



Review of Approaches to the Measurement of Public Sector Productivity Statistics

Mary O'Mahony, Mechelle Viernes and Martin Weale

ESCoE Technical Report TR-26

March 2025

ISSN 2515-4664

TECHNICAL REPORT

Abstract

Productivity is generally defined as a measure of the amount of output generated per unit of input. In many countries, public sector productivity has been assumed to be zero in the national accounts. The output of the government sector has been measured as equal in value to the total value of inputs. This output=input convention has increasingly come under scrutiny in recent years. The challenge is to devise alternative estimates based on output measurement in a public sector context – where collective services are provided and where there is, in most instances, no market transaction in services provided to individuals (Boyle, 2006).

This report aims to provide a summary of approaches to defining and measuring public sector productivity. The first part of the paper begins with a review of international practices for a few national statistical institutes (NSIs) and Productivity Commissions where information was readily available from websites. This is followed by more general discussions of issues that impact on public sector productivity. We first consider the organization of public sector production which could feed into measurement. This draws heavily on a report reviewing measurement for the police, undertaken by The Productivity Institute (TPI). We then consider the question of weighting of outputs and quality adjustments. Next, we look at approaches that are based on outcomes rather than outputs and how to incorporate preventative measures. The final part contains some recommendations for future research.

Keywords: public sector productivity, input, measurement of output, quality adjustment, health, education, social security

JEL classification: D24, H10, H11, H30, H40, H51, H52, H53, H55, H56

Mechelle Viernes, King's College London and ESCoE
mechelle.viernes@kcl.ac.uk

Published by:
Economic Statistics Centre of Excellence
King's College London
Strand
London
WC2R 2LS
United Kingdom
www.escoe.ac.uk

ESCoE Technical Reports describe research in progress by the author(s) and are published to elicit comments and to further debate. Any views expressed are solely those of the author(s) and so cannot be taken to represent those of the Economic Statistics Centre of Excellence (ESCoE), its partner institutions or the Office for National Statistics (ONS).

Review of Approaches to the Measurement of Public Sector Productivity Statistics

Mary O'Mahony, Mechelle Viernes and Martin Weale
Economic Statistics Centre of Excellence, King's College. London

March 2025

Table of Contents

I. Introduction	3
II. Review of the International Best Practices in the Production of the Public Sector Productivity Statistics.....	3
A. Measurement of Public Sector Productivity.....	5
B. Public Sector Productivity Measures: Meso-level.	7
1.1. Health Sector: United Kingdom, Denmark and New Zealand	9
1.2. Education Sector: United Kingdom, New Zealand, Denmark, Malaysia and the Netherlands 13	
1.3. Social Protection: Denmark.....	21
1.4. Libraries: The Netherlands	22
1.5. Public Order and Safety: UK	24
1.6. Fire Department: The Netherlands ¹⁸	28
1.7. Courts: The Netherlands and New Zealand	30
1.8. Public Prosecution Service: The Netherlands ²⁰	32
1.9. Police Service: The Netherlands and New Zealand	35
C. Public Sector Productivity Measures of Asian Productivity Organization (APO) Member Countries	37
1.10. Tax Collection Service: Bangladesh, Indonesia, Philippines, Sri Lanka, Thailand.....	39
1.11. Passport Service: Bangladesh, Indonesia, Sri Lanka	41
D. The BLS Productivity Measures.....	43
1.12. Correctional Services	43
1.13. State Employment Security Services	44
E. Non-market output measures in the Australian National Accounts	45
III. Issues in defining and measuring public sector productivity.....	48
A. The organization of public sector production.....	48
B. Weighting	51
C. Outcome based approaches	51
D. Prevention.....	52
IV. Final Observations.....	53
V. References:	55

I. Introduction

Productivity is generally defined as a measure of the amount of output generated per unit of input. In many countries, public sector productivity has been assumed to be zero in the national accounts. The output of the government sector has been measured as equal in value to the total value of inputs. This output=input convention has increasingly come under scrutiny in recent years. The challenge is to devise alternative estimates based on output measurement in a public sector context – where collective services are provided and where there is, in most instances, no market transaction in services provided to individuals (Boyle, 2006).

This report aims to provide a summary of approaches to defining and measuring public sector productivity. The first part of the paper begins with a review of international practices for a few national statistical institutes (NSIs) and Productivity Commissions where information was readily available from websites. This is followed by more general discussions of issues that impact on public sector productivity. We first consider the organization of public sector production which could feed into measurement. This draws heavily on a report reviewing measurement for the police, undertaken by The Productivity Institute (TPI). We then consider the question of weighting of outputs and quality adjustments. Next, we look at approaches that are based on outcomes rather than outputs and how to incorporate preventative measures. The final part contains some recommendations for future research.

II. Review of the International Best Practices in the Production of the Public Sector Productivity Statistics.

This section presents the methodologies adopted by some NSIs and Productivity Commissions to measure public sector output and productivity. A desk review was conducted on the methodologies adopted by NSIs and Productivity Commissions, listed below, together with a review of documents produced by some international organizations which helped inform the choice of services to include. The main criterion for inclusion was the availability of documentation on the methodologies employed in public facing websites. Therefore, it does not cover any experimental measures that might be being developed by NSIs but are not at a stage where these are yet available for public scrutiny.

1. Office for National Statistics, UK
2. Statistics Netherlands
3. Statistics New Zealand
4. New Zealand Productivity Commission
5. Statistics Denmark
6. Statistics Finland
7. Australia Bureau of Statistics

8. Australia Productivity Commission
9. US Bureau of Labor Statistics (BLS)
10. Malaysia Productivity Corporation
11. Development Academy of the Philippines
12. Organization for Economic Cooperation and Development (OECD)
13. World Bank
14. Asian Productivity Organization (APO)

Although most of this represents ongoing work by the institutions listed above, there are some exceptions. Statistics Finland discontinued releasing statistics on central and local government productivity in 2013 with the data concerning 2012¹. Likewise, the BLS released public sector productivity measures until it was terminated in 1994². However, the BLS continued releasing labour productivity estimates for the state and local government in selected industries and government services. While government labour productivity was no longer published, measures of public productivity data, particularly of the federal government and state and local government were included in the industry-level production.

This review is limited to the following:

- a. methodologies used by the NSIs and Productivity Commissions on measuring meso-level public sector productivity, which focused on the productivity of the various policy sectors such as health and social care, education and training, or courts, among others;
- b. methodologies used by the NSIs and Productivity Commissions on measuring meso-level non-market output adjusted for quality; and
- c. The BLS and Statistics Finland methodology on measuring public sector productivity were still documented in this report, although this work has been discontinued.

This review focused on productivity measures within the public sector, specifically the general government, which encompasses central government, local government, and other public bodies involved in delivering services like healthcare, education, defense, and social security. It also covers public corporations that are owned or controlled by the government. While these corporations may operate in market environments and offer goods or services similarly to private companies, they remain government-controlled, such as public transportation services and utilities like water supply.

Although the review primarily focused on the general government and public corporations under government control, it also addressed non-market output. These entities provide goods and services without seeking profit, with their output typically funded through taxation or other public funding sources.

¹ https://www.stat.fi/ajk/tiedotteet/2014/uutinen_040_2014-12-18_en.html

² <https://www.bls.gov/productivity/articles-and-research/previous-research.htm#gov>

Further, some aspects of country practices in measuring public sector productivity correspond to national accounts methodologies, while others were outside the national accounts practices. National accounts primarily capture public sector productivity through output measures aligned with the System of National Accounts (SNA), often using cost-weighted activity indicators. However, many countries also employ broader performance indicators—such as service quality, efficiency metrics, and outcome-based assessments—that are not directly incorporated into national accounts but are crucial for policy evaluation. These external indicators provide additional insights into the effectiveness of public services, complementing the national accounts framework.

A. Measurement of Public Sector Productivity

The Atkinson Review (2005) pointed out the need to measure public sector output and productivity based on set of principles and that it should be measured consistently with private sector output. The Review also stressed the need to measure public service output not only based on quantity but also adjusted for quality, which can focus on how the outcome was achieved as a direct result of the public service provided.

It is recognized that measuring public sector productivity remains a challenge since the public goods and services that the government delivered to the citizen or firms are often free or highly subsidized. Some of these challenges were on the measurement of outputs and inputs – availability and reliability of the prices of public goods and services; linking or attributing the outputs to specific inputs particularly for services that are collective in nature; and the adjustments for quality (Dunleavy, 2017; Lau et. al., 2017; Foxten et al., 2019).

While challenges on measuring productivity of the public sector remain, there were several recommendations provided to address these measurements issues and concerns. Lau et. al (2017) outlined approaches that can be adopted on measuring public sector productivity across countries, including estimation at the macro-level public productivity statistics, meso-level which focuses on the productivity of different sectors, or at the micro-level which focuses on the productivity of the institutions or key activities and functions. Schreyer (2010) recommended the use of implicit and explicit adjustments in adjusting the changes in the quality of output of public services.

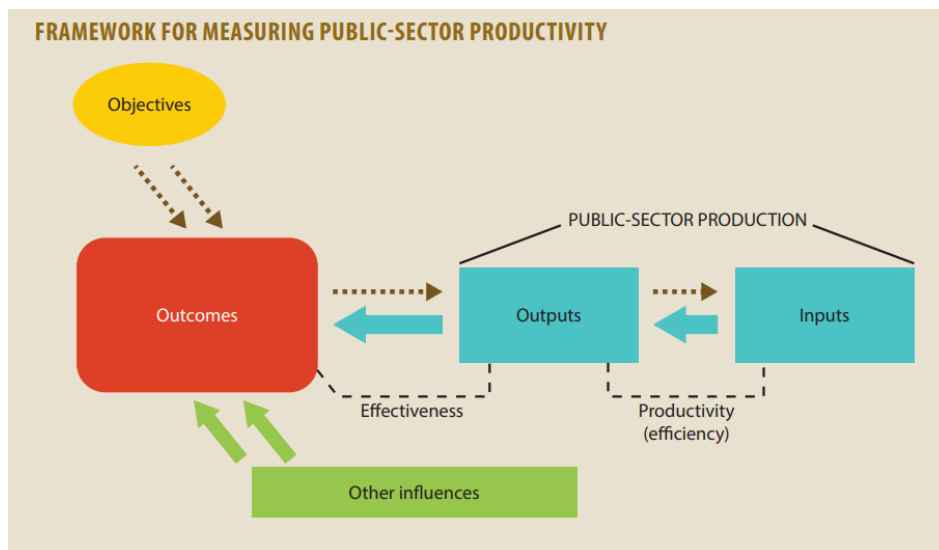
The degree to which an outcome can be attributed to a public service is a key consideration when adjusting for quality. Quality adjustment is used to show the impact of a service provided on specific successful outcomes of the service. In essence, it stresses that when evaluating the effectiveness or productivity of public sector activities, the assessment should focus on the marginal contribution of these activities to the outcomes, rather than considering the total outcome or an average contribution.

