Administration productivity measure to pre-2018.

Social Security Administration

- **Recommendation 100:** The ONS should undertake further work to identify the best treatment of the Coronavirus Job Retention Scheme.
- **Recommendation 101:** The ONS should review whether data exists to better understand, for Social Security Administration, whether non-administrative costs for concessionary fares and Housing Benefit are correctly recorded, and whether they are sufficiently impactful to merit further work.
- **Recommendation 103:** The ONS should seek to acquire data to incorporate Tax Credits, Housing Benefit and Child Benefit in both output and inputs measures of Social Security Administration, while making appropriate adjustments to the "Other" grouping of inputs.

- **Recommendation 106:** The ONS should work with the Department of Work and Pensions to replace stock measures of fraud and error with more appropriate flow measures to better reflect in-period performance, in Social Security Administration quality adjustment.
- **Recommendation 107:** The ONS should investigate data availability accounting for benefit administration in Northern Ireland, as the current measure does not include that activity.
- **Recommendation 108:** The ONS should further investigate whether administrative costs associated with benefits delivery are being captured in Social Security Administration.
- **Recommendation 109:** The ONS should continue to work to identify the most appropriate route to capture user satisfaction in Social Security Administration across the UK.
- **Recommendation 110:** The ONS should continue to collaborate with the Department of Work and Pensions to develop data on timeliness for Social Security Administration.
- **Recommendation 111:** Weighting mechanisms for baskets of quality adjustment factors, such as in Social Security Administration, should be considered and if necessary, data collected to ensure quality adjustment estimates reflect societal preferences and that these are aligned with how other relevant areas are treated, such as health waiting lists.

Local Government, including Adult and Children's Social Care

- **Recommendation 112:** The ONS should continue to work with the Ministry of Housing, Communities and Local Government to determine a strategy to capture the remaining local services, with a particular focus on ensuring alignment of treatment with Environmental Services, noting the potential impact of the United Nations Classification of the Functions of Government Review.
- **Recommendation 113:** Local government data across the devolved governments should be sought to enable UK-wide local government estimates to be prepared.
- **Recommendation 114:** The ONS should seek to separate elements of Children's Social Care where output measures cannot be identified, with particular reference to Early Years Childcare, and in the absence of direct output measures estimate output in this area through the 'inputs = outputs' methodology until further measures can be developed.
- **Recommendation 115:** The ONS should periodically investigate the methodology into Children Social Care in relation to non-government sectors in order to make further improvements to the measure.
- **Recommendation 116:** The ONS should continue to keep the Adult Social Care system under review with the aim, utilising ongoing data developments, to address areas of this service which continue to use the 'inputs = outputs' methodology.

- **Recommendation 117:** The ONS should consider the user need for a separate Adult Social Care, England publication in the suite of public service productivity publications.

Environment

- **Recommendation 118:** The ONS should work with relevant departments to recommend a definition of 'Environmental Services' to the upcoming consultation on the international Classification of the Functions of Government (COFOG) definitions to encourage the creation of a wider COFOG sector covering 'Environmental Services', for use in developing measures, using the work of this Review to inform developments in this space.
- **Recommendation 119:** An improved Classification of the Functions of Government structure for 'Environmental Services' should differentiate between 'Environmental Protection', 'Natural Resource Management' and 'Climate Change and Net Zero', with further detail embedded below this structure.
- **Recommendation 120:** The ONS should undertake research to identify appropriate deflators for inputs into, and outputs from, Environmental services.

Annex A: Summary for each service examined by the review

ANNEX A:

Summary for each service examined by the review

Healthcare

What percentage of total government expenditure?

41.5% of total government spend in 2021.

What has the Review done?

Inputs:

- Small changes to data, now include legal and audit services.
- Labour inputs weights updated to align with Full-Time Equivalent staff.

Outputs:

- Inclusion of preventative services commissioned to non-NHS providers.
- Incorporated certain screening services.
- Improved measures for primary care.
- Equivalised unit costs for equivalent treatments across different modes of provision.
- Removal of excess bed days activity.

Quality adjustment:

- Expanded measure of GP patient outcomes.
- Quality and Outcomes Framework used to weight indicators.
- Extend the application of the health gain quality adjustment from elective to non-elective procedures.

What are the recommended incremental next steps?

Continue to monitor:

- Data on NHS-funded services contracted from the independent sector.
- Quality of National Cost Collection activity data for selected screening services.
- Development of patient satisfaction surveys.

Work with devolved governments to:

- Improve the coverage of Healthcare services not currently measured in the output for the devolved governments.
- Assess the feasibility of including selected screening services in the output measure.
- Explore the feasibility of applying the improvements on handling equivalent treatment across different modes of provision made for England to the devolved governments.

What are the remaining big issues for research?

- Evaluate the benefits and costs of switching to the person-level data provided by the Hospital Episode Statistics for measuring hospital output in England.
- Further research to continue to explore the feasibility of improving and expanding the existing Healthcare quality adjustment for waiting times.
- Explore how to improve the measurement of preventative services in Healthcare output, including reviewing potential data sources to ensure consistent application across the range of preventative treatments.
- Investigate whether aspects of the incentives in NHS Payment Scheme can be incorporated in the relative weighting of different services in Healthcare productivity.

Education

What percentage of total government expenditure?

15.8% of total government spend in 2021.

What has the Review done?

Inputs:

- Improved methodology for weighting labour inputs and new imputation for any missing salary categories.

Outputs:

- Capturing the shift in the compulsory education age to 18 years.
- Better reflecting the conversion of primary schools to academies.
- Capturing the increase in activity from the increase in funding entitlement for pre-primary.
- Updating the treatment of healthcare and teacher training to reflect changes in provision.

Quality adjustment:

- Inclusion of student wellbeing.
- Improved accounting for attainment during the coronavirus (COVID-19) pandemic.
- Inclusion of Further Education attainment.

What are the recommended incremental next steps?

Continue to review:

- If full-time salary data for relevant staff categories for education are being captured in the Annual Survey of Hours and Earnings (ASHE) post-2020, and whether alternative sources may be needed for occupations no longer available in ASHE.
- If alternative sources of attainment during the coronavirus (COVID-19) pandemic are available.

Work with devolved governments to:

- Improve data sources as far as possible, particularly Northern Ireland and their attainment statistics.

What are the remaining big issues for research?

- Continue to explore the links between education, health expenditure and human capital acquisition, and particularly labour market returns, to better understand the output of these sectors.
- Grading policies change with changing Government Administrations. More research is needed on how this is reflected in productivity statistics.

Defence

What percentage of total government expenditure?

9.0% of total government spend in 2021.

What has the Review done?

General:

- The ONS inputted to the United Nations Review of Classifications of Functions of Government to recommend a further breakdown of the Defence categories.

Inputs:

- A more granular direct measure of labour inputs has been implemented.
- Improvements have been made to the intermediate consumption and capital deflators.

What are the recommended incremental next steps?

- Work with Ministry of Defence to explore their new readiness measure when available and its potential for underpinning output metrics.

What are the remaining big issues for research?

- Continued research into the measurement of what constitutes Defence output.
- Subsequent research on quality adjustment measures when an output becomes available.

Policing

What percentage of total government expenditure?

Policing and Immigration and Citizenship combined = 4.5% of total government spend in 2021.

What has the Review done?

Inputs:

- More granular police salary data.
- Account for impact of the Police Uplift Programme.
- Reviewed impact of reduced capacity.
- Improved coverage for the devolved governments.
- Improved deflators.

Output:

- Investigated potential data sources and methods for producing a direct output measure.

What are the recommended incremental next steps?

- Continue to explore with Home Office the potential to run an England and Wales Police Activity Survey as part of the ONS's survey portfolio, and investigate the potential to source similar data for Scotland and Northern Ireland.
- Continue to develop direct output measure for policing.

What are the remaining big issues for research?

- Further investigate data sources to fill existing data gaps in police 'Public Safety and Welfare' activities.
- A literature review of the impact of crime prevention activities which fall under the police 'criminal prevention' activity category.
- Further research into introducing a combined police crime and police 'Public Safety and Welfare' output measure by critically assessing alternative activity and weighting sources.
- Investigation of suitable quality adjustment measures.

Immigration and Citizenship

What percentage of total government expenditure?

Policing and Immigration and Citizenship combined = 4.5% of total government spend in 2021.

What has the Review done?

General:

- The ONS inputted to the United Nations review of the Classification of Functions of Government to recommend separating Immigration and Citizenship activities from Policing.

Inputs:

- Identification of inputs measure needs.
- Improved deflators.

Outputs:

- Identification of some sources of activities and associated cost weights

What are the recommended incremental next steps?

- Implement separate inputs for Immigration and Citizenship from 2004.
- Continue to monitor the deflators used for input to ensure robustness.
- Further research into sources of activities and associated cost weights.

What are the remaining big issues for research?

- Defining suitable activity measures for the output of Immigration and Citizenship services.
- Research into quality change in this service.

Public Order and Safety (POS)

What percentage of total government expenditure?

3.1% of total government spend in 2021.

What has the Review done?

Inputs:

- Implemented intermediate consumption deflators to all POS services, apart from fire.
- Creation of a direct measure for fire service labour inputs.

Outputs:

- Improved output measures for criminal courts, legal aid, probation and prisons.

Quality adjustment:

- More granular data used for timeliness, in line with improvements to outputs.

- Re-introduced the coronavirus (COVID-19) pandemic impacted re-offending quality adjustment.
- Improvement of weighting of quality adjustments.

What are the recommended incremental next steps?

- Further development and implementation of improved inputs and output measures, including expanding to cover devolved governments.
- Further research on expenditure weights for the different services in POS output.

What are the remaining big issues for research?

- Research whether legal aid should be considered an intermediate stage in the provision of justice as opposed to an output in itself.
- Identification of appropriate data to develop a direct measure of labour inputs for each of prison services and court services.

Tax Administration

What percentage of total government expenditure?

0.6% of total government spend in 2021.

What has the Review done?

General:

- Tax Administration has been removed from the 'Other' grouping and is now treated as its own services. This should be treated as 'Official statistics in development' whilst undergoing further testing.

Inputs:

- HMRC 'Cost of Collection' data, adjusted by Government expenditure data, identified as suitable for use.

Outputs:

- Development of experimental revenue adjusted output measure for Taxation.

Quality adjustment:

- Identification of potential data sources for activities and associated cost weights.

What are the recommended incremental next steps?

- Obtain feedback on experimental output measure.
- Look to integrate further tax data, including custom duties, activities to reduce tax evasion and non-compliance.
- Look to further increase coverage of taxes administered locally and in devolved governments.

What are the remaining big issues for research?

- Investigate the feasibility of extending the time period for Tax Administration productivity measure to pre-2018.
- Investigate the feasibility of Tax Administration quality adjustments.

Social Security Administration (SSA)

What percentage of total government expenditure?

1.7% of total government spend in 2021.

What has the Review done?

Outputs:

- Created a new benefit-weighted index for the administration output of Universal Credit and legacy benefits. This should be treated as 'Official statistics in development' whilst undergoing further testing.
- Created a cost weighted index for remaining benefits.
- Using total administrative expenditure to weight together the two output indices.
- A method to account for the changes in composition of Universal Credit claims.

Quality adjustment:

- Inclusion of fraud and error rates as a quality adjustment.

What are the recommended incremental next steps?

- Investigate sources to incorporate data on Child and Working Tax Credits, Housing Benefit and Child Benefit in both output and inputs measures of SSA, while making appropriate adjustments to the 'Other' grouping of inputs.
- Investigate data availability accounting for benefit administration in Northern Ireland.
- Further refinement of the fraud and error quality adjustment.

What are the remaining big issues for research?

- Local authority administrative costs for Housing Benefit and concessionary fares need to be fully investigated.
- Continue work towards fully covering SSA across central government departments, local authorities and devolved governments.
- Customer satisfaction and timeliness, as well as their weighting, as Quality Adjustment factors needs further research.

Local Government, including Adult and Children's Social Care

What percentage of total government expenditure?

Local Government is currently a part of 'other' grouping. Adult and Children's Social Care make up 5.7% and 2.7% of total government spend in 2021, respectively.

What has the Review done?

Local Government:

- Early identification of potential data sources for inputs and outputs.

Adult and Children's Social Care:

- Not prioritised for review given the significant improvements implemented in 2019.
- Further potential improvements identified.

What are the recommended incremental next steps?

Children's Social Care:

- Separate elements where output measures cannot be identified, and estimate output in this area through the "inputs = outputs" methodology until further measures can be developed.
- Periodically investigate the methodology into Children's Social Care in relation to non-government sectors in order to make further improvements to the measure.
- Investigate new source of data for inputs and output for the devolved governments.

Adult Social Care:

- Investigate the feasibility of using new person-level data collected by NHS England to improve output measures.
- Investigate the user need for an Adult Social Care specific publication.

What are the remaining big issues for research?

- A strategy is needed to capture local services that do not sit with other services, reflecting potential change in scope of Environmental services.
- Local government data across the devolved governments should be investigated to create a UK-wide local government services measure.

Environmental Services

What percentage of total government expenditure?

Environmental services is currently part of 'other' grouping.

What has the Review done?

- Developed a draft categorisation of Environmental Services to underpin future measurement.
- The ONS inputted to the United Nations (UN) review of the Classification of Functions of Government (COFOG) to recommend the creation of a COFOG sector covering Environmental Services, for use in developing measures in future.

What are the recommended incremental next steps?

- Continue to influence the UN Review of COFOG to encourage the creation of a COFOG sector covering Environmental Services.

What are the remaining big issues for research?

- Implement improved Environment measures, in line with new international standards, if they are felt to meet UK needs

ANNEX B:

Roadmap into the UK National Accounts

As referenced in Chapter 6 in preparation for the System of National Accounts (SNA) 2025 [see link 3, Annex I], the Review has developed a high level 'roadmap' to incorporate quality adjustments into the UK National Accounts. This has been compiled as part of the Office for National Statistics (ONS) planning for implementation of SNA25 more widely. This roadmap will need to be maintained and updated in alignment with the final SNA25 implementation plans.

