International approaches to police performance measurement

Emma Zürcher, Lana Eekelschot, Annalena Wolcke, Lucy Strang
Preface

RAND Europe was commissioned by the Dutch government’s Research and Documentation Centre (WODC), at the request of the Dutch Ministry of Justice and Security, to conduct an international review of approaches to police performance measurement. The goals of the study were to examine how police performance is measured in a selection of countries and assess which starting points these approaches may offer for improving police performance measurement in the Netherlands.

This report summarises the findings from the study, drawing on information gathered from ten countries (including the Netherlands), an in-depth evidence review and stakeholder interviews in five case study jurisdictions, as well as the Netherlands, a review of the available literature, and information solicited through expert interviews.

It may be of interest to policymakers and professionals responsible for law enforcement organisations, and to researchers with an interest in police performance. The primary audience is the Directorate General for the Police and Safety Regions of the Ministry of Justice and Security in the Netherlands.

RAND Europe is an independent not-for-profit research institute whose mission is to help improve public policy and decision making through objective research and analysis. Part of the RAND Corporation, RAND Europe has over 30 years of experience in conducting policy research. This report has been peer-reviewed in accordance with RAND’s quality assurance standards.

For more information about RAND Europe or this document, please contact Lucy Strang (lstrang@randeurope.org) or Stijn Hoorens (hoorens@rand.org).
Executive summary

Police performance measurement is a complex activity that encompasses considerations of the social, legal, institutional and political contexts in which a police force operates. The definitions and methods used need to be regularly updated and adapted to reflect the constant changes in these contexts. This study follows from a broader desire, articulated by the national police force and the Ministry for Justice and Security, to improve how police performance is measured in the Netherlands. Indeed in 2014, Tilburg University and KU Leuven conducted a study to develop a new framework for productivity measurement of the Dutch police. While this study provided useful insights, approaches to performance measurement adopted by police forces abroad may also offer valid examples for the Netherlands and provide opportunities for general learning.

Research design

The study had two main goals:

1) To gather insights into how different police jurisdictions have approached performance measurement.

2) To assess what lessons these approaches can offer for improving police performance measurement in the Netherlands.

A series of research questions were developed to meet these goals (see Section 1.1.2.). These questions focused on the methods and indicators used to measure performance in a selection of ten policing jurisdictions, including the Netherlands; the stated purposes of the performance measurement; the reliability of the approaches; identified or potential adverse side effects of measuring performance; and examples of good or innovative practice in these areas. In addition, several research questions focused on the methods and indicators used to gain insight into the performance of the Dutch police; the aspects of policing not currently captured in this framework; the lessons that approaches to performance measurement in other jurisdictions may offer the Netherlands; and the transferability of these approaches to the Netherlands.

To address these research questions, the study team conducted a targeted literature review and expert and stakeholder interviews in each of these jurisdictions, including the Netherlands. The team then selected five case study jurisdictions for more in-depth data collection and analysis: England and Wales, Israel, North Rhine-Westphalia (Germany), Seattle (United States) and Sweden. Using the resulting findings, the study team focused on extracting current practices from the case study jurisdictions that had potential to offer...
starting points for improving police performance measurement in the Netherlands. These are presented in the main body of this report. For the countries not selected as case studies (Australia, Canada, Finland, New Zealand and the United States), the study team prepared high-level summaries of current approaches and highlighted several relevant practices that may be of interest to policy makers for the Dutch National Police. These are presented in Annex A.

To provide focus and enhance the feasibility of the research, this study looks exclusively at how police performance measurement has been approached by stakeholders who can directly influence policing policy, such as the police force and national- and local-level government stakeholders. For each of the case study jurisdictions, the research team has analysed the practices used by the relevant police force(s) and the administrative and/or political authority under which they sit. We are confident that the findings will therefore be of particular interest to decision-makers for the Dutch National Police, including police leadership at both national and regional levels and the Ministry of Justice and Security, and provide leads for improving how police performance is currently measured.

**Key findings**

In the following section, we discuss the findings of the study according to the first eight research questions (some of which have been grouped together) which concern relevant practices in the case study jurisdictions. The insights that these findings offer for the Netherlands (research questions 9, 10, 11 and 12) are addressed as part of the recommendations outlined at the end of this executive summary.

*How did the methodology for the performance measurements come about? Were stakeholders involved? If so, which ones?*

In the Netherlands, the key stakeholders involved in police performance measurement are the national and regional police units, the Ministry of Justice and Security, and local mayors and the Public Prosecution Service. At the national level, performance measurement is guided by the indicators and targets defined in the Ministry of Justice and Security’s Security Agenda. This framework is supplemented by regional and local performance agreements between municipalities, the Public Prosecutor and the police. The Court of Audit and the Inspectorate of Justice and Security also monitor the functioning of the national police and routinely conduct studies that analyse its performance.

This study found that in most case study jurisdictions, police performance measurement is organised at the national or state/regional level, while leaving flexibility for local units to complement this framework with additional measures, indicators, objectives and/or targets appropriate for the context in which they operate. The only exception to this is Seattle, where the framework for performance measurement was designed entirely at the local level.

*For what purposes are the performance measurements used?*

An important first consideration in developing a police performance measurement framework is to establish a common understanding of why performance should be measured and what the results will be used for. An organisation’s objectives and performance indicators, as well as the associated targets, will be determined by whether it is seeking to improve the quality of its services, provide strategic direction for its work, show
accountability to the public, or demonstrate a return on investment (or a combination of several or all of these factors).

Interviews conducted for this study indicate that the purpose of police performance measurement is conceived differently across stakeholders in the Netherlands, leading to differences in approaches. Due to their varying responsibilities and concerns, the government, certain groups within the police and academics have all developed different ideas about why it is important to measure the performance of the Dutch police and, consequently, about what should be measured. Opinions expressed include a need to demonstrate value for money by emphasising quantitative targets, increasing police accountability and legitimacy through consultation with stakeholders, and ensuring the provision of a high-quality service through benchmarking.

This report highlights that some of these motivations for measuring police performance also underlie the frameworks adopted in the case study countries. In England and Wales, for example, interviewees noted that some stakeholders at the national policy level believe that police performance should be measured in terms of crime reduction. The Seattle Police Department, on the other hand, has chosen to make accountability and the improvement of the quality of services the focus of its performance framework. In Israel, a new framework for performance measurement was developed specifically to increase public trust in the police and to support police stations in addressing local needs.

**How is police performance measured in the case study countries?**

Police performance is typically measured against a set of performance indicators and targets. The choice to use certain indicators over others tends to reflect the broader strategic priorities of law enforcement and/or government bodies. As such, police performance indicators can be context and jurisdiction specific. In all study forces, including the Netherlands, the indicators used to look at performance were primarily quantitative or quantitatively measured qualitative constructs. Some forces also employ qualitative indicators. For example, the Seattle Police Department considers qualitative feedback to measure the quality of citizen police interactions.

Moreover, stakeholders in all jurisdictions heavily rely on reported crime data gathered by the forces themselves to inform part or all of the performance measurement. While interviewees reported that such data were useful in measuring performance in relation to crime-related outcomes as it is relatively accessible and perceived by some stakeholders to be easy to interpret, using these data was reported to pose significant risks and challenges, as described in the next key finding. Data relating to response times and clearance rates were also frequently used. Seattle reported using more innovative data sources such as location data from police vehicle location tracking systems and automated transcriptions from interactions between police and members of the public recorded on body-worn surveillance cameras.

Data on rates of satisfaction with police services among citizens, often collected via survey instruments, are used for performance measurement in some jurisdictions such as England and Wales and Sweden. In the Netherlands, data on public perceptions of police performance are collected via several citizen surveys, but do not appear to be used systematically by the Dutch police to measure performance. Indeed, citizen perception of the police is not included as an indicator in the Security Agenda and is only mentioned by some regional units in their performance measurement.
What are the (further) advantages and disadvantages of performance measurements? Are any (side) effects of the performance measurements known? If so, which ones?

This study identified several disadvantages and challenges with the frameworks for police performance measurement developed in both the Netherlands and the case study countries.

As noted above, the performance of the Dutch police is measured primarily using quantitative indicators or qualitative constructs that are measured quantitatively. Previous studies have shown that this may increase the risk of perverse incentives, meaning that the indicators of police performance inadvertently encourage undesirable results, such as leading police officers to decline to register cases they feel may negatively impact their quality indicators. A dependence on exclusively quantitative indicators may also mean that some aspects of police work, particularly tasks related to the quality of provided services, are not captured. Interviewees also noted that the Dutch police face several challenges common to police performance measurement, namely the difficulty of establishing causality and measuring the effectiveness of crime prevention activities.

Furthermore, findings from the case studies demonstrate that securing and maintaining engagement from officers in how performance is measured can prove challenging. Interviewees from Israel indicated that this may be especially difficult when the process for implementing new aspects of a framework for performance measurement is complicated and requires a large degree of organisation change. Interviewees from Seattle and Sweden likewise mentioned the importance of ensuring buy-in among force stakeholders. In particular, it was noted that efforts must be made both top-down (by police leadership) and bottom-up (for instance by training officers early on) to ensure that all levels of officers are engaged in the process from the very start so that data collection (and analysis) can be operationalised effectively.

How reliable, valid and representative are the performance measurements?

This study found that there are challenges associated with using any type of data in measuring police performance; however, interviewees identified police recorded crime data as carrying the most risk in terms of validity, reliability and representativeness. Across all case study jurisdictions, interviewees expressed concerns about the quality of crime data from police management systems. In the Netherlands, interviewees reported that data collected through incident registration systems of the police, such as the Basisvoorziening Handhaving (BHV), may sometimes be inconsistent and inaccurate. Similar concerns were raised by interviewees in England and Wales and Sweden. The drivers of poor-quality data were also relatively consistent across the jurisdictions and tended to spring from the reliance on police staff to manually input crime data. Drivers included complex data-input processes increasing human error, a lack of resources to ensure that data inputters are appropriately trained and have sufficient time to enter correct and complete information, and under-reporting by members of the public about their crime victimisation.

Which (parts of) foreign performance measurements can be labelled as ‘qualitatively good’, ‘innovative’ and/or ‘promising’?

This study has identified several innovative and promising practices in measuring police performance in the case study jurisdictions, although there is limited evidence to assess their effectiveness.

In both England and Wales and North Rhine-Westphalia, interviewees positively highlighted the introduction of a process of continuous engagement that had created a feedback loop reported to enhance
stakeholders’ understanding of the forces and their performance. In certain forces, performance measurement appears to play an important role in supporting internal learning and improvement. In most jurisdictions, measurements are not (primarily) used to punish forces for poor performance, but rather to trigger reflection and actions to address performance issues and provide support to local and national stakeholders in introducing improvements. For example, the Seattle Police Department has integrated new equity measures to evaluate performance into its professional learning and development structures. Similarly, in Sweden a broad set of new indicators that cover additional dimensions of police work have been suggested to provide a more holistic understanding of performance and enable authorities to allocate resources more effectively.

**Recommendations for the Netherlands**

This study highlights several leads for improving performance measurement in the Netherlands. The researchers recommend that the Dutch police and policymakers first of all consider several fundamental questions about the type of performance measurement framework they would like to implement. Findings from the study showed that there are two key areas on which there should first be agreement before a new framework for police performance measurement can be developed: the purpose of measuring the performance of the police; and the level (national, regional and/or local) at which objectives, indicators, measures and/or targets should be set.

**The purpose of measuring police performance**

Careful consideration needs to be given to the purpose for tracking performance before any of the practices discussed in the case study chapters can be applied in the Netherlands. As noted above, an organisation’s objectives and the indicators, as well as the targets associated with them and data sources used, will be determined by what it is seeking to achieve by measuring performance. However, this study found that there is currently no agreement among Dutch stakeholders on why performance should be measured. Consensus on this issue may be reached through consultation among key stakeholders, such as the Ministry of Justice and Security and national and regional police units. It may also be useful to gather meaningful insights from other stakeholders, such as the Public Prosecution Office and the Court of Audit. The consultation strategy could include round table discussions, workshops, conferences and staff surveys that facilitate discussion on the organisational objectives of the police and how performance measurement can further these objectives.

**The level(s) at which the framework for performance measurement should be set**

The level at which performance objectives, indicators, measures and/or targets should be set and by whom needs to be agreed. Currently, the strategic direction from the Ministry of Justice and Security, laid out in the Security Agenda, is supplemented by both regional and local agreements on what should be measured, and against which targets. The study found that key national stakeholders in the Netherlands consequently do not have a clear picture of the performance indicators and targets currently used across the country, as the design of these metrics is partly left to regional and local stakeholders.

Some of the case study countries have chosen a different approach in this regard. For instance, a more local approach was introduced in Israel to address local needs and encourage trust in the police, while in Sweden
most strategic priorities for the national police authority are set by the Ministry of Justice and apply to all police units. This means that some of the practices adopted in each jurisdiction may not be transferable to the Netherlands. Indeed, before any of these practices can be integrated into the Dutch framework for police performance, careful consideration should be given as to whether the current structure for stakeholder involvement is the preferred approach.

The decisions on these two areas for agreement, the purpose of performance measurement and the level(s) at which the framework should be set, will give greater coherence to the Dutch strategy for police performance measurement and determine what practices for measuring performance from the case study jurisdictions might be transferable to the Netherlands.
# Table of contents

Preface ....................................................................................................................... i

Executive summary .................................................................................................. ii

Figures .................................................................................................................... x

Tables ..................................................................................................................... xi

Boxes .................................................................................................................... xii

Abbreviations .......................................................................................................... xiii

Acknowledgements .................................................................................................... xv

1. **Introduction**............................................................................................................. 1

2. **Methodological approach**.................................................................................. 10
   2.1. Research activities: Phase 1 ................................................................................. 10
   2.2. Research activities: Phase 2 ................................................................................ 14

3. **Police performance measurement in the Netherlands**........................................ 21
   3.1. Organisational context, stakeholders and evolution of performance measurement .......................................................... 22
   3.2. Current practices in measuring police performance ............................................. 27
   3.3. Challenges and complexities in measuring police performance ......................... 35
   3.4. Concluding comments ....................................................................................... 38

4. **Case study 1: England and Wales**...................................................................... 40
   4.1. Organisational context, stakeholders and evolution of performance measurement .......................................................... 41
   4.2. Current practices in measuring police performance ............................................. 42
   4.3. Challenges and complexities in measuring police performance ......................... 51
   4.4. Perceived innovative and/or promising practices in measuring police performance .......................................................... 53
   4.5. Concluding comments ....................................................................................... 54

5. **Case study 2: Israel**............................................................................................ 55
   5.1. Organisational context, stakeholders and the evolution of performance measurement .......................................................... 56
   5.2. Current practices in measuring police performance ............................................. 58
   5.3. Challenges and complexities in measuring police performance ......................... 60
   5.4. Perceived innovative and/or promising practices ................................................ 63
   5.5. Concluding comments ....................................................................................... 63

6. **Case study 3: North Rhine-Westphalia**.............................................................. 65
   6.1. Organisational context, stakeholders and the evolution of performance measurement .......................................................... 66
Figures

Figure 1: Evidence review protocol......................................................................................................... 15
Figure 2: Organisational structure of the Dutch police since 2013.......................................................... 24
Figure 3: The EMUN dashboard ........................................................................................................... 60
Figure 4: The three EAQ components.................................................................................................... 81
Figure 5: Organisation of the Swedish Police Authority................................................................. 88
Figure 6: Sub-indicators designed by the Ministry of Justice............................................................. 92
Figure 7: Australian Federal Police: Performance criteria, targets and outcomes 2021-2022. .......... 132
Figure 8: Return on investment calculation: Transnational crime efforts of the AFP ..................... 133
Figure 9: New Zealand Police districts ................................................................................................. 148
Tables