World Bank (2021) prepared the challenges in measuring public sector productivity and recommended approaches in implementing its measurement such as exploring the combination of multiple measures of public sector productivity given the complex nature of the public sector production function. It also recommended to conceptualize the service-delivery productivity

chain and examine measures at each level of the productivity chain – starting from identifying the indicators to measure the performance, explore possible data source that can be supplied in each level of the chain, and to identify consequences that may arise.

The interest in public sector output is not confined to developed countries. APO (2019), drawing on the work by the World Bank, referred to the general framework set out in different research efforts (Figure 1) in identifying the public sector inputs, outputs and outcomes that would be used in measuring productivity. It was noted that there are range of different services provided by the public sector to achieve a variety of outcomes. These different services should be aggregated to estimate the overall output measure. It was stressed though that while service counts of outputs is a good strategy in productivity measure, this can only be done in services provided to individuals such as health, education, or social welfare and that different estimation procedures are required if the services are collective such as police, defense, or fire service. Recently, several APO member countries were exploring ways to incorporate public sector productivity in their national accounts framework, referring to Dunleavy and Carrera (2013). The Development Academy of the Philippines (DAP), the APO's Center of Excellence for Public Sector Productivity is currently refining the PSP Measurement Framework in the context of the Philippine public sector, specifically incorporating the value chains for efficient and effective public service delivery. Currently, DAP is conducting case studies on productivity measurement using this refined framework with two national government agencies and ten local government units in the country. The refined framework and methodology for the Philippine PSP is currently being discussed for approval and adoption. Some of these initiatives of the APO member countries are discussed below.

Figure 1. General Framework for Measuring Public Sector Productivity



Source: Asian Productivity Organization

B. Public Sector Productivity Measures: Meso-level.

Significant progress has been made by the NSIs and Productivity Commissions on the development of measures for public sector productivity with focus on different sectors. The table below shows the summary of the available meso-level public sector productivity measures.

Table 1. Meso-level Estimation of Public Sector Productivity

	UK	The Netherlands	New Zealand	Denmark	Finland	Malaysia
Health Care	✓	✓	✓	✓		
Education*	✓	✓	✓	✓		✓
Early Childhood Development			✓			
Universities/Tertiary			✓			
Military Defense	✓					
Police	✓	✓	✓			
Public sector and safety (Fire, courts, probation, prisons)	✓					
Fire		✓				
Courts		✓	✓			
Prosecution Service		✓				
Children's Social Care	✓					
Adult's Social Care	✓					
Social Protection				✓		
Social Development			✓			
Social security Administration	✓					
Other Government Services	✓					
Library		✓				
Central government					✓	
Local government					✓	

*Productivity measure on education in Denmark includes primary, secondary, higher and further education

Statistics Finland measures overall public sector productivity for the central government and the local government. The output measure for the central government was identified by each government agency and not by the statistical agency. These agency output measures were aggregated using a Tornqvist index. Niemi (1997) suggested that the output indicator method can be used to measure the volume of output for a productivity index. These output measures include teaching hours for educational services, numbers of library visitors for library services, number of bed days and customers or visits for social services.

The following presents the details of the meso-level PSP measure adopted by the NSIs and Productivity Commissions. Each entry involves a brief discussion of the institutional setting, inputs, outputs and quality adjustments.

1.1. Health Sector: United Kingdom, Denmark and New Zealand

1.1.1. United Kingdom³

Institutional Setting	ONS incorporated methodological changes in PSP on healthcare services in 2021. Estimates of the volume of healthcare output included the following components: 1. General practice; 2. Dental services; and 3. Ophthalmic services.
Inputs	Inputs were estimated based on the following three components: <i>Labour Input:</i> Directly measured using a Laspeyres cost-weighted labour index (CWLI), which tracks the growth in the number of full-time equivalent staff, weighted by their associated costs. <i>Intermediate Consumption:</i> Intermediate consumption of goods and services used in the healthcare provision, such as gloves and syringes (including agency staff), legal and audit services, calculated by deflating expenditure using relevant deflators. <i>Capital:</i> Consumption of fixed capital, covering the cost of depreciation of capital goods (items that are anticipated to be in use for a year or more, such as buildings and vehicles) over time.
Outputs	The volume output was divided into components: a. measured using data on the number of activities undertaken and its unit costs by service type in cost-weighted activity index (CWAI) through the Laspeyres approach: i. hospital and community health services (HCHS), which includes: • hospital services, • community care, • mental health, • ambulance services ii. primary care services, formerly known as Family Health Services (FHS) which includes:

³ ONS-UK (2021)

- general practice,
 - i. General practitioners (GP) appointments split by mode of consultation (surgery consultations, home visits, telephone consultations, and video or online consultations); and
 - ii. Non-GP or other practice staff led consultations (including practice nurses, counsellors, health visitors, physiotherapists, and other professionals), which was the new measure replacing the previous indicator of consultations led by practice nurse only.
- dental practice services, activities are weighted based on the complexity using dental bands, and
- ophthalmic services, include the sight tests at an optician, home-based sight tests (individual test), home-based sight tests (group tests), optical vouchers issued.
- iii. community prescribing which includes GP prescribing and prescription drugs dispensing in the community.
- iv. Non-NHS provision, includes healthcare funded by the government but provided by the private or third sector which is indirectly measured using the “output-equals-inputs” approach
 - local authority commissioned treatments for alcohol and drug misuse (excluding NHS providers) measured using the number of psychosocial interventions
 - local authority commissioned smoking cessation service measured using the number of quit attempts

Quality-adjusted Output

The quality adjustment applied to the volume of activity index to generate estimates of healthcare output were the following:

1. Hospital procedures:

- 1.1. short-term post-operative survival rates (derived from Hospital Episodes Statistics (HES))
- 1.2. estimates of health benefit from procedures (derived from research studies, ONS Life Tables and Patient Reported Outcome Measures)
- 1.3. waiting times (from HES)

- 1.4. Care management of patients on a GP list who have been diagnosed with hypertension, coronary heart disease, chronic kidney disease or as having had a stroke
2. aggregate data on clinical measures recorded on GP practice computers (from the Quality and Outcomes Framework)
3. patients experience surveys: *The GP Patient Survey (GPPS) results is used to estimate the Overall Practice Experience Score (OPES)*

1.1.2. Denmark⁴

Institutional Setting

The Public Danish Healthcare Services focused on the hospital activity, dental treatments, and on social provisions with and without institutional care. (Statistics Denmark, 2014)

Inputs

Input measure was estimated using the general government production which includes:

Labour input:

Compensation of employees, which measures the current value of the flow of services of labour input

Intermediate consumption:

Intermediate consumption measures current value inputs of goods and services in production

Capital:

Consumption of fixed capital refers to the deterioration and technical obsolescence of the stock of fixed assets (incl. research and development)

Other production taxes less subsidies:

Other production taxes less subsidies that the government pays because of engaging in production.

Outputs

The volume of output of health services was measured as the quantity of health services provided to individuals adjusted for new products and services. It covered the following relevant areas:

⁴ Statistics Denmark (2014)

1. Hospital activity

The volume index for general hospital activity was based on the Diagnosis Related Group (DRG) database, which provides information on the cost per type of treatment and the number of treatments

2. Treatment by dentists

The volume index for dental services was estimated using the number of treatments distributed according to type of treatment and unit costs.

3. Social provisions with and without institutional care

The volume index for residential care related to social services for adults and elderly was estimated by applying the number of people receiving care and the unit cost.

Note that the health care activity of the residential care is already included in the calculations of health care volume.

Quality-adjusted Output

The Danish health care output was adjusted for quality based on the successful delivery of services and the extent to which service corresponds to user's requests. The quality adjustment incorporates the following:

1. Death rate of patients admitted with an acute disease, which measures success in delivering health gain
 - a. AMI 30-day mortality rate
 - b. Hemorrhagic stroke 30-day mortality rate
 - c. Ischemic stroke 30-day mortality rate
 - d. Cervical cancer five-year relative survival rate
 - e. Breast cancer five-year relative survival rate
 - f. Colorectal cancer five-year relative rate
 - g. Asthma mortality rate

Note that the since quality indicators were only available for the DRG groups listed above, quality adjustment was only partially adjusted to these DRGs.

2. Waiting time for treatment
3. Average lifetime can be use as quality output for preventive treatment; an increase in average lifetime may partly be due to the preventive health care arrangements.

4. Patients experience based on the satisfaction survey.

Aggregation

The weights used were based on subjective assessment, using the judgement of experts, specifically based on the perceived importance of the indicator. To illustrate, the survival rate of the DRGs listed above was given with a weight of 0.6 as this was regarded as more important than waiting time and patient experience. Total weights used were not equal to 1 given the assumption that there are still unmeasured factors that cannot be measured or observed.

1.2. Education Sector: United Kingdom, New Zealand, Denmark, Malaysia and the Netherlands

1.2.1 United Kingdom⁵

Institutional Setting

ONS UK included in the measurement of PSP the following education sectors that were publicly funded were:

1. pre-schools
2. publicly funded Private, Voluntary and Independent (PVI) pre-school places
3. primary schools including academies.
4. secondary schools including academies and city technology colleges.
5. special schools
6. Further Education (FE)

Further education was included in the volume of output measure in public service productivity although its funding was classified in the non-profit institutions serving households (NPISH) sector which was not included in national accounts' measures of general government final consumption expenditure.

On the other hand, Non-State funded schools and Higher Education, other than the training of teachers and some health professionals were excluded in the productivity index estimation.

Inputs

The inputs index included the aggregate of labour, goods and services, and capital, broken down by the accountable level of government.

Labour input:

⁵ ONS-UK (2013)

The education labour input included full-time equivalent teaching and support staff in local authority-maintained schools from school census data (measured directly) plus local and central government administrative staff (measured indirectly).

The change in the number of hours-adjusted full-time-equivalent staff for each category was calculated and weighted using expenditure shares.

Intermediate Consumption:

Intermediate consumption inputs were measured indirectly using comprehensive annual current price expenditure data submitted provided by government departments to HM Treasury and figures from national accounts is deflated using the consumer, producer and service sector price indices.

Capital:

Education capital services input is measured using the consumption of fixed capital series, which represents the annual flow of services provided by the capital stock (includes buildings) within the education system each year. This was sourced from Volume Index of Capital Services from the ONS.

Outputs

The education output measure was measured using cost-weighted activity indices of full time equivalent (FTE) student activity data gathered on the following publicly funded educational sectors (includes academies in England) :

- Pre-primary school; and
- Primary from their local authority (LA)-maintained counterparts; and
- Secondary from their local authority (LA)-maintained counterparts; and
- Special schools from their local authority (LA)-maintained counterparts; and
- Further education (FE): colleges, sixth form, apprenticeships.