Roadmap footnotes

1. The column 'Preparing for implementation' reflects an interval of unspecified length. The precise length of this phase will depend on a range of factors including resourcing, the methods approval process of proposed changes, systems development and other factors impacting delivery.
2. National Statistician's Committee for Advice on Standards.
3. A resourcing assessment will be carried out to determine the additional business-as-usual requirements for processing improvements in the UK National Accounts.
4. This refers to developing statistical processing systems to produce cost weighted activity indices which form the basis of implied deflators for non-market output in the national accounts.
5. UK National Accounts software development refers to changes that are required to existing national accounts processing systems to accommodate elements of the proposed improvements to non-market output volumes.
6. The proposed changes may be implemented over a series of annual 'Blue Book' publications dependent on readiness and priority.
7. The introduction of quality adjustment is one of a number of improvements associated with the implementation of the new System of National Accounts framework (SNA25) that will impact general government output volumes. The extent to and time frame under which ONS adopts and implements all parts of SNA25 is under review and will be communicated in the future. For further information, see UK Statistics Authority Chair Sir Robert Chote's keynote address Reflections on the forthcoming System of National Accounts revision [see link 99, Annex I].
8. The development of the new Statistical Business Register (SBR) is intended to be a reliable and definitive register of UK businesses and will be used as the main sampling frame for business surveys conducted by ONS and other government departments. It will replace the current Inter-departmental Business Register (IDBR).

9. Decisions on how to follow SNA 2025 are discussed by the National Statistician's Committee for Advice on Standards for Economic Statistics (NSCASE), whose recommendations are then submitted to the National Statistician for final approval.
10. SIC refers to the Standard Industrial Classification, and COFOG refers to the Classification of the Functions of Government.

ANNEX C:

The Current Classification of Functions of Government (COFOG) Structure

C.1 The Classification of Functions of Government: ONS Recommendations

There are currently several areas of the United Nations (UN) Classification of Functions of Government (COFOG) [see link 4, Annex I] that the Review thinks could be improved to better support accurate measurement of Government expenditure. This would enable improvements to the granularity and quality of economic statistics derived from HM Treasury's system for recording government expenditure, the Online System for Central Accounting and Reporting (OSCAR). These expenditure data are fundamental to the measurement of public services productivity.

This Annex summarises key potential areas for COFOG improvement that the Office for National Statistics (ONS) has recommended to the UN review of COFOG. That review is expected to report in 2027.

C.2 Tax Administration

As part of the Public Services Productivity Review, an experimental productivity statistic for Tax Administration has been developed. This relies on the availability of government spending data to construct accurate input metrics, as well as the availability of measures for output. Government spending on Tax Administration is currently entirely captured within financial and fiscal affairs, COFOG 1.1.2, which also includes management of public funds and public debt as well as other functions. It is therefore not independently identifiable or measurable, despite being an essential service undertaken in some form by almost all governments, often by a distinct body with responsibility for tax collection alone (rather than wider public finance management).

Current structure

Division: 01 – General public services

- Group: 01.1 – Executive and legislative organs, financial and fiscal affairs, external affairs
  - Classes
    - 01.11 Executive and Legislative Organs
    - 01.12 Financial and fiscal affairs (including tax admin)
      - Administration of financial and fiscal affairs and services; management of public funds and public debt; operation of taxation schemes; operation of the treasury or ministry of finance, the budget office, the inland revenue agency, the customs authorities, the accounting and auditing services; production and dissemination of general information, technical documentation and statistics on financial and fiscal affairs and services.
      - Includes: financial and fiscal affairs and services at all levels of government. Excludes: underwriting or flotation charges and interest payments on government loans (01.7.0); supervision of the banking industry (04.1.1).
      - 13 External Affairs (CS)

To enable accurate input measures to underpin a productivity statistic, the Review recommends developing an extra class within group 01.1 which separately identifies spending on Tax Administration.

Alternative approach

- 01.11 Executive and legislative organs
- 01.12 Financial and fiscal affairs
  - Administration of financial and fiscal affairs and services; management of public funds and public debt; operation of the treasury or ministry of finance, the budget office, the accounting and auditing services; production and dissemination of general information, technical documentation and statistics on financial and fiscal affairs and services.
- 01.13 Administration of taxation
  - Operation of taxation schemes; operation of the inland revenue agency, the customs authorities; public information schemes and communication related to taxation. Includes: activities related to counter-fraud and prosecution of tax evasion and smuggling.
- 01.14 External affairs

C.3 Defence

Currently, military Defence spending is categorised by a single COFOG group (2.1.0). To improve the way the inputs for Defence are measured from a productivity perspective, it would be advantageous to be able to split this out further, to cover at least the operation of land, air and sea forces individually. This would also reflect the more straightforward nature of recording this spend incomparison to other areas of Defence.

It may also be useful to see nuclear, cyber and space disaggregated too at the lower levels, however this Review's overall priority is to at least have the military defence classes split by land, sea and air. Whether this division is made at the second level or third level of the classification, the Review is happy to discuss further. Initially, the Review suggests being led by the demand for this change within the UN.

Current structure

Division: 02 – Defence

- Group: 02.1 – Military defence
  - Class: 02.1.0 – Military defence
    - Administration of military defence affairs and services; operation of land, sea, air and space defence forces; operation of engineering, transport, communication, intelligence, personnel and other non-combat defence forces; operation or support of reserve and auxiliary forces of the defence establishment.
    - Includes: offices of military attachés stationed abroad; field hospitals. Excludes: military aid missions (02.30); base hospitals (07.3); military schools and colleges where curricula resemble those of civilian institutions even though attendance may be limited to military personnel and their families (09.1), (09.2), (09.3) or (09.4); pension schemes for military personnel (10.2).

Alternative approach

Division: 02 – Defence

- Group: 02.1 – Military defence
  - Class: 02.1.1 – military defence of which contains operations of land forces
  - Class: 02.1.2 – military defence of which contains operations of air forces
  - Class: 02.1.3 – military defence of which contains operations of sea forces

C.4 Preventative Health Spending

Currently only COFOG 7.4 explicitly covers prevention activities as "public health services", however it also includes the administration of public health services, epidemiological data, family planning etc, while many preventive care activities are likely to be delivered by a provider that fall under other COFOG categories, such as hospitals (COFOG 7.3). This does not allow the separate identification of preventative public health spending, and therefore assessment of its long-term impact on improving the productivity of the Healthcare public service.

Current structure

Division: 07 – Health

- Group: 07.4 – Public Health Services

07.40 PUBLIC HEALTH SERVICES (IS) Provision of public health services; administration, inspection, operation or support of public health services such as blood-bank operation (collecting, processing, storing, shipping), disease detection (cancer, tuberculosis, venereal disease), prevention (immunisation, inoculation), monitoring (infant nutrition, child health), epidemiological data collection, family planning services and so forth; preparation and dissemination of information on public health matters.

Alternative approach

To better represent spending on preventive care activities, the COFOG classification could report expenditure based on the System of Health Accounts (2011) framework's definition of Healthcare function (ICHA-HC), either at the second- or third-digit level. This would represent a shift away from the current reporting practice, which is more tailored towards recording spend by provider organisation (for example hospitals as COFOG 7.3, outpatient providers as COFOG 7.2). This would be advantageous to identifying preventive Healthcare services, as these services are not just provided by specific public health bodies (for example the UK Health Security Agency) but also within other Healthcare settings such as general practice, dentistry, hospitals, etc. One advantage of using the IHCA-HC definitions of preventive care spending as a basis for further disaggregation within COFOG is that this is an already well-established set of internationally standardised definitions.

Based on the Health accounts framework that identifies primary and secondary preventive care (not tertiary prevention), a proposed alternative COFOG structure is:

Division: 07 – Health

- Group: 07.4 – Public Health Services
  - Class 07.4.1 Information, education and counselling programmes (e.g. healthcare information campaigns, smoking cessation programmes, antenatal classes)

- Class 07.4.2 Immunisation programmes
- Class 07.4.3 Early disease detection programmes (e.g. screening programmes)
- Class 07.4.4 Healthy condition monitoring programmes (e.g. health check-ups, antenatal care, etc)
- Class 07.4.5 Epidemiological surveillance and risk and disease control programmes
- Class 07.4.6 Preparing for disaster and emergency response programmes
- Class 07.4.7 All other Public Health Services

C.5 Education

Currently it is challenging to determine what expenditure is attributed to the Further Education (FE) sector. Education expenditure is captured under COFOG 9 and follows the structure:

Current structure

Division: 09 – Education

- Group: 09.1 – Pre-primary and primary education
- Group: 09.2 – Secondary education
  - Class: 9.2.2: Upper Secondary Education
- Group: 09.3 – Post secondary non-tertiary education
- Group: 09.4 – Tertiary education
- Group: 09.5 – Education not definable by level
- Group: 09.6 – Subsidiary services to education
- Group: 09.7 – R&D Education
- Group: 09.8 – Education N.E.C

Currently, expenditure for FE is captured under COFOG 9.2.2 'Upper Secondary Education'. COFOG 9.2.2 captures expenditure relating to the second (final) stage of secondary education which prepares students for tertiary education or providing skills relevant to employment across a range of subject options and streams. COFOG 9.2.2 disaggregated figures from COFOG 9.2 are not currently available, but this is likely due to the approach by which OSCAR records the data.

This COFOG structure could benefit from structural changes that allow for visualisation of FE inputs, and to make it clearer where expenditure is related to FE. For example, in the UK, the FE space is typically termed as education for 16 to 19-year-olds, but the current naming 'Upper Secondary Education' does not make this immediately clear. Other countries will not follow this 16 to 19-year-old referral, but it may be worthwhile considering an internationally recognised term for FE to be clearly stated within the COFOG structure. For example, 'Post-Secondary Education' could potentially be more appropriate.

Furthermore, granular figures for FE are not available within the current structure, therefore the system may benefit from restructuring in such a manner that FE inputs are clearly visible for users.

It must also be acknowledged that the method of recording the data is also relevant in this case. The approach by which the data are recorded in OSCAR may be a mechanism that does not allow for obtainment of disaggregated figures from COFOG 9.2.2. Therefore, some structural changes, in addition to the method of recording, should be considered when evaluating education. Such changes might include:

Alternative approach

Division: 09 – Education

- Group: 09.1 – Pre-primary and primary education
- Group: 09.2 – Secondary education
- Group: 09.3 – Post Secondary education
- Group: 09.4 – Post secondary non-tertiary education
- Group: 09.5 – Tertiary education
- Group: 09.6 – Education not definable by level
- Group: 09.7 – Subsidiary services to education
- Group: 09.8 – R&D Education
- Group: 09.9 – Education N.E.C

This would allow FE to be included in its own sub-COFOG (0903 in the example).

C.6 Net Zero and climate change mitigation government expenditure

Under current measures, spending on climate change mitigation measures are not adequately captured. COFOG division 5 covers environmental protection but is mainly focused on broader environmental activities such as waste management, nature conservation and pollution reduction. While these will cover aspects of climate change adaptation, mitigation activities will largely be included within 4.3 Fuel and Energy, and 4.5 Transport, along with other activities associated with housing and communities that are captured in other COFOG categories.

As Governments worldwide respond to, and mitigate, climate change, having a dedicated Net Zero and climate change mitigation COFOG would allow for closer measurement of expenditure related to this area, in turn facilitating the evaluation of its effectiveness. This might include:

- Investment in renewable energy (currently bundled up with in 4.35 Electricity)
- R&D linked to climate change mitigation projects
- Transition assistance for households and businesses (eg subsidies for insulation or heat pumps)

- Electrification of transport vehicles and networks
- Flood defences

C.7 Policing and immigration

Policing and Immigration expenditure are currently both covered by Police services (COFOG 3.1). Substantial changes in expenditure for either service make spending look volatile overall. Disaggregated public expenditure data are available using 'HM Treasury-defined sub-functional divisions' and the Review proposes that this is substantively reflected in the structure through the creation of a second level COFOG to cover 'Immigration and Citizenship' (i.e. 3.2 – immigration). COFOG 3.1 would remain and be limited to Police services. This would support developing separate productivity series for both policing and immigration.

Current structure

Division: 03 – Public order and safety

- Group: 03.1 – Police services

Alternative approach

Division: 03 – Public order and safety

- Group: 03.1 – Police services

Includes: anything in the current definition not related to immigration

Excludes: Immigration

- Group: 03.2 – Immigration

Includes: all immigration services removed from 3.1

ANNEX D:

Quality adjustment of education

To generate quality adjusted attainment estimates for Education, the Office for National Statistics (ONS) takes account of educational performance, in terms of qualifications and grades achieved by students. This annex will discuss in greater depth two topics. Firstly, the wider UK National Accounts implications of this and secondly, how the Review has improved the method of applying qualification data to different cohorts.

D.1 UK National Accounts implications

It should be noted that the methods proposed in this Review look to align to UK National Accounts conventions, as per the Atkinson Report [see link 1, Annex I]. The area this is most immediately pertinent is in relation to human capital.

Human capital is recognised in the national accounts through the spending on Education and Healthcare which creates it, and the wages which result from it. However, human capital itself is not recognised as a capital and the flows which relate to it are not considered as investment or earnings from capital: they are considered recurrent spending and compensation of employees. As described in Canry (2019) [see link 100, Annex I], Dunn (2022) [see link 101, Annex I] and Mubarak Heys, Meirinhos, and Taylor (2024) [see link 102, Annex I], this constrains efforts to incorporate human capital into national accounts and hence into this Review.

If human capital was recognised, the return to Healthcare and Education spending could be 'monetised' through a rate of return calculation treating the discounted sum of additional future earnings as the value created by this expenditure. This would have two effects: it would significantly increase the value added of these services, but it would also require substantial parts of this expenditure to be reclassified as Gross Capital Formation by the government sector.