Table 1: Study research questions

Table 2: Key terms and definitions

Table 3: Overview of jurisdictions included in the study

Table 4: Search strings used for the targeted review

Table 5: Key informant interviews

Table 6: Criteria used for the selection of case study jurisdictions

Table 7: Case study selection

Table 8: Databases used for evidence review

Table 9: Search strings for evidence review

Table 10: Inclusion and exclusion criteria

Table 11: Number of sources screened and relevant sources reviewed

Table 12: Phase 2 stakeholder and expert interviews

Table 13: Summary of the 2023–26 Security Agenda

Table 14: Additional indicators and targets formulated by the regional unit for Rotterdam

Table 15: Police-related indicators – Security Monitor 2021

Table 16: Home Office National Crime and Policing Measures

Table 17: HMICFRS PEEL assessment framework

Table 18: Suggested new indicators

Table 19: Typology of performance indicators used across case study jurisdictions

Table 21: Canadian Police Performance Metrics Framework: Crime and victimisation metrics

Table 22: Canadian Police Performance Metrics Framework: Police activities and deployment metrics

Table 23: Canadian Police Performance Metrics Framework: Police resources metrics

Table 24: Canadian Police Performance Metrics Framework: Trust and confidence metrics

Table 25: Examples of indicators per strategic theme of the Finnish government

Table 26: Examples of measures from the Operational Performance Framework per strategic objective

Table 27: HMICFRS PEEL Assessment Framework
Boxes

Box 1: Key findings for the Netherlands ................................................................. 22
Box 2: Key findings in England and Wales ............................................................ 40
Box 3: Key findings in Israel .................................................................................. 55
Box 4: Key findings in North Rhine-Westphalia (NRW) ........................................... 65
Box 5: Key findings in Seattle ................................................................................ 74
Box 6: CompStat ................................................................................................... 76
Box 7: Assessing disproportionality in ‘Terry Stops’ using the equity pillar .......... 80
Box 8: Key findings in Sweden ............................................................................. 87
Box 9: The Swedish National Crime Survey ............................................................ 94
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACRO</td>
<td>Criminal records office (UK)</td>
</tr>
<tr>
<td>ACT</td>
<td>Australian Capital Territory</td>
</tr>
<tr>
<td>ADR</td>
<td>Annual data return</td>
</tr>
<tr>
<td>AFP</td>
<td>Australian Federal Police</td>
</tr>
<tr>
<td>ANZPAA</td>
<td>Australia and New Zealand Policing Advisory Agency</td>
</tr>
<tr>
<td>ARMS</td>
<td>Active risk management system</td>
</tr>
<tr>
<td>AVL</td>
<td>Automatic vehicle locator</td>
</tr>
<tr>
<td>BHV</td>
<td>Basisvoorziening Handhaving</td>
</tr>
<tr>
<td>Brå</td>
<td>Brottsförebyggande rådet (Swedish National Council for Crime Prevention)</td>
</tr>
<tr>
<td>BOSCAR</td>
<td>New South Wales Bureau of Crime Statistics and Research</td>
</tr>
<tr>
<td>CBS</td>
<td>Central Bureau of Statistics</td>
</tr>
<tr>
<td>CCHI</td>
<td>Cambridge Crime Harm Index</td>
</tr>
<tr>
<td>CFMP</td>
<td>Canadian Forces Military Police</td>
</tr>
<tr>
<td>CSEW</td>
<td>Crime Survey for England and Wales</td>
</tr>
<tr>
<td>DHI</td>
<td>Drug harm index</td>
</tr>
<tr>
<td>DoJ</td>
<td>Department of Justice</td>
</tr>
<tr>
<td>EAQ</td>
<td>Equity, Accountability and Quality</td>
</tr>
<tr>
<td>EBPC</td>
<td>Evidence-Based Policing Centre</td>
</tr>
<tr>
<td>EFR</td>
<td>Estimated financial return</td>
</tr>
<tr>
<td>GIDS</td>
<td>Geïntegreerde Databank voor Strategische bedrijfsvoering</td>
</tr>
<tr>
<td>GSS</td>
<td>General Social Survey (Canada)</td>
</tr>
<tr>
<td>HMICFRS</td>
<td>His Majesty’s Inspectorate of the Constabulary and Fire &amp; Rescue Service</td>
</tr>
<tr>
<td>HQ</td>
<td>National Headquarters (Israel)</td>
</tr>
<tr>
<td>INP</td>
<td>Israeli National Police</td>
</tr>
<tr>
<td>IOM</td>
<td>Integrated Offender Management</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>-------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>KPBs</td>
<td>Kreispolizeibehörden (NRW police units)</td>
</tr>
<tr>
<td>LAFP NRW</td>
<td>Landesamt für Ausbildung, Fortbildung und Personalangelegenheiten der Polizei NRW (State Office for Training, Continuing Education and Personnel Affairs of NRW)</td>
</tr>
<tr>
<td>LKA NRW</td>
<td>Landeskriminalamt NRW (State Criminal Office of NRW)</td>
</tr>
<tr>
<td>LZPD NRW</td>
<td>Landesamt für Zentrale Polizeiliche Dienste NRW (State Office for Central Police Services of NRW)</td>
</tr>
<tr>
<td>NCRS</td>
<td>National Crime Recording Standard</td>
</tr>
<tr>
<td>NIBRS</td>
<td>National Incident-Based Reporting System (US)</td>
</tr>
<tr>
<td>NRW</td>
<td>North Rhine-Westphalia</td>
</tr>
<tr>
<td>NYPD</td>
<td>New York City Police Department</td>
</tr>
<tr>
<td>NZ CHI</td>
<td>New Zealand Crime Harm Index</td>
</tr>
<tr>
<td>NZCVS</td>
<td>New Zealand Crime and Victims Survey</td>
</tr>
<tr>
<td>OMI</td>
<td>Office of Management Information</td>
</tr>
<tr>
<td>ONS</td>
<td>Office for National Statistics (UK)</td>
</tr>
<tr>
<td>OPF</td>
<td>Operational Performance Framework</td>
</tr>
<tr>
<td>PAS</td>
<td>Police Administration Survey (Canada)</td>
</tr>
<tr>
<td>PCC</td>
<td>Police and crime commissioners</td>
</tr>
<tr>
<td>PDCA</td>
<td>Plan-Do-Check-Act</td>
</tr>
<tr>
<td>PEEL</td>
<td>Police efficiency, effectiveness and legitimacy assessments</td>
</tr>
<tr>
<td>PEQF</td>
<td>Policing Education Qualifications Framework (UK)</td>
</tr>
<tr>
<td>PKS</td>
<td>Police Crime Statistics (Germany)</td>
</tr>
<tr>
<td>PMIU</td>
<td>Planning, Management and Information Unit</td>
</tr>
<tr>
<td>PNC</td>
<td>Police national computer</td>
</tr>
<tr>
<td>PVP</td>
<td>Protecting vulnerable people</td>
</tr>
<tr>
<td>RCMP</td>
<td>Royal Canadian Mounted Police</td>
</tr>
<tr>
<td>ROI</td>
<td>Return on investment</td>
</tr>
<tr>
<td>SOC</td>
<td>Serious and organised crime</td>
</tr>
<tr>
<td>SPD</td>
<td>Seattle Police Department</td>
</tr>
<tr>
<td>UCR</td>
<td>Uniform Crime Report (US)</td>
</tr>
<tr>
<td>WODC</td>
<td>Wetenschappelijk Onderzoek- en Documentatiecentrum (Dutch government’s Research and Documentation Centre)</td>
</tr>
<tr>
<td>ZEVA</td>
<td>Die Zentralstelle Evaluation (NRW)</td>
</tr>
</tbody>
</table>
Acknowledgements

The authors wish to acknowledge the extensive contributions of a number of organisations and individuals to this study. First, we are grateful to the police forces, authorities, academic experts and other stakeholders who engaged with the study team throughout the two phases of the research.

In addition, we warmly acknowledge the intellectual contributions and support of the WODC’s scientific advisory committee, whose members were Prof. Gabriele Jacobs (Erasmus University Rotterdam, Erasmus School of Social and Behavioural Science), Prof. Joris Knoben (Tilburg University, Tilburg School of Economics and Management), Dr Peter Kruize (Bureau Ateno), Fons van Gessel (Ministry of Justice and Security) and Frank Willemsen (WODC).

We would also like to Stijn Hoorens and Dr Mafalda Pardal for their constructive comments during the peer review process as part of RAND’s quality assurance procedures, as well as Famke van der Meer and Sara Treneman for their research assistance.

Finally, we would like to thank all the interviewees who participated in this study.
1. Introduction

This chapter describes the background to the study, the research questions, and the wider academic and policy context in which the study was carried out. It highlights previous research on the development of comprehensive performance indicators and considerations, including the quality and availability of relevant data and the broad range of work that police undertake, including, but not limited to, crime prevention. It then discusses the different ways in which police performance has been defined both historically and more recently, as well as existing methodologies, current approaches and the challenges that may obstruct efforts to measure performance effectively.

1.1.1. Study background

Measuring police performance is a complex activity that considers the social, legal, institutional and political contexts in which a law enforcement agency operates. The definitions and methods employed must regularly be updated and adjusted to reflect the constant changes in these contexts. This study analyses the various ways in which different police jurisdictions have approached this issue, focusing on the different methods for measuring police performance used internationally. Furthermore, it assesses what insights these approaches offer for measuring police performance in the Netherlands.

There has been a broad desire over the past decade among Dutch stakeholders, such as the national police force and the Ministry for Justice and Security, to improve how police performance is measured in the Netherlands. In 2014, the Research and Documentation Centre (WODC) commissioned a study to develop a new framework for productivity measurement of the police. In the resulting report, Zouridis et al. set out a range of possible performance objectives and indicators, as well as methods for data collection. However, approaches to performance measurement adopted by police forces abroad may also offer valid examples to improve current processes and provide opportunities for general learning. As such, the aim of the present study is to provide further insights into the frameworks that a selection of jurisdictions abroad have adopted to measure police performance and to find points of departure for measuring police performance in the Netherlands.

---

1.1.2. Study context

Research questions

The study addressed 12 research questions, outlined in Table 1.

Table 1: Study research questions

<table>
<thead>
<tr>
<th>Research questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. How is police performance measured abroad?</td>
</tr>
<tr>
<td>2. What aspects of policing do the measurements cover?</td>
</tr>
<tr>
<td>3. For what purposes are the performance measurements used?</td>
</tr>
<tr>
<td>4. How did the methodology for the performance measurements come about? Were</td>
</tr>
<tr>
<td>stakeholders involved? If so, which ones?</td>
</tr>
<tr>
<td>5. How reliable, valid and representative are the performance measurements?</td>
</tr>
<tr>
<td>6. What are the (further) advantages and disadvantages of the performance</td>
</tr>
<tr>
<td>measurements?</td>
</tr>
<tr>
<td>7. Are any (side) effects of the performance measurements known? If so, which ones?</td>
</tr>
<tr>
<td>8. Which (parts of) foreign performance measurements can be labelled as '</td>
</tr>
<tr>
<td>qualitatively good', 'innovative' and/or 'promising'? According to what criteria?</td>
</tr>
<tr>
<td>9. Which methods and/or indicators are used in the Netherlands to gain insight into</td>
</tr>
<tr>
<td>the performance of the police? Which aspects of policing are or are not measured?</td>
</tr>
<tr>
<td>10. Which 'high-quality', 'innovative' and/or 'promising' foreign performance</td>
</tr>
<tr>
<td>measurements (see research question 7) are relevant to Dutch police work and</td>
</tr>
<tr>
<td>performance, and are not (yet) applied here?</td>
</tr>
<tr>
<td>11. Can these relevant foreign performance measurements also be applied in the</td>
</tr>
<tr>
<td>Netherlands? If so, in what way and under what conditions?</td>
</tr>
<tr>
<td>12. What (direct) insights does the research provide with regard to improving the</td>
</tr>
<tr>
<td>measurement of police performance in the Netherlands?</td>
</tr>
</tbody>
</table>

The understanding of police performance measurement has evolved over time

Today, police responsibilities in the jurisdictions discussed in this report cover a broad range of areas, such as crime prevention and detection, traffic regulation and maintaining public order. Forces also aim to build strong relationships with local communities and often fulfil important social roles as mediators. Understanding how police perform in each of these different areas is important and can be a tool in helping local and national governments track whether public resources are being spent efficiently. Performance measurement can also help police assess and respond to internal and external criticism. Moreover, the

---

4 Davis (2012).
5 Davis (2012).
International approaches to police performance measurement

evaluation of existing practices can facilitate organisational learning and growth as the results may offer actionable insights. At the same time, the analysis of police performance that considers the views of the public can help build community trust, as people may feel that their concerns are being heard and can consequently be addressed.6 Performance measurement also provides stakeholders with a chance to reflect on the relationship between everyday police work and complex concepts such as police legitimacy, accountability, procedural justice and community safety.7

Defining police performance

Police performance is commonly measured against a set of performance indicators (see Table 2). The choice to use certain indicators over others tends to reflect the broader strategic priorities of law enforcement and/or government bodies.8 As such, police performance indicators can be context and jurisdiction specific.

Historically, the performance of police forces in developed countries has been measured primarily through crime-focused indicators, such as crime rates, clearance rates, response times, and productivity or workload statistics.9 However, in recent decades there has been a move away from the sole use of such indicators towards integrating a more expanded set of performance metrics that better reflect the increasing breadth of the policing mission, including the many social roles that the police perform alongside their crime-related work.10 In line with this development, it has become more common to deploy performance indicators such as ‘accuracy’ (the frequency with which determinations or judgments made by police are eventually shown to be correct) and ‘citizen satisfaction’ (the degree to which the public is content with police performance),11 and the fair distribution of police services across people and communities.12

Table 2: Key terms and definitions

<table>
<thead>
<tr>
<th>Key terms</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance</td>
<td>Performance is the ability of an organisation to achieve its objectives and/or targets.13</td>
</tr>
<tr>
<td>Performance measure</td>
<td>Performance measures are the quantitative and qualitative constructs that can be used to assess how an organisation is performing against its targets14:</td>
</tr>
<tr>
<td></td>
<td>• i.e. crime rates, number of arrests, client satisfaction and clearance rates.</td>
</tr>
<tr>
<td>Performance indicator</td>
<td>Performance indicators are qualitative and quantitative measures looked at during a specific timeframe to evaluate performance against pre-defined targets15:</td>
</tr>
<tr>
<td></td>
<td>• i.e. the percentage of cases solved annually.</td>
</tr>
</tbody>
</table>

6 Davis (2012).
7 Telep (2018); Aviv and Weisburd (2016).
8 Davis (2012).
9 Davis (2012); Hodgkinson, Caputo and McIntyre (2019).
10 Mastrofski (1999); Moore and Braga (2003); Hodgkinson, Caputo and McIntyre (2019); Davis (2012).
12 Aviv and Weisburd (2016).
13 Blake (2020).
14 Maslov (2016).
15 Parsons, Gokey and Thornton (2013).
<table>
<thead>
<tr>
<th>Key terms</th>
<th>Definition</th>
</tr>
</thead>
</table>
| Performance objective  | A performance objective is a short, medium or long term end goal designed to improve current results and potentially give strategic direction to police work\(^{16}\):  
  - i.e. improving road safety.                                                                                                                   |
| Performance target     | A performance target is a measurable expected or desired success level for a specific performance measure or indicator in the short, medium or long term:  
  - i.e. a 30% reduction in property crimes in one year as compared to the previous year.                                                        |
| Input                  | The materials that provide the basis for police work\(^{17}\):  
  - i.e. funding for a certain intervention or the number of staff.                                                                               |
| Output                 | Quantitative tangible and intangible products that result from undertaken police work:  
  - i.e. the achieved average response times to calls over a year.\(^{18}\)                                                                      |
| Outcome                | The qualitative benefits that police work (or a specific intervention) is designed to deliver. Outcomes are often achieved by means of outputs:  
  - i.e. an increased feeling of safety in society.\(^{19}\)                                                                                     |

\(^{16}\) U.S. Department of Defense (n.d.).  
\(^{17}\) Parsons, Gokey and Thornton (2013).  
\(^{18}\) Parsons, Gokey and Thornton (2013).  
\(^{19}\) Parsons, Gokey and Thornton (2013).  
\(^{20}\) SCRCSSP (1999).  
\(^{21}\) Davis (2012).