Quality-adjusted Output

The education output measure in primary and secondary schools were adjusted using the:

1. average point score (APS) per student in GCSE level examinations which are normally taken during the student's eleventh year of schooling;

- During the years impacted by the coronavirus pandemic, the National Reference Test (NRT) is used as a quality adjustment metric for education. This serves as a suitable measure of academic performance at the GCSE level. The NRT is a brief assessment designed to mirror the type of questions typically encountered in GCSE Mathematics and English examinations
- 2. disadvantage pupil attainment gap index; and
- 3. bullying quality adjustment from British Household Panel Survey, which was seen to have an impact on the mental, social and emotional disorders over life and can have devastating effect on pupils learning and fulfilling their potential.

1.2.2 Denmark⁶

Institutional Setting Statistics Denmark (2014) treated primary and secondary education differently with higher and further education in the productivity index estimation. Public non-market education was produced in the following levels of education:

1. Primary education
2. Secondary vocational education
3. Higher education
4. Adult education

Inputs The following were the input measures used in the productivity estimation:

Labour input:

Compensation of employees, which reflects the labour intensity in the general government production

Intermediate Consumption:

Purchase of goods and services

Capital:

Consumption of fixed capital (ICT equipment, other machinery and equipment, building capitals, intellectual

⁶ Statistics Denmark (2014)

property products), the most capital-intensive activity identified was on public administration

Outputs

In the education sector, non-market production was derived from the following groups:

- 1) Primary and lower secondary education
- 2) Post-secondary and other mid-length education and preparatory schools for tertiary level
- 3) Tertiary education
- 4) Education not definable by level
- 5) Education NEC

Volume indices will be calculated for each level.

- The volume of primary education is estimated using student-hours.
- Although student-hours a good output measure, there is no available data on this at the secondary education, thus, number of students (based on full-time equivalent) was used as the measure of output for this education level.
- At the higher and further education, the number of students (FTE) was used as an output measure considering that this level of education was organized differently, number of lessons provided to students is smaller and may vary depending on the area of education

Volume data for the four indices were all sourced from Denmark's Education register.

While market prices on the educational services is not available, the government subsidized individual educational institutions published each year in the annual Danish budget. These fees were the best indication of prices for educational services and was used as weights for all types of education, except for public primary and lower secondary education.

Quality-adjusted Output

The volume of output was adjusted for quality using the following:

- 1) Class size (student/teacher ratio), which maximizes student's ability to learn as it is believed that small classes give more benefits than larger ones.

- 2) Teacher/student time provides information on time spent by teachers with each student providing feedback.
- 3) Teachers' qualifications, proportion of teachers with a relevant education
- 4) Project/teamwork, share of time students spend on projects and teamwork.
- 5) Marks or examination results
- 6) PISA competencies on reading, math and science
- 7) Dropout rate

1.2.3 Education (Malaysia)

Institutional Setting Education (Kementerian Pendidikan) public services include the following levels of education:

1. Administrative Education Services
 - a. Public pre-primary education
 - b. Public primary education
 - c. Public General Secondary Education
2. Education
 - a. Technical and Vocational Education below the level of higher education
 - b. Other education

In the estimation of the education sector productivity, the public service programmes were mapped with the activities and the output then quality adjustment for the output measure was identified.

Inputs *Labour input:*

The labour (number of teachers and staff) for which number of employees was identified as the direct input while emolument as indirect input.

Intermediate Consumption:

Intermediate consumption including consumption on supplies and services.

Capital:

Consumption of fixed capital

The volume input index was computed using relative cost weights such that weights were based on the respective share of each component to the aggregate expenditure.

Outputs The output indicator identified were student enrolment in pre-primary, primary and secondary education, matriculation, and institute of teacher education.

Output index was computed using the chain-link or Laspeyres Index method.

Quality-adjusted Output The volume for output index was adjusted for quality using:

1. Student examination performance in Primary School Achievement Test, also known as Ujian Pencapaian Sekolah Rendah (primary school); and
2. Student examination performance in Malaysian Certificate of Education examination or the Sijil Pelajaran Malaysia (secondary school)

1.2.4 Education and Training: (New Zealand)⁷

Institutional Setting The school productivity measure was limited to the primary and secondary education (State and State Integrated Schools). A separate productivity measure was done for the early childhood education.

Inputs *Labour inputs:*
Teachers FTEs were used as the labour input. Teacher categories include Principal, Management, Teacher, Resource Teachers, Community Education, Guidance and Therapists.

Adjusted Labour inputs:
The real salaries was used to adjust the labour input for quality. Adjustment of the labour input to reflect the variations in the hours of work (e.g., full and part time) and the skill composition of the labour force (to account for the changes in the composition by sub-categories such as specialist teachers and teacher aides).

Total input:
The school total revenue was used as an indicator for total inputs. Revenue was based on the Core Crown expenses⁸ and the percentage share of revenue from Government deflated to real terms using the full CPI.

⁷ Gemmell, N., Nolan, P. & Scobie, G. (2017)

⁸ Core Crown Expenditure cover roll-based operations funding to schools, teacher and management salaries, support costs and supplementary funding programmes.

Outputs	Output measure was defined as the number of pupils.
Quality-adjusted Output	The number of pupils as output measure was adjusted for pupil attainment or performance of New Zealand students in international testing:

1. Aggregate Programme for International Student Assessment (PISA)⁹ points, which was obtained by multiplying the average score by total number of pupils.
2. Aggregate Trends in International Mathematics and Science Study (TIMMS)¹⁰ points, which was obtained by multiplying the average score by total number of pupils.

The following were some caveats noted in the use of the attainment measures as adjustment for quality:

- a. The PISA and TIMMS attainment scores are used to adjust total school education output but may not be representative of attainment across the broader school curriculum as this are limited to 15-year-olds.
- b. Scores may be attributed with other factors such as demographic and socio-economic background, parents' education, and their engagement to education among others.
- c. PISA and TIMMS cannot provide progress of the student on different assessments.

The measures were also adjusted for final outcomes to consider the school leavers in the labour market. The adjustment includes:

1. Domestic attainment of students; and
2. The average real expected income for students based on this attainment was then estimated and multiplied by the number of students in each category.

1.2.6 Early Childhood Education: (New Zealand)¹¹

Institutional Setting	Early Childhood Education is provided through playcenters and te kohanga reo, ¹² which are predominantly under the volunteers and home-based services. New Zealand's ECE is funded by the government based on child hours.
------------------------------	---

Te Kohanga Reo was excluded in the estimation given the constraint on the availability of available data on teachers since this only rely predominantly on volunteers.

⁹ PISA scores refer to the scores on reading, mathematics and science for 15-year olds

¹⁰ TIMMS scores refer to the average score for mathematics in Year 5 and so are based on performance at primary school.

¹¹ Green, Nicholas (2017).

¹² Kohanga Reo is an early childhood education and care (ECE) centre where all education and instruction is delivered exclusively in te reo maori (Maori language). Kohanga reo is a home-based service where parents are involved in the children's' development and learning as well as the management, operation and everyday decision making for the Kohanga Reo.

Inputs	<i>Labour inputs:</i> - Teachers Numbers index: Part-time staff numbers for home-based * 0.5 + part-time staff numbers for teacher-led services * 0.5 - Playcentre adults' numbers index: $\frac{\text{Playcentre adults numbers} \times \text{average weekly hours of duty}}{35 \text{ hours}}$ - Aggregated labour inputs: teachers numbers index + Playcentre adults numbers index <i>Financial Inputs:</i> Total government expenditures on ECE deflated using the CPI: 1) Vote¹³ education subsidy and equity funding 2) Vote education review office 3) Vote Social Development
Outputs	Output measure was defined as the funded child hours by service type: 1. Education and care 2. Kindergarten 3. Home-based 4. Playcentre
Quality-adjusted Output	Suggested Quality adjustment measures are: 1) Child development resulting to ECE participation – this requires assessment on the child's cognitive, attitudinal and social competencies before and after their entry to school. 2) Parental employment resulting to ECE participation (employment status of the primary carer) – data matching between parental (or primary caregiver) employment status and child's ECE participation is required.

1.2.7 Universities: The Netherlands

Outputs	Output measure was the equivalent full-time students (EFTS) reflecting both teaching and student time.
Quality-adjusted Output	Case-mix adjusted approach was used to adjust quality of the output based on the quality of the inputs and/or the nature of services.

¹³ Vote is a grouping of one or more appropriations for output class, capital contributions or benefits and grants that are the responsibility of a Minister of the Crown and are administered by the Departments.

Government subsidies for teaching and domestic and international student fees were used as proxies for the complexity of courses taught by the university.

Domestic price weight= (single-university average domestic EFTS fee & subsidy/ - university average

International price weight = (single-university average international EFTS fee/all-university average)

Case mix weight= (domestic EFTS * domestic price weight) + (1-domestic EFTS) *internal price weight

Case mix-adjusted EFTS = EFTS * case weight

1.3. Social Protection: Denmark¹⁴

Institutional Setting Danish social protection focused on residential care activities and social work activities without accommodation (Statistics Denmark, 2014).

Inputs

The following were the input measures used in the productivity estimation:

Labour input:

Compensation of employees, which reflects the labour intensity in the general government production

Intermediate Consumption:

Purchase of goods and services

Capital:

Consumption of fixed capital (ICT equipment, other machinery and equipment, building capitals, intellectual property products), the most capital-intensive activity identified was on public administration

Outputs

The volume index was calculated by the total number of people receiving care and the unit cost drawn from COFOG.

Quality-adjusted Output

The output measure was adjusted for quality using the following quality indicators:

Quality indicator for elderly care: quality of help based on the result of the satisfaction survey.

Quality adjustment for special institutions and homes:

1) Children per adult

¹⁴ Statistics Denmark (2014)

- 2) Children in foster care and orphanages and their highest obtained degree and their exam results over time.
- 3) Personal, social, physical, emotional and behavioral development over time
- 4) Labor market commitment how well are they performing in the labor market over time.
- 5) Socioeconomic and demographic development.

1.4. Libraries: *The Netherlands*¹⁵

Institutional Setting In 2015, a legal framework for the public libraries was passed through the Public Library Facilities System Act, which enumerates the libraries functions from merely making knowledge and information available and providing opportunities for learning and a venue to promote reading to a more expanded function of performing other social functions such as organizing meetings and adoption of digitalization through the national digital library.

Inputs Statistics Netherlands used the COFOG in the identification of the inputs for the estimation of productivity. The following nominal inputs were calculated using the following transactions sourced from the Statline at CBS:

- a. Intermediate consumption deflated using the price index of the material consumption of the government.
- b. Compensation of Employees, price of labour was estimated using direct measures of employees' volume.
- c. Taxes on production deflated using CPI.
- d. Depreciation of fixed capital deflated using the price index of gross government investments; and
- e. Less subsidies on production deflated using CPI.

Aggregation: Geometric mean aggregation where exponents are simply the average of their share in the total cost for the current and past year

Note: The Index of Price Labour can be further refined using the education and skill level as weights to reflect the specific characteristics of the workforce.