Alongside this issue, there is a further issue related to a core Atkinson principle, around attribution. Lifetime earnings can go up with improved qualifications because of either improvement in skill levels as a result of gaining the qualifications, or because of a signalling effect where gaining qualifications reveals the individual as having higher innate ability. Because the latter cannot be attributed to government spending, there is an argument this metric should not be used to quality adjust public services. In addition, extra qualifications may be gained because of privately-funded education or private spending on tutors, materials, books or other educational factors. Again, this mitigates against using human capital as a quality adjustment.

Further, human capital will not be addressed in the System of National Accounts (SNA) 2025 for technical reasons and a broad agreement that insufficient international research has gone into developing the necessary methods. ([See link 3, Annex I] for more details on SNA 2025). Mubarak, Heys, Meirinhos, and Taylor (2024) attempts to close this gap but this Review encourages national statistics institutes around the world to undertake the necessary research, so the opportunity offered by future SNA revisions is not missed again. Nevertheless, developments at the ONS around its new Inclusive Income and Wealth Accounts, which deliberately extend the asset boundary to include human capital, provide a potential mechanism to showcase the value of improving the understanding of human capital formation.

Finally, it is important to note that whilst human capital is the largest example of this problem, the same issues arise for spending on the environment as natural capital, spending on museums, galleries and national monuments as heritage capital, and defence, justice and other mechanisms to ensure the rule of law and defence of property rights, which could be conceptualised as social capital. In fact, large swathes of government spending could be reconceptualised as types of capital expenditure in this fashion. Obviously, such a sweeping change would call for wholesale revision of the methods proposed in this Review and currently used. Such a piece of work is outside the scope of this Review, but it is important to be aware of the alternatives and their potential impacts.

D.2 Applying quality adjustments

Attainment

The ONS does not directly apply growth rates in exam scores in the year they are achieved as this would imply that GCSE exam achievement is solely attributable to performance during year 11. Instead, the model considers education as a cumulative experience, where each year of education builds on existing knowledge and skills, and contributes in aggregate towards the final academic achievement.

For example, GCSE attainment data published for the academic year ending 2018 reflects the effectiveness of the teaching from Reception year (in this case, academic year ending 2007) to Year 11 (academic year ending 2018). As such, the “cohort split” approach applies weighted contributions from the new attainment data back to previous years.

The coronavirus (COVID-19) pandemic compelled reconsideration of this model because it is methodologically incorrect to attribute the effect of the pandemic into the back series. For example, if attainment dropped by 10% during the pandemic (academic year ending 2021), there is no justification for reducing the attainment in academic year ending 2018 for students who were in school at that time.

The coronavirus pandemic did not affect previous time periods: if the coronavirus pandemic had not occurred the Review would have expected students to have otherwise demonstrated a similar performance to their immediate predecessors, that is the cohort who completed their examinations in, for example, academic year ending 2019.

Therefore, a key aim for Education during the second year of the public service productivity (PSP) Review was to revise and re-develop the cohort-split model to robustly account for the coronavirus pandemic's influence on attainment. This development can be broken into two elements: the first is sourcing evidence to scale the impact of the coronavirus pandemic, and the second is how to augment the methodology to cope with this shock.

In the Public service productivity annual publication of April 2023 [see link 103, Annex I], prior to the beginning of the Review, the ONS used learning loss' metrics [see link 104, Annex I] published by the Department for Education (DfE) and Education Policy Institute (EPI) to inform attainment from academic year ending 2020. These measures only applied to England schools, however in the absence of similar metrics for the devolved governments, the English learning loss data were applied to all school phases across the UK.

The learning loss measures represented the average months of learning lost in English and Mathematics for Primary and Secondary schools during the academic year ending 2020 and 2021, in comparison with pre-pandemic performance. Although these data served as appropriate proxy indicators of attainment, upon reflection and discussion with experts on this topic, the Review has concluded that this measure is not the best for its purpose.

Learning loss is not strictly a measure of academic performance, it is used to gauge how in front or behind one cohort is performing in comparison to another. This renders the measure inconsistent to what would typically be used in the model.

Learning loss potentially double-counted movements that were being accounted for in the quantity output model, which was designed to account for lost teaching hours during the pandemic. The quantity output model was modified to account for the impacts of challenges that arose during the pandemic such as remote learning and increased absence rates.

Learning loss was used to adjust the raw attainment index in academic year ending 2020, but it had no further impact on the subsequent 'cohort-split' model. Nevertheless, a break in the cohort model had to be applied at academic year ending 2019 as a result so previous years were not impacted by the loss in academic year ending 2020. The learning loss impact applied in academic year ending 2020 meant this year was 'fixed' and hence caused a break in the model whereby attainment in future years could not be 'cast-back' into academic year ending 2020. This led to a break in the calculations at this point.

The study commissioned by the DfE and EPI did not have data for academic years ending 2021 onwards. Although alternative studies are available which have assessed learning loss during the pandemic, the points raised in this annex emphasises the need for actual attainment data to inform quality adjustments for Education.

Following conversations with DfE and the Office of Qualifications and Examinations Regulation (Ofqual), the National Reference Test (NRT) was identified as an alternative indicator of academic performance ([see link 43, Annex I] for more information on the NRT). The NRT, commissioned by Ofqual in 2017, was implemented to monitor GCSE-level performance over time, and is a short English and maths exam which is delivered to Year 11 students in England who are due to take their GCSE's that academic year. The questions reflect the kinds of questions that will be taken in a formal GCSE examination. A nationally representative 300 of the 3,452 secondary schools are selected to take the test.

The outcomes of the first NRT in 2017 were benchmarked against GCSE results in 2017 to establish a baseline for following years. In 2017, the test booklets were marked, and the proficiency of the students were obtained; the proficiency estimates were benchmarked against the proportions of students achieving grade 7 and above, grade 5 and above, and grade 4 and above at GCSE. From 2018, the proportions of students at these proficiency levels are compared to the 2017 baseline, which allows for changes in academic performance over time to be observed.

The NRT has key advantages in informing attainment over learning loss:

- The NRT is a direct indicator of GCSE-level performance and has a strong link to academic achievement that doesn't apply to the learning loss measures.
- The NRT is an appropriate metric for the cohort model, as the years prior to the pandemic will have had some bearing on NRT outcomes, and it is consistent with data typically used in the model. This means that the cumulative nature of education can continue to be accounted for in ongoing PSP estimates for Education.
- The NRT is independent of teacher assessed grades and avoids debates around whether grade improvements were a result of better education or grade inflation.
- Data for the NRT are available in a consistently measured time series back to 2017, and it covers the academic years affected by the pandemic.

On the other hand, there are some limitations that do need to be acknowledged when using the NRT, such as:

- The NRT is only undertaken in England, but no similar measures have been uncovered for the devolved governments.
- The NRT only assesses performance in Maths and English, but not other subjects. This is unlike, for example, the Attainment 8 score (which was typically used before the pandemic) which considers a pupil's average grade across 8 subjects.

- Although the NRT is an indicator of GCSE-level performance, it is not a formally recognised GCSE, therefore it is not entirely consistent with previous measures used to inform quality adjustment for Education.

Upon consideration of the strengths and limitations, the NRT is the only viable source of secondary school attainment during the pandemic. Therefore, for 2020 and due to the absence of alternative data, the NRT was also used to inform attainment for England and the devolved governments. GCSE attainment data has returned to pre-pandemic arrangements in academic year 2023 to 2024.

In terms of how the cohort-split model needed development to absorb this data, this falls into two parts: the first is how to cope with structural breaks around the coronavirus pandemic to cope with adjustments made only during the pandemic years, and secondly this work provoked a deeper review of the weights applied in the model to allocate attainment across all the years each cohort is in education.

To adapt the Cohort model in the years of the coronavirus pandemic, the Review had to add some methodological adjustments. The key elements, discussed in the following section were:

- How to address the years prior to the coronavirus pandemic for students in school during the pandemic.
- How to apply residual balancing adjustments in the light of the solution to previous bullet.
- Whether in the light of the Review the weights used in the cohort-split model needed to be updated.
- How to handle missing data in this new model.

Previous Years:

The existing model works to retrospectively apportion attainment outcomes across all years spent in education instead of taking it all from one single year's attainment score. The ONS made an additional adjustment to the academic years affected by the coronavirus pandemic, namely academic years ending 2021 and 2022. Prior to the coronavirus pandemic, the attainment index score for a given year would be back-cast and apportioned between year groups in the back-series e.g. X% of the score goes to Year 7, Y% of the score goes to Year 10 etc. However, during the pandemic years, the attainment index score has not been back-cast in the typical manner, but rather year groups prior to academic year ending 2021 are now calculating using the average performance of respective year groups over the previous five years. The Review considers this provides the most reasonable estimate of likely attainment at this time in the absence of final attainment data unaffected by the coronavirus pandemic for these year groups.

However, this imposes a constraint on the system: by effectively fixing the value for one or more years, this means the total contributions from within a cohort affected by the pandemic will not automatically equal the attainment value

achieved when that cohort eventually sit their GCSEs. It should be noted that the NRT factor described above similarly 'fixes' the values in academic year ending 2021, implying that residual adjustments will apply to all year groups who were in school during the pandemic, but did not sit their GCSEs in that year.

Residual Adjustments:

Therefore, residual adjustments need to be made to year groups affected during the coronavirus pandemic to balance the sum of annual contributions to the final achieved attainment value for that year group. Residual adjustments are only applied from academic year ending 2021 onwards. For the first cohort affected by the coronavirus pandemic (Year 11 in academic year ending 2021), the Year 11 score needs to be adjusted so that the contributions of Year 7 to Year 11 equal the attainment index value. To prevent arbitrary application of residual adjustments in different cohorts in the same year, the Year 11 residual adjustment in that year is applied to the other year groups. The same process is then followed for the subsequent years taking previous residual adjustments as given. These adjustments will automatically drop out of the system when the last cohort who were in school during the coronavirus pandemic sit their GCSEs.

Cohort Weights:

Having developed this method, in discussion with experts a follow-up question emerged, which the Review has also addressed, namely the value applied to the weights placed on each year in the cohort-split model.

Within the model, primary and secondary schools were adjusted separately, each with their own corresponding weights and these adjustments were performed for all four nations of the United Kingdom. Whilst the seven years of primary education (Reception plus Years 1 to 6) received equal weights of 14.3% each, secondary schools cohort scores diminished with distance from sitting GCSEs, (Table 6), so most weight was given to those years closest tothe achieved scores. Equal weights (25%) were also applied to Further Education (FE) when this was added into the model for age groups 16-19. FE attainment will be factored into PSP estimates for the first time in Spring 2025.

The Review considered the secondary school weighting for revision because the model appeared inconsistent with primary and FE, and because attainment was apportioned back to Reception, when these years had already been captured in the primary attainment. For these reasons the new weights laid out below are proposed for implementation into the cohort-split model. Following discussion with DfE, equal weights (20% for each year) was given to the years 7 to 11 because this approach was felt to provide a reasonable split across the secondary education in absence of robust literature review or additional data.

Table 6: Weights for each school year's contribution to attainment

Secondary Schools (previous)		Secondary Schools (proposed)	
Year Group	**Contribution to attainment (%)**	**Year Group**	**Contribution to attainment (%)**
Reception	2.0%	Reception	
Year 1	2.0%	Year 1	
Year 2	2.0%	Year 2	
Year 3	2.0%	Year 3	
Year 4	2.0%	Year 4	
Year 5	2.0%	Year 5	
Year 6	3.0%	Year 6	
Year 7	5.0%	Year 7	20.0%
Year 8	10.0%	Year 8	20.0%
Year 9	15.0%	Year 9	20.0%
Year 10	25.0%	Year 10	20.0%
Year 11	30.0%	Year 11	20.0%

The review proposes this setup based on the following criteria:

- It allows for consistency in the weighting design between all school phases, in that weights are allocated to year and age groups specific to that sector.
- The Review acknowledges there is a paradigm shift between school phases, which supports the notion that achievements in the previous phase serve as a gateway to the next phase and new knowledge begins accumulating from this point.
- The main outcome of Primary schooling are exams sat at Year 6 (for example, SATs in England and Wales), therefore primary year groups weighted contributions need to build exclusively towards this.

Missing Data:

It must be acknowledged that the coronavirus pandemic caused unprecedented, widespread disruption to education, which has severely limited capacity to accurately measure output. This problem affects different settings in the different devolved governments.

For primary schools:

- England: No data have been published for academic year ending 2020 and 2021.
- Wales: No data have been published since academic year ending 2019.
- Northern Ireland: No data have been published since academic year ending 2019 – The Northern Irish Department of Education have confirmed they are reviewing the design of key stage assessments following the coronavirus pandemic, and that data for key stages 1 to 3 will likely not be available while the Review takes place.
- Scotland: No data have been published for academic year ending 2020.

For further education:

- England Level 2: The index between academic year ending 2020 and 2025 will need to be kept constant due to inconsistent grading practices during this time, and attainment can be factored back into the model from academic year ending 2026.
- England Level 3: The index between academic year ending 2020 and 2023 will need to be kept constant due to inconsistent grading practices during this time, and that attainment can be factored back into the model from academic year ending 2024.

There is no 'perfect' approach to the data gaps issue, and there are caveats regardless of the approach.

Whilst the NRT was sourced from secondary schools and hence could be directly applied, the same issues applied to primary and FE phases, but thorough discussions with DfE and other experts could not reveal NRT-equivalent data for these phases. Despite the difficulties related to applying the NRT outside of secondary school, as the NRT is a secondary school measure and analysis conducted by the ONS indicated that the trends in the NRT do not show synergy with trends observed in primary schools, in the absence of alternative data, using the NRT to indicate attainment trends in other school phases remains the only option.