1.1.3. Approaches and challenges to measuring police performance

Police performance measurement has received some attention in academic and grey literature in recent years, particularly with respect to the various methodologies that may be employed and the challenges that can arise. This section summarises some of the main findings from previous work on the topic.

**Existing methodologies to measure police performance**

As noted above, police performance is often measured against performance indicators, which are qualitative and quantitative measures collected and analysed over time to evaluate performance against pre-defined targets. A performance target is a measurable expected or desired success level for a specific performance measure or indicator in the short, medium or long term, the achievement of which usually contributes to the fulfilment of broader strategic objectives (see Table 2 above).

Once these indicators have been developed, data can be collected and used to measure how the police force is performing against these indicators. A combination of methods is usually required to look at both outcomes and outputs. As set out in Table 2, outcomes reflect the qualitative benefits for society that the police organisation seeks to achieve, and outputs are quantitative indicators of the steps that need to be undertaken to achieve these desired outcomes.\(^{20}\)

Previous RAND research has identified several different methods of measuring police performance.\(^{21}\) These include:

---

\(^{16}\) U.S. Department of Defense (n.d.).  
\(^{17}\) Parsons, Gokey and Thornton (2013).  
\(^{18}\) Parsons, Gokey and Thornton (2013).  
\(^{19}\) Parsons, Gokey and Thornton (2013).  
\(^{20}\) SCRCSSP (1999).  
\(^{21}\) Davis (2012).
International approaches to police performance measurement

1) **Analysis of administrative records.** Administrative records can be used to measure policing outputs. They contain quantitative data that relates to topics such as crime rates, number of arrests, number of citizen complaints and clearance rates.22

2) **Surveys.** These can be used to extract qualitative and quantitative data from different stakeholders to advance knowledge of police performance. Several types of surveys are available, each of which serves a particular purpose23:
   a. **Staff surveys** are intended for internal use and ask officers about their work. They allow for an assessment of the departmental environment that conditions officers.
   b. **Community surveys** are used to question a representative sample of the community on satisfaction with law enforcement or perceptions of police effectiveness. They give an indication of the overall support for police among the public.
   c. **Contact surveys** are intended for people who have been in contact with the police, either as (possible) offenders or victims. They assess the quality of the interaction between law enforcement and the public.

3) **Direct observation.** This is a type of field research that can be used to gather data on police performance by observing the conduct of officers as they go about their work and interact with the public. It can be used to establish both outputs and outcomes, as trained observers are able to measure how often certain phenomena occur and analyse their impact on socially desirable outcomes.24

Internationally, a combination of these methodologies usually forms the basis of efforts to evaluate police performance.25 Other methods for collecting data, such as citizen or officer consultation through focus groups or social media tracking by means of web scraping, may also be used to provide additional information.

**Challenges of measuring police performance**

The retrospective development of police performance frameworks may only reflect current strategies, rather than wider organisational objectives.

Many challenges have been identified in the literature regarding the effective measurement of police performance. These challenges can emerge at the development stage of a performance measurement framework. For example, earlier studies have recommended that jurisdictions consider the overarching goals of a policing system during its development.26 However, this is a complex task. Performance measures are often developed retrospectively – that is, they reflect the work the police currently do, such as for accountability or control purposes.27 However, criminologist Cody Telep suggests that it is more beneficial...

---

22 Davis (2012).
23 Davis (2012); Maslov (2016).
24 Davis (2012); Maslov (2016).
25 Davis (2012); Maslov (2016).
to consider wider organisational objectives in the development of police performance measures, such as what the role of police in a modern society should be.\textsuperscript{28} He suggests that the development of a performance measurement system should go together with broader organisational reforms. Developing a performance measurement system in this outcome-oriented way can support police in maintaining a proactive position in a world in which society’s demands for police work are continuously evolving.\textsuperscript{29}

**Comparing the performance of different police units within and across jurisdictions can be complex.**

No two forces, or even units within forces, operate in entirely the same manner as they work within different contexts, face different challenges and use different data collection methods. This means that comparisons between their outputs and between their share in achieving outcomes are not always meaningful. Furthermore, comparing a police unit’s performance to its previous performance alone can skew the understanding of performance, as exogenous factors may cause crime rates, for example, to vary significantly from one year to the next. Comparison between different units, although complex, is therefore crucial.\textsuperscript{30}

When comparing one dataset against another, it is important that methods of risk adjustment, which allow statistical data to be controlled for variations, are used to mitigate this complexity. Such methods include stratification (dividing police forces into peer groups of similar size, type, jurisdiction and workload) and adjusting measures statistically (for instance through regression analysis).\textsuperscript{31}

**Collecting data to measure police performance may change how officers behave.**

The mere activity of collecting data to populate performance indicators may affect individual and organisational behaviour. For instance, officers may respond differently to (survey) questions when they are aware that their performance is being evaluated ('response bias') or reject or ignore questions all together ('non-response bias'). Furthermore, they may engage differently with the public or feel incentivised to increase certain easily quantifiable measures artificially if performance indicators relating to police outputs are used ('social desirability bias'). Officers may, for example, increase the number of traffic checkpoints to boost the number of traffic tickets given. These biases can influence the data being collected and may, therefore, mislead assessments of performance.

**Performance measurement can lead to unintended consequences.**

Measuring police performance may also encourage units to under-perform in an attempt to receive more resources\textsuperscript{32}; make officers feel pressured to misclassify crimes or suppress crime reports to keep recorded crime numbers low\textsuperscript{33}; or cause units to re-focus resources from pressing local issues to those on the agenda of national or state performance measurements.\textsuperscript{34} To minimise the negative effects of such changes in behaviour, it is important to employ a range of methods when measuring performance.\textsuperscript{35}

\textsuperscript{28} Telep (2018).
\textsuperscript{29} Hodgkinson, Caputo and McIntyre (2019).
\textsuperscript{30} Sparrow (2015); Neyroud (2008); Guilfoyle (2015).
\textsuperscript{31} Maguire (2004).
\textsuperscript{32} Davis (2012).
\textsuperscript{33} Sparrow (2015); Hodgkinson, Caputo and McIntyre (2019).
\textsuperscript{34} Davis (2012).
\textsuperscript{35} Maguire (2004); Davis (2012).
International approaches to police performance measurement

Proving causality between police inputs and desired outcomes can be complicated, as performance indicators can often be interpreted in different ways.

For example, if a higher number of arrests is recorded, it could mean either that police forces are more effective in their enforcement and are therefore detecting more offenders, or that they are less effective in their enforcement and that criminality rates have risen. Similarly, if the number of reported crimes decreases in a year, it may not be clear whether fewer crimes were indeed committed, or whether the public merely reported fewer crimes due to a lack of trust or confidence in the police. In general, it is a complex task to demonstrate causality between police work and crime reduction, as police work is only one of many factors influencing crime rates.

An over-reliance on crime-related outcome metrics may misrepresent police performance.

Such metrics tend to be tailored to measure reactive police work (i.e. where police engage after a crime was committed), while neglecting proactive policing (i.e. work meant to prevent a crime from being committed). Furthermore, the dark figure of crime (i.e. unreported crime) complicates the focus on reported crime in performance measurement. Pressure to reduce crime numbers under a performance measurement system can stand in the way of lowering the dark figure of crime, as there may be an incentive for officers to report as few crimes as possible, rather than uncovering a higher number.

Crime-related outcome metrics alone do not accurately reflect how police spend their time. Although police are responsible for crime control, activities such as engaging with communities for non-crime related matters, responding to emergencies, mediating disputes and providing the public with information are not commonly measured, despite constituting important elements of everyday police work. At the same time, little is known about the extent to which police engage in such activities, precisely because they rarely form part of performance measurement.

Many police performance measurement frameworks fail to consider the relationship between police and the community.

In many jurisdictions, it is no longer considered sufficient to measure administrative police data relating to crime reduction efforts, such as crime rates, clearance rates and the number of arrests. It is also increasingly considered important to measure whether communities feel safer. However, public perception is often not captured by the existing, primarily quantitative, measures that police employ to measure performance.

---

36 Davis (2012).
37 Telep (2018).
38 Sparrow (2015); Hodgkinson, Caputo and McIntyre (2019).
39 Hodgkinson, Caputo and McIntyre (2019); Lum and Nagin (2017).
41 Sparrow (2015).
42 Telep (2018).
44 Telep (2018).
45 Davis (2012); Sparrow (2015); Lum and Nagin (2017); Telep (2018).
46 Davis (2012); Davis et al. (2015).
Although some jurisdictions have recently started to consider public perception in their performance measurement, researchers continue to highlight the need for a greater emphasis on indicators that focus on the public opinion of the police and their work.\(^{48}\)

The case study chapters of this report will provide further insights from jurisdiction-specific literature concerning current police performance measurement systems, their challenges and their promises for innovation.

**Gaps and complexities in the literature reviewed**

Research in this area often does not address the measurement of police performance across different geographical areas.

Most literature on measuring police performance stems from the United States. However, lessons from US-based police performance measurement systems may not be applicable in jurisdictions that are organised differently, that operate under different hierarchical structures and that have different priorities regarding the tasks that the police carry out. As such, this report seeks to contribute to understanding the range of approaches to police performance adopted by different police jurisdictions with different organisational structures and strategic objectives.

There is limited research on exactly how performance data can be used or presented to promote police accountability.

There is a gap in the literature concerning the connection between measuring police performance and police accountability, and limited research examining how exactly performance data can be used or presented to promote police accountability. A recent study conducted in Israel found that the mere sharing of police performance data had only minimal effects on citizens’ perceptions of the police.\(^{49}\) The authors concluded that their ‘findings question the usefulness of performance reporting mechanisms as a major tool for strengthening accountability and raise serious doubts about the possible ways to improve these mechanisms and make them more effective.’\(^{50}\) These findings do not suggest that collecting police performance data and sharing it with the public is a futile exercise. Rather, they invite us to think about how police performance data can be used to achieve the greatest possible impact on police work and police perception. This report helps illuminate this connection by presenting examples of how and why different jurisdictions have chosen to use their police performance data.

\(^{48}\) Lum and Nagin (2017); Telep (2018).

\(^{49}\) Mizrahi and Minchuk (2019).

\(^{50}\) Mizrahi and Minchuk (2019).
1.1.4. **Report structure**

The remainder of this report presents the findings from the study. Chapter 2 presents the research questions and the methodologies used by the research team to answer these questions. Chapter 3 presents a high-level overview of the current state of police performance measurement in the Netherlands. The subsequent chapters present findings from the five case studies that examine police performance measurement in England and Wales (Chapter 4), Israel (Chapter 5), North Rhine-Westphalia in Germany (Chapter 6), Seattle in the United States (Chapter 7) and Sweden (Chapter 8). These chapters provide an overview of police performance measurement across these jurisdictions with respect to research questions one to eight. Following a brief description of the organisational context in which these police forces operate, as well as the key stakeholders engaged in measuring police performance, the chapters highlight current approaches to performance measurement, challenges and complexities in measuring police performance, and perceived innovative practices and recent developments.

The report concludes by reflecting on key findings and cross-cutting themes from the case studies in Chapter 9 and suggesting lessons for the Netherlands in Chapter 10. Annex A presents research findings from Phase 1 for five countries (Australia, Canada, Finland, New Zealand and the United States).
2. Methodological approach

This chapter outlines the methodological approach adopted by the study team. The research questions are presented, the approach to the two phases of literature review and expert interviews are outlined, and there is a discussion of the limitations of the study.

2.1. Research activities: Phase 1

During Phase 1 of the study, the research team conducted a high-level review of approaches to measuring police performance in ten countries/jurisdictions, including the Netherlands. The ten countries/jurisdictions pre-selected to be included in the study were: North Rhine-Westphalia (Germany), the United Kingdom, Sweden, Canada, Australia, New Zealand, Finland, Israel, the United States and the Netherlands. The choice was made to focus on North Rhine-Westphalia rather than the whole of Germany because of its proximity to the Netherlands, both geographically and in terms of population size. It was deemed appropriate to include at least one jurisdiction that met these criteria. Moreover, the research team proposed changing the United Kingdom to solely England and Wales, as Scotland and Northern Ireland have different policing systems. The final selection of countries/jurisdictions was as follows: Australia, Canada, England and Wales, Finland, Germany (North Rhine-Westphalia), Israel, New Zealand, Sweden and the United States (Table 3). For the purpose of answering research questions 9, 10, 11 and 12 we also conducted a high-level review of approaches to police performance measurement in the Netherlands.

Table 3: Overview of jurisdictions included in the study

<table>
<thead>
<tr>
<th>Country</th>
<th>Population (2022)</th>
<th>Jurisdictional organisation</th>
<th>Number of police officers</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Netherlands</td>
<td>17.56 million</td>
<td>National police force</td>
<td>62,942&lt;sup&gt;53&lt;/sup&gt;</td>
</tr>
<tr>
<td>Australia</td>
<td>26.37 million</td>
<td>State and territory police agencies</td>
<td>66,000&lt;sup&gt;54&lt;/sup&gt;</td>
</tr>
<tr>
<td>Canada</td>
<td>38.70 million</td>
<td>State police agencies</td>
<td>70,566&lt;sup&gt;55&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

<sup>51</sup> Unless otherwise stated, these data have been retrieved from UN population estimates and projections. United Nations (2022).

<sup>52</sup> The terms used here follow those used by the relevant police agencies.

<sup>53</sup> De cijfers van (2019).

<sup>54</sup> Police Federation of Australia (2023).

<sup>55</sup> Statistics Canada (2023).
Phase 1 comprised a targeted review and key informant interviews. Both are described in more detail below.

2.1.1. Targeted review

The first task of the study consisted of a targeted review of relevant documents including academic work, grey literature and policy documents. Other sources such as government or departmental websites highlighting relevant information (e.g. policing structures or performance measurement initiatives) were also considered. To start identifying relevant literature, a basic search string was used (see Table 4) to conduct searches in Google, Google Scholar and Scopus. Searches were furthermore limited to publications in English and the national language of the relevant country, and only if published in 2014 or later.