Outputs The following output indicators were used to assess the expanded functions of the public libraries:

1. To make knowledge and information available the indicators on loans (both physical and digital loans), and studying facility were identified.

¹⁵ Maroni (2021)

Loans

Physical Collection (x1000)

Physical Loans (x1000)

Loans, download and use of digital materials (x1000)

Digital Collection (x1000)

Studying Facility

Annual Visits (x1000)

2. To assess productivity on the performance of social functions, a number of activities organized was identified as indicator.

Organizing Activity

Number of informing activities

Number of learning activities

Number of reading activities

Number of meeting and debates

Number of art and culture activities

Quality-adjusted Output

Loans data was adjusted by their quality taking into consideration the changes in the variety of physical collections and active licenses.

Aggregation of the Quality adjusted volume of loans using shares in the total number of annual loans

$$L_i = S_i^p \left(PL_i \frac{COLL_i}{NLib_i} \right) + S_i^d (DL_i * LIC_i)$$

S_i^p = share of physical loans to total loans

S_i^d = share of digital loans to total loans

PL_i = annual volume of loans

$COLL_i$ = total physical collection

$NLib_i$ = number of active libraries

DL_i = annual digital loans

LIC_i = active digital license

Aggregation

With the limitation on the available cost structure that can be used to better understand how libraries invest their resources or users' marginal valuation, determining the right weight to attach to each indicator became a challenge.

Sensitivity analysis was done using different combinations of weights for the different dimensions considering the functions of the libraries. In The Netherlands, each year, different combinations of weights were employed to obtain output estimations, which are then divided by the index representing inputs to calculate productivity.

The following scenarios were used in the aggregation of the output measures:

Scenario 1: (w_Loans – 80%, w_visits- 10%, w_activities-10%)

Scenario 2: (w_Loans – 60%, w_visits- 10%, w_activities-30%)

Scenario 3: (w_Loans – 40%, w_visits- 20%, w_activities-40%)

Substantial differences in the productivity level were observed using these three scenarios. Comparing 2015 and 2018, productivity level increased by 24% using Scenario 1, while productivity level increased by 59% using Scenario 3.

1.5. Public Order and Safety: UK¹⁶

Institution- al Setting	The public order and safety service area covers the following: 1. fire 2. courts, which itself has five further sub-components: magistrates' courts, county courts, Crown Courts, Crown Prosecution Service, legal aid 3. probation 4. prisons
------------------------------------	--

The police service was excluded in the public order and safety area, as it has a separate productivity measurement.

Inputs	The public order and safety volume of inputs series is a weighted combination of fire, courts (including probation) and prison (chain-linked using the UK National Accounts expenditure weights).
---------------	--

Labour inputs:

- *courts and probation:* Measured using current price expenditure on labour deflated by the Index of Labour Costs per Hour (ILCH)
- *fire and prisons:* Measured using deflators constructed by combining salary changes for key occupations, weighted by their share of total staff headcount or full-time equivalents, based on salary data from the Annual Survey of Hours and Earnings (ASHE)

Intermediate consumption inputs:

- Measured using current price expenditure on goods and services, deflated by a GDP-implied deflator.

Capital inputs:

- Measured using the current price expenditure on the consumption of fixed capital, deflated by a constructed for local and central government public order and safety capital.

¹⁶ ONS (2022)

Outputs

Fire output activities

Output of the fire and rescue service (FRS) which encompasses fire response, fire prevention, and fire special services is estimated using direct output methods

Direct output measures are used for all the categories under the. Cost-weighted activity index (CWA) is constructed for all the components of the public order and safety service area, using on the Economic Cost of Fire estimates as weights.

Fire response services: This includes fire response involving dwellings, commercial premises, vehicle, chimney, and false alarms.

- Number of incidents attended from Home Office data

Fire prevention: Activities include inspections, investigation, and community safety such as example fitting fire alarms.

- Number of hours of workload from Home Office data.

Special services response: Covers responses to includes the road and non-road incidents

- Number of incidents attended from Home Office data.

Courts Output activities

Direct output measures and forecasting are used to measure the output of the courts.

Separate cost-weighted activity indices for different areas of the courts system are constructed and then further aggregated based on expenditure shares.

Magistrates' courts: This output is measured based on the completed proceedings in 14 case types, namely (1) adult indictable and triable either way offences, (2) all breaches and revocations of sentences, (3) non-motoring offences, (4) motoring offences, (5) all youth crime, indictable and summary, including breaches (6) Section 8 orders in Children's Act (private law), (7) emergency protection orders on children, (8) Children's Act care proceedings (public law), (9) all other family proceedings and Children's Act cases, for example, adoptions and financial cases, (10) licensing applications heard by licensing committees, (11) licensing applications heard by magistrates in petty sessions, (12) any other civil complaints, made to obtain an order, for example, dangerous dogs, (13) all means enquires with defendant present, and (14) all applications for legal aid granted or refused. The weights are calculated from large samples of cases and reflect the average time required to complete each type of case to provide an overall unitary value of caseload.

- Number of magistrates caseloads from the Ministry of Justice

County courts: Growth rates in the county courts are predicted from the activity series within the civil and family courts data. Predictions are informed by historic case data that is no longer available.

Crown Courts: Crown court cases encompass the actual committals for trials, actual cases for sentence, and actual appeals. Historic data on unit costs are used as weights for the output index.

- Number of crown court cases from the Ministry of Justice

Family Courts: This covers private, public, divorce and adoption cases. Hearings and final orders are used to produce a weighted caseload for courts.

- Number of family court caseloads from the Ministry of Justice

Crown Prosecution Service: The output measure for the crown prosecution service (CPS) is estimated by predicting the growth rate in this area using the magistrate's courts and crown courts. Indices are aggregated based on their expenditure shares to approximate the growth in activity for the CPS.

Legal aid: To construct the cost weighted activity index for legal aid services, the data on legal representation expenditure and fees such as lower standard fees, higher standard fees, non-standard fees and exempt cases, and second claims for deferred sentencing are used. The data for the output are sourced from the legal aid agency in England and Wales only.

- Prisoner population, and
- Number of people under supervision caseload

Probation Output Services

The probation output services include (1) offenders supervised by the probation service at end of period, under court orders and pre- and post-release supervision, and (2) a person counted only once in the total even if they were subject to several types of sentences at the year-end. This only covers England and Wales.

- Number of offenders supervised by the Probation Service

Prisons Output

- Average number of prisoners in UK prisons

Quality-adjusted Output

Of the four components within the public order and safety service area, fire protection is not quality-adjusted, while court services are partially quality-adjusted, with county courts being the only component excluded from the adjustment. These two services are not quality-adjusted because they primarily handle civil cases.

The criminal justice components of the court such as magistrates courts, crown courts, and crown prosecution are quality-adjusted based on:

- (1) *recidivism adjustment* approximates the effect the CJS has on reducing the volume and severity of further crimes being committed by those who have gone through it – this being an important social outcome for the system
- (2) prison safety adjustment relates to the number of incidents of assaults, self-harm and deaths that occur in prison custody
- (3) custody escapes adjustment is measured based on changes in the difference between the number of escapes and a baseline of 0.05% of the England and Wales prison population – a historic target used by the Ministry of Justice, and
- (4) courts' timeliness adjustment are different for each courts:
 - a. Magistrate court adjustment measure is based on the mean average time of charge and laying of information to completion
 - b. Crown Court adjustment measure captures the average waiting times experienced by all defendants and the mean time from main hearing to completion.

Legal aid is quality adjusted based on case type and fee types.

Probation is quality adjusted based on the number of serious offences committed by offenders on probation

Aggregation To aggregate all components of public order and safety, including those without quality adjustments, the components are cost-weighted to create a composite index of quality-adjusted output for public order and safety. For components with multiple adjustments, the growth factors are combined using a weighted average approach, as illustrated below:

Quality adjustment weights

	Re-offending weight	Prison safety weight	Prison escapes weight	Courts timeliness weight
Fire-protection Services	Non-quality adjusted			
Magistrates Courts	50.0%			50.0%
County Courts	Non-quality adjusted			
Crown Courts	50.0%			50.0%
Crown Prosecution Service	100.0%			
Legal Aid	100.0%			
Probation	100.0%			
Prisons	29.2%	37.5%	33.3%	

Source: Office for National Statistics

1.6. Fire Department: The Netherlands¹⁸

Institutional Setting The Fire Department follows a decentralized structure and works in cooperation with other organizations to perform public order and safety services, fire services, disaster and crisis management and medical assistance. This nature made it challenging to measure the productivity of the sector.

Inputs Statistics Netherlands used the COFOG in the identification of the inputs for the estimation of productivity. The following nominal inputs were calculated using the following transactions sourced from the Statline at CBS:

- a. Intermediate consumption deflated using the price index of the material consumption of the government.
- b. Compensation of Employees, price of labour was estimated using direct measures of employees' volume.
- c. Taxes on production deflated using CPI.
- d. Depreciation of fixed capital deflated using the price index of gross government investments; and
- e. Less subsidies on production deflated using CPI.

Issue on direct measure of labour given the case of volunteers' contribution.

Outputs The output of the fire department relates to fire prevention, fire intervention, and other interventions.

The prevention function of the fire department aims to proactively reduce the number of incidents and damages such as periodic checks to some detectors or sprinkler systems in public facilities and houses, however there is lacking information on this area. Further, the quantification of the effectiveness of the prevention measure may not be immediately apparent or might become evident over the long term. Since the fire prevention activity with an objective to reduce the number of fire events is interconnected with the nature of fire interventions, productivity measure can focus on the latter.

FIRE INTERVENTION and OTHER INTERVENTIONS

1. fire incident (no.)
2. other type of incidents requiring the intervention of a fire brigade (no.)
3. average response time (min)
4. Fire fatalities (no.)
5. Fire injuries (no.)
6. Average fire claim to insurers (Euro)

Quality-adjusted Output To account for the quality adjustments, impact on monetary impact on life losses, injuries and economic damages on real estate properties were considered.

Quality adjustment on Fire interventions (QAFI):

This was calculated as the ratio of the fire intervention volume indicator (FI) divided by the annual average monetary fire damages (DMG). The latter was a combination of the average claim for fire incidences, the annual number of fire victims times a monetary value, and the annual health cost of fire injuries.

$$QAFI_i = \frac{FI_i}{DMG_i}$$

Quality adjustment on other interventions index (QAOI):

A monetary approach to account for quality changes for other interventions was a challenge given the diverse intervention on other activities of the fire department. Thus, the quality adjusted index was just the ratio of the volume indicator (OI) and the average response time (RT).

$$QAOI_i = \frac{OI_i}{RT_i}$$

It was assumed that both dimensions on fire intervention and other dimensions of fire services have the same value to the citizens. Thus, the two dimensions' volumes were used as weights to combine the two indices:

$$wFI_i = \frac{FI_i}{FI_i + OI_i}$$
$$wOI_i = \frac{OI_i}{FI_i + OI_i}$$

To estimate a single output measure (Y_i), a weighted sum of the quality-adjusted fire intervention and quality adjusted other interventions was computed using the equation below:

$$Y_i = wFI_i * QAFI_i + wOI_i * QAOI_i$$

Aggregation Total output measure was aggregated by weighting the intervention-related side and the prevention activities of the fire department.

$$TY_i = wI * Y_i + wP * P_i$$

Since the weights for the prevention and intervention are unknown, sensitivity analysis was used where different combinations of weighting factors were applied to the functions of the fire department with the limitations on the cost information.