The Review therefore proposes to include the NRT to fill in data gaps for remaining primary schools across all four nations of the UK. There is also the question of whether FE also needs to be treated in a similar manner. Unlike Primary and Secondary education, FE does not have as much of a focus on academic-orientated subjects and does include more technical-orientated learning.

Attainment for Further Education

The Review has also expanded the quality adjustment to FE, to deliver coverage of all years of compulsory schooling (ages 3 to 19 years), which has been mandatory since 2013. Through consulting with the DfE, the Review was able to identify the National Qualifications Framework (NQF) [see link 105, Annex I], which is a structured and consistent foundation for reporting the different qualifications sat by 16 to 19-year-olds in FE. The NQF also has good sustainability for future research, as the further introduction of new qualifications will be compatible with the existing framework.

The Review identified Level 2 and Level 3 attainment at ages 16 to 25 [see link 106, Annex I] as a data source, this is based on the percentage of students who achieve the minimum grades at each level, with coverage from academic year ending 2004. Level 2 (L2) qualifications refer to GCSE level qualifications or equivalent technical qualifications. In addition, students may re-sit Mathematics and English GCSEs throughout FE, allowing this relevant output of FE to be captured in the estimates. Level 3 (L3) qualifications refer to A level qualifications and vocational qualifications, such as NVQs.

The growth rate in attainment from this source serves as the basis of the ONS' FE attainment measure, using the percentage of students who achieve the minimum required grades for L2 and L3 by age 19 only. This is because:

- Students will have sat their final exams and moved through the FE system by age 19.
- L3 attainment at age 16 is low (it is uncommon for students to complete L3 qualifications by age 16), and L3 attainment at age 17 drops significantly in academic year ending 2016 with the exclusion of the contribution of AS levels to final attainment at L3.

- L2 attainment by age 19 accounts for students who resit their GCSE Mathematics and English examinations throughout FE.

The review proposes to treat FE as a standalone component in the cohort-split model; in other words, FE attainment will be retrospectively apportioned to age groups exclusive to FE (age 16 onwards) and not to previous school phases (Primary and Secondary schools), based on consultation with DfE, Ofqual, and the Institute for Fiscal Studies. The reason for this is because:

- Students often move to a different sixth form or college to their secondary education, and FE has different funding mechanisms and organisational frameworks to secondary education. Therefore, FE is more distinctly isolated from previous school phases, and this makes it difficult to link FE performance to previous years of schooling in a consistent manner.
- Students may not sit academic qualifications at FE, instead they may pursue vocational certificates which do not draw on prior knowledge to the same degree as academic qualifications.
- GCSEs serve as a gateway to entering FE, therefore knowledge at the FE level starts at age 16 and builds from this point.
- Attainment at age 19 is reported by age group, as opposed to academic year groups (e.g. Years 12 and 13) which are reported by Primary and Secondary schools. Consequently, there would be difficulties in integrating age groups in FE to academic year groups in previous school phases.

Upon further discussions with experts, it was also agreed that each age group being accounted for in the cohort model (ages 16, 17, 18, 19) will bear equal weight (25% each). Currently, there is no compelling or robust enough evidence to clearly indicate that one age group has more bearing on achieved attainment than others. Therefore, until such evidence emerges, equal weighting will be applied to these age groups.

As two separate attainment indices (L2 achievements by age 19, and L3 achievements by age 19) will be informing the FE attainment measure, there is a question on how these will be weighted and aggregated into a single index measure for FE attainment following cohort model adjustments for each. The Review explored avenues such as weighting by the relative cost of delivering L2 and L3 qualifications, enrolment figures for each qualification type, or the employability potential of each qualification. Following workshop discussions on the topic with DfE, it was agreed that the enrolment figures for students completing L2 and L3 qualifications by age 19 each year were deemed the most appropriate and empirical metric to determine the weights. It is typical across PSP to weight components based on their cost of delivery, however the total costs for providing education according to qualification type will be directly linked to enrolment figures. The figures vary on an annual basis, but the proportions of L2 and L3 qualifications achieved each year are roughly 60% and 40%, respectively.

Student Well-being

In addition to including attainment in the quality adjustment, the ONS has historically included bullying rates as an additional measure. The Review has identified that this does not fully represent the wider impact of schools on their students beyond academic performance. For this reason, it has been replaced by a more inclusive measure of well-being in schools, which is seen to better reflect different aspects of the development of young people into being capable members of society. Student well-being has also been identified as a more appropriate quality adjusted measure than bullying because of its association with academic achievement and school engagement (Gutman and Vorhaus, 2012) [see link 107, Annex I]. An increasing number of policies focusing on well-being were also implemented in the education service after the coronavirus pandemic, reflecting a growing focus in this area.

The Review identified in the Understanding Society's harmonised UK Household Longitudinal Survey (UKHLS) [see link 105, Annex I] and its predecessor, the British Household Panel Survey (BHPS), as its preferred data sources. The BHPS ran from 1991 to 2009, while the UKHLS commenced in 2009 and continued sampling from households that were included in the BHPS. These surveys represent the entire United Kingdom. The student wellbeing index is created based on weighted responses on the following questions: how do you feel about your 'school?' and how do you feel about your 'schoolwork'?. The possible answers range from 1 (completely happy) to 7 (completely unhappy), with the midpoint 4 representing neither happy nor unhappy. Those who provide positive answers (from 1 to 3) to both the 'school' and schoolwork' questions are flagged as 'happy' in each year from 2003 onwards to derive a wellbeing index.

Because well-being is only one component of the total quality adjustment for Education, the ONS need to derive an appropriate weight for well-being compared with the other components. It is common across service areas to weight components based on their delivery cost. However, in the absence of an alternative metric, the Review proposes using this approach.

The challenge is that it is not straightforward to identify expenditure that is solely attributable to the delivery of activities directed at well-being in education. For example, expenditure allocated to breakfast clubs has a link to well-being, but delivery of these activities also correlates with improved performance and attainment. However, the National Funding Formula (NFF) was identified as the most relevant data source to provide cost-associated weights for well-being.

The NFF is the process by which the DfE decides how much core funding to allocate to mainstream state-funded schools in England. Specifically, the ONS is using planned expenditure share (%) allocated towards funding pupil deprivation to proxy for expenditure targeted at pupil well-being. The association between ameliorating pupil deprivation and improving satisfaction and emotional experiences for students implied that expenditure allocated to this space was an appropriate means to weight the well-being measure, in the absence of explicit expenditure targeted at "well-being programmes". The NFF was introduced in academic year ending 2019, so there are only observable weights from this period.

For the years 2003 to 2017 (the well-being measure commences in 2003), the ONS has held the weights constant with the earliest known annual weight. Due to weights being provided on an academic year basis, the ONS is required to spline the expenditure shares to generate figures in calendar year terms (consistent with periodicity of the annual statistics). The calculated weights assign roughly 9% to well-being; it must be acknowledged that well-being serves as a precursor to the primary outcome of education, which is attainment, and there are factors beyond the control of schools which can impact a student's response to the survey. Therefore, the well-being measure cannot be assigned too much weight, and it was agreed between the ONS and DfE that the weights provided by the NFF for well-being are appropriate and fit for purpose.

ANNEX E:

Defence outputs research

This Annex presents a summary of research work to develop a suitable output estimate for Defence services.

The reason this topic is so challenging is because output should reflect the activities being 'undertaken', but this definition is one where great caution is required. The Ministry of Defence (MOD) [see link 108, Annex I] defines its purpose as being "to protect the nation and help it prosper" with priorities to "... protect the UK, its Crown Dependencies, and its Overseas Territories; pursue a campaigning approach to counter the threats from state and non-state actors; promote our national interests globally; and secure strategic advantage, achieve greater economic and industrial resilience, and contribute to national prosperity."

Given this definition of MOD's objectives, deterrence is clearly fundamental: one of its objectives is for UK military forces to present an image of strength and readiness to react in order to deter any attack and 'protect the nation'. In a perfect world, deterrence therefore implies that active deployment of the armed forces is to be avoided if possible. Clearly, a scenario where there was no use of Defence services would seem to imply no output, but this would be incorrect. For example, defence capital assets, such as aircraft carriers and ships, are designed both to perform their function in rare combat situations, and to act as a deterrent to prevent conflicts from arising or escalating. However, estimating outcomes such as 'wars avoided' is highly speculative, as is measuring threats such as 'potential military attacks'.

Deriving a suitable 'unit' of deterrence is challenging and multiple attempts have been made over time. Many rely on estimating the capability of the military to deliver against its stated objectives but even this has proved challenging. This is because the levels and types of capability of the armed forces vary over time, dependent on the type of threat, the arena in which deployment may be necessary, technological advances (e.g. the emergence of drones as a fighting platform), and military strategy. The 2024 House of Commons Defence Committee report "Ready for War" [see link 109, Annex I] provides additional context on this point.

Nevertheless, active deployment clearly is an output which needs to be considered and taken into account. The Review therefore starts with a basic formula to shape its conceptual work:

Figure 8

$$Output = \alpha Active\ Deployment + \beta Deterrence$$

The Review has concluded that producing direct outputs measures for Defence is a three-part problem:

- developing activity measures relating to a directly observed measure of active deployment
- considering models for estimating deterrence, where this can be characterised as the deterrence value delivered by non-active deployment
- the calculation of a weighting model to combine the two into an overall measure of output

In addition, the Atkinson Principles states to only consider those matters under the government's control to determine productivity is pertinent. As an increase in healthiness to the health service would not be counted if it was caused by individuals independently deciding to take up going to the gym, should the actions of potential hostile state and non-state actors be a determination of capability? If a hostile foreign state doubled spending on military hardware, including long distance missiles, would the capability, assuming the type and level of spending stayed the same, halve? Would it offer less of a deterrent? Or is the deterrence determined by the stock of assets held by that foreign actor, which the long-range missiles could destroy? Again, growth in foreign National Balance Sheets or Gross Domestic Product is not under the control of the UK Government and hence should it be in scope?

Compounding the theoretical challenges, there are issues of pragmatism too: there is a considerable degree of confidentiality with regard to the activities and capabilities of the military for reasons of national security, thus rightly limiting data availability. Therefore, even where methods may be feasible in theory they may not be in practice. Furthermore, there is a question of ethics when considering the frameworks for Defence output.

If the UK's preference is towards deterrence, then this implies that it has a greater weight than active deployment. However, this would also imply that, in the event of deterrence failing, a shift to active deployment would cause measured output, and hence productivity, to fall. Conversely, if active deployment is weighted more highly and the same inputs swap from being used for active deployment to being a deterrent at the end of a conflict, output and productivity would again fall.

It is clear this is not a simple argument. If the UK was a small country whose defence objectives were only domestically focussed, it would be easier to make a clear distinction between the two.

However, as an advanced military power, there are three further complexities to consider:

1. Defence activity often takes place through multi-national groupings, for example developing joint new capabilities such as through the AUKUS (a trilateral security partnership between Australia, the United Kingdom, and the United States) submarine programme, sharing intelligence such as through the Five Eyes alliance (an Anglosphere intelligence alliance comprising Australia, Canada, New Zealand, the United Kingdom, and the United States), and projecting power overseas in line with international law, for example as peacekeepers, all of which suggest an additional objective related to wider foreign policy as an outcome for Defence services.
2. Membership of alliances such as The North Atlantic Treaty Organization (NATO) augments UK military power through agreements around collective defence and mutual aid, such as Article 5 of the North Atlantic Treaty. This means that the level of UK 'homeland defence' is not simply a factor of UK defence capability.
3. Active deployment can serve as evidence of deterrence capability, or enhancement of the deterrence value of military forces through participation in training exercises. In other words, active deployment which makes the strength of UK and allied forces visible may act to increase the deterrence value of these forces.

This third point opens the possibility of double counting by trying to assess these two measures independently of each other.

E.1 Output in the form of active deployment

The measurement of active deployment output can be approached through two methods: using a single 'macro' indicator which acts as a proxy for the whole set of activities, or compiling a 'micro' driven approach, whereby the Review aggregates different activities which Defence services may deliver under active deployment. An example could be the person-days on active deployment. Clearly the exact composition of these would be key to successful implementation, but this is the model used in other services and the Review considers the 'micro' approach as best practice to be applied if a comprehensive set of measurable outputs can be achieved.

However, as with other areas, simply measuring the number of person-days on active deployment is clearly flawed because this quantity measure fails to reflect the dimension of quality. Defence is an area where technology, strategy, and implementation methodologies all matter and change quickly. For example, returning to the point about the value of alliances, UK operational capability has often been enhanced by a fundamental alignment of approach with the USA:

> "...the 1991 Gulf War confirmed to British officers the importance of maintaining interoperability with developing US warfighting concepts, especially given the relative prominence the Americans afforded the British Army compared with other less modernized allied contingents."

Source: Chapter 16 (Tactics and Trade-Offs: The Evolution of Manouvre in the British Army) in this book: Advanced Land Warfare: Tactics and Operations – Mikael Weissmann (ed.) and Niklas Nilsson (ed) (April, 2023) [see link 110, Annex I].

As such, some method to 'quality adjust' any directly observed output metrics relating to active deployment is clearly required, particularly as in many cases these outputs may appear little different from input measures of labour.

The Review looked to enhance the existing 'inputs = output' approach, through allowing the addition of capital to augment labour inputs. If it can be assumed that labour does not show quality improvements, and therefore does not change through time – that is a soldier in 2008 putting aside the technology available to them, is of equal capability as a soldier in 1998 or 2018 then capital investment can be treated as a proxy for how much more output each unit of labour delivers. Whilst basic and not particularly elegant, this model would deliver a starting point to impute outputs distinct from the valuation of inputs. However, despite the concept appearing to work in theory, it was not possible to practically implement.

This is because the Review explored the potential to use evidence from significant capital projects and the investment cases which are produced to deliver these. In many areas of government these contain estimates of the total costs and benefits (both discounted to present day terms), and the ratio of benefits to costs, which could be used as a proxy for the degree to which these investments enhance labour inputs. However, due to the difficulty of defining outputs, benefits cannot be monetarised, which makes deriving the ratio of benefits to costs very difficult. Additionally, there were security concerns about including data on active deployments.