### Table 4: Search strings used for the targeted review

<table>
<thead>
<tr>
<th>Language</th>
<th>Search strings</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>Police Policing Law enforcement AND 'Performance measur*' 'Measur* performance' Measur* Quality 'Performance indicator*' AND [Name of relevant police jurisdiction]</td>
</tr>
</tbody>
</table>

---

56 Office for National Statistics (2022). The next population update is due to be published in the summer of 2023.
57 Home Office (2023).
58 Ministry of the Interior Finland (2023).
60 New Zealand Police (2023b).
62 Polizei NRW (2023d).
63 OSCE POLIS (2023b).
64 National Law Enforcement Officers Memorial Fund (2023).
The aim of this exercise was to identify key sources that could be used for a high-level analysis of measuring police performance in the selected countries, and to gather initial information on their practices. The review was also used to produce a working definition of the key concepts for this study, refine the study scope and approach, identify and analyse key sources from the selected countries, and identify key stakeholders to be approached for interviews. The review primarily focused on gathering information on the frameworks for measuring police performance in the countries, including their purposes and the aspects of policing they cover; what is known about their reliability, validity and representativeness; the advantages and disadvantages – including unintended consequences or side effects – of using the measurements; and an additional focus on gathering data from the Netherlands.

### 2.1.2. Key informant interviews

After conducting the targeted review for each of the ten countries/jurisdictions, the study team conducted interviews with key informants from those countries/jurisdictions. Using the documentation gathered during the targeted review, police stakeholders with experience of police performance measurement and academic experts with extensive experience in researching police performance measurement were selected and invited to participate. In total, 69 key informants were invited for an interview during this phase of the study, of which 25 agreed to be interviewed (see Table 5).

<table>
<thead>
<tr>
<th>Jurisdiction</th>
<th>Informants contacted</th>
<th>Informants interviewed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>Canada</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>England and Wales</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Finland</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>Israel</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>New Zealand</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Germany - North Rhine-Westphalia</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>Sweden</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>United States</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>69</strong></td>
<td><strong>25</strong></td>
</tr>
</tbody>
</table>

Topic guides were prepared in advance of the interviews and covered questions on the frameworks for measuring police performance in the countries, including their purposes and the aspects of policing they cover; what is known about their reliability, validity and representativeness; advantages and disadvantages, including unintended or side effects of using the measurements; and an additional focus on gathering data from the Netherlands. Interviewees were also asked to share relevant literature not previously identified in the targeted review. The interviews were semi-structured to leave room for elaboration or additional questions raised by interviewees. All key informant interviews were conducted virtually and were recorded. Prior to the interview, each interviewee was informed about the purpose of the study, their consent to
participate voluntarily and the confidentiality of the interview. Researchers also sought consent to video- and audio-record. Interviews were conducted in English, Dutch and German, as appropriate. They were fully transcribed and analysed qualitatively with the help of a data extraction tool that the research team developed based on the topic guide. Findings from the interviews informed the selection of case study jurisdictions (see Section 2.1.3), helped the design of the study tools for Phase 2 of the study and are included in this report. The interviews gave greater breadth and qualitative depth to the findings from the targeted review and provided further insight into the design and operation of particularly innovative and promising practices in specific police jurisdictions within the selected countries.

2.1.3. Case study selection
The study team developed five criteria to select the case study jurisdictions for Phase 2. These criteria were informed by 1) the research questions; 2) the findings from the targeted review and stakeholder interviews; and 3) a discussion with the study’s scientific advisory committee. Considerations included the potential of the case study jurisdiction to provide possible lessons for police performance measurement in the Netherlands, whether elements of operational transferability could be identified, whether there were similarities in terms of the policy context, and whether it could reasonably be assumed that sufficient additional data could be gathered to answer the research questions. Table 6 gives an overview of all six selection criteria.

Table 6: Criteria used for the selection of case study jurisdictions

<table>
<thead>
<tr>
<th>Selection criteria</th>
<th>Clarification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Evidence of innovative performance measures in the selected jurisdiction</td>
<td>The selection should include jurisdictions where innovative and/or evidence-based measures have been identified that may provide leads for improving performance measurement practices in the Netherlands.</td>
</tr>
<tr>
<td>Elements of operational transferability from the jurisdiction in question to the Netherlands have been identified</td>
<td>The selection should include jurisdictions where the police operate in such a way that it is reasonable to expect that there is some organisational overlap with the Netherlands, meaning that it may be possible to replicate (elements of) potentially innovative and/or interesting practices in the Dutch context.</td>
</tr>
<tr>
<td>Similarities to the Netherlands in terms of policy context</td>
<td>The selection should include jurisdictions with similarities in terms of policy context, such other EU member states and jurisdictions with a population size similar to that of the Netherlands.</td>
</tr>
<tr>
<td>Further evidence concerning police performance measurements in the jurisdiction in question is likely to be available (in academic as well as grey literature)</td>
<td>The selection should include jurisdictions for which it is reasonable to assume that further evidence from academic and grey literature is available to ensure that sufficient additional data on relevant policies and practices can be gathered to answer the research questions.</td>
</tr>
</tbody>
</table>
Selection criteria | Clarification
--- | ---
**Potential to conduct several additional interviews with stakeholders and experts in the jurisdiction in question** | The selection should include jurisdictions for which it is reasonable to assume that further interviews can be conducted to ensure that sufficient additional data on relevant policies and practices can be gathered to answer the research questions.

For some of the countries analysed in Phase 1 of the study, police performance measures are a local or regional rather than national matter, with different policing jurisdictions within the countries adopting their own approaches. Therefore, the researchers used the findings from the targeted review and the stakeholder interviews to further select specific police jurisdictions as case studies. For example, overarching approaches to measuring police performance in the United States were captured in the high-level analysis conducted in Phase 1 (see A.5). Evidence from the interviews conducted for the United States and relevant literature indicated that the framework for performance measurement recently introduced by the Seattle Police Department was of particular interest in terms of the above-mentioned selection criteria, and was therefore selected as a case study for further analysis in Phase 2 (see Chapter 7).

Based on the findings of Phase 1 and discussion with the scientific advisory committee, the following case studies were selected: England and Wales, Israel, North Rhine-Westphalia, Sweden and Seattle (US). Table 7 below provides an overview.

**Table 7: Case study selection**

<table>
<thead>
<tr>
<th>Criteria</th>
<th>AUS</th>
<th>CAN</th>
<th>ENG &amp; WAL</th>
<th>FIN</th>
<th>ISR</th>
<th>NRW</th>
<th>NZ</th>
<th>SWE</th>
<th>US (SEA)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Innovative and/or evidence-based practices</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Elements of operational transferability</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Similarities to the Netherlands in terms of policy context</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td></td>
</tr>
<tr>
<td>Further evidence likely to be available in the literature</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
<tr>
<td>Potential to conduct further interviews</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
<td>✔️</td>
</tr>
</tbody>
</table>

Note: AUS = Australia; CAN = Canada; ENG & WAL = England & Wales; FIN = Finland; ISR = Israel; NRW = North Rhine-Westphalia; NZ = New Zealand; SWE = Sweden; US = The United States; SEA = Seattle.

**2.2. Research activities: Phase 2**

To develop a better understanding of the approaches to measuring police performance in the five selected jurisdictions, the research team conducted an in-depth analysis for each of the case studies that comprised an evidence review and stakeholder and expert interviews.
2.2.1. Evidence review

Building on the literature review conducted as part of a 2014 study by Zouridis et al. on the topic of police performance measurement, the study team conducted an in-depth review of the available evidence from 2014 onwards. Documentation was included from both academic and grey literature, as well as relevant policy documents for the five case study jurisdictions and the Netherlands.

Selection of relevant documentation

The evidence followed seven steps, as outlined in Figure 1.

Figure 1: Evidence review protocol

The research team first conducted a series of searches through online databases (see Table 8) of academic and grey literature.

Table 8: Databases used for evidence review

<table>
<thead>
<tr>
<th>Database</th>
<th>Link</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scopus</td>
<td><a href="https://www.scopus.com/">https://www.scopus.com/</a></td>
</tr>
<tr>
<td>Web of Science</td>
<td><a href="https://www.webofscience.com/">https://www.webofscience.com/</a></td>
</tr>
<tr>
<td>WorldCat</td>
<td><a href="https://www.worldcat.org/">https://www.worldcat.org/</a></td>
</tr>
</tbody>
</table>

Around ten different search strings were trialled using the terms set out in Table 2, as well as various synonyms, before the search strings were finalised (see Table 9). Additional search strings were tailored and/or translated for specific jurisdictions.

Table 9: Search strings for evidence review

<table>
<thead>
<tr>
<th>Language</th>
<th>Search strings</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>‘Police performance mea$’ OR ‘Police performance indicator$’ AND [Name of relevant police jurisdiction]</td>
</tr>
<tr>
<td>Dutch</td>
<td>‘Police performance’ AND (Indicator$ OR mea$)</td>
</tr>
<tr>
<td>German</td>
<td></td>
</tr>
<tr>
<td>Swedish</td>
<td></td>
</tr>
</tbody>
</table>

The purpose of these searches was to gather and include in the review any documents that met the criteria detailed in Table 10.

Table 10: Inclusion and exclusion criteria

<table>
<thead>
<tr>
<th></th>
<th>Inclusion criteria</th>
<th>Exclusion criteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>England and Wales, Israel, the Netherlands, North Rhine-Westphalia, Seattle, Sweden</td>
<td>Other jurisdictions</td>
</tr>
<tr>
<td>Language</td>
<td>English, German, Dutch, Swedish</td>
<td>Other languages</td>
</tr>
<tr>
<td>Topic</td>
<td>Police performance measurement</td>
<td>Other topics</td>
</tr>
<tr>
<td>Timeframe</td>
<td>From 2014 onwards</td>
<td>Before 2014</td>
</tr>
<tr>
<td>Study type</td>
<td>Peer-reviewed journal publications (including systematic and non-systematic literature reviews); published reports from governments, non-governmental organisations or independent research organisations; PhD and MA theses; policy documents; national legislation; editorials; opinion pieces; comments/letters</td>
<td>Sub-MA-level theses; documents without clear individual or organisational authorship; videos and podcasts</td>
</tr>
</tbody>
</table>

Searches were conducted in English, Dutch, German and Swedish. Due to constraints in the language skills of the study team, the searches pertaining to Israel were conducted exclusively in English.

To supplement the limited documents identified in the evidence review, the research team also asked the WODC and study interviewees for access to additional relevant literature, and snowballed citations from the materials identified in Phase 1, meaning that the reference list of these articles were used to identify additional documents.
International approaches to police performance measurement

**Data extraction**

Table 11 presents an overview of the number of relevant documents identified for each jurisdiction after an initial screening of results. These include documents identified through the literature search outlined above, as well as literature identified through snowballing.

**Table 11: Number of sources screened and relevant sources reviewed**

<table>
<thead>
<tr>
<th>Country/jurisdiction</th>
<th>Number of screened results</th>
<th>Included academic studies</th>
<th>Included grey literature</th>
</tr>
</thead>
<tbody>
<tr>
<td>England and Wales</td>
<td>25</td>
<td>6</td>
<td>12</td>
</tr>
<tr>
<td>Israel</td>
<td>159</td>
<td>5</td>
<td>8</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>465</td>
<td>12</td>
<td>11</td>
</tr>
<tr>
<td>North Rhine-Westphalia</td>
<td>20</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Sweden</td>
<td>102</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Seattle</td>
<td>77</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>848</strong></td>
<td><strong>30</strong></td>
<td><strong>44</strong></td>
</tr>
</tbody>
</table>

The results of the literature search were captured in data extraction matrices. Papers were first organised by title and abstract. Once screened for relevance, the following data were extracted from the sources reviewed: contextual information on how the measures were developed, with whose input and to what purpose(s); what is known about their reliability, validity and representativeness; advantages and disadvantages – including unintended consequences or side effects – of using the measurements; approaches that may be assessed as innovative, promising and/or good and according to what criteria; and information on methods and indicators used to measure performance in the Netherlands.

### 2.2.2. Stakeholder and expert interviews

The research team supplemented the evidence review with semi-structured stakeholder and expert interviews. These interviews were a crucial step in the data collection for this study, as the available literature on the topic of police performance measurement for the case study countries is limited (see Table 11).

The research team used a purposive sampling strategy to select the interviewees. Due to its flexible nature, quota sampling – a form of purposive sampling – was used, in which minimum quotas per jurisdiction were determined (four interviewees per case study), which 'ensures that key groups are represented in the sample, while providing flexibility in the final sample composition.' We interviewed policing representatives whose portfolios include measuring police performance; academic researchers with subject matter expertise on measuring police performance; and policymakers and representatives from law enforcement professional

---

66 As Robinson (2014, p.32) summarises: 'The rationale for employing a purposive strategy is that the researcher assumes, based on their a-priori theoretical understanding of the topic being studied, that certain categories of individuals may have a unique, different or important perspective on the phenomenon in question and their presence in the sample should be ensured.'

67 Robinson (2014, p.34.)
development and performance organisations. In total, the team interviewed 29 additional stakeholders and experts during this phase of the study (see Table 12).

Table 12: Phase 2 stakeholder and expert interviews

<table>
<thead>
<tr>
<th>Jurisdictions</th>
<th>Stakeholders contacted</th>
<th>Stakeholders interviewed</th>
</tr>
</thead>
<tbody>
<tr>
<td>England and Wales</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Israel</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>North Rhine-Westphalia</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Sweden</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>Seattle</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>34</td>
<td>29</td>
</tr>
</tbody>
</table>

Topic guides in the different target languages were prepared in advance of the interviews, with topics covering the main research questions. Preliminary data collected in Phase 1 fed into the development of these topic guides. In general, the topic guides covered questions concerning how performance measures were developed in the police jurisdictions, with whose input and to what purpose(s); what is known about their reliability, validity and representativeness; advantages and disadvantages, including unintended or side effects of using the measurements; approaches that may be assessed as innovative, promising and/or good and according to what criteria; and information on methods and indicators used to measure performance in the Netherlands.

As in Phase 1 of the study, all interviews were conducted virtually and audio recorded. Prior to the interview, each interviewee was informed about the purpose of the study, their consent to participate voluntarily and the confidentiality of the interview. Researchers also sought consent to audio-record. Interviews were conducted in English, Dutch and German. Interviews were fully transcribed and analysed qualitatively with the help of a data extraction tool that the research team developed based on the topic guide. Anonymised data from the interviews are presented in this report.

2.2.3. Expert workshop

After the conclusion of the evidence review and stakeholder interviews, the research team organised a virtual expert workshop with stakeholders from the Netherlands. This took place on 13 July 2023. The aim of the meeting was to explore in greater depth whether the findings from the case studies could provide relevant leads for improving how the Dutch police measures its performance. Feedback was also gathered on the potential implications and necessary considerations for adopting some of the practices highlighted throughout the report in the Netherlands.

Workshop participants included stakeholders from the police, the public prosecution service, and the Inspectorate of Justice and Security, as well as academic experts. In total, 21 individuals were invited and 6 attended the meeting. One additional stakeholder provided written feedback.
At the outset of the meeting, the research team provided a short overview of the study and presented the findings from each of the case studies, highlighting practices of particular interest (such as innovative, good-quality and promising practices) and their scope, limitations, target audiences, and effectiveness. The participants were subsequently invited to share their views on whether these approaches could potentially provide meaningful leads for police performance measurement in the Netherlands, and if so, why. Furthermore, the study team asked participants to provide additional detail on the available evidence and/or perspectives on designing and implementing the identified approaches. The workshop also offered an opportunity to refine analysis and identify additional sources, and to reflect on how and which findings should be presented in the reporting.

The results of this discussion are captured in Chapter 10.

2.2.4. Scope of the study

Policing, and in particular discussions around the quality of police policies, practices and decision making, has received significant attention in academic and grey literature. As such, many different organisations collect data on the police, from forces themselves to national agencies for statistics, research centres and community interest organisations.