There were five scenarios that were explored to estimate the total output measure for fire services:

- Scenario 1 - Considers only the fire intervention tasks given the lack of data on fire prevention.
- Scenario 2- considers only fire intervention task, however, number of fatalities was estimated as the average of the fatalities of the year of interest and two years before
- Scenario 3 – assumes that quality-adjusted prevention activities output decreases by 2% per year. It also has three weight scenarios such that prevention can have 25% or 50% or 75% share to total output of fire activities.
- Scenario 4 - assumes that their stable number of inspections per year. It also has three weight scenarios such that prevention can have 25% or 50% or 75% share to total output of fire activities.
- Scenario 5 - assumes that there is increasing adjusted prevention intervention at 2% per year. It also has three weight scenarios such that prevention can have 25% or 50% or 75% share to total output of fire activities.

Using these different scenarios TFP registered high variability and change from -5% to 8% from 2013 to 2019. It was also argued that number of fires may decrease over time due to the increasing effort on prevention, which would make the Scenario 5 more likely.

1.7. Courts: The Netherlands and New Zealand

1.6.1 The Netherlands¹⁷

Institutional Setting The Courts are responsible for truth-finding and conducting trials to settle conflicts. There are several levels of courts that exist in The Netherlands with different functions in terms of the delivery of justice. The following are the different levels of court system:

1. District courts, which handle cases concerning civil law, criminal law, cantonal matters and administrative law in the first instance.
2. Courts of Appeal, which deals with appeals in criminal, civil, cantonal and tax matters.
3. Supreme Courts, which is the highest judicial instance in The Netherlands that has the power to determine that a case is closed or has to be reopened.
4. Board of Appeals of Business¹⁾ handles cases concerned with business matters such as the Competition act.
5. Raad van State ('Council of State') is the highest instance for administrative cases.
6. Centrale Raad van Beroep ('Central Board of Appeals') is responsible for higher appeal cases concerning matters such as study financing, social security law and civil servant's law.

¹⁷ Gerlach (2021)

Inputs	Using the COFOG classification, the following were the identified input variables. Intermediate consumption (M euro) Compensation of employees (M euro) Other current expenses (M euro) Capital expenditures (M euro) Total employees (headcount) FTE
Outputs	The output measures was obtained by observing the number of cases handled per case category and can be aggregated using the cost weights through the corresponding monetary compensation, which was sourced from the annual report of the courts. Cases Category: 1) Criminal cases 2) Civil cases 3) Administrative Cases 4) Immigration Cases 5) Cantonal Cases 6) Tax Cases Appeals: 1) Civil Law higher appeal 2) Criminal law higher appeal 3) Tax cases higher appeal 4) Central Board appeal
Quality-adjusted Output	Implicit quality adjustment was done by stratifying case categories. To adjust the quality for the courts, the number of held appeals for each category was used to reflect the quality of verdicts.
Aggregation	With the availability of cost information, weights used in the aggregation was based on the realized prices that accurately reflect the handling times, workload and seriousness of each case category. The volume measures of handled cases of the different case categories were aggregated using the previous year's cost share in the total cost.

1.6.2. New Zealand Case¹⁸

Institutional Setting	New Zealand has four levels of courts, the Supreme Court has been the highest court in New Zealand, the Court of Appeal, the High Court, and the District Court.
Inputs	The following were the inputs in the public sector productivity measurement: 1. Labour input, which includes the people directly involved in the production such as the judges and the people who contribute to the functioning of the court system such as the administrative and security staff. 2. Capital expense in the production of the output. 3. Intermediate goods and services: other goods or services consumed as inputs in the production of the output, for example electricity.
Outputs	Output measure was the case disposals or the final outcome of the case, excluding cases for appeals.
Quality-adjusted Output	Implicit quality adjustment was done by stratifying the case disposals by type of court.

1.8. Public Prosecution Service: The Netherlands²⁰

Institutional Setting	The Dutch public prosecution office engages in several activities such as prosecuting criminal offences and bringing suspects to a criminal court, dealing with criminal offences without the intervention of a judge, and leading the police in detecting criminal offences, which should all be considered to measure the output of the institution. The TFP analysis only covered the following tasks of the public prosecution service: 1) bringing suspects before a court 2) handling cases without the intervention of a judge 3) investigating behaviours that could harm the integrity of the government or suspected criminal offenses conducted by other civil servants (Rijksrecherche investigations) 4) Imposing mandatory care on someone who is possible endangering himself or other (Wet Bopz activity due to the law)
------------------------------	--

¹⁸ Genet, Terry (2017)

Inputs Using the COFOG classification, the following were the identified input variables:

Intermediate consumption (M euro)
Compensation of employees (M euro)
Other current expenses (M euro)
Capital expenditures (M euro)
Total employees (headcount) FTE

Outputs Public prosecution service ensures that criminal offenses are detected, prosecuted and offenders will be charged. To measure the output of the prosecution service, the number of cases brought to trial, cases solved, investigations and other duties of the prosecution service will be aggregated.

Types of cases

- 1) Number of felony cases at court
- 2) Number of Administrative cases at court
- 3) Number of higher appeals at court
- 4) Number of traffic fine as appeals at court.
- 5) Number of felony cases handled without intervention of judge.
- 6) Number of administrative cases handled without judge.
- 7) Number of Investigations
- 8) Other duties

Challenges:

- 1) Possible time lag between the inflow of cases and the date of trial.
- 2) Lack of available data for certain outputs particularly on the missing cost structures

Quality-adjusted Output To adjust the output for quality, the ratio of acquittals to the total cases brought to trial was used to consider not just the volume of cases. The quality adjusted measures were limited to felony and administrative offenses considering the limitations on the availability of data.

Ratio of acquittal to the total cases brought to trial.

$$QCases_t = C_t * \left(1 - \frac{AC_t}{C_t}\right)$$

C_t = number of felony/admin cases brought to the court

AC_t = number of acquittals

Aggregation of the cases

Case 1. Aggregated cases brought to the court with intervention from the judge:

Felony cases will be weighted using the price received for criminal case and the compensation received by the courts.

Administrative offences and appeals will be weighted based on the price received for these cases.

Weights will be weighted based on the share of each case categories' cost to the total cost in the previous year.

$$TA_t = \sum_i^N QAC_{it} * wQAC_{it-1} + \sum_j^N Appeals_{jt} * wAppeals_{jt-1}$$

TA_t = aggregated activities task

QAC_{it} = quality-adjusted case i in year t, that is, i=felony or administrative cases

$wQAC_{it}$ = realized prices received per case category i in the previous year t-1, that is, i=felony or administrative cases

$Appeals_{jt}$ = number of higher appeal case j in year t, that is j=higher appeals or appeals on traffic fines

$wAppeals_{jt}$ = realized compensation received per case category j in the previous year t-1, that is, j= higher appeals or appeals on traffic fines

Case 2. Aggregated cases brought to the court without intervention from the judge:

To aggregate the cases without intervention from the judge, the number of felony cases, administrative cases and traffic fines that does not need judge intervention will be added and weighted using the compensation received by the courts for the respective case category in total costs.

$$OMt = Ft * wFt-1 + At * wAt-1 + TFt * wTFt-1$$

OMt = aggregated index of cases without judges' intervention

Ft = number of felony cases without judge intervention

At = number of administrative cases without judge intervention

TFt = number of traffic fines without judge intervention

Aggregation

With the limitation on the availability of data on the marginal valuation of the different outputs by the consumers and the cost shared to the different activities, weights used were identified arbitrarily, such that public prosecution function on bringing suspects before the court (Task 1) and handling cases without the intervention of judge (Task 2) were given with the largest weights while the two remaining tasks – criminal investigations and other duties received small weights.

The following weighting scenarios were used in the TFP analysis:

1) Scenario 1 :

$$w(\text{Task1}) = 60\%, w(\text{Task 2})=30\%, w(\text{Task 3})=5\%, w(\text{Task 4})=5\%$$

2) Scenario 2:

$w(\text{Task1}) = 30\%$, $w(\text{Task 2})=30\%$, $w(\text{Task 3})=30\%$, $w(\text{Task 4})=10\%$

Scenario 1 results to smaller decrease in TFP as compared to the decrease in Scenario 2, attributed to the increase in investigation activities from 2015 to 2019.

1.9. Police Service: The Netherlands and New Zealand

1.8.1 The Netherlands²⁰

Institutional Setting The Dutch Police is organized in ten regional units and one central unit tasked to provide a safe and pleasant environment for Dutch citizens to live in, protect the democracy and enforce Dutch laws.

Inputs Using the COFOG classification, the following were the identified input variables concerning the police:

Intermediate consumption (M euro)
Compensation of employees (M euro)
Other current expenses (M euro)
Capital expenditures (M euro)
Total employees (headcount) FTE

Outputs Measuring output of the police sector has been a challenge considering that the task of the police on ensuring public order and safety and eliminating criminal activity depends on several factors and not solely on police output. Thus, it is a challenge to identify the direct contribution of the police tasks to the changes in criminal activities.

Statistical Netherlands highlighted several studies such as Lind and Lipsky (1971), Zouridis (2015) underscored the difficulty to identify the citizen valuation on the outputs of the police work because citizens derived utility from the outcome – crime reduction and keeping a safe environment for everyone.

Thus, output measure covered indicators related to the police's tasks can be primarily on (1) crime control (2) crime prevention (3) emergency management and (4) maintaining public order.

Crime control

Number of registered crimes
Number of solved crimes
Number of investigations

Crime prevention

Police to population ratio – ratio of the number of patrol officers and the Dutch population by 5,000 inhabitants

Number of actionable emergency Calls – this corresponds to the number of police operations in response to the emergency calls.

Maintaining Order

Presence in activities such as demonstrations and festivals

Quality-adjusted Output

Quality-adjustment:

There were different weighting procedures was identified in aggregating the cases:

1. Weighting based on the share of cases of each category in total cases/convictions, higher number of cases or convictions will have higher weights.
2. Weighting based on the social cost of crime, example, monetary value of the stolen property, cost to health, among others, although there is a limitation on the availability of social cost for other categories.

Implicit quality adjustment was adopted by weighting the different crime categories instead of using explicit quality adjustments due to challenges in conceptualizing the outputs of the police. The discussion of quality is conceptually complex on the police particularly with regards to the changes in societal outcomes caused by the police and the inability to link changes in outcomes directly and unambiguously to police work.