However, work at MOD may shortly lead to publication of statistical data on a measure of readiness. The definition of Readiness, which is broadly aligned with NATO's recommendations, is based on:

- Availability: the number of platforms (e.g. tanks, frigates, drones etc) within a particular time frame.
- Capability: what the platforms can do.
- Sustainability: how long these activities can be sustained for.

As work to measure readiness is in progress, there may not be scope to include data for the entire back series from 1997. This is due to the difficulties in retroactively applying qualitative methodology to historic shifts in policy focus. For example, it may not be possible to determine 'Readiness' of the armed forces in 2003, due to the lack of availability of the databases which are used to inform the current statistic in development. As another potential limitation, due to the nature of the data that inform the Readiness metric (from MOD strategic risk assessments), the Office for National Statistics (ONS) may not be able to access and scrutinise the data behind the numbers; only aggregated Readiness numbers may be available.

Nevertheless, if MOD began to routinely publish outputs based on Readiness at an aggregated level, then Inputs multiplied by Readiness could give a conceptual framework for looking at outputs. The Review has recommended the ONS and MOD should continue to work to explore the potential of these data.

Readiness therefore provides an estimate of the military preparedness to respond to emerging operations; that is, to deploy onto active service. If an aircraft carrier is immobilised, that is clearly a case of output not being delivered for the input resources used. In this case the deterrence attributable to and actionable by the military is also weakened as potentially hostile actors can factor in such limitations in the UK's defence capability into their decision making. As such, the concept of readiness has the potential to also proxy for the deterrence value of Defence services: the readiness of military forces to react and engage is clearly a feasible definition of deterrence.

Therefore, whilst the Review has considered deterrence models, Chapter 9 highlights the need for future work in this area.

E.2 Output in the form of deterrence activities

What is 'deterrence' and how should it be best thought of? Deterrence is the deployment of the threat of force to deter other nations or actors from considering that it would be beneficial to use force against the UK. How best to measure this? The easiest way to consider this is that the UK has means of estimating its 'value', and that an attack would diminish this value through destruction of people and places. The RAND Study (2021) [see link 48, Annex I] serves as a basis for exploring conceptual models, but noted it would be difficult to gauge whether the public feel more "protected" for having a Defence service,

in the absence of any imminent international threat. This ties in with wider arguments that there are challenges in determining society's optimal Defence output (see Understanding the contribution of defence to UK prosperity [Link 48, Annex I]).

The Review assumes that, recognised by the public or not, the value of the demand for deterrence can be considered as the reduction in the probability of an attack multiplied by the value of the assets being protected multiplied by the proportion of these assets which could be destroyed in any such incident:

Figure 9

$$Demand\ for\ UK\ Deterrence = Value\ of\ UK\ Assets \times Economic\ cost\ of\ deterred\ attack \times probability\ of\ incident\ occurring$$

Mathematically, demand for deterrence would therefore increase if:

a. the value of the UK's assets increased

b. the military capability of potentially hostile actors increased (that is if the destructive capacity of their weapons increases the proportion of UK assets which could be destroyed in an incident)

c. the likelihood of potentially hostile actors of any particular size (represented by spend) attacking the UK increased (that is the probability of an incident occurring increases)

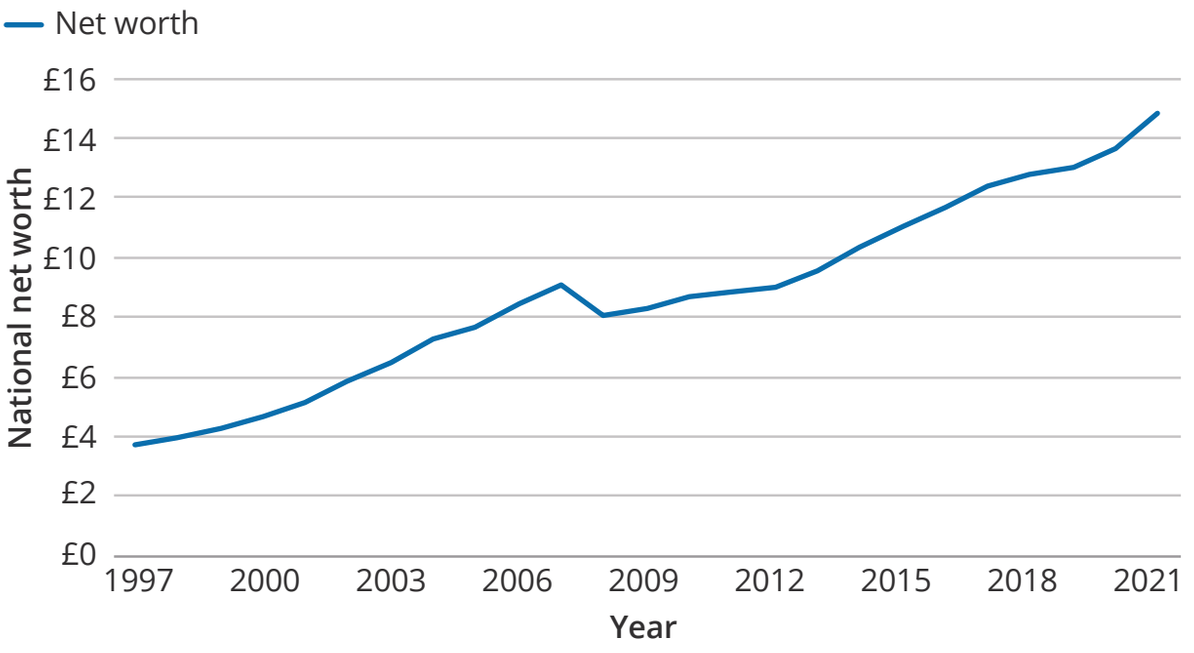
In this model the Review assumes that UK defence capability is akin to an insurance policy, where for output to be held constant, the inputs (over time) are calibrated to the relative value of UK deterrence desired.

Taking a particular time as a reference point, if UK defence capability (proxied by spending in the first instance) increases by more than the value of UK deterrence then one could argue output has increased (that the UK has purchased more insurance relative to the value of the risk), and conversely if the value of UK deterrence demanded increases by more than UK defence capability, one could argue output has fallen (that the UK has purchased less insurance relative to the risk). Therefore, if one can find data sources for the formula components, one can attempt to estimate Defence deterrence output.

Components in Demand for UK Deterrence

The value of UK assets: The UK National Accounts records the Net Worth of the UK. Whilst this estimate of the value of the capital assets of the UK omits human and natural capital, it presents a starting value which can be used. These data are readily available in monetary terms.

Figure 10: UK National Net Worth, 1997 to 2021



Economic cost of a deterred attack: Clearly an attack would not destroy all economic value in the UK. Assuming that all attacks are not equal, the measure needs to adjust this factor over time. Obviously, the Review cannot expect to foresee the precise scale of any attack and hence an expected or average scale of attack in that year needs to be derived which allows a value to be computed. However, given the fact there have been no substantive incidents of physical damage in a prolonged period, there is a limited degree of sophistication possible. For simplicity, the Review proposes this factor would need to be held constant and will be assumption-based, on the basis of the capability of weapon systems available to potentially hostile actors, but clearly additional evidence would be beneficial in this area.

Additionally, the very concept of economic disruption has evolved, with cyber-attacks providing an alternative type of attack which should be factored in.

The probability of incident occurring: Four methods for considering this component have been considered:

- military knowledge
- national threat levels
- spend by potentially hostile state actors
- the Global Risk Index

Military knowledge: One could assume that there is military knowledge and intelligence which could provide relevant information, but this is unlikely to be accessible for security reasons. For this reason this source was discounted as a viable option.

National Threat Levels: The National Threat levels are monitored and updated by the Joint Terrorism Analysis Centre, which involves the police and security services. This source was discounted as a viable option for the following reasons.

- These could be conceptually too far removed from the international military threat, covering terrorism (including domestic terrorism) etc.
- Although the threat levels are regularly updated, they do not change much over time; for example, the threat level staying at "severe" for several years might be too static for comparing the productivity of the military over time.
- There are standard challenges of converting cardinal scales such as the Threat Level into ordinal values; particularly the assumptions of linearity and homogeneity. Linearity relates to the assumption the movement from one category to another is the same all the way up the scale (the movement from 1 to 2 is the same as the movement from 6 to 7 etc). In this case, is the increase in threat to move from 'severe' to 'critical' the same as to move from 'low' to 'moderate'? Homogeneity relates to whether a threat level recorded as 'severe' in one period would be consistently recorded as 'severe' in all other time periods. Assuming this system is being consistently applied and through a robust structure these assumptions should hold, but even so, how to reflect this information?

Spend by potentially hostile state actors: Initial discussions with the MOD in December 2023 highlighted a number of practical challenges which suggested this approach would not be appropriate. There are difficulties in acquiring data for the military capabilities for aggressive actors. Firstly, these are hard to quantify, and secondly it would not be appropriate to share data on this, due to national security concerns. Spend is available, for example via the World Bank, but it would seem inappropriate to assume that inputs work as a strong proxy for other countries if the calculations are deliberately assuming that the value of the output of UK defence differs from the value of its inputs.In short, capability does not necessarily equal spend and therefore spend is not a good metric.

Further considerations

However, assuming different countries are gaining capability at a broadly similar rate, relative spend may give sufficient signal about deterrence to meet the review's need. That is, if potential aggressors quadruple their military spending, the current spend of the UK, all other things being equal, would become less of a deterrent, and would need to increase by a similar factor to deliver the same amount of deterrence. There are, however, three key issues to consider:

- What to do with allies? The deterrence from attacking the UK is not just a factor of the UK's military capability. The UK is a member of a number of military alliances, particularly NATO which mean that an attack on one is perceived as an attack on all. How does the Review take account of the spend of allies? How does the Review take account of UK's spend on alliances that also benefit others? The analysis does not include allied spending at the present time, but clearly this would be a key next step.

- Who should be included within potential threats? What makes a country or actor a threat to the UK? Is this just a question of doctrine or policy? Or is it also a question of capability? Striking a balance of these two suggests that the Review is looking at a subset of those nations who have 'expeditionary capability' – that is the ability to project power beyond their boundaries or immediate neighbours – who have doctrinal or other clashes with the UK or western norms. Whilst this may provide a baseline, if the countries which meet these criteria were to change, this may have a material 'step-change' impact on the output measure, assuming no change in UK defence capability. Whether this would feel a fair reflection of Defence services is a valid question.
- Is relative spend a sufficient proxy? Does the Review consider that the different actors will experience similar paces of technological change? If it is viewed that potential aggressors may have slower rates of technological advancement or different price deflators then relative spending over time may become more or less effective relative to UK spending.

Having interrogated World Bank data, the Review found the results to deliver extremely volatile and erratic trends, strongly suggestive that different countries had experienced very different inflation or technology paths to the UK, or there may be some mis-reporting of spend data. Results were so extreme the Review was unable to recommend their use at this time for this purpose.

The Global Risk Index

An alternative approach is that based on the Caldara and Iacoviello (American Economic Review, 2022) [see link 111, Annex I]) Geopolitical Risk (GPR) Index, which is a measure of adverse geopolitical events and associated risks. The index is based on automated text-search results of the number of articles related to adverse geopolitical events using the electronic archives of ten US and UK newspapers. The GPR Index spikes around the two world wars, the Korean War, the Cuban missile crisis, and following the 9/11 attacks in the US. A higher geopolitical risk is associated with the higher probability of economic downturns and risks to the global economy.

The historical GPR Index starts in 1900, and the recent index starts in 1985, and country-specific indexes are constructed for 44 different countries. The text-search is organised in eight categories: war threats (1), peace threats (2), military build-ups (3), nuclear threats (4), terror threats (5), beginning of war (6), escalation of war (7), and terror acts (8). There are two sub-indexes, geopolitical threats based on categories (1) to (5), and geopolitical acts based on categories (6) to (8).

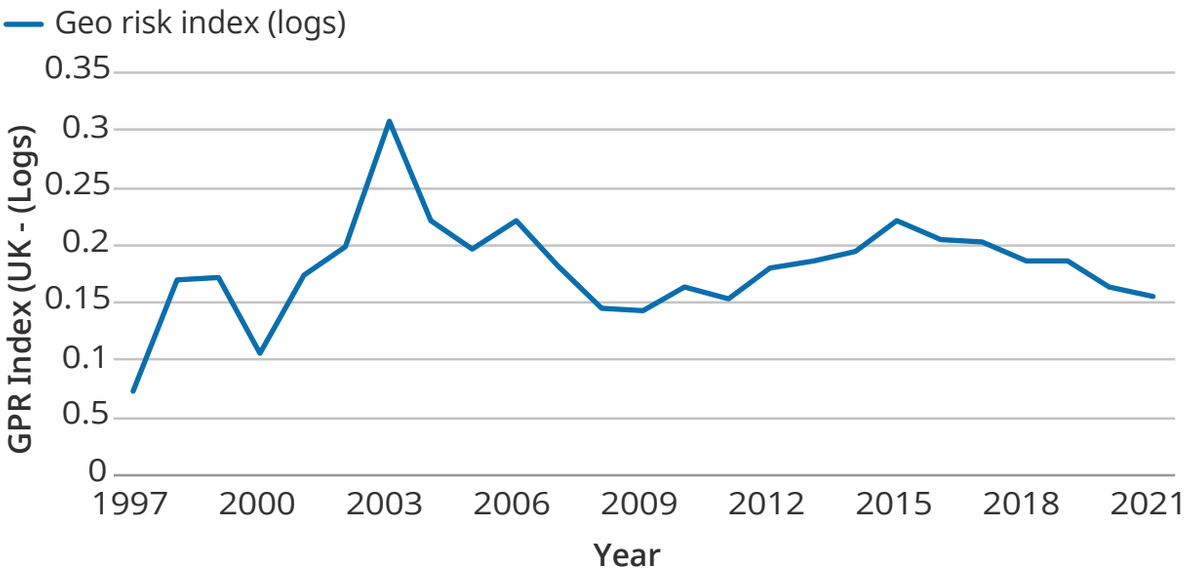
This dataset is produced on a standardised basis for a number of countries, including the UK, and provides the strongest dataset the Review has been able to source.