However, the collection of data related to policing alone does not constitute police performance measurement. As set out in Table 2, performance is the ability of an organisation to achieve its objectives and/or targets. As such, the collection of data that could be used to measure how the police are performing, but which is not currently being used to track the progress of a particular force against a pre-defined set of objectives and/or targets, falls outside the scope of this study.

Furthermore, police performance is of interest to, and can be evaluated by, a wide range of stakeholders. For the research design and data collection of this study, we have distinguished between ‘key’ and ‘peripheral’ stakeholders in police performance measurement, with key stakeholders those in a position to use performance measurement to shape policy in the relevant jurisdiction, and peripheral stakeholders those interested in police performance for other reasons. For example, depending on the jurisdiction, the police force and national- and local-level government stakeholders may be described as key actors in police performance measurement as they can directly use relevant findings to shape policy in the area of policing. Peripheral stakeholders may include actors such as civil society groups and academics who use the results for other purposes, such as community interest or academic research. Their findings may still be used to influence policy, but only via one of the key stakeholders.

To provide focus and enhance the feasibility of the research, this study looks exclusively at how police performance measurement has been approached by key stakeholders. For each of the case study jurisdictions, the research team has analysed the practices in use by the relevant police force(s) and the administrative and/or political authority under which they sit. We are confident that the findings will therefore be of particular interest to this category of stakeholders in the Netherlands, including police

---

68 See, for example, the studies published by the RAND Center for Quality Policing. As of 14 November 2023: https://www.rand.org/well-being/justice-policy/centers/quality-policing.html

69 Blake (2020).
leadership at both national and regional levels, and the Ministry of Justice and Security, and provide ideas for improving how police performance is currently measured.

2.2.5. **Limitations of the study**

Relevant literature that describes police performance measurement at the jurisdictional level is limited. Consequently, data collection for this study has relied heavily on interviews with stakeholders and experts, which presents some risks:

- As Table 12 demonstrates, the interview sample for both the key stakeholder and expert interviews was relatively limited, as agreed beforehand with the WODC. As such, the data collected, and thus the findings of this study to some extent, may not be representative. There is also a potential risk of selection bias and non-response bias (see Section 1.1.3). To mitigate these risks, the team sought to conduct interviews with a variety of stakeholders with different backgrounds, roles and perspectives in each jurisdiction to triangulate and validate the data. However, individual interviewees may have gaps in their knowledge and/or understanding of the topic. Where this has posed an issue for the analysis, it has been clearly flagged in the report.

- Research was conducted in English, German, Dutch and Swedish where appropriate. The evidence review and interviews for Finland and Israel were conducted exclusively in English, which may have limited the interview sample and documentation used.

It should be noted that the findings set out in this report are not intended to provide a comprehensive review of police performance measurement in the ten jurisdictions under study, as the number of countries included renders this infeasible within the confines of a single report. Furthermore, some of the countries, such as the United States, have many different local, regional and federal police forces with distinct frameworks for performance measurement which were not included in the Phase 2 analysis. Providing a complete overview of all approaches to police performance measurement in these countries would be challenging.

Instead, the study team has focused on extracting current practices for each of the case study jurisdictions that may offer starting points for improving police performance measurement in the Netherlands. For those countries not selected as case studies, the study team has prepared high-level summaries of current approaches and highlighted several relevant practices that may be of interest for stakeholders in the Netherlands (see Annex A).

---

70 Bureau of Justice Statistics (2022).
This chapter provides an overview of current approaches to measuring police performance in the Netherlands. It begins with a discussion of the evolution of performance measurement within the Dutch police, its organisational structure and the role of key stakeholders (see Section 2.2.4). Relevant practices are then highlighted, paying particular attention to the indicators, targets and data used at different levels and by different stakeholders, as well as varying opinions on the purpose of police performance measurement. The chapter closes with a reflection on the challenges and complexities that complicate efforts to measure the performance of the Dutch police force in a meaningful and representative manner.
Box 1: Key findings for the Netherlands

- The key stakeholders involved in measuring police performance in the Netherlands are the national and regional police units, the Ministry of Justice and Security, and mayors and public prosecutors.
- Additional, peripheral stakeholders that assess the performance of the Dutch police are the Inspectorate of Justice and Security, the Court of Audit and the Office of Management Information.
- Evidence suggests that there are different ideas concerning the purpose of police performance measurement, as the practice serves different functions for different stakeholders in the Netherlands.
- At the national level, performance measurement is guided by the indicators and targets defined in the Ministry of Justice’s Security Agenda. This framework is supplemented by regional and local performance agreements.
- The Court of Audit and the Inspectorate of Justice and Security also monitor the functioning of the national police, and routinely conduct studies that analyse its performance.
- Data used to measure police performance are primarily gathered using the police force’s incident registrations systems, such as the Basisvoorziening Handhaving (BHV).
- Interviewees indicated that data collected through the registration systems of the police may sometimes be inconsistent and inaccurate.
- Data on public perceptions of police performance are collected via citizen surveys, but this information does not appear to be used systematically by the Dutch police to measure organisational performance.
- Beyond issues with data quality, interviewees noted that the Dutch police face several challenges common to police performance measurement, namely the difficulty of proving causality and measuring prevention.
- Existing literature indicates that the reliance of the Dutch police on quantitative indicators may increase the risk of perverse incentives and may mean that meaningful qualitative insights are not captured.
- Key national stakeholders do not have a clear picture of the performance measurement indicators and targets currently used across the country, as the design of these metrics is partly left to stakeholders at the national and regional levels.
- Reporting on police performance in relation to certain topics, such as preventative efforts and the quality of services provided, remains limited at both national and regional levels.
- Interview data collected for this study indicate a desire among staff within the Commissioner’s Office to expand the collection of data on digital activity relating to perceptions of the police.

3.1. Organisational context, stakeholders and evolution of performance measurement

It was not until the early 21st century that both academics and the police in the Netherlands began to pay attention to police performance measurement. This is a relatively late development given the critical attention given to the concept in international literature in the 1980s and 1990s. Literature on the development of police performance measurement in the Netherlands indicates that this may be partly

---

71 Van Waarden and Boutykova (2009).
because levels of trust in Dutch public institutions, including the police, have historically been high, meaning that systematic performance measurement was not felt to be strictly necessary.\textsuperscript{72,73}

When attention did start to shift towards police performance measurement, ‘performance contracts’ were introduced through the Dutch Police National Framework 2003-2006 (2003).\textsuperscript{74,75} This was agreed upon by the Ministry of the Interior, the Ministry of Justice and the Dutch police organisation. At this time, the Dutch police was not yet strictly a national force, instead comprising a national unit (\textit{Korps landelijke politiediensten}) and 25 regional units (\textit{regiokorpsen}).

The performance contracts stipulated several targets for both national and regional units that were highly focused on output and quantitative indicators.\textsuperscript{76} For instance, they required Dutch police forces to issue 180,000 additional fines per year and to submit 40,000 extra cases to the Public Prosecutor each year.\textsuperscript{77} For the very first contracts (also referred to as the 1st round of contracts), performance results were monitored and recorded at the national level.\textsuperscript{78}

Around the same time that regional police units were integrated into one national force in 2013, police and policy stakeholders gradually moved away from the exclusive use of performance contracts. However, as this chapter will set out, the move to a national police force and shift away from performance contracts does not appear to have been accompanied by a shift towards a purely national approach to performance measurement within the police force.\textsuperscript{79}

The key stakeholders in measuring police performance in the Netherlands are the national and regional police units, which are responsible for annually reporting on their performance.

Since 2013, the force has been led by the Commissioner’s Office (\textit{korpsleiding}), headed by one Commissioner (\textit{korpschef}), and the Commissioner’s staff (\textit{staf korpsleiding}). The force further consists of ten regional units, one national unit, the Police Academy, as well as the Police Service Centre (Figure 2).\textsuperscript{80} In 2021, the police force employed more than 63,000 people, including over 50,000 operational officers.\textsuperscript{81}

- The regional units (\textit{regionale eenheden}) consist of districts, each of which is headed by a police chief (\textit{politiechef}) and divided into frontline teams (\textit{basis teams}). Each team is responsible for providing basic policing in a municipality (or in a cluster of smaller municipalities). A team may constitute a constable, neighbourhood police officers, detectives and team chiefs. Together they aim to ensure that the region is a safe and pleasant environment to live in. This includes answering emergency calls, conducting street patrols, conducting basic investigative activities and assisting the public.

\textsuperscript{72} Van Waarden and Boutylkova (2009).
\textsuperscript{73} Terpstra, Duchatelet, Jansens, Van Ryckeghem and Versteegh (2015).
\textsuperscript{74} Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (2003).
\textsuperscript{75} Jochoms (2014).
\textsuperscript{76} Jochoms (2014).
\textsuperscript{77} Ministerie van Binnenlandse Zaken en Koninkrijksrelaties (2003).
\textsuperscript{78} Van Sluis, Cachet and Ringeling (2008).
\textsuperscript{79} INT4_NL; INT5_NL.
\textsuperscript{80} Government of the Netherlands (2023a).
\textsuperscript{81} Politie Nederland (2021).
The **National Unit** (*Landelijke Eenheid*) is guided by the Chief Constable. The unit’s tasks partly support the regional units and consist of monitoring and coordinating major operations, combating organised crime, providing security support in the fight against crime, deploying mounted police and forensic expertise, providing security services to private individuals, organising specialised policing (for instance in the air), and combatting violence and terrorism.

Both the National Unit and the regional units are responsible for reporting on their performance annually against pre-defined targets (see Section 3.2).

The **Police Academy** (*Politieacademie*) is the education, knowledge and research institute for the Dutch police. It carries out research projects with the aim of improving and innovating police practices. Its researchers have produced several evaluations of earlier efforts to measure police performance in the Netherlands.  

The **Police Services Centre** (*Politiedienstcentrum*) is responsible for operational management services, such as IT, communications and human resources.

---

**Figure 2: Organisational structure of the Dutch police since 2013**

Other key stakeholders in police performance measurement are the Ministry of Justice and Security, which formulates the strategic objectives for the Dutch police, and mayors and public prosecutors, who may set additional objectives for the police at a local level.

The police are under the administrative and political responsibility of the **Ministry of Justice and Security**, which sets the strategic objectives for the force at a national level (see Section 3.2.1). However, authority

---

82 See, for example, Jochoms (2014).
83 Government of the Netherlands (2023b).
International approaches to police performance measurement

over the police lies with the **mayor of the relevant municipality** (even after the move towards a national police force in 2013), who is also concerned with public order and safety, and with the **public prosecutor**, *(Officier van Justitie)*, who is concerned with criminal investigation and detection. Together with the local police commissioner or head of unit, these stakeholders are responsible for determining how the national framework set by the Ministry of Justice and Security is supplemented by local performance agreements and priorities (see Section 3.2.2).

**The Inspectorate of Justice and Security plays an important role in police performance measurement as it is responsible for conducting research on the quality of the services provided by the police.**

The Inspectorate of Justice and Security *(Inspectie Justitie en Veiligheid)*, which has operated in its current form since 2012, is responsible for monitoring and evaluating the performance of each executive body that falls under the administrative responsibility of the Ministry of Justice and Security, including the police. It conducts independent studies into the quality of the processes of the Dutch police force regarding detection, enforcement and support in emergencies and the effect of these processes on citizens. Through its publications in these areas, it seeks first and foremost to increase the capacity of the police force to assess which aspects of its work require changes, and subsequently to improve its practices.

**Additional peripheral stakeholders that assess the performance of the Dutch police are the Court of Audit and the Office of Management Information.**

The Dutch **Court of Audit** *(Algemene Rekenkamer)* monitors national government spending for efficiency and legality. In this role, it also scrutinises the functioning of the national police and routinely carries out studies that analyse its performance.

The data analysts of the **Office of Management Information** *(OMI)* regularly generate data insights from the various police systems used in the Netherlands. Based on these strategic insights, they produce dashboards for police leadership.

**Data collected for this study indicate that ideas concerning the purpose of police performance measurement are multifaceted in the Netherlands, as the practice serves different functions for different key stakeholders.**

Due to their varying responsibilities and concerns, key stakeholders have developed different ideas about why it is important to measure the performance of the Dutch police and, consequently, about what should be measured.

The government appears to have remained largely focused on the importance of proving efficiency, i.e. measuring the services the police can provide for a certain amount of investment, even after the move away from performance contracts. Two interviewees reported that the Ministry of Justice and Security uses

---

84 Government of the Netherlands (2023b).
85 Inspectorate of Justice and Security (2023a).
86 Inspectorate of Justice and Security (2023b).
87 Rijksinspecties (2023).
88 Algemene Rekenkamer (2016).
89 INT6_NL.
90 INT3_NL.
performance measurement primarily to demonstrate the return on the government’s annual financial investment in the police. This is in line with the Governments Account Act 2016, which outlines that ministers are responsible for the ‘efficiency, regularity, orderliness and auditability of the acquisition, management and disposal of material resources’. Furthermore, a recent report by the Bosman Committee, a temporary committee set up to investigate the causes of various problems within public organisations, found that public institutions, including ministerial agencies, remain wedded to the principles of New Public Management, an approach to performance measurement that focuses on the importance of measuring performance to demonstrate quantitative factors such as cost-effectiveness and value for money. The Bosman Committee argued that in many cases no suitable alternative has been identified and/or adopted. The Ministry of Justice and Security uses its framework for performance measurement, the Security Agenda, to give strategic direction to the Dutch police. As set out in Section 3.2.1, this document likewise primarily emphasises quantitative indicators.

Nevertheless, there has been a separate movement within the Dutch police organisation that emphasises the need for measuring performance not only on the basis of whether the organisation has met pre-determined quantitative targets, but also on qualitative feedback gathered from officers and the public through consultations. This project, Rijker Verantwoorden or ‘richer accountability’, starts from the position that the purpose of police performance measurement is accountability to the public, and that quantitative targets alone do not allow the public to get a full picture of how the police are functioning.

This approach was first introduced in 2014, and has since been developed further as different police units have experimented with various forms of stakeholder consultation. The intention of such sessions is to improve organisational reflection on the objectives and results of police work with the aim of steering future efforts and increasing police legitimacy in the eyes of the public. It should be noted that the latest budget and management plan published by the national police outlines a plan to explore how the principles of Rijker Verantwoorden could be integrated into the national Security Agenda (Veiligheidsagenda), for instance in the context of the force’s work against human trafficking.
International approaches to police performance measurement

At the same time, there has also been criticism regarding how Rijker Verantwoorden defines the objective of performance measurement.\textsuperscript{104} One academic expert interviewed for this study noted that measuring performance should be desired by police stakeholders to ensure that the force is providing a high-quality service, not just accountability to the public.\textsuperscript{105} The same interviewee argued that the approach taken by Rijker Verantwoorden causes police stakeholders to focus more on how the public perceive the performance of the organisation, rather than on what it is actually delivering. This leaves the police force in a vulnerable position. Indeed, by neglecting to use quality benchmarking, the practice of comparing current efforts to best practices in the field of police performance, the organisation cannot itself determine whether its performance is adequate and make appropriate adjustments if needed.\textsuperscript{106}

Data collected for this study indicate that there are different ideas in the Netherlands about why and how the performance of the Dutch police should be measured, with opinions including a need to demonstrate value for money by means of an emphasis on quantitative targets, increasing police accountability and legitimacy through consultation with stakeholders, and ensuring the provision of a high-quality service through benchmarking.

3.2. Current practices in measuring police performance

Performance indicators and targets are formulated at both regional and national levels by different key stakeholders, i.e. those in a position to influence policing policies (see Section 2.2.4), as outlined below. At the national level, the Dutch police appear to primarily use data extracted from their registration systems to assess their performance. However, regional units and additional peripheral stakeholders, such as the Central Bureau for Statistics, use data from citizen surveys and consultations to measure levels of satisfaction with the police and demonstrate the performance of the force.