Aggregation TFP Sensitivity Analysis was using different combinations of weighting factor considering the functions of the police. There were identified panels of analysis in terms of measuring the TFP using different weights:

1) Panel A considers the crime control indicators only. The output series were computed using the social cost adjusted weights. The different crime categories are aggregated according to their previous year's cost shares in social costs.

2) Panel B presents the TFP using the 2 combinations of dimensions, that is, (a) combination of the crime control and prevention activities dimension and (b) combination of crime control and emergency management dimension. Weights were identified with arbitrariness such that one combined dimension can have 25%, 50% or 75% weights and the other combined dimension can have the complement to add to 100%.

3) Panel C covers the combination of the three dimensions (a) solved crimes, (b) prevention, and (c) emergency management. There were different weight combinations used to calculate TFP using this Panel:

a. Option 1 gives equal weights to all three dimensions.

b. Option 2 assumes crime control has the bigger weight in the police tasks with 50% weight while prevention and emergency management have 25% weight share each.

c. Option 3 assumes that 15% of the police tasks was on solving crime while prevention and emergency management were each allotted with 42.5% weight.

In Panel C, productivity of police decreases but large decreased was observed using Option 2 contributed by the large decrease in the crime control dimension because of the shift in focus of the police from crime control to other dimensions.

1.8.2 New Zealand¹⁹

Institutional Setting	The productivity measure for New Zealand's police was estimated partially on a specific area of police activity particularly the responses to mental health incidents as policing services for people with mental health conditions was identified as a particular area of focus of the New Zealand Police in their current 4-year plan.
Inputs	Number of police hours responding to mental health, threatened and attempted suicide incidents.
Outputs	Total number of mental health, threatened and attempted suicide incidents weighted using the average number of hours spent on mental health, threatened or attempted suicide incidents.
Quality-adjusted Output	No quality adjustment

C. Public Sector Productivity Measures of Asian Productivity Organization (APO) Member Countries

APO member countries measured public sector productivity for tax collection and passport issuance employing Dunleavy and Carrera (2013) as the main reference point to ensure consistency across different countries. Dunleavy and Carrera (2013) stressed that there is a need to cover the full range of activities performed by an agency to better estimate productivity. Organizational charts of tax collection and passport issuance was also used by countries to better understand the procedure of the public service delivery. Table 2 presents the countries that measured productivity in these two public services. Note, however, that

¹⁹ Terry Genet and Mike Hayward (2017). The New Zealand Productivity Commission Research Note 2017/09: Productivity measurement case study: Police

there are countries with limitations on estimating productivity for the passport services because of the restrictions on availability of data brought about by security concerns.

Table 2. APO Member Countries that Estimated Public Sector Productivity

Country	Tax Collection	Passport Issuance
Bangladesh	✓	✓
Indonesia	✓	✓
Iran	✓	
Pakistan	✓	
Philippines	✓	
Sri Lanka	✓	✓
Thailand	✓	

Source: APO (2016)

In the productivity estimation, countries identified the range of different services delivered by the sector to the public to achieve the outcomes. These services were grouped together based on the unit cost of production to generate average unit cost of different services and service groups. The growth in these services were then aggregated to estimate the growth in total output weighted based on the relative costs of providing a unit of each service type.

APO member countries estimated public sector productivity focusing on the tax collection and passport issuance based on the Dunleavy and Carrera's framework in which total factor productivity is measured by volume of output divided by the volume of total inputs. Some countries such as Philippines, Thailand and Vietnam faced challenges on the estimation of total factor productivity for passport services due to unavailability of data. The following table summarises the estimation strategies adopted by different member countries of APO.

1.10. *Tax Collection Service: Bangladesh, Indonesia, Philippines, Sri Lanka, Thailand*

Country	Institutional Setting	Input	Output	Agency
Bangladesh ²⁰	The main responsibility of the NBR is to collect domestic revenue (primarily, import duties and taxes, VAT, and income tax) for the government. Other responsibilities include administration of all matters related to taxes, duties, and other tax producing fees.	Salaries of full-time equivalent employees (in taka)	Tax collection (in taka) Key taxes: 1) Tax on income and profit 2) Value Added tax 3) Import duty 4) Export duty 5) Excise duty 6) Supplementary duty 7) Other taxes and duty	National Board of Revenue
Indonesia ²¹	The functions of tax administration include the facilitating taxpayers in meeting their obligations, ensuring taxpayers meet their tax obligations, and enforcing law to those that are non-compliant. The agency manages central government taxes such as the income tax, value-added tax, and mining and forestry tax.	Number of full-time employment	Total amount of collected taxes (deflated by inflation)	Directorate General of Taxes (DGT), Ministry of Finance
Philippines	The agency has the following major functions:	Deflated administrative costs (total	Deflated total revenue collection (total amount of	Bureau of Internal Revenue

²⁰ Report written by Dr. Md. Nazrul Islam, Director of National Productivity Organisation

²¹ Report written by Dr. Anwar Sanusi, General Secretary of the Ministry of Villages, Development of Disadvantage Regions, and Transmigration and Dr. Septiana Dwiputrianti, Associate Professor of the School of Administration, National Institute of Public Administration

Country	Institutional Setting	Input	Output	Agency
	1. assess and collect taxes 2. Enforce law to prevent and prosecute tax evasions and all other illegal economic activities,	cost of operations including personal services, maintenance and other operating expenses, capital outlay using GDP deflator) obtained from annual reports	taxes collected for the different types of taxes: taxes on net income and profit, value added tax, excise taxes, percentage taxes and other taxes) No cost weighting used due to the difficulty in identifying the personnel cost for type of tax collected	
Sri Lanka	The following were the agency's functions: 1. Collecting taxes in terms of relevant tax and other related laws, 2. encouraging voluntary compliance 3. deterring tax evasion and tax avoidance	total number of full time equivalent staff allocated to tax processing in the head office and the island wide regional offices	total tax revenue from Income Tax, corporation tax; capital gains tax; inheritance tax; VAT; excise duties; other indirect taxes Cost-weighted using the Unit costs for each tax above, estimated Department of Inland Revenue performance reports	Department of Inland Revenue
Thailand	Three main tax collection agencies under the Ministry of Finance, the Revenue Department, Excise Department, and Customs Department.	Operation cost	Total amount of tax collected	Ministry of Finance, the Revenue Department, Excise Department, and Customs Department

Country	Institutional Setting	Input	Output	Agency
	Revenue Department is responsible for the collection of income tax from personal and corporate, petroleum tax, value added tax, and duty stamp. This is the highest tax collection agency in Thailand. The Excise Department collects excise tax from specific goods and services. The Customs Department has the main responsibility for international tax collection.			

1.11. Passport Service: Bangladesh, Indonesia, Sri Lanka

Country	Institutional Setting	Input	Output	Agency
Bangladesh	The following were the functions of the agency: 1) Issuance of passports 2) Blacklisting of nationals and foreign nations at passport facilities 3) Revocation of passports 4) Granting of visa and permits to	Total number of employed persons	Number of passports issued	Department of Immigration and Passports

	foreign nationals 5) Printing, procurement, supply and distribution of passports 6) Communication for government decisions relating to issuance of passports and visas			
Indonesia	The functions of the Directorate of Immigration: 1. Formulation of policies; 2. Implementation of policies; 3. Preparation of norms, standards, procedures, and criteria in the field of immigration; - and 4. Providing technical guidance and evaluation in the field of immigration.	Total number of employed persons	Number of passports issued	Directorate of Immigration
Sri Lanka	The following were the functions of the agency: 1. Issuing visas to foreign nationals and travel documents to the Sri Lanka nationals 2. Maintain detention places approved by the minister; 3. conduct investigations on violations of	Deflated staffing, outsourcing, procurement, and capital costs published in Administrative Reports of the Department of Immigration and Emigration	Total number of passports issued, obtained from annual reports of the Department of Immigration and Emigration and its predecessor Cost-weighted using the unit costs for different types of passports obtained from Administrative Reports of the	Department of Immigration and Emigration

	provisions made by or under the act;		Department of Immigration and Emigration	
--	--------------------------------------	--	--	--

D. The BLS Productivity Measures²²

The Bureau of Labor Statistics generated labour productivity indices for important state and local government services. There were 10 services examined for state and local government. Five of these services were enterprise-related services – electric power, natural gas, drinking water, mass transit and alcoholic beverages sales. As these are normally outside the boundaries of the public sector productivity measures, we do not include these in the review below. The relevant services that were included in the review were the three services in the correction field – prisons, jails and juvenile institutions and two employment-related services— unemployment insurance and employment service.

1.12. Correctional Services

1.11.1 State Adult Correctional Inputs

Institutional Setting State adult correctional services operate with facilities that are general confinement prisons, operated reception, diagnostic centers, farms, road camps and medical treatment centers.

The productivity measure focused on the State adult correctional institutions. State juvenile operations, adult parole, pardon and probation operations, Federal facilities, and most local correctional operations are excluded.

Inputs The labour index reflects the number of full-time-equivalent employees.

Outputs The output measure used for the adult correctional institutions is the number of inmates incarcerated differentiated by security level.

1.11.2 Local Jails

Institutional Setting Local jails operations in the US include detention centers, county prisons, workhouses, or houses of correction excluding the temporary lockups that house individuals for less than 48 hours.

Inputs The labour index reflects the number of full-time-equivalent employees.

Outputs The output index used was for local government's number of inmates

²² Note: There were no quality adjustments applied to these measures.

1.13. State Employment Security Services

1.12.1 Unemployment Insurance

Institutional Setting The unemployment program was instituted through the Social Security Act of 1935 which aims to provide aid to temporary unemployed persons. It sets out three programmes – for the regular State workers, former service workers and unemployed for Federal workers.

The Unemployment programmes focuses on providing aid to qualified unemployed persons and collection of taxes from employers to finance the programmes.

Inputs The labour input used was the full-time equivalent employee index

Outputs Five measures were used for the UI output:

- 1) Claims processed Index, measures the number for unemployment insurance applications processed by UI staff.
- 2) Weeks compensated Index, measures the aggregated number of weeks for which unemployment insurance is paid during a given period.
- 3) Benefits Index measures the number of individuals who draw unemployment insurance benefits.
- 4) Finance index measures the number of employers for which taxes were collected from.
- 5) Total unemployment insurance programme Index, measures the output associated with service to the unemployed and the finance operations using staff year labour weights to produce a weighted index

1.12.2 Employment Service

Institutional Setting Employment service has three major functions:

1. Labour market information function includes the collection of data on employment, hours and earnings, local area unemployment, occupational employment, insured employment, and wage statistics.
2. The compliance and enforcement function are the responsibility of the national office staff, but State employees perform the services including certification of housing and pay of migrant workers, and documentation and certification of the need for alien workers.
3. Labour exchange services are provided by State Employment Service staff.