Deterrence as an insurance model: This approach takes forward successful aspects from previous methods and works through an insurance supply and demand model for UK Defence output.

In this case, the Review treats domestic deterrence capability as a function of defence spending, given the size of protected assets and probability of attack. The UK's insurance demand is the value of assets to be insured given the estimated risk to those assets. Domestic assets are given by national accounts estimates of UK non-financial assets (ONS, [see link 112, Annex I]); whilst the risk to those assets is given by the UK's estimated geo-political risk index, as described in Caldara and Iacoviello, 2022 [see link 113, Annex I]. Insurance supply ('input') is simply the UK's total defence spending, supplied by the World Bank [see link 114, Annex I].

From this point, an insurance 'output' is then derived which is the ratio of insurance supply to insurance demand, and 'productivity', which divides the output by the 'input' (insurance supply).

Figure 11: The UK Geopolitical Risk Index (1997 to 2021)

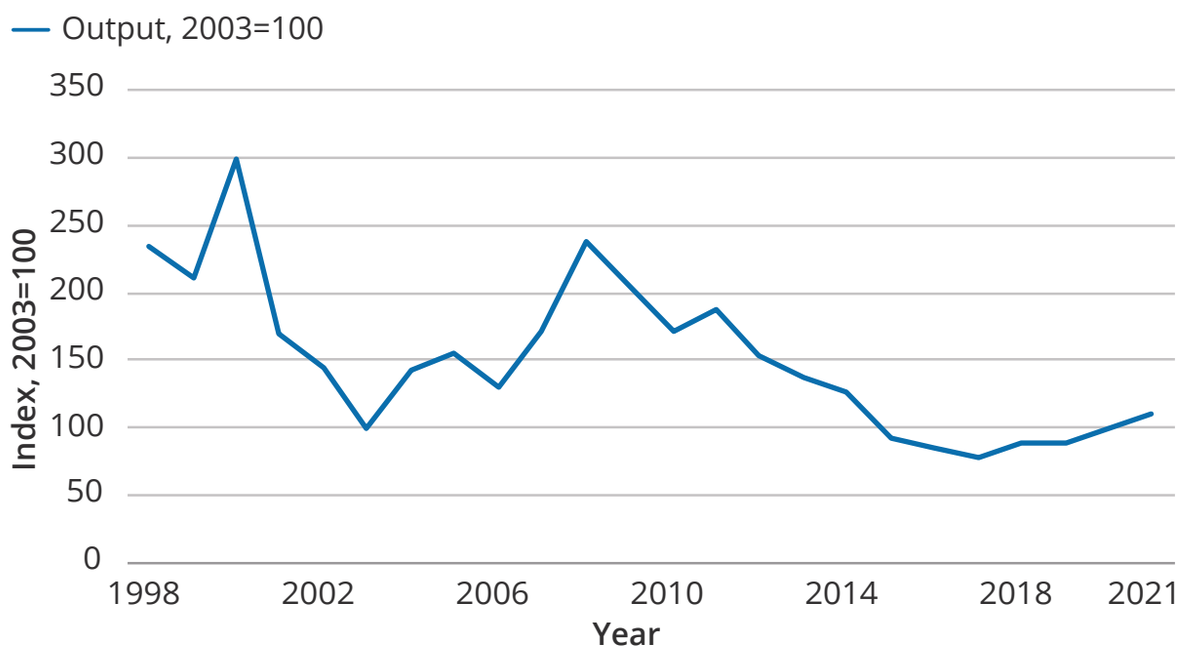


Reviewing this dataset shows distinct break-points. The 1990s are characterised by values which reflect the evolving resolution of the Cold War, during which large military forces were still being retained and 'peace dividends' had not yet been cashed. The 9/11 attacks mark a clear breakpoint in the series. In the absence of further research, the Review anticipates that this may form such a substantive break in the series that it may not be possible to take comparable analyses back before 2002.

Nevertheless, it appears that using this data source does deliver a broadly believable path for Defence output. The increase in deterrence capability up to 2009 is driven primarily by falling demand for insurance and the insurance supply becoming relatively more proficient. From 2009 (to 2015), the series is falling as the risks exposure increases but spending remains fairly constant.

The UK therefore became less-well insured. It should be noted, however, that the range of the movements appears unfeasible: did military output really increase by 60 index points between 2003 and 2008, and then fall by an equivalent amount from 2008 to 2015?

Figure 12: UK Defence output, utilising the Geopolitical Risk Index (1998 to 2021)



E.3 Conclusions

It is noted that in this initial estimation used, the UK defence spend was used as a proxy for UK defence capability. Again, the new data on readiness may provide a means to make the necessary adjustments and provide a single, unified measure of Defence output, recognising that measuring both active deployment and deterrence may result in double-counting, and presents challenges to weight these two outputs together. The latest method developed, which combines the two aspects (including formula mentioned in figure 13, 14 and 15 into a single metric is as follows:

Figure 13

$$Defence\ Capability = Defence\ Spending \times Readiness$$

Figure 14

$$Geopolitical\ Risk\ Index = Economic\ cost\ of\ a\ deterred\ attack \times Probability\ of\ incident\ occurring$$

Figure 15

$$Output_t / Output_0 = {Defence\ Capability_t / Demand\ for\ UK\ deterrence_t} \Big/ {Defence\ Capability_0 / Demand\ for\ UK\ deterrence_0}$$

Where year 0 = 2002

Beyond issues of sourcing data and understanding the volatility of results, there remain with this model some conceptual issues: for example, in a scenario where Defence output increases – that is defence capability increases relative to the value of deterrence – could this represent 'over-insurance'? Is buying 'too much' insurance a positive output? The UK National Accounts principles – that all goods and services produced are output – would suggest that the answer is yes, but the opportunity cost of this spending may suggest a need to quality adjust.

Secondly and similarly, if the level of Defence outputs fall – i.e. defence capability falls relative to the value of deterrence – and no attack takes place, has the UK produced the desired outcome at a lower cost? Is this a good thing? Does this suggest the Review should 'invert' the series so a fall in the ratio is perceived as an improvement? This is rather like the question of whether you can save money by not buying house insurance: most of the time you do, but there are occasions you lose very substantially. The struggle is this model has no place for an actualised attack. Looking back to the run-up to previous conflicts – for example the example of investment in the development of Spitfire and Hurricane aircraft prior to the commencement of World War II suggests that the calculation would not wish to disincentivise such activity by making more spending appear 'worse' in output terms prior to an actualised attack, if one was ever to occur. These questions suggest the need for continued methods debate with academics and MOD experts to reach an informed conclusion.

ANNEX F:

Direct and Indirect Labour Input Measurement

The overriding methodological justification for the improvements to the labour measurement is to account for changes in the quality of labour, as these excerpts from the OECD Manual on Measuring Productivity [see link 115, Annex I] and the Atkinson Review (2005) [see link 1, Annex I] explain.

"Labour input reflects the time, effort and skills of the workforce. While data on hours worked capture the time dimension, they do not reflect the skill dimension. When total hours worked are the simple sum of all hours of all workers, no account is taken of the heterogeneity of labour.

For the estimation of productivity changes, the question is whether, over time, the composition of the labour force changes, i.e. whether there is an increase or decrease in the average quality of labour input."

OECD Manual, paragraphs 85 and 86

"Principle F recommends that labour inputs be measured using both direct (number of hours worked, with different skill categories being weighted) and indirect (deflation of pay by a labour cost index) methods."

Atkinson Review, paragraph 5.62

F.1 Indirect approach

The indirect approach simply deflates current price (CP) expenditure by a relevant price deflator:

Figure 16

$$Indirect\ Labour\ Volume\ growth_y = Total\ Current\ Price\ Labour\ Expenditure_y\ x\ Implied\ inputs\ deflator_y$$

where CP labour expenditure, classified by European Systems of Accounts 2010 [see link 116, Annex I] code D1, is the total compensation of employees and includes remuneration in excess of gross pay.

F.2 Direct approach

The direct approach uses data on full-time equivalent employees (FTEs) and relative salaries for employment ranks:

Figure 17

$$Direct\ Labour\ Volume\ growth\ rank\ i_y = \sum_{i=1}^{n} \frac{FTE_{iy}}{FTE_{iy-1}} \times \frac{Gross\ Mean\ Pay_{iy-1} \times FTE_{iy-1}}{Total\ Implied\ Expenditure_{all\ ranks_{y-1}}}$$

The direct approach equation estimates the labour growth contribution of staff in employment at each rank or grade. To get a total direct labour estimate the following process is followed:

- sum the growth contributions of all relevant rank or grade; i to n
- each rank is weighted within the calculation by their own implied expenditure relative to the total implied expenditure (for staff of all ranks)

This methodology aligns with that which is proposed, or already in place, for other areas within public service productivity.

While direct labour measurement is ideally based on the number of hours worked for all employees, the lack of data means that it is rarely possible, and the next best metric to use is the number of FTE staff. To account for the skill-level of different groups of labour input, FTE data are typically weighted by the average salary data of these constituent groups.

ANNEX G:

Social Security Administration

G.1 Calculation of Universal Credit output given composition complexities

As Universal Credit is calculated on an additive-component basis where additional payments are made for additional entitlements, the Review can calculate the per household composition adjustment factor using a Laspeyres index:

Figure 18

$$L_t = L_{t-1} \times \frac{\sum_i p_{t,i} \times c_{t-1,i}}{\sum_i p_{t-1,i} \times c_{t-1,i}}$$

Where:

I = index value of per household compositional adjustment

p = proportion of UC claims including component i

c = average contribution of component to overall UC benefit payment

t = year

i = component of UC

This approach uses a limited composition adjustment factor based on the first five components of the Department for Work and Pensions (DWP) Universal Credit formula that is used for calculating payments. These components are as follows (with the categories the Review used in brackets):

- standard allowance (Single or Couple)
- entitlement for children (Have 0 children, 1 child, 2+ children)
- entitlement for childcare costs if working (claim childcare for 0 children,1 child, 2+)
- entitlement for disability (capacity for work or Limited capacity for work)
- entitlement for disability (Carer or not a carer)
- housing element (housing element or no housing element)

The Review has utilised Stat-Xplore (the online DWP data portal for benefit statistics [see link 117, Annex I]) for entitlement data, alongside experimental, unpublished data for the number of children and childcare elements. However, due to data limitations, this approach can only be used to measure Social Security Administration output growth for the period 2016 onwards.

G.2 Fraud and error quality adjustments

The quality adjustment derived from DWP's fraud and error rates has been constructed as follows:

- a total fraud and error rates series is constructed back to 2008 by adding splined rates of overpayments and underpayment according to DWP's Fraud and error in the benefit system annual report [see link 118, Annex I]
- because these rates are small (at around 3-5%), the Review inverted the total fraud and error rates by taking it away from a value of 100 to essentially give a 'correctness index'. This avoids the construction of a volatile index based on small rates.
- the growth rate of this correctness index is the Social Security Administration (SSA) quality adjustment, which is then added to the growth rate of the non-quality adjusted SSA output index

Error rates are calculated in Figure 19 as follows:

Figure 19

$$ErrorRate_t\ (\%) = \frac{ErrorValue_t\ (£)}{Expenditure_t\ (£)}$$

Therefore, after swapping error for correctness, the combined DWP and HM Revenue and Customs (HMRC) correctness rate would be as per Figure 20:

Figure 20

$$TotalCorrectnessRate_t\ (\%) = \frac{\sum_{DWP}^{HMRC} CorrectnessRate_t \bullet Expenditure_t}{\sum_{DWP}^{HMRC} Expenditure_t}$$

The proposed method for producing a "correctness" index for the combined correctness rate DWP payments is to base the series at 100 in 2007 and calculate a based index. This can be shown in Figure 21.

Index=n=TotalCorrectnessRatet=nTotalCorrectnessRatet=0×100Index=n=Total CorrectnessRatet=nTotalCorrectnessRatet=0×100

This is valid until 2018, when DWP made a number of changes to their methodology (see Fraud and Error in the Benefit System: Background information and methodology [Link 119, Annex I]). Because DWP have published statistics on both methods for this year, the growth rate can becalculated for 2018 to 2019 based on correctness rates from their new method. The index then continues from here as shown in Figure 21 and Figure 22.