3.2.1. Indicators and targets

At the national level, performance measurement is guided by the indicators and targets defined in the \textit{Ministry of Justice and Security’s Security Agenda}.

Every four years, the Ministry of Justice and Security sets the strategic priorities for the police in the Security Agenda. This document highlights several themes that the government considers most urgent, which in its most recent iteration are organised crime, human trafficking, cybercrime, social cohesion and societal unrest, and intelligence and international cooperation (set out in Table 13). These themes are linked to specific agreements between the ministry and the police, all of which are linked to quantitative or qualitative targets (likewise set out in Table 13).\textsuperscript{107} While the quantitative targets are linked to specific quantitative performance indicators, no qualitative indicators have been formulated to accompany the qualitative targets. This demonstrates that while the ministry is interested in qualitative targets, it has not developed any tangible means of measuring performance against them.

\textsuperscript{104} Kruisbergen (2022).
\textsuperscript{105} INT4_NL.
\textsuperscript{106} INT4_NL.
\textsuperscript{107} Rijksoverheid (2022).
Results are subsequently reported on by the police in its Annual Account (*Jaarverantwoording*), which refers to the objectives, indicators and targets laid out in the Security Agenda and which is sent to the Dutch Senate and House of Representatives.108

**Table 13: Summary of the 2023–26 Security Agenda**

<table>
<thead>
<tr>
<th>Strategic theme</th>
<th>Objective</th>
<th>Indicator</th>
<th>2023 target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Undermining practices and organised crime</td>
<td>Tackling criminal syndicates</td>
<td>The annual number of criminal syndicates dealt with through the criminal justice system.</td>
<td>1,530 criminal syndicates dealt with through the criminal justice system.</td>
</tr>
<tr>
<td></td>
<td>Tackling criminal financial flows</td>
<td>The annual value of confiscated assets (in million euros).</td>
<td>€190,000,000.</td>
</tr>
<tr>
<td>Human trafficking</td>
<td>Further development of online detection capabilities</td>
<td>N/A</td>
<td>To provide national coordination and central signalling and intervention in various forms of exploitation.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N/A</td>
<td>To strengthen cooperation with partners regarding the detection of exploitation taking place online.</td>
</tr>
<tr>
<td></td>
<td>International cooperation</td>
<td>N/A</td>
<td>To contribute to a dynamic international picture of the phenomenon of human trafficking.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>N/A</td>
<td>To develop an international programme of targeted interventions.</td>
</tr>
<tr>
<td></td>
<td>Prosecution of trafficking suspects</td>
<td>The annual number of human trafficking suspects delivered to the Public Prosecutor.</td>
<td>220 suspects delivered to the Public Prosecutor.</td>
</tr>
<tr>
<td></td>
<td>Commitment to cross-unit cooperation</td>
<td>N/A</td>
<td>The police to develop a cross-unit prioritisation strategy and work across units.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The annual number of cross-unit investigations per year.</td>
<td>14 cross-unit investigations per year.</td>
</tr>
<tr>
<td>Cybercrime and digital crime, including online child sexual abuse</td>
<td>Cybercrime</td>
<td>The annual number of regular cybercrime suspects.</td>
<td>310 cybercrime suspects.</td>
</tr>
</tbody>
</table>

---

108 Politie Nederland (2022b).
## International approaches to police performance measurement

<table>
<thead>
<tr>
<th>Strategic theme</th>
<th>Objective</th>
<th>Indicator</th>
<th>2023 target</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Digital crime</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The annual number of suspects of regular digital crimes registered.</td>
<td>2,200 suspects registered.</td>
<td></td>
</tr>
<tr>
<td></td>
<td>The annual number of studies into criminal phenomena related to digital crime and offender groups.</td>
<td>5 studies.</td>
<td></td>
</tr>
<tr>
<td><strong>Online sexual child abuse</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The annual number of interventions.</td>
<td>600 interventions (of which at least 130 into producers and abusers, 20 into key players and networks).</td>
<td></td>
</tr>
<tr>
<td><strong>Societal connectivity and social unrest</strong></td>
<td>Yearly conversations about the deployment and effectiveness of the police.</td>
<td>N/A</td>
<td>Regional units to discuss the deployment and effectiveness of police with different societal stakeholders, including neighbourhoods and target groups.</td>
</tr>
<tr>
<td></td>
<td>Development of a toolbox of best practices and integrated intervention strategies</td>
<td>N/A</td>
<td>To develop a toolbox of best practices and integrated intervention strategies.</td>
</tr>
<tr>
<td></td>
<td>Development of an information strategy for dealing with civil unrest</td>
<td>N/A</td>
<td>To develop information strategy for dealing with civil unrest.</td>
</tr>
<tr>
<td><strong>Intelligence and international cooperation</strong></td>
<td>Intelligence</td>
<td>N/A</td>
<td>To contribute on all themes mentioned in the Security Agenda in the form of intelligence reports.</td>
</tr>
<tr>
<td></td>
<td>International cooperation</td>
<td>N/A</td>
<td>To strengthen international cooperation, especially to tackle undermining practices.</td>
</tr>
</tbody>
</table>

The framework for national performance measurement established by the Security Agenda is supplemented by regional and local performance agreements.
The national framework outlined above is supplemented by regional and local agreements and priorities, which often determine what aspects of regional police force work are measured and in what manner. At a regional level, separate strategic agendas and frameworks for performance measurement are formulated to complement the national Security Agenda (see Table 13 above). These agendas are agreed upon by the ‘regional authority triangle’, which is a network of the regional mayor (regioburgemeester, the representative of all mayors in the region), the chief public prosecutor for the region and the police chief of the regional unit. Each regional unit also produces a separate annual report in which it reports on its performance in relation to objectives, indicators and/or targets outlined in both the national and regional strategic priorities.\textsuperscript{109} The regional unit for Rotterdam, for instance, has formulated additional indicators and targets to accompany those set out in the Security Agenda, as detailed in Table 14.

Table 14: Additional indicators and targets formulated by the regional unit for Rotterdam

<table>
<thead>
<tr>
<th>Theme</th>
<th>Indicator</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Police for everyone</td>
<td>Levels of satisfaction with the functioning of the police in general, measured by means of the Police Monitor and compared to the last measurement.\textsuperscript{110}</td>
<td>An improvement in levels of satisfaction with the functioning of the police in general, measured by means of the Police Monitor and compared to the last measurement.</td>
</tr>
<tr>
<td></td>
<td>Levels of satisfaction with the functioning of the police in neighbourhoods, measured by means of the Police Monitor and compared to the last measurement.</td>
<td>An improvement in levels of satisfaction with the functioning of the police in neighbourhoods, measured by means of the Police Monitor and compared to the last measurement.</td>
</tr>
<tr>
<td>Citizen central, victim first</td>
<td>Response times (category 1: urgency, use of siren and lights).</td>
<td>At least 82% of category 1 incidents responded to within 15 minutes.</td>
</tr>
<tr>
<td></td>
<td>Response times (category 2: less urgent, usually no use of siren or lights).</td>
<td>At least 77% of category 2 incidents responded to within 30 minutes.</td>
</tr>
<tr>
<td></td>
<td>Levels of satisfaction with last contact with the police when filing a report, measured by means of the Police Monitor and compared to the last measurement.</td>
<td>An improvement in levels of satisfaction with last contact with the police when filing a report, measured by means of the Police Monitor and compared to the last measurement.</td>
</tr>
<tr>
<td>Capacity for police deployment</td>
<td>Staffing of neighbourhood officers expressed in actual full-time equivalents (FTEs) as a percentage of available FTEs.</td>
<td>At least 97% of available FTEs for neighbourhood officers filled.</td>
</tr>
<tr>
<td></td>
<td>Sick leave taken, expressed as a percentage of the total working days in a calendar year.</td>
<td>&lt;6.8% of total working days.</td>
</tr>
</tbody>
</table>

\textsuperscript{109} See, for example, Politie Eenheid Rotterdam (2022).

\textsuperscript{110} The Police Monitor is a survey instrument developed to measure the quality of interactions between citizens and the police, and its impact on levels of trust in and satisfaction with the police. It was introduced for the frontline teams in the Rotterdam region in March 2021. Politie Eenheid Rotterdam (2022).
<table>
<thead>
<tr>
<th>Theme</th>
<th>Indicator</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>The percentage of officers who have completed the ‘Profcheck’, a course that tests and trains the officer’s knowledge of relevant legislation and skills.</td>
<td>100% of officers have completed the ‘Profcheck’.</td>
</tr>
</tbody>
</table>

Source: Politie Eenheid Rotterdam (2022).

There appears to be limited consistency in the form that the regional frameworks for performance measurement take. While the unit for Rotterdam has formulated specific additional indicators and targets, other regional units, such as The Hague and Netherlands East, appear to have only formulated additional objectives, without clear indicators or targets, which means that these units are unable to measure how they are performing in an objective and reliable manner.111,112 This discrepancy in approaches to performance measurement means that there is limited scope for comparison between different regional units.

At the local level, additional strategic priorities and targets are formulated and agreed upon by the ‘local authority triangle’ (lokale gezagdriehoek), a network of the relevant mayor, the public prosecutor, and the local police commissioner or head of team.113,114 However, the targets and indicators set at this level are not usually reported on publicly or at a national level.115

**Interviewees suggested that key national stakeholders do not have a clear picture of the performance measurement indicators and targets currently used across the country, as the design of these metrics is partly left to regional and local stakeholders.**

Data from five interviews suggest that neither the Commissioner’s staff, the Ministry for Justice and Security, nor the Court of Audit have a clear sense of how the broad strategic aims of the police are currently being linked to indicators across the different police organisations, nor what measurement methods are currently being used across the country.116 This is partly due to the fact that, as described above, performance measurement of the Dutch police has to some extent been decentralised, with regional units setting additional targets to complement those laid out in the Security Agenda.

Previous literature on the topic of police performance in the Netherlands paints a similar picture. One report identified more than 1,700 different indicators, many of which were found to be challenging for an evaluative framework as they were formulated as policy directions or objectives, rather than as measurable indicators.117 For example, ‘working more efficiently’ and ‘operating with more authority in public spaces’.118

---

111 Eenheid Oost Nederland (2022).
113 INT5_NL.
114 Regioburgemeesters (2023).
115 INT5_NL.
116 INT1_NL; INT2_NL; INT4_NL; INT5_NL; INT6_NL.
Reporting police performance in relation to certain topics, such as preventative efforts and the quality of services provided, remains limited at both national and regional levels.

The national unit’s Annual Account only tracks the force’s performance against the targets formulated in the Security Agenda.\textsuperscript{119} Given that the Security Agenda primarily sets crime-related targets (see Table 13), this means that there is limited reporting on results relating to other aspects of policing work, such as prevention and provision of services. While the Annual Account gives a brief overview of work in this area, it does not provide objective insights into what these developments can tell us about police performance by tracking efforts against predefined targets.

These aspects of policing appear to be measured against tangible targets by some, but not all, regional units. As highlighted above, while the regional unit for Rotterdam has included indicators and targets related to the quality of services provided (see Table 14), other regional units, such as The Hague and the Netherlands East, appear to have only formulated additional objectives.

Nevertheless, interviewees confirmed that data on such policing tasks are collected continuously.\textsuperscript{120} Some of these additional data sets, for instance on response times, are available via the Dutch police’s online data portal.\textsuperscript{121} However, this information is limited and there is no indication of the annual targets for these measures and whether they have been met.

These findings are in line with conclusions by the Inspectorate of Justice and Security, which periodically publishes large-scale reviews of specific aspects of police work, such as detection,\textsuperscript{122} enforcement,\textsuperscript{123} support in emergency situations\textsuperscript{124} and the functioning of the national unit.\textsuperscript{125,126} The most recent iterations of these publications have expressed profound criticism at the force’s capacity for quality assessment and performance measurement. In 2019, the inspectorate concluded that:

\begin{quote}
…the police currently have no system for quality management and conduct limited evaluations. As a result, the organisation’s ability to learn from practice is not sufficiently secured and the necessary conditions for a learning organisation\textsuperscript{127} are lacking.\textsuperscript{128}
\end{quote}

In response to these findings, the inspectorate launched a separate study into the force’s current quality management practices, which is currently under way.\textsuperscript{129}

\textsuperscript{119} Politie Nederland (2021).
\textsuperscript{120} INT5_NL; INT6_NL.
\textsuperscript{121} Politie Nederland (2023). ‘Landelijke Indicatoren.’
\textsuperscript{122} Inspectorate of Justice and Security (2019).
\textsuperscript{123} Inspectorate of Justice and Security (2019).
\textsuperscript{124} Inspectorate of Justice and Security (2021a).
\textsuperscript{125} Inspectorate of Justice and Security (2021b).
\textsuperscript{126} Inspectorate of Justice and Security (2022a).
\textsuperscript{127} The concept ‘learning organisation’ is widely used in the Netherlands to denote an organisation that seeks to enhance the capacities of its members to achieve institutional objectives. The term was first coined by Peter Senge (1990).
\textsuperscript{128} Inspectorate of Justice and Security (2019).
\textsuperscript{129} Inspectorate of Justice and Security (2022b).
3.2.2. Data collection

To report on performance at both national and regional levels, data are collected using police registration systems.

Since 2009, the primary incident registration system for frontline police officers is the Basisvoorziening Handhaving (BHV), which can be used to register incidents, process reports and create criminal records. BHV information is accessed in de Geïntegreerde Databank voor Strategische bedrijfsvoering (GIDS) and/or Cognos, the police management information systems.\textsuperscript{130} Data from the BHV relating to the number of incidents per crime type is used to report on police performance in the Annual Account,\textsuperscript{131} and to determine whether the various targets for the number of registered suspects laid out in the Security Agenda have been met.\textsuperscript{132}

However, according to two interviewees, one of the main complexities with collecting data is the fact that other sections within the Dutch police use different registration systems.\textsuperscript{133} This means that an incident may be recorded in four or five different systems before it reaches the public prosecutor. However, none of the systems were designed to complement each other and each system has its own particular data model.\textsuperscript{134} Moreover, these systems are not geared towards generating management information that would allow for data insights on performance to be easily gathered.\textsuperscript{135} These are challenges that the Office of Management Information (OMI) seeks to overcome by gathering and triangulating data from the various police systems in use in the Netherlands to produce dashboards for police leadership. One interviewee noted that they believe the OMI is mostly successful in this regard, and that a transition towards one single system across different sections of the police organisation is therefore not a priority.\textsuperscript{136}

Data on public perceptions of police performance are collected via several citizen surveys, but this does not appear to be used consistently by the Dutch police to measure its performance.

Data are also collected by means of the biennial Security Monitor (Veiligheidsmonitor), an online citizen survey commissioned by the Ministry of Justice and Security and the Central Bureau for Statistics (CBS). However, the study team did not find any evidence that the data gathered are used by the police to track performance against particular indicators and/or targets.