Inputs	The productivity index for the employment service was limited to the labor market information, particularly the data collection efforts. The labour input used was the full-time equivalent employee index
Outputs	Service-based index was used to measure output. This includes individual services that were aggregated: 1) Job orders, 2) intake applications, 3) counselling, 4) testing, and 5) referrals.

E. Non-market output measures in the Australian National Accounts

The Australian Bureau of Statistics is currently putting effort into enhancing the measurement of non-market economic activities, focusing on the output measures of major public services such as health and education industries. This initiative to enhance the non-market output measures would help in the improvement of the productivity measures for non-market industries.

ABS (2020) discussed the current progress made on the different non-market economic activities.

a. Hospital services

The output measure focuses on the hospital services that are mainly provided by public hospitals such as the emergency department activity and non-admitted patient care. The interest was in capturing the shifts across different types of care to assess efficiency gain, that is, from admitted care to non-admitted care. The Australian Institute of Health and Welfare generates data on cost weights for these hospital services which can be used in constructing indexes.

However, with the limited information from the hospital systems, there was an identified challenge in terms of tracking patients as this may pose double counting in the source data – those who were treated as an admitted patient and was subsequently treated but as a non-admitted patient should represent one treatment.

In this case, it is currently being studied if it is better to construct separate indices for the different hospital services – emergency department care and non-admitted patient care, aggregated using the admitted care index.

Another challenge identified was the consistency of the scope and coverage of cost data as well as the classifications and coding systems of the hospital reports across time. To address these concerns on the emergency department activity, the study investigates the possibility

of stratifying the Urgency Related Group classifications or regrouping data by triage and International Classification of Diseases (ICD) that is similar to the coding standards and classifications of the admitted care.

On the data for non-admitted patients, two stratification strategies are being explored, that is the use of Tier 2 Independent Hospital Pricing Authority (IHPA) classification or to map it with ICD.

There was also an identified limitation on the availability of expenditure data matched to the level of detailed activities prior to 2009, which posed a challenge in capturing shifts across different types of hospital services prior to these years.

The challenges on the difference in the source data scope and coverage are currently being dealt with by the ABS in their studies focusing on the stratification strategies identified above and developing an output index covering emergency department care and non-admitted patient care aggregated using the admitted patient care index.

b. School education

The output measure for the school education will be limited to all streams of school-based education – preschool, primary, secondary and special needs education delivered by the government schools. The output measure for school education does not include tertiary education as this has separate output measures.

The output index for school education will involve student enrolment data on a full-time equivalent basis and the school expenditure data. The data on the student enrolment by school type, by jurisdiction and by affiliation, which will be used as a quantity variable, is available in the ABS except for the special needs education school type. On the other hand, the data on the school expenditure will be used as the cost weights.

There is a wide range of data sources available in Australia producing expenditure data on school education, however, there is a challenge in terms of the comparability of these data sources, especially for the non-government schools. Thus, ABS explores the possibility of repurposing its survey data particularly in capturing expenditure data for non-government schools.

Another challenge identified is the availability of data on how much each individual year of school education costs relative to other years. This challenge will be experienced in schools that provide continuous education from kindergarten to year 12, for which school education cost would be difficult to break down for primary and secondary education.

In the current study being done by the ABS, modelling of expenditure data by individual school year is being explored to address the issue on the availability of appropriate data. Further, adjusting enrolment data using student absentee rates is being explored to improve the accuracy of the school enrolment data.

c. Tertiary education

Output measures for the tertiary education will focus on the public universities and will exclude private sector universities which are adult, vocational and community education providers.

The volume indicator identified for this industry is headcount of the number of students enrolled at university, with the aim to refine the measure by stratifying by academic discipline; post-graduate students separately from undergraduates; and valuing the non-tuition output of universities, such as research papers and books.

The R&D partnership activities of the Universities are assumed to grow at the same rate as the tuition and research output growth. In this case, the tuition output can be measured using the student load statistics from the Department of Education and the research output can be measured using the count of published research output.

The cost weights for the output index for tuition and research will use the total cost of the university for teaching and research component with the assumption that the ratio of teaching to research staff reflects the university's ratio of teaching to research costs. Stratification by discipline and by programs (postgraduate and undergraduate programs) will be explored by the ABS in their future initiatives.

III. Issues in defining and measuring public sector productivity

The review of NSIs' and Productivity Commissions practice presented in Part II highlights a number of issues which we briefly set out in the next section, as areas for further discussion.

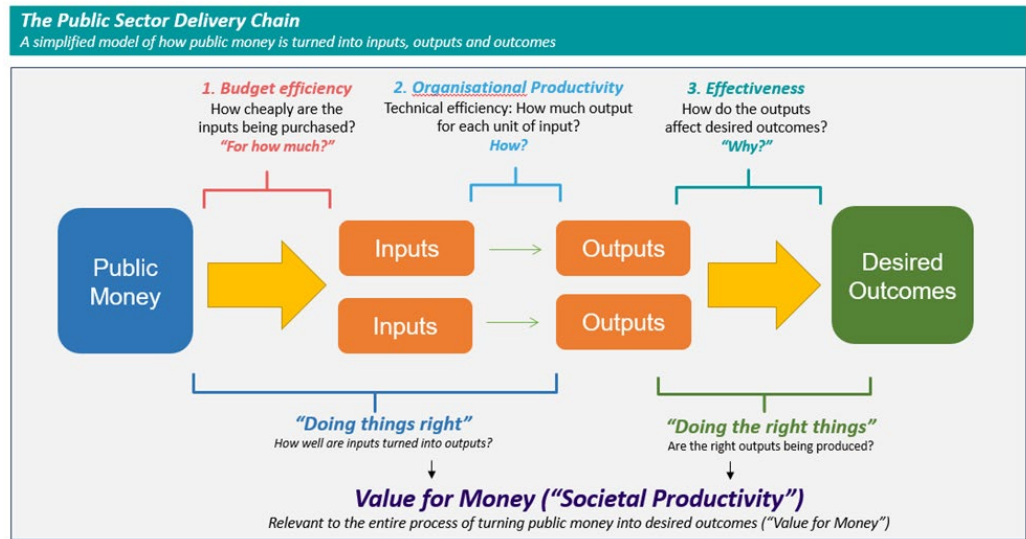
A. The organization of public sector production.

The outputs, inputs and quality adjustments incorporated by NSIs and Productivity Commissions vary by service in the countries examined. This is to be expected as the provision and mode of delivery of services varies across countries. However, it is unclear if these differences also reflect data availability, especially as regards quality adjustments. This suggests the need for a more systematic approach.

In a recent review of productivity in the UK policing service, Van Ark and Hoskins (2024) argue that a first step is to understand the police delivery chain. This is consistent with the approach adopted by the World Bank and the APO, mentioned above. The authors suggest that one of the challenges that make the assessment of the sector's productivity performance difficult relate to the 'multi-outputs' and 'multi-outcomes' nature of police activity. They argue that describing the delivery chain is one way of describing the mix of activities being produced to achieve outcomes. This divides the delivery chain of a public sector activity into discrete functions as illustrated in Figure 2.

Van Ark and Hoskins (2024) argue that "the delivery chain concept facilitates the development of targeted strategies, improvement in the allocation of resources correctly, and the identifying of productivity-enhancing opportunities". However, from a measurement perspective a delivery chain can be useful to visualize how different parts of a service interact to achieve outcomes, and the latter can be used to adjust for quality of outputs.

Figure 2. The Public Sector Delivery Chain.



Source: From van Ark and Hoskins (2024)

Using the concept of the delivery chain, the authors identify core and ancillary functions of the police, as shown in Table 3.

Table 3: Core and Ancillary Functions of the Police

3.a. Core Functions

<i>Core Functions</i>	Input	Output	Outcome
Emergency Response	Rapid Response Vehicles, Conflict management skills, Communication technology	Arrests, emergencies responded to.	Lives saved
Crime Investigation	Investigation skills, Information technology	Arrests, evidence	Criminals arrested, crime deterred
Neighbourhood Policing	Relational skills	Visible Patrols	Legitimacy
Crime prevention	Collaboration with property owners, Information and Communication technology	Arrests, safer behaviour from members of the public	Lower crime rate

3.b. Ancillary Functions

<i>Ancillary functions</i>	Input	Output	Outcome
Crowd Control	Riot gear, co-ordinated control, horses, Information and Communication technology	Police control barriers	Avoided calamities
Missing Person Investigation	Investigation skills, Information technology	Investigations	Missing individuals found
Anti-terror	Intelligence, anti-personnel equipment, Information technology	Arrests and apprehension of terrorists	Reduced terror rate
Organised Crime	Intelligence, Information technology	Disrupted supply chains, arrests.	Dissuaded recruits, reduced economic viability of operations

Source: Core and ancillary functions from The Police Foundation (2022). Input, output and outcomes from van Ark and Hoskins (2024).

The above discussion is illustrative of how an understanding of the delivery chain can inform work on measuring productivity. While data may not be readily available for all parts of the chain, it gives an ideal to work towards and also might feed into the next discussion on weighting. The alternative is measuring inputs, outputs and quality adjustments in a piecemeal and fragmented way, that is driven by what data are available rather than the importance of the particular output or quality adjustment being measured.

Van Ark and Hoskins (2024) also note that policing performance is affected by underfunding and productivity weakness in other public services. The police delivery chain is dependent on the performance of other parts of the criminal justice system, for example, the courts or the prison service. But there are also important challenges for the police dealing with societal issues which are not within the immediate scope of the criminal justice system itself, but nevertheless can use up a significant amount of police time. The most widely reported example is the rising number of mental health incidents, with the need for police to safeguard individuals with mental health problems until they can be taken care of by other professionals. Another example is the well known ‘bed blocking’ issue in hospitals, due to the lack of adult social services in some localities. Therefore, overall public sector productivity may be depressed because some services are produced by bodies which are not designed for this.

B. Weighting

There are a number of weighting issues involved in measure public sector productivity, for specific service areas. On the output side many NSIs and Productivity Commissions use cost weights as these tend to be readily available. There are arguments that value weights should be used instead, for outputs that are not sold in competitive markets. However, the methods to include these are controversial and the data are rarely available. If outcomes are used in measurement, as discussed below, then value weights are appropriate. The use of cost shares is less controversial for inputs as these are generally purchased under competitive conditions. There are likely to be some deviations from this, e.g. the NHS is close to a monopsony for some inputs such as drugs, but arguably this simply offsets the market power of drug companies.

Quality adjustments also require weighting and there is much less of a consensus across NSIs and Productivity Commissions on this. There are two main issues on the weighting of the quality adjustments. The first issue is how to construct the weights. It seems that many NSIs and Productivity Commissions resort to expert judgement, but it is unclear how systematic this is. The second is what overall weight should be on the quality adjustment. If this is one, e.g. as used for dividing between attainment and other adjustments in UK education, it assumes that all quality dimensions are covered which is a very strong assumption. In the Danish healthcare case above the quality adjustment is less than one, to allow for other unknown factors that affect quality. This may then lead to higher jumps in growth rates when new quality adjustments are introduced and a need to revise backwards.