Figure 21

$$Index_{t=n} = \frac{TotalCorrectnessRate_{t=n}}{TotalCorrectnessRate_{t=0}} \times 100$$

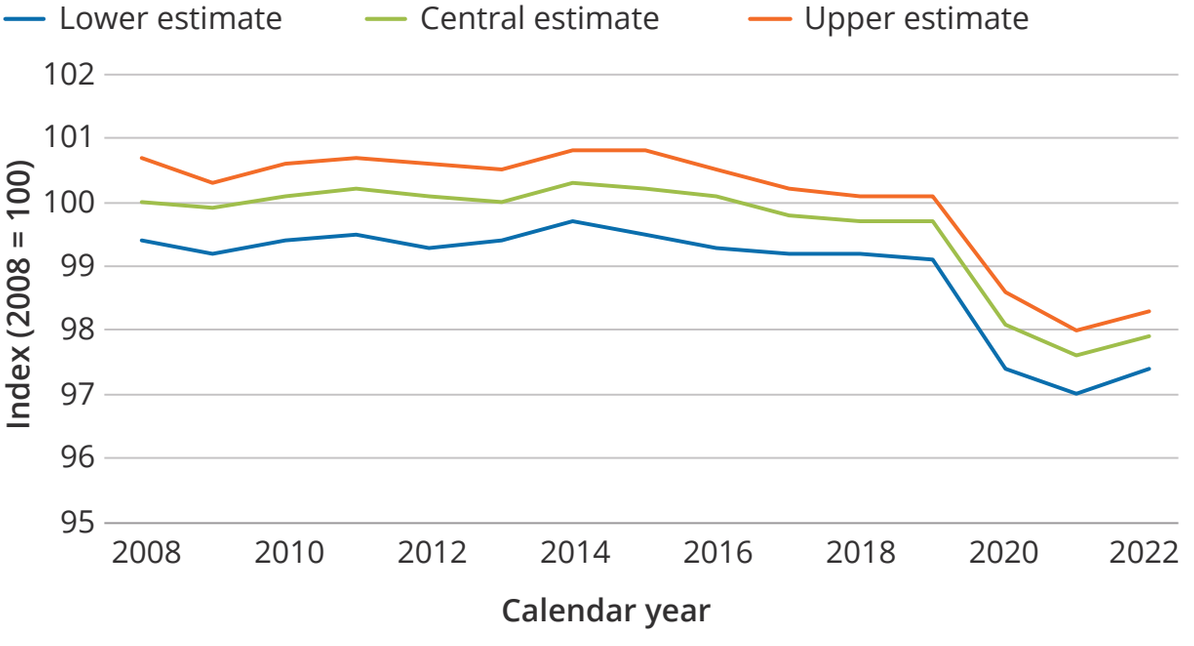
Figure 22

$$Index_{t>2018} = \frac{TotalCorrectnessRate_{t}}{TotalCorrectnessRate_{t-1}} \times Index_{t-1}$$

Total fraud and error rates were relatively stable between 2008 to 2019 at around 3%, resulting in small quality adjustments between plus or minus 0.3 percentage points throughout this period. However, total fraud and error rates increased to around 5% from 2020, and therefore adjustments are larger in these years. The trend in the chart is partially due to complex cases being transferred to Universal Credit over time, which contribute to higher error rates within the system and may also reflects the introduction of "Trust and Protect" easements in response to the coronavirus pandemic (See DWP Annual Reports and Accounts 2020/21,[Link 120, Annex I]). Nevertheless, the adjustment results in only small changes to both SSA's quality adjusted output and productivity indices.

Figure 23: Fraud and error quality adjustment indices (DWP only), central estimate, upper and lower confidence intervals (relative bases), 2008 to 2022

'Correctness' rate of administered benefits quality adjustment indices (based on DWP fraud and error rates), central estimate, upper and lower confidence intervals (relative bases), 2008 to 2022



ANNEX H:

Proposed Environmental Services COFOG classification for Natural Resource Management, and Climate Change

Table 7 outlines the existing COFOG categories which are proposed for inclusion into Natural Resource Management (referred to as E.2). This covers activities related to the use and provision of farming, forestry, fisheries, food, animals and plants.

Table 7: Proposed COFOG for Natural Resource Management (referred to as E.2)

E.2: Natural Resource Management	Share of lines of spend within this category to be brought into E.2 (% of category likely aligned to E.2 activities)	Share of Total Spend to be brought into E.2 (% of category likely aligned to E.2 activities)
1.6.0 General public services n.e.c.	11.6%	0.0%
4.2.1 Market support under CAP	97.7%	69.4%
4.2.2 Other agriculture, food and fisheries policy	11.2%	10.4%
4.2.3 Forestry	5.9%	-4.9%*
4.3.0 Fuel and energy	1.1%	0.0%
4.5.3 Local public transport	21.1%	5.7%
4.5.4 Railway	3.2%	2.3%
6.5.0 R&D housing and community amenities	60.0%	60.4%
8.2.0 Cultural services	10.6%	0.9%

*Negative values reflect earned income exceeding costs)

Table 8 outlines the existing COFOG categories which are proposed for inclusion into Climate Change (referred to as E.3). This covers activities relating to climate change including adaptation, mitigation, net zero and impacts.

Table 8: Proposed COFOG for Natural Resource Management (referred to as E.3)

E.3 Climate Change	Share of lines of spend within this category to be brought into E.2 (% of category likely aligned to climate change activities)	Share of Total Spend to be brought into E.2 (% of category likely aligned to climate change activities)
4.2.2 Other agriculture, food and fisheries policy	2.4%	1.9%
4.3.0 Fuel and energy	24.7%	75.1%
4.5.3 Local public transport	21.1%	5.7%
4.5.5 Other transport	4.4%	24.8%
4.8.0 R&D economic affairs	1.3%	11.9%
5.3.0 Pollution abatement	27.3%	13.3%
5.5.0 R&D environment protection	1.4%	0.6%
5.6.0 Environment protection n.e.c.	29.2%	50.0%
6.6.0 Housing and community amenities n.e.c.	6.9%	0.7%
9.2.0 Secondary education	0.6%	0.1%

The COFOGs captured in all three proposed categories (E.1, E.2 and E.3) are not currently captured by other productivity service areas, and (apart from COFOG 9.2.0) are currently included in the 'Other' grouping. There would therefore be no impact to the other services. This could also be a potential route to improving coverage of productivity calculations across a wider proportion of government spend.

ANNEX I:

References

Link 1: The Atkinson Review: Final Report: Measurement of Government Output and Productivity for the National Accounts
www.semanticscholar.org/paper/The-Atkinson-Review%3A-Final-Report%3A-Measurement-of-Atkinson/aa515465288bbe305ccf2c79b18450344626e78e

Link 2: Independent review of UK economic statistics: final report
www.gov.uk/government/publications/independent-review-of-uk-economic-statistics-final-report

Link 3: Updating the System of National Accounts (SNA)
unstats.un.org/unsd/nationalaccount/towards2025.asp

Link 4: Manual on sources and methods for the compilation of COFOG statistics
ec.europa.eu/eurostat/documents/3859598/10142242/KS-GQ-19-010-EN-N.pdf/ed64a194-81db-112b-074b-b7a9eb946c32?t=1569418084000

Link 5: Public service productivity, UK Articles
www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/articles/publicserviceproductivityuk/previousreleases

Link 6: Public service productivity, quarterly, UK Statistical bulletins
www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/bulletins/publicserviceproductivityquarterlyuk/previousreleases

Link 7: Developing nowcast methodologies for public service productivity, UK
www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/articles/developingnowcastmethodologiesforpublicserviceproductivityuk/2024-12-11

Link 8: Public service productivity, quarterly, UK: July to September 2024
www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/bulletins/publicserviceproductivityquarterlyuk/julytoseptember2024

Link 9: National Statistician's Independent Review of the Measurement of Public Services Productivity
uksa.statisticsauthority.gov.uk/publication/national-statisticians-independent-review-of-the-measurement-of-public-services-productivity/

Link 10: Public service productivity: total, UK, 2021
www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/articles/publicservicesproductivityestimatestotalpublicservices/totaluk2021

Link 11: System of National Accounts 2008
unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf

Link 12: Review of approaches to the measurement of public sector productivity statistics (ESCoE TR-26)
www.escoe.ac.uk/publications/review-of-approaches-to-the-measurement-of-public-sector-productivity-statistics/

Link 13: International comparisons of the measurement of non-market output during the coronavirus (COVID-19) pandemic
www.ons.gov.uk/economy/economicoutputandproductivity/output/methodologies/

Link 14: New medical treatments, quality improvements and public health measures in the National Accounts: The effects of the Diabetes Prevention Programme (ESCoE DP 2024-12)
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/

Link 15: A Satellite Account for Health in the United States
www.aeaweb.org/articles?id=10.1257/aer.20201480

Link 16: Composite indices: Reflections on recent UK experience and lessons for public service productivity measurement (ESCoE DP 2024-18)
www.escoe.ac.uk/publications/composite-indices-reflections-on-recent-uk-experience-and-lessons-for-public-service-productivity-measurement/

Link 17: Wellbeing - Science and Policy
https://www.cambridge.org/core/books/wellbeing/B64B0CDF6266FB37FBE263B9F4A1BA57?pageNum=1&searchWithinIds=B64B0CDF6266FB37FBE263B9F4A1BA57&productType=BOOK_PART&searchWithinIds=B64B0CDF6266FB37FBE263B9F4A1BA57&productType=BOOK_PART&sort=mtdMetadata.bookPartMeta._mtdPositionSortable%3Aasc&pageSize=30&template=cambridge-core%2Fbook%2Fcontents%2Flistings&ignoreExclusions=trueinternationalcomparisonsofthemeasurementofnonmarketoutputduringthecoronaviruscovid19pandemicinternationalcomparisonsofthemeasurementofnonmarketoutputduringthecoronaviruscovid19pandemicinternationalcomparisonsofthemeasurementofnonmarketoutputduringthecoronaviruscovid19pandemicinternationalcomparisonsofthemeasurementofnonmarketoutputduringthecoronaviruscovid19pandemic

Link 18: Green Book supplementary guidance: wellbeing
www.gov.uk/government/publications/green-book-supplementary-guidance-wellbeing

Link 19: The Economic Theory of Price Indices
ideas.repec.org/b/eee/monogr/9780122577505.html

Link 20: Sources and methods for public service productivity estimates - 9. Difference between annual and quarterly statistics
www.ons.gov.uk/economy/economicoutputandproductivity/publicservices productivity/methodologies/sourcesandmethodsforpublicserviceproductivity estimates#difference-between-annual-and-quarterly-statistics

Link 21: A guide to quality adjustment in public service productivity measures
www.ons.gov.uk/economy/economicoutputandproductivity/publicservices productivity/articles/aguidetoqualityadjustmentinpublicserviceproductivity measures/2019-08-07

Link 22: Code of Practice for Statistics - Cross-cutting themes
code.statisticsauthority.gov.uk/the-code/cross-cutting-themes/

Link 23: OSR Insight: Coherence
osr.statisticsauthority.gov.uk/wp-content/uploads/2025/07/OSR-Insight-Report-Coherence.pdf

Link 24: Public service productivity, UK: 1997 to 2022
www.ons.gov.uk/economy/economicoutputandproductivity/publicservices productivity/articles/publicserviceproductivityuk/latest

Link 25: Time use in the public sector, Great Britain Statistical bulletins
www.ons.gov.uk/economy/economicoutputandproductivity/publicservices productivity/bulletins/timeuseinthepublicsectorgreatbritain/previousreleases

Link 26: JEEA-FBBVA Lecture 2013: The New Empirical Economics of Management
academic.oup.com/jeea/article-abstract/12/4/835/2318424?login=false

Link 27: Management and Expectations Survey
www.ons.gov.uk/surveys/informationforbusinesses/businesssurveys/ managementandexpectations

Link 28: World Management Survey
worldmanagementsurvey.org/

Link 29: Public Sector Management Practices Survey pilot, UK: 2023
www.ons.gov.uk/economy/economicoutputandproductivity/publicservices productivity/bulletins/publicsectormanagementpracticessurvey pilotuk/2023

Link 30: Public sector managers' views on management practices, Great Britain: August to September 2024
www.ons.gov.uk/economy/economicoutputandproductivity/publicservices productivity/articles/publicsectormanagersviewsonmanagementpractices greatbritain/augusttoseptember2024

Link 31: Use of artificial intelligence
www.nao.org.uk/wp-content/uploads/2024/03/use-of-artificial-intelligence-in-government-summary.pdf

Link 32: Time use in the public sector, Great Britain
www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/datasets/timeuseinthepublicsectorgreatbritain

Link 33: GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models
arxiv.org/abs/2303.10130

Link 34: NSCASE: Decisions made by National Statistician
uksa.statisticsauthority.gov.uk/wp-content/uploads/2023/08/National-Statistician-Decision-on-NSCASE-Advice.pdf

Link 35: Productivity of the English National Health Service: 2021/22 update
www.york.ac.uk/media/che/CHE196.pdf

Link 36: Independent report: Policing Productivity Review (accessible)
www.gov.uk/government/publications/policing-productivity-review/policing-productivity-review-accessible

Link 37: Public service productivity estimates: healthcare QMI
www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/methodologies/publicserviceproductivityestimateshealthcareqmi

Link 38: National Cost Collection for the NHS
www.england.nhs.uk/costing-in-the-nhs/national-cost-collection/

Link 39: Planning to Safely Reduce Avoidable Conveyance
www.england.nhs.uk/publication/planning-to-safely-reduce-avoidable-conveyance/

Link 40: The NHS Long Term Plan
www.england.nhs.uk/wp-content/uploads/2022/07/nhs-long-term-plan-version-1.2.pdf

Link 41: Experiences of NHS healthcare services in England: December 2024
www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthcaresystem/bulletins/experiencesofnhshealthcareservicesinengland/latest

Link 42: Improved methods for total public service productivity: total, UK, 2021
www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/methodologies/improvedmethodsfortotalpublicserviceproductivitytotaluk2021

Link 43: Background report: National Reference Test Information
https://www.gov.uk/government/publications/national-reference-test-your-information-and-your-privacy/background-report-national-reference-test-information

Link 44: Understanding Society: Waves 1-14, 2009-2023 and Harmonised BHPS: Waves 1-18, 1991-2009
beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=6614

Link 45: The measurement of public goods: Lessons from 10 years of Atkinson in the United Kingdom
www.econstor.eu/handle/10419/309833

Link 46: Challenges of measuring defence output
www.escoe.ac.uk/challenges-of-measuring-defence-output/

Link 47: Measuring Defence productivity: a first step
quarterly.blog.gov.uk/2019/10/30/measuring-defence-productivity-a-first-step/

Link 48: Understanding the contribution of defence to UK prosperity
www.rand.org/randeurope/research/projects/2021/defence-contribution-to-uk-prosperity.html

Link 49: Fixing the foundations: public spending audit 2024-25
www.gov.uk/government/publications/fixing-the-foundations-public-spending-audit-2024-25/fixing-the-foundations-public-spending-audit-2024-25-html