The survey consists of a large national sample (n=173,000 in 2021). In addition, individual municipalities can opt for more surveys among residents of their municipality (oversampling).\textsuperscript{137} The resulting data sets are publicly available and allow for responses to be compared across the 10 regional units, 43 districts and 167 frontline teams on a range of themes.\textsuperscript{138} Quantitative data are collected on a broad range of topics, such as safety, liveability and experiences of crime victimisation, as well as the extent to which citizens are satisfied

\textsuperscript{130} Centrum voor Criminaliteitspreventie en Veiligheid (n.d.).
\textsuperscript{131} Politie Nederland (2021).
\textsuperscript{132} Politie Nederland (2021).
\textsuperscript{133} INT5_NL; INT6_NL.
\textsuperscript{134} INT6_NL.
\textsuperscript{135} INT6_NL.
\textsuperscript{136} INT6_NL.
\textsuperscript{137} CBS (2021).
\textsuperscript{138} Veiligheidsmonitor (2022).
with the quality of their interactions with the Dutch police. Table 15 illustrates the indicators regarding
citizen satisfaction used in the 2021 iteration of the Security Monitor.

Table 15: Police-related indicators – Security Monitor 2021

<table>
<thead>
<tr>
<th>Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>Satisfaction with contact with the police at municipal level in the context of enforcement.</td>
</tr>
<tr>
<td>Satisfaction with contact with the police at municipal level in the context of reporting an incident or crime.</td>
</tr>
<tr>
<td>Satisfaction with contact with the police at municipal level in the context other reasons for contact.</td>
</tr>
<tr>
<td>Satisfaction with police functioning at neighbourhood level.</td>
</tr>
<tr>
<td>Satisfaction with police functioning in general.</td>
</tr>
<tr>
<td>Satisfaction with police visibility at neighbourhood level.</td>
</tr>
</tbody>
</table>


Recent versions of the monitor have also included questions concerning online crime, prevention,
discrimination and transgressive behaviour. In 2021, a pilot study was conducted to explore whether
additional questions could be added to the Security Monitor based on equity indicators that, for example,
measure perceived disproportionality in interactions with police officers. This has yet to become a regular
feature of these surveys.

However, it is unclear from the findings of the study how the data from the Security Monitor relate to the
broader framework for performance measurement used by the Dutch police, as set out in the Security
Agenda. As Table 13 demonstrates, none of the indicators set out in the Security Agenda relate to citizens’
perceptions of how well the police carries out its remit, or indeed the public’s confidence in the Dutch
police, nor are the survey results mentioned in the Annual Account. Furthermore, interviewees were
unable to say in what capacity the results from the Security Monitor are used to assess the performance of
the police. As such, it appears that the monitor is not used by the police organisation in its framework for
performance measurement.

Nonetheless, the results from the monitor are used outside of the police organisation to measure police
performance in terms of citizen satisfaction. Following the 2021 iteration of the survey, the Central Bureau
of Statistics published a report that includes a chapter in which results relating to the indicators set out in
Table 15 are analysed. Moreover, data from the monitor have also been used by both the Inspectorate of

139 CBS (2021).
142 INT3_NL; INT5_NL; INT6_NL.
143 Akkermans et al. (2022).
International approaches to police performance measurement

Justice and Security and academic researchers in the Netherlands to speak about the performance of the Dutch police regarding quality of services and interactions with the public.\textsuperscript{144,145,146,147}

In recent years, quantitative data regarding citizens’ perceptions of police performance have also been gathered using a ‘Confidence and Reputation Monitor’ (\textit{Vertrouwen- en Reputatiemonitor}), a separate online survey commissioned by the Dutch police that asks a sample of citizens (\textit{n}=3,000 in 2021) how they view the police, the organisation’s role in society and the quality of interactions with the police.\textsuperscript{148} However, the study team was also unable to establish if the data from this survey are used for organisational performance measurement. The results are not used in relation to police performance in the Annual Account\textsuperscript{149} of the national unit nor in any of the regional annual accounts, but are instead published in a stand-alone report.\textsuperscript{150}

Finally, while the performance framework for the regional unit of Rotterdam indicates that the results of a ‘Police Monitor’ are tracked against pre-defined targets (see Table 14), the study team has been unable to establish whether other regional units make use of similar survey instruments. As such, it appears that data on levels of trust in and satisfaction with the police are not currently used systematically at either the national or the regional level to measure performance.

3.3. Challenges and complexities in measuring police performance

The reliance of the Dutch police on quantitative indicators may increase the risk of perverse incentives.\textsuperscript{151,152,153}

The data gathered by means of the instruments described above, notably the registration systems, concern purely quantitative measurements, such as the number of suspects delivered to the Public Prosecutor and the number of high-tech crime investigations carried out, as well as qualitative constructs measured quantitatively, such as levels of satisfaction with the functioning of the police as measured through the Police Monitor (in Rotterdam). It appears that relatively little qualitative data, such as written or oral feedback from citizen consultations or case studies, are currently being collected and reported on (a point also made by the \textit{Rijker Verantwoord})\textsuperscript{154} project).\textsuperscript{154} Existing literature on the topic of police performance measurement indicates that a reliance on primarily quantitative data may increase the risk of perverse incentives, meaning that indicators of police performance inadvertently encourage undesirable results.\textsuperscript{155,156}

For example, as set out in Table 13, the Security Agenda includes ‘the annual number of suspects of regular

\textsuperscript{144} Inspectorate of Justice and Security (2019).
\textsuperscript{145} Inspectorate of Justice and Security (2021a).
\textsuperscript{146} Eysink Smeets and Baars (2016).
\textsuperscript{147} Stals and Van der Walle (2020).
\textsuperscript{148} Kantar Public (2022b).
\textsuperscript{149} It should be noted that the latest iteration of the Annual Account mentions that a new ‘Police Monitor’ is currently being developed to improve the force’s ability to measure how satisfied citizens are with the services that the police provide. However, the document does not specify how the results from this new survey instrument will be used to track and report on performance, or if particular targets will accompany its introduction. See Politie Nederland (2022b).
\textsuperscript{150} Kantar Public (2022b).
\textsuperscript{151} INT3_NL.
\textsuperscript{152} Faull (2010).
\textsuperscript{153} Davis (2012).
digital crimes registered’ as a performance indicator to which specific targets are attached. However, evidence demonstrates that measures relating to the number of registered crimes may not always be representative. One study based on a survey (n=1,480) of members of the Dutch Police Federation (Nederlandse Politiebond, NPB), published in the independent weekly news magazine, *De Groene Amsterdammer*, showed that Dutch police officers have in the past declined to register cases they feel may not result in prosecutions, and which could therefore negatively impact their quality indicators.\textsuperscript{154} Conversely, Table 13 also includes the indicator, ‘the annual number of suspects delivered to the Public Prosecutor’, as previous international studies on the topic have shown, using this figure as an indicator of performance may encourage officers to maximise the number of cases passed on to the Public Prosecutor by transferring cases even when there is insufficient evidence.\textsuperscript{155,156,157} While this may improve clearance rates, it may also lead to relatively more cases being dismissed.\textsuperscript{158} Looking at the ‘accuracy’ rates of cases passed to the Public Prosecutor, meaning the frequency with which decisions made by the police are eventually shown to be correct, could provide valuable insights in this regard.

### 3.3.1. Issues with the reliability, validity and representativity of data

The reliance of the Dutch police on quantitative data to measure performance may mean that some aspects of police work, particularly tasks related to the quality of provided services, are not captured. As noted above, the data on which police performance measurement in the Netherlands currently relies are primarily quantitative. Beyond the risk of creating perverse incentives, this may also mean that valuable information is not captured, as generalised questions on satisfaction with the police may not allow for reporting on precisely what it is that citizens are satisfied or unsatisfied with.\textsuperscript{159}

Similarly, the strong focus on numerical expressions of performance gathered through the registration systems means that some tasks that police officers in the Netherlands perform daily may be ignored as they are not easily captured by such measures. This is particularly the case for service-related activities, such as the conversations a neighbourhood police officer may have with members of the public to resolve a conflict and de-escalate a situation, or to provide information.\textsuperscript{160} Such tasks often go unregistered and are thus not captured by performance metrics.\textsuperscript{161} Instruments such as interviews, focus groups, contact surveys with a qualitative element or direct observations, which are not currently regularly used, may provide meaningful additional data in this regard, as well as actionable insights to facilitate organisational learning.\textsuperscript{162}

As has been noted above, equity indicators, such as those regarding perceived disproportionality in interactions with police officers, are not routinely applied in designing the surveys currently used to collect

\textsuperscript{154} Van de Beld et al. (2019).
\textsuperscript{155} Zouridis (2016).
\textsuperscript{156} Baughman (2020).
\textsuperscript{157} Leipold (2021).
\textsuperscript{158} Leipold (2021).
\textsuperscript{159} Maslov (2016).
\textsuperscript{160} INT3_NL.
\textsuperscript{161} Zouridis (2016).
\textsuperscript{162} INT3_NL; INT5_NL.
International approaches to police performance measurement

data on police performance. The systematic inclusion of such indicators may allow for the development of a more comprehensive picture of the performance of the Dutch police.

**There is a desire to expand the collection of data on digital activity relating to citizen perceptions of the police.**

Interviewees noted that the police organisation could do more to integrate analysis of digital activity and social media into its efforts to measure performance. Currently, data analysts within the police make some use of ‘web scraping’ (the process of data extraction from websites) to gauge societal sentiment towards the police, but there is a desire to expand these efforts. To this end, analysts are working on the development of a ‘Sentiment Monitor’ in collaboration with the CBS that would collect further data to improve efforts to detect emerging risks and gather information on how the public views police performance.

**Data collected through police registration systems may sometimes be inconsistent and inaccurate.**

In the Netherlands, as in most police jurisdictions, the quality of the data entered is largely dependent on the officer who enters it. Interviewees have indicated that this can cause issues with consistency and accuracy, as different individuals may have a different understanding of what should be recorded and how it is recorded. This may be aggravated by the fact that the Basisvoorziening Handhaving has been criticised by the police itself for not being user-friendly, increasing the likelihood of errors. Moreover, as previously mentioned, different sections within the Dutch police use different registration systems, which, according to interviewees, often leads to further inconsistencies in the data. The Office of Management Information seeks to mitigate this by using data science techniques to clean and triangulate data from the various police systems in use to produce dashboards for police leadership.

### 3.3.2. Other challenges or complexities

Beyond issues with data quality, the Dutch police face several challenges common to police performance measurement, namely the difficulty of proving causality and measuring prevention.

Two interviewees indicated that the outcome of policing, i.e. the societal benefits derived from actions taken by police officers, is often not used to assess police performance in the Netherlands as it is considered challenging to do so meaningfully. They highlighted that while throughput (the execution of police work) and output (the measurable results of that work) are usually directly controlled by the police and its partners (the Ministry of Justice and Security, mayors and public prosecutors), outcomes may be influenced by a range of societal factors. This renders the demonstration of causal links between factors such as

---

163 INT5_NL; INT6_NL.
164 INT6_NL.
165 INT6_NL.
166 INT5_NL; INT6_NL.
167 Politie Nederland (2017).
168 INT5_NL; INT6_NL.
169 INT6_NL.
170 INT5_NL; INT6_NL.
171 Ministerie van Financiën (2017).
throughput, output and outcome particularly challenging. One interviewee noted that the Dutch police has not yet found a way of overcoming this complexity to allow outcomes to be considered when measuring performance. The difficulty of demonstrating causality has proven to be a challenge not just in the Netherlands, but also in other jurisdictions, as has been well documented in literature on the topic. This issue of demonstrating effect was brought up especially in the context of measuring the success of preventative efforts by the police. A 2016 WODC study recognised this challenge and proposed a ‘prevention index’ that would comprise ‘the number of warnings and recommendations issued to prevent crimes’, ‘the number of “disruptions” to suspected crimes’, ‘the number of young offenders whose cases are decided on by a consultative case evaluation’ and ‘the number of targeted surveillances’. However, these indices were never adopted. According to one interviewee involved in the report’s decision-making process, this was partly because the proposed framework relied too heavily on quantitative indicators.

3.4. Concluding comments

Several factors complicate police performance measurement in the Netherlands. First, there are different ideas among key stakeholders on the purpose of measuring police performance, and what ‘good’ performance looks like for the organisation. The Ministry of Justice and Security has emphasised that the police first and foremost need to be able to provide detailed insights into how the budget is spent and a comprehensive overview of the quality of services provided, as well as accountability to the public. However, some stakeholders within the police force argue that accountability should be the primary objective. Others have put forward that police performance should be measured to provide a comprehensive overview of the quality of services the organisation provides.

Second, at both a national and a regional level, police performance measurement in the Netherlands appears primarily dependent on quantitative data collected using the organisation’s registration systems. This could mean that meaningful insights are not captured. A further complexity is that different sections within the Dutch police use different registration systems, leading to inconsistencies in the data.

Third, data on public confidence in and satisfaction with the police are not systematically used across the national, regional and local units to measure police performance. As detailed in subsequent chapters, the use of survey instruments in jurisdictions such as Seattle (Chapter 7) and Sweden (Chapter 8) demonstrate that such data can provide valuable information on police performance.

172 Ministerie van Financiën (2017).
173 INT6_NL.
174 Davis (2012).
175 INT3_NL; INT6_NL.
177 INT5_NL.
178 INT5_NL.
179 Written comments from an employee at the Ministry of Justice and Security.
International approaches to police performance measurement

Finally, beyond issues with the quality of the data collected, the Dutch police also face several challenges common to police performance measurement elsewhere, such as the difficulty of proving causality. As detailed in Chapter 6, this was also highlighted as a complexity in North Rhine-Westphalia.

The subsequent chapters of this report detail the approaches adopted by a variety of police jurisdictions around the world. Chapter 10 sets out how these practices may offer actionable insights for police performance measurement in the Netherlands to mitigate some of these challenges.
4. Case study 1: England and Wales

Box 2: Key findings in England and Wales

| How did the methodology for the performance measurements come about? Were stakeholders involved? If so, which ones? |
| - At the national level, the Home Office and His Majesty’s Inspectorate of the Constabulary and Fire & Rescue Service (HMICFRS) have somewhat overlapping roles in measuring the performance of police forces. At the local level, each force’s police and crime commissioner (PCC) is responsible for assessing performance. |

| For what purposes are the performance measurements used? |
| - The three key stakeholders in this area set their own frameworks for measuring police performance, reflecting differing views on the police’s role and approaches to assessing their work. |

| How is police performance measured in England and Wales? What aspects of policing do the measurements cover? |
| - Differences in how aspects of police work are perceived and prioritised between national and local policing stakeholders are reflected in different approaches to measuring police performance. |
| - The Home Office’s National Policing and Crime Measures set out six national priority areas where forces’ performance is scrutinised. They also detail the metrics by which a force’s performance is measured in each area, and the data sources for each metric. |
| - The data sources used to assess performance under these measures are primarily quantitative, including crime data collected from forces and survey data on crime victimisation. |
| - HMICFRS uses an intelligence-led continuous assessment model, with ongoing engagement from dedicated HMICFRS force liaison leads assigned to particular forces, culminating in the production of an annual report on forces’ performance. |
| - While data sources such as recorded crime are used to inform police performance measurement in all inspections, the type of data and how they are used varies greatly by inspection. |
| - The third major stakeholder in measuring police performance is the force PCC, who has responsibility for, among other things, preparing the force’s police and crime plan and holding the chief officer accountable for performance under that plan. |
| - Police and crime plans are typically set out around five priority areas, with indicators assigned to measure the force’s performance in that area. |

| What are the (further) advantages and disadvantages of the performance measurements? Are any (side) effects of the performance measurements known? If so, which ones? |
| - Challenges for England and Wales include that data sources vary widely by forces, reflecting the differences in approaches to how performance is measured locally. |

| How reliable, valid and representative are the performance measurements? |
| - Interviewees expressed concerns with the reliability, validity and representativeness of recorded crime data. |

| Which (parts of) foreign performance measurements can be labelled as ‘qualitatively good’, ‘innovative’ and/or ‘promising’? |
| - The relationship between HMICFRS and forces was generally viewed as constructive by both sides, particularly since the inspection approach changed to ongoing, intelligence-led engagement, with dedicated HMICFR staff assigned to individual forces. |
4.1. Organisational context, stakeholders and evolution of performance measurement

There are 43 territorial police forces in England and Wales, as well as three specialist police forces (the British Transport Police, the Ministry of Defence Police and the Civil Nuclear Constabulary). Territorial forces cover areas that vary significantly in terms of population and geographic size, which means that force size, structure, priorities and resourcing also differ drastically. For example, the Metropolitan Police Service serves most of London (with a population of almost nine million people) and has almost 35,000 officers. By contrast, Lincolnshire Police is staffed by around 1,200 officers and covers a largely rural area of around 750,000 people.

Literature on policing in England and Wales often refers to the ‘British model of policing’, which – although not formally defined – typically refers to three foundational and interlinked concepts:

1. The Office of the Constable: Grants powers to all police officers to detect, prevent and investigate crime.

2. Operational independence: Governance arrangements for each force rests on the understanding that officers are under the ‘direction and control’ of the force’s chief officer, usually called the Chief Constable. Chief officers are independently responsible for all operational matters within their force. However, operational independence is not legally defined and is ‘fluid and context-driven’. As a result, chief officers are not exempt from political scrutiny or influence.

3. Policing by consent: Derived from instructions issued to new officers in the Metropolitan Police Service from 1829. These include instructions to ‘recognise always that the power of the police to fulfil their functions and duties is dependent on public approval of their existence, actions and behaviour and on their ability to secure and maintain public respect’ and to ‘seek and preserve public favour […] by constantly demonstrating absolutely impartial service to law’.

Closely related to the principle of policing by consent is the expectation that forces should operate with transparency. Reporting on and engaging with reviews of their performance is perceived as central to this expectation. This includes sharing data on their work with the Home Office (referred to as the ‘annual

---

180 Greater London Authority (2023).
181 Metropolitan Police (2023).
182 Lincolnshire Police (2022).
183 Lincolnshire County Council (2023).
184 Bacon (2013).
188 College of Policing (2023).
data requirement’, discussed further below), and publicly disclosing information on policing matters in their force area.

At the national level, the Home Office and HMICFRS have somewhat overlapping roles in measuring the performance of police forces, while each force’s PCC is responsible for assessing performance at the local level.

The **Home Office** is a government department with a range of powers that include matters relating to policing. Regarding police performance measurement, it produces the Strategic Policing Requirement, which is a statutory document that identifies priority national crime threats and sets out how forces should respond to them. The Home Office also prepares the annual funding grants for forces and funding available for national policing priorities, for approval by Parliament. It has a range of formal and informal powers to review police performance and require improvements where necessary. It publishes ad hoc strategy documents, including the recent Beating Crime Plan, and in 2021 set out the National Crime and Policing Measures, discussed in Section 4.2. The Home Office has also recently started working with HMICFRS to publish performance tables comparing forces’ performance in their response times to emergency calls.

**His Majesty’s Chief Inspector of Constabulary and Fire & Rescue Services (HMICFRS)** has a statutory duty to conduct independent inspections of all 43 territorial police forces in England and Wales. These inspections are governed by an assessment framework, described in Section 4.2, and entail continuous performance assessment that feeds into an annual force performance score. As an inspectorate, rather than a regulator, HMICFRS has the power to make recommendations to forces, but not to enforce them; that is the responsibility of the Home Office. It can, however, support and advise forces in developing and implementing improvements to their performance.

All but 4 of the 43 forces have **police and crime commissioners**, who are elected by the local community within the police force area and responsible for setting an annual budget for their force; developing the force’s five-year police and crime plan; appointing the force’s chief officer; scrutinising the force’s performance; and holding the chief officer accountable for the force’s performance under the police and crime plan.

### 4.2. Current practices in measuring police performance

The three main stakeholders in this area all set their own frameworks for measuring police performance, reflecting differing views on the police’s role and approaches to accurately assessing their work.

Interviewees noted that while these stakeholders consult each other and share some data sources, their work lacks coordination and alignment.
4.2.1. The Home Office

Between 2010 and 2020, the Home Office did not measure the performance of police in England and Wales, with the then home secretary expressing the view that ‘targets don’t fight crime; targets hinder the fight against crime’.193 This position was subsequently revised by her successor on the basis that the Home Office expected ‘improved outcomes’ following the government’s investment in police officer recruitment.194 In 2021, the Home Office published its Beating Crime Plan,195 a strategy document that sets out the government’s priorities and actions around cutting crime, increasing confidence in the criminal justice system and supporting victims of crime. The National Crime and Policing Measures were introduced alongside this plan.

The measures set out six national priority areas where forces’ performance is scrutinised. They also detail the metrics by which each force’s performance is measured in each area, and the data sources for each metric.

The stated aims of these measures are to ‘help focus effort on key national priorities, allow performance to be measured and help to demonstrate value for money in policing’.196 The Home Office has stated that these measures are not numerical targets but are ‘directional’ and intended to show forces’ progress against indicators, with the expectation that forces will embed new ways of working and show clear improvements against the indicators within three years.197 The measures also state that the Home Office has sought to:

… reflect, as far as possible, the breadth of crime activity with which the Police are faced, reducing the risk of significant areas being deprioritised. Where possible the measures are based on data that is collected independently, lowering the risk of perverse incentives and ‘gaming’.198

These measures, set out in Table 16, focus primarily on forces’ work on crime reduction.

Table 16: Home Office National Crime and Policing Measures

<table>
<thead>
<tr>
<th>Priority area</th>
<th>National metrics</th>
<th>Data sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduce murder and other homicide</td>
<td>Homicides</td>
<td>Police recorded crime</td>
</tr>
<tr>
<td>Reduce serious violence</td>
<td>Hospital admissions of under 25s for assault with a sharp object</td>
<td>National Health Service</td>
</tr>
<tr>
<td></td>
<td>Offences involving discharge of a firearm</td>
<td>Police recorded crime</td>
</tr>
<tr>
<td>Disrupt drugs supply and county lines</td>
<td>Drug-related homicides</td>
<td>Police recorded crime</td>
</tr>
<tr>
<td></td>
<td>Police referrals into drug treatment</td>
<td>Public Health England199</td>
</tr>
</tbody>
</table>

The data sources used to assess performance under these measures are primarily quantitative, including crime data collected from forces and survey data on crime victimisation.

Police recorded crime data are provided directly from all 43 forces by the Home Office as part of the ‘annual data return’ (ADR). While this information was once supplied annually, it is now supplied by forces on a weekly, monthly or quarterly basis, according to the type of data. A wide range of data are collected, including recorded crime, arrests, charges, use of stop and search powers, and demographic information on offenders and victims and police workforce. The Office for National Statistics (ONS) prepares and publishes these data in aggregate form on a quarterly and annual basis. The ADR is also used for a range of other research and statistical purposes, including HMICFRS police inspections, described below. Quantitative health-related data on the number of referrals of offenders by police into drug treatment services are supplied by an agency of the Department for Health and Social Care, and information on hospital admissions collected from the National Health Service is used to inform the measurement of reduction in serious violence.

Data from the Crime Survey for England and Wales (CSEW) are also used to understand crime victimisation and the satisfaction of victims with their experience with the police. The CSEW is an annual victimisation survey of about 50,000 households. It asks residents about their experiences of crime victimisation in the preceding 12 months, including crimes not reported to the police. The survey is conducted by a private research company on behalf of the ONS. The CSEW is considered a particularly useful source of insight into unreported crime, which previous survey results have shown amount to around

---

199 PHE was an executive agency of the Department of Health and Social Care and has been replaced by the Office for Health Improvement and Disparities.
201 INT3_EW.
International approaches to police performance measurement

60 per cent of all crime in England and Wales, and into public perceptions of the police. Survey data from the Department for Digital, Culture, Media and Sport are also used to understand the rates of cyber attacks or breaches. Regarding ‘tackling cybercrime’, the Cyber Aware Tracker mentioned in the measures is not currently in operation and no information is publicly available about if or when it will be introduced.

In June 2023, HMICFRS announced that it had consolidated data on the performance of all forces in England and Wales against three of the measures – reducing homicide, reducing serious violence and reducing neighbourhood crime – and published the data in a public-facing Digital Crime and Performance Pack. This information was already published by the ONS, but in a potentially difficult to interpret format. HMICFRS worked with the National Police Chiefs’ Council, the Association of Police and Crime Commissioners, the College of Policing, the National Crime Agency and the Home Office to produce the pack and plans to update it every quarter.

4.2.2. HMICFRS

HMICFRS has conducted various forms of police force inspection in England and Wales for over 160 years. It first introduced what are commonly referred to as PEEL (police efficiency, effectiveness and legitimacy assessments) in 2014, which were further developed in 2018, with forces assessed annually according to three ‘pillars’ of efficiency, effectiveness and legitimacy. In consultation with stakeholders that included the Home Office, the National Police Chief’s Council (the national coordination body for police leaders in the UK) and the College of Policing (the professional training and development body for police in England and Wales), HMICFRS again revised its approach to conducting inspections in 2021.

HMICFRS now uses an intelligence-led continuous assessment model, with ongoing engagement from dedicated HMICFRS force liaison leads assigned to individual forces, culminating in the production of an annual report on forces’ performance.

In the annual report, forces are scored on their performance against eight areas of policing, called ‘core questions’. Several indicators by which forces are assessed are set out for each core question, as well as several metrics to be considered by the inspector in judging the forces’ performance for each indicator. Forces receive one of the following scores for each core question:

- Outstanding: The force has substantially exceeded the characteristics of good performance.
- Good: The force has demonstrated substantially the characteristics of good performance.
- Adequate: An appreciable number of areas where the force should make improvements has been identified.
- Requires improvement: A sufficiently substantial number of areas where the force needs to make improvements has been identified.

203 Kantar Public (2022a).
205 INT2_EW.
206 INT3_EW.
• Inadequate: There are causes for concern and the inspector has made recommendations to the force to address them.²⁰⁷

Broadly, the aspects of policing covered in the core questions are the quality of forces’ engagement with victims of crime, vulnerable people and the broader community; its effectiveness in preventing and deterring crime; its effectiveness in investigating crime; its effectiveness in managing offenders; its effectiveness in disrupting serious and organised crime; and workplace culture and organisational management.

Table 17 sets out the core questions and associated indicators. The numerous and detailed metrics for each indicator are included in Annex B of this report.

**Table 17: HMICFRS PEEL assessment framework**

<table>
<thead>
<tr>
<th>Core question</th>
<th>Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. How good is the force’s service for victims of crime?</td>
<td>1.1 The force manages incoming calls, assesses risk and prioritises the police response well.</td>
</tr>
<tr>
<td></td>
<td>1.2 The force deploys its resources to respond to victims and incidents in an appropriate manner.</td>
</tr>
<tr>
<td></td>
<td>1.3 The force’s crime recording can be trusted.</td>
</tr>
<tr>
<td></td>
<td>1.4 The force has effective arrangements for the screening and allocation of crimes for further investigation and these take into account vulnerability.</td>
</tr>
<tr>
<td></td>
<td>1.5 The force carries out a proportionate, thorough and timely investigation into reported crimes, with senior level governance providing robust scrutiny.</td>
</tr>
<tr>
<td></td>
<td>1.6 The force makes sure that it follows national guidance/rules for deciding the outcome it gives for each report of crime.</td>
</tr>
<tr>
<td>2. How good is the force at engaging with the people it serves and treating them fairly, appropriately and respectfully?</td>
<td>2.1 The force engages with all its diverse communities to understand and respond to what matters to them.</td>
</tr>
<tr>
<td></td>
<td>2.2 The workforce understands why and how to treat the public with fairness and respect.</td>
</tr>
<tr>
<td></td>
<td>2.3 The workforce understands how to use stop and search powers fairly and respectfully.</td>
</tr>
<tr>
<td></td>
<td>2.4 The force understands and improves the way it uses stop and search powers.</td>
</tr>
<tr>
<td></td>
<td>2.5 The workforce understands how to use force fairly and appropriately.</td>
</tr>
</tbody>
</table>

²⁰⁷ HMICFRS (2022).
<table>
<thead>
<tr>
<th>Core question</th>
<th>Indicators</th>
</tr>
</thead>
</table>
| 3. How good is the force at preventing and deterring crime, anti-social behaviour and vulnerability? | 2.6 The force understands and improves the way in which it uses force.  
3.1 The force prioritises the prevention of crime, anti-social behaviour and vulnerability.  
3.2 The force uses problem solving and works in partnership to prevent crime, anti-social behaviour and vulnerability.  
3.3 The force understands the demand facing neighbourhood policing teams and manages resources in line with that demand. |
| 4. How good is the force at responding to the public?                         | 4.1 The force identifies and understands risk effectively at initial contact.  
4.2 The force provides an appropriate response to incidents, including those involving vulnerable people.  
4.3 The force understands the demand faced by officers responding to calls for service and manages its resources to cope with that demand.  
4.4 The force has a good understanding of the wellbeing needs of its contact management staff and officers initially responding to emergency calls. |
| 5. How good is the force at investigating crime?                              | 5.1 The force understands how to carry out quality investigations on behalf of victims and their families.  
5.2 The force understands the crime demand it faces and what resources it needs to meet it effectively.  
5.3 The force provides a quality service to victims of crime.  
5.4 The force manages the wellbeing of staff involved in investigations. |
| 6. How good is the force at protecting vulnerable people?                     | 6.1 The force understands the nature and scale of vulnerability (this includes all types of vulnerability e.g. older people, hate crime, mental ill health, domestic abuse, children, child sexual exploitation, modern day slavery and human trafficking).  
6.2 The force provides ongoing safeguarding and support for vulnerable people, including those at risk of criminal exploitation.  
6.3 The force works effectively and proactively with partners to reduce vulnerability and repeat victimisation.  
6.4 The force has a good understanding of demand and resources, including when working with other agencies. |
<table>
<thead>
<tr>
<th>Core question</th>
<th>Indicators</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>7. How good is the force at managing offenders and suspects?</strong></td>
<td>6.5 The force maintains and improves the wellbeing of staff involved in protecting vulnerable people and understands the effect of the action it is taking.</td>
</tr>
<tr>
<td></td>
<td>7.1 The force is effective in apprehending and managing suspects and offenders to protect the public from harm.</td>
</tr>
<tr>
<td></td>
<td>7.2 The force effectively manages the risk posed to the public by the most dangerous offenders.</td>
</tr>
<tr>
<td></td>
<td>7.3 The force has an effective Integrated Offender Management (IOM) programme.</td>
</tr>
<tr>
<td></td>
<td>7.4 The force understands the demand and has the resources it needs to manage suspects and offenders effectively.</td>
</tr>
<tr>
<td><strong>8. How good is the force at disrupting serious and organised crime?</strong></td>
<td>8.1 The force makes good use of all available intelligence to identify, understand and prioritise serious and organised crime (SOC) and inform effective decision making.</td>
</tr>
<tr>
<td></td>
<td>8.2 The force has the right systems, processes, people and skills to tackle SOC and keep the public safe.</td>
</tr>
<tr>
<td></td>
<td>8.3 Disruptive activity reduces the threat from SOC (Pursue).</td>
</tr>
<tr>
<td></td>
<td>8.4 The force prevents people from engaging or re-engaging in organised crime (Prevent).</td>
</tr>
<tr>
<td></td>
<td>8.5 Communities, organisations and individuals are resistant and resilient to the impact from serious and organised crime (Protect and Prepare).