C. Outcome based approaches

The outcomes approach to measuring productivity takes a more holistic approach than normally considered in the national accounts or even satellite accounts. For example, in the case of healthcare, the real societal benefits are not the treatments typically measured by statistical offices, but healthier and longer lives. By not accounting for these benefits, healthcare productivity growth may be substantially understated, and our understanding of its driving forces severely limited.

In a pioneering study of healthcare for the United States, Cutler et al. (2022) propose a condition-based approach which has some similarities with a delivery chain approach. They contrast the change in quality-adjusted life expectancy (QALY) of those with a particular condition with the change in the healthcare costs related to this condition. This does raise issues in handling the link between outputs and outcomes, required in a national accounts setting. A possible way forward is that outputs and outcomes could be measured separately so that quality can be derived as the ratio. See also Dauda et al. (2022) for an alternative outcome based measure for the US.

An outcome based measure has also been implemented for education by Corrado et al. (2021), using the increment to lifetime earnings attributed to education (see also O'Mahony and Stevens (2009) for an earlier study. Corrado et al. (2021) argue that this measure can be thought of as an inventory held in the education sector until students graduate, and so is in the spirit of the national accounts. This method might also be used in a similar way for health, at least for those of working age, although most healthcare is delivered to people who have retired.

More recently, Australia's Productivity Commission issued a report in April 2024 that used an outcome measure to estimate productivity for treating a range of conditions (Productivity Commission, 2024). They estimate the change in healthcare-attributable health outcomes per case of disease per year between 2011-12 and 2017-18 in quality-adjusted life years (QALYs). The diseases they study - treatment of cancers, cardiovascular diseases, blood and metabolic disorders, endocrine disorders and kidney and urinary diseases – account for about one third of healthcare spending in Australia. They assume that healthcare affects health outcomes for patients diagnosed with a disease, but that the number of cases of a disease are affected by other factors. As an example, the report considers lung cancer, where cases have been falling over time due to declining smoking rates, the main cause of lung cancer. They state that:

*“Comparing health outcomes due to lung cancer per capita in 2011-12 and 2017-18 would be misleading because of the confounding influence of changes in smoking rates. But if we assume that changes in smoking rates (among other non-healthcare factors) influence the prevalence of lung cancer but not outcomes among prevalent cases, we can isolate the contribution of changes in healthcare by examining changes in health outcomes due to lung cancer **per case of lung cancer**.”*

Productivity Commission (2024) estimate multi-factor productivity by comparing the value of the changes in health outcomes (per case of disease) with changes in the average dollar cost of treating each case of disease.

Cutler et al. (2022) find that their measure of healthcare productivity rises at a much more rapid pace than healthcare productivity in official statistics. Productivity Commission (2024) find an average productivity growth of 3% per annum over the period 2011-12 and 2017-18, which is much higher than the rate calculated by the Australian Bureau of Statistics in the national accounts. Corrado et al. (2021) shows greater productivity growth in the UK education sector than the measure currently in the national accounts. These findings suggest that the outcome based approach might be a useful tool for triangulating against other measures.

D. Prevention

The Productivity Commission (2024) report reviewed in the previous section assumes that healthcare affects health outcomes among patients with a disease but not the number of cases of that disease. This may understate the impacts of healthcare in preventing these diseases (e.g. hypertension medication can prevent cardiovascular disease). This highlights an important area of public sector activity on which there is very little research, the impact of preventative

initiatives. There is a concern that the benefits of preventative treatment are not shown in national income statistics.

Weale (2024) discusses a particular case of a preventative public health measure – the UK Diabetes Prevention Programme – a training programme offered to people with pre-diabetes where blood sugar levels are above normal but not at a level regarded as diabetic. In this paper the author sets out a framework for managing the introduction of new services whose benefit may be several times their cost. He shows that the proposed treatment is completely consistent with existing approaches to managing quality improvement and cost reduction. A quality improvement can be seen as the replacement of a low quality product by a high quality product. Or, with an index of quality, it can be represented as a price reduction and (with spending unchanged) a volume increase. The two methods give the same answer if the low quality service is assumed, once it has been replaced, to have a reservation price reflecting its poor quality. In the particular case studied, the author argues for the use of the money value of the discounted QALY gain, less the discounted costs of future treatment of those exposed to the public health measure as the reservation price. Then use of a symmetric index (Tornqvist or Fisher) will record the same increase in output with both approaches. Similarly, a straight fall in costs (prices) will give the same result as replacement of an expensive service by one which is cheaper but otherwise similar. Although Weale (2024) finds that the benefits of the diabetes prevention programme are small, his general framework is useful for examining public health prevention initiatives.

IV. Final Observations

This literature review aimed to identify the methodologies used by NSIs and Productivity Commissions on the estimation of public sector productivity. The discussion has been based on publicly available documents, drawn from across the public and private sector and from the UK as well as from abroad.

This literature reviews points at some key areas of attention to drive improvements in the productivity performance of the public sector:

1. Quality-adjusted measures were considered by the NSIs and Productivity Commissions in the development or improvement of the public sector productivity measures; however, adjusting the measure for quality remains a challenge in some countries due to data availability. The UK has been at the forefront in updating the methodology for public sector productivity, but there have been some important developments in other countries also.
2. There is a need for clearly mapping out the service delivery chains for each function in the public services. There is a need for better understanding of the institutional settings, budgets, inputs, output and outcomes to better identify what drives productivity. There is a need for a greater understanding of linkages between services.
3. Weighting, especially in taking account of quality change is an important issue which requires further attention.

4. Outcomes based approaches, in particular in healthcare and education, appear to be a fruitful avenue for further work.
5. Incorporating preventative measures provided by the public sector, such as campaigns to reduce smoking rates, are also important to gain a comprehensive understanding of public service provision and productivity.

Although much more research is needed, and importantly better data, this report highlights the current considerable effort being devoted to this topic worldwide. It is acknowledged that improvements in the public sector productivity measures, both in the UK and internationally, are expected in the coming years. These developments, driven by continuous research and policy innovation, may further refine our understanding and measurement of productivity in public services. Readers should consider this as part of a dynamic and evolving landscape.

V. References:

- Asian Productivity Organization. (2016). Measuring Public Sector Productivity in Selected Asian Countries. Report of the Research on Performance Management for Public-sector Organizations
- Atkinson, Anthony B. (2005) The Atkinson review: final report. Measurement of government output and productivity for the national accounts. Palgrave Macmillan, Basingstoke, England.
- Bureau of Labor Statistics. (1998). Measuring State and Local Government Labour Productivity: Examples from Eleven Services
- Corrado, Carol, O'Mahony, Mary and Samek, Lea (2021), Measuring Education Services Using Lifetime Incomes, ESCoE DP 2020 02.
- Abidin, Rauzah Zainal (2016). Measuring Public Sector Productivity. Malaysia Productivity Corporation.
- Dauda, S, Dunn, A and Hall, A 2 22, 'A systematic examination of quality-adjusted price index alternatives for medical care using claims data', Journal of Health Economics, vol. 85, p. 102662.
- Dunleavy P. and Carrera L (2013). Growing the Productivity of Government Services. Cheltenham, UK: Edward Elgar; 2013.
- Dunleavy, P. (2017). Public Sector Productivity - Measurement Challenges, Performance Information and Prospects for Improvement, OECD Journal on Budgeting, 17 (1). pp.1-28. ISSN 1608-7143.
- Foxton, F., J. Grice, R. Heys and J. Lewis (2019), "The Welfare Implications of Public Goods: Lessons from 10 years of Atkinson in the UK", ESCoE Discussion Paper 2019-11.
- Gemmell, N., Nolan, P. & Scobie, G. (2017). Public sector productivity: Quality adjusting sector-level data on New Zealand schools. New Zealand Productivity Commission Working Paper 2017/2
- Genet, Terry (2017). Productivity measurement case study: Courts. The New Zealand Productivity Commission Research Note 2017/04
- Genet, Terry and Hayward, Mike (2017). Productivity Measurement Case Study: Police. The New Zealand Productivity Commission Research Note 2017/09
- Gerlach, Brianna (2021). Assessing Public Sector Productivity: The case of the courts, public prosecution service and the police in The Netherlands
- Green, Nicholas (2017). Productivity measurement case study: Early Childhood Education. The New Zealand Productivity Commission Research Note 2017/05

Lau Edwin, Lonti Zzuzsanna, Schultz Rebecca (2017). Challenges in the Measurement of Public Sector Productivity in OECD Countries. OECD Working Paper Series.

O'Mahony, Mary and Stevens, Philip. (2009). Output and productivity growth in the education sector: Comparisons for the US and UK. *Journal of Productivity Analysis*, 31:177–94.

Maroni, Enrico (2021). Productivity in the Dutch Public Sector: The case of libraries and fire services.

Moore, Sandra and Hayward, Mike (2017). Productivity measurement case study: Ministry of Social Development. The New Zealand Productivity Commission Research Note 2017/07:

New Zealand Productivity Commission (2018). Measuring State Sector Productivity.

New Zealand Productivity Commission. Te Kōmihana Whai Hua o Aotearoa (2017). Productivity measurement case study: Police.

Niemi, M. (1998), 'Measuring government sector output and productivity in Finland - application of the output indicator method', agenda item 1, OECD meeting of national accounts experts, Statistics Directorate, STD/NA (98)4, 22-25 September, Paris: OECD

Police Foundation. (2022). A New Mode of Protection: Redesigning Policing and Public Safety for the 21st Century. The Police Foundation.

Productivity Commission 2024, Advances in measuring healthcare productivity, Research paper, Canberra.

Schreyer, P. (2010). Towards measuring the volume output of education and health services: A handbook. OECD Statistics Working Papers 2010/02. Paris, France: OECD. Retrieved 20 November 2017 from <http://dx.doi.org/10.1787/5kmd34g1zk9x-en>

Statistics Denmark (2014). General Government Output and Productivity 2008 – 2014.

UK Office for National Statistics (ONS) (2013) ONS website, A QMI report, public service productivity estimates: Education

UK Office for National Statistics (ONS) released 29 March 2023, ONS website, public service productivity, healthcare, England: financial year ending 2021.

UK Office for National Statistics (ONS) released 11 May 2022, ONS website, Sources and methods for public service productivity estimates.

UK Office for National Statistics (ONS) released 8 March 2024, ONS website, methodology, Improved methods for total public service productivity: total, UK, 2021

van Ark, B. (2022), Making Public Sector Productivity Practical. The Productivity Institute and Capita.

van Ark, B. and Hoskins, J. (2024), Probing Police Productivity: A Literature Review. The Productivity Institute.

Weale, M.R. (2024). “Public Health Measures and the National Accounts: the Case of Diabetes Prevention. ESCOE Deliverable submitted to ONS.

World Bank (2021). “Public-Sector Productivity (Part 1) Why is it important and how can we measure it?” EFI Insight-Governance. Washington, DC: World Bank.