Link 50: Ron P Smith
www.escoe.ac.uk/people/ron-smith/

Link 51: Economic Statistics Centre of Excellence
www.escoe.ac.uk/

Link 52: Issues in Measuring Defence Output: A Review (ESCoE DP 2024-06)
www.escoe.ac.uk/publications/issues-in-measuring-defence-output-a-review/

Link 53: Economic and Labour Market Review
escoe-website.s3.amazonaws.com/wp-content/uploads/2020/09/04203700/Economic-Labour-Market-Review-January-2009.pdf

Link 54: Public Service Productivity, Measuring Defence Labour
webarchive.nationalarchives.gov.uk/ukgwa/20151014062937/http:/www.ons.gov.uk/ons/rel/psa/public-service-productivity/measuring-defence-labour/index.html

Link 55: Quality adjustment of public service criminal justice system output: experimental method
www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/articles/qualityadjustmentofpublicservicecriminaljusticesystemoutputexperimentalmethod/2017-10-06#next-steps

Link 56: Independent report: Policing Productivity Review
www.gov.uk/government/publications/policing-productivity-review

Link 57: Police officer uplift, England and Wales, quarterly update to 31 March 2023
www.gov.uk/government/statistics/police-officer-uplift-quarterly-update-to-march-2023/police-officer-uplift-england-and-wales-quarterly-update-to-31-march-2023

Link 58: Independent Review of Police Officer and Staff Remuneration and Conditions
assets.publishing.service.gov.uk/media/5a7c4561e5274a1b00422a33/8325_i.pdf

Link 59: Police Remuneration Review Body
assets.publishing.service.gov.uk/media/62d6cb23e90e071e7cbe1eda/PRRB_2022_report_for_publication.pdf

Link 60: What to report
www.police.uk/pu/contact-us/what-and-how-to-report/what-report/

Link 61: Police recorded crime and outcomes open data tables
www.gov.uk/government/statistics/police-recorded-crime-open-data-tables

Link 62: What is Action Fraud?
www.actionfraud.police.uk/what-is-action-fraud

Link 63: PEEL assessments data
hmicfrs.justiceinspectorates.gov.uk/police-forces/data/peel-assessments/

Link 64: Missing Persons Statistical Bulletins
www.nationalcrimeagency.gov.uk/find-support/downloads

Link 65: Definition of policing by consent
www.gov.uk/government/publications/policing-by-consent/definition-of-policing-by-consent

Link 66: Policing in the UK
researchbriefings.files.parliament.uk/documents/CBP-8582/CBP-8582.pdf

Link 67: Victims commissioner
victimscommissioner.org.uk/our-work/reviews-and-reports/

Link 68: Improving police productivity: a response to the recommendations of the Policing Productivity Review
www.gov.uk/government/publications/policing-productivity-review-government-response/improving-police-productivity-a-response-to-the-recommendations-of-the-policing-productivity-review-accessible

Link 69: Immigration system statistics quarterly release
www.gov.uk/government/collections/immigration-statistics-quarterly-release

Link 70: Migration transparency data
www.gov.uk/government/collections/migration-transparency-data

Link 71: Immigration system statistics data tables
www.gov.uk/government/statistical-data-sets/immigration-system-statistics-data-tables

Link 72: Activities, spending and productivity in the Fire and Rescue Services since 2009
www.ons.gov.uk/economy/governmentpublicsectorandtaxes/publicspending/articles/activitiesspendingandproductivityinthefireandrescueservicessince2009/2019-05-30

Link 73: Prison population statistics
www.gov.uk/government/collections/prison-population-statistics

Link 74: Offender management statistics quarterly
www.gov.uk/government/collections/offender-management-statistics-quarterly

Link 75: Criminal court statistics quarterly: October to December 2023
www.gov.uk/government/statistics/criminal-court-statistics-quarterly-october-to-december-2023

Link 76: Criminal Court Statistics publication: Pivot Table Analytical Tool for England and Wales
view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fassets.publishing.service.gov.uk%2Fmedia%2F6758548541183cab7aeb64b1%2Fcc_waiting_hearing_tool.xlsx&wdOrigin=BROWSELINK

Link 77: Legal aid statistics
www.gov.uk/government/collections/legal-aid-statistics

Link 78: Criminal court statistics quarterly: July to September 2024
www.gov.uk/government/statistics/criminal-court-statistics-quarterly-july-to-september-2024

Link 79: HM Prison and Probation Service workforce quarterly: September 2024
www.gov.uk/government/statistics/hm-prison-probation-service-workforce-quarterly-september-2024/hm-prison-and-probation-service-workforce-quarterly-september-2024#:~:text=As%20at%20September%202024%2C%20there,4%20%2F%20prison%20officer%20(incl.

Link 80:
uksa.statisticsauthority.gov.uk/about-the-authority/uk-statistical-system/types-of-official-statistics/

Link 81: Types of official statistics
unstats.un.org/unsd/classifications/cofog/revision

Link 82: Classification of the Functions of Government
www.gov.uk/government/statistics/measuring-tax-gaps

Link 83: UK National Accounts, The Blue Book: 2024
www.ons.gov.uk/economy/grossdomesticproductgdp/compendium/unitedkingdomnationalaccountsthebluebook/2024

Link 84: Annual reports and accounts published by HMRC
www.gov.uk/government/collections/hmrcs-annual-report-and-accounts

Link 85: HMRC annual report and accounts: 2023 to 2024
www.gov.uk/government/publications/hmrc-annual-report-and-accounts-2023-to-2024

Link 86: Public service productivity Articles
www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/articles/publicservicesproductivityestimatestotalpublicservices/previousreleases

Link 87: Numbers of taxpayers and registered traders
www.gov.uk/government/statistics/numbers-of-taxpayers-and-registered-traders

Link 88: HMRC tax receipts and National Insurance contributions for the UK
www.gov.uk/government/statistics/hmrc-tax-and-nics-receipts-for-the-uk

Link 89: Background information: Fraud and error in the benefit system statistics, 2023 to 2024 estimates
www.gov.uk/government/statistics/fraud-and-error-in-the-benefit-system-financial-year-2023-to-2024-estimates/background-information-fraud-and-error-in-the-benefit-system-statistics-2023-to-2024-estimates

Link 90: Northern Ireland Omnibus Survey
www.nisra.gov.uk/northern-ireland-omnibus-survey

Link 91: Continuous Household Survey
www.nisra.gov.uk/statistics/find-your-survey/continuous-household-survey

Link 92: Improved methods for total public service productivity: total, UK, 2019
www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/methodologies/improvedmethodsfortotalpublicserviceproductivitytotaluk2019#children-s-social-care

Link 93: Public service productivity: adult social care, sources and methods, 2019 update
www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/methodologies/publicserviceproductivityadultsocialcaresourcesandmethods2019update

Link 94: Public service productivity: total, UK, 2022
www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/articles/publicservicesproductivityestimatestotalpublicservices/latest

Link 95: Public service productivity, adult social care, England: financial year ending 2021
www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/articles/publicserviceproductivityadultsocialcareengland/latest

Link 96: Experimental estimates of green jobs, UK: 2024
www.ons.gov.uk/economy/environmentalaccounts/bulletins/experimentalestimatesofgreenjobsuk/2024

Link 97: Expenditure on UK and international biodiversity
jncc.gov.uk/our-work/ukbi-biodiversity-expenditure/

Link 98: The initial list of issues for consideration in the Classification of the Functions of Government (COFOG)
unstats.un.org/unsd/classifications/cofog/revision#:~:text=Introduction&text=The%20update%20of%20COFOG%20will,be%20completed%20by%20March%202027.

Link 99: Reflections on the forthcoming System of National Accounts revision
uksa.statisticsauthority.gov.uk/news/reflections-on-the-forthcoming-system-of-national-accounts-revision/

Link 100: Why and How Should Human Capital be Measured in National Accounts?
www.insee.fr/en/statistiques/fichier/4770152/05-ES_517-518-519_Canry_ENWeb.pdf

Link 101: Measuring Human Capital in the UK Economic Accounts: An experimental satellite account (ESCoE DP 2022-12)
www.escoe.ac.uk/publications/measuring-human-capital-in-the-uk-economic-accounts-an-experimental-satellite-account/

Link 102: Measuring the Economic Flows of UK Human Capital (ESCoE DP 2024-05)
www.escoe.ac.uk/publications/measuring-the-economic-flows-of-uk-human-capital/

Link 103: Public service productivity: total, UK, 2020
www.ons.gov.uk/economy/economicoutputandproductivity/publicservicesproductivity/articles/publicservicesproductivityestimatestotalpublicservices/2020

Link 104: EPI research for the Department for Education on pupil learning loss (March 2022)
epi.org.uk/publications-and-research/epi-research-for-the-department-for-education-on-pupil-learning-loss-march-2022/

Link 105: Quality adjustment of education
uksa.statisticsauthority.gov.uk/publication/national-statisticians-independent-review-of-the-measurement-of-public-services-productivity/pages/25/

Link 106: Level 2 and 3 attainment age 16 to 25
explore-education-statistics.service.gov.uk/methodology/level-2-and-3-attainment-age-16-to-25

Link 107: The Impact of Pupil Behaviour and Wellbeing on Educational Outcomes
assets.publishing.service.gov.uk/media/5a747ef340f0b604dd7ae609/DFE-RR253.pdf

Link 108: Ministry of Defence
www.gov.uk/government/organisations/ministry-of-defence/about

Link 109: Ready for War?
committees.parliament.uk/publications/43178/documents/214880/default/

Link 110: Tactics and Trade-Offs: The Evolution of Manoeuvre in the British Army
academic.oup.com/book/45784/chapter/400601525?login=false

Link 111: Geopolitical Risk (GPR) Index
www.matteoiacoviello.com/gpr.htm

Link 112: National balance sheet estimates for the UK: 2024
www.ons.gov.uk/economy/nationalaccounts/uksectoraccounts/bulletins/nationalbalancesheet/2024#uk-net-worth

Link 113: Measuring Geopolitical Risk
www.matteoiacoviello.com/gpr_files/GPR_PAPER.pdf

Link 114: Military expenditure (current USD) - Russian Federation, China, United Kingdom
data.worldbank.org/indicator/MS.MIL.XPND.CD?end=2021&locations=RU-CN-GB&start=1997

Link 115: Measuring Productivity - OECD Manual
www.oecd.org/en/publications/measuring-productivity-oecd-manual_9789264194519-en.html

Link 116: European system of accounts ESA 2010
ec.europa.eu/eurostat/documents/3859598/5925693/KS-02-13-269-EN.PDF

Link 117: Stat-Xplore
stat-xplore.dwp.gov.uk/webapi/jsf/login.xhtml

Link 118: Fraud and error in the benefit system Financial Year Ending (FYE) 2023
www.gov.uk/government/statistics/fraud-and-error-in-the-benefit-system-financial-year-2022-to-2023-estimates/fraud-and-error-in-the-benefit-system-financial-year-ending-fye-2023

Link 119: Fraud and Error in the Benefit System: Background information and methodology
assets.publishing.service.gov.uk/media/5ed0c138d3bf7f4604912084/fraud-and-error-statistics-background-and-methodology-revised-29-may-2020.pdf

Link 120: Annual Report & Accounts 2020-21
assets.publishing.service.gov.uk/media/60fa97498fa8f5043411d83d/dwp-annual-report-and-accounts-2020-2021-web-ready.pdf

Link 121: 2023/24 Annual Data Requirement from Police Forces in England and Wales
https://www.npcc.police.uk/SysSiteAssets/media/downloads/publications/disclosure-logs/[...]coordination-committee/2023/274-2023-adr-notice-2023-24.pdf

Acknowledgements

This Review has been made possible through close collaboration with a large number of stakeholders across multiple disciplines. The National Statistician would like to recognise and express gratitude to the UK government bodies, devolved governments, academics, and wider sector experts for their contribution. In particular to Professor Martin Weale CBE who acted as an independent reviewer of the Review Report.

UK government departments

Cabinet Office (CO)

Department for Education (DfE)

Department for Environment, Food and Rural Affairs (Defra)

Department for Transport (DfT)

Department of Health and Social Care (DHSC)

Department for Work and Pensions (DWP)

His Majesty's Revenue and Customs (HMRC)

His Majesty's Treasury (HMT)

Home Office (HO)

Ministry of Defence (MoD)

Ministry of Housing, Communities and Local Government (MHCLG)

Ministry of Justice (MoJ)

Arm's length bodies, devolved governments and public sector bodies

Centre for Police Productivity (CPP)

College of Policing

His Majesty's Inspectorate of Constabulary and Fire and Rescue Services (HMICFRS)

National Audit Office

National Police Chiefs' Council (NPCC)

NHS England

Northern Ireland Statistics and Research Agency (NISRA)

Office for Local Government (OFLOG)

Office of Qualifications and Examinations Regulation (Ofqual)

Scottish Government

Welsh Government

Academic Institutions

Alan Turing Institute

The Bennett Institute, Cambridge

Birkbeck, University of London

Economic Statistics Centre of Excellence

Imperial College London

Kings College London

The Productivity Institute

University of Groningen

University of Manchester

University of York

Think tanks, Independent Bodies and Individuals

Centre for Social Justice

Chartered Institute of Financial Accountants

Crest Advisory

Energy Security

Institute for Fiscal Studies (IFS)

Institute for Government

National Statistics Institutes

Australian Bureau of Statistics

Bureau of Economic Affairs (USA)

Bureau of Labor Statistics (USA)

Italian National Institute of Statistics (ISTAT)

Statistics Canada

The National Institute of Statistics and Economic Studies (INSEE)

Contact details

If you would like any further information or to provide feedback on any elements mentioned in the National Statistician's Independent Review of the Measurement of Public Services Productivity, please email the Office for National Statistics' Public Service Productivity team at **psp.review@ons.gov.uk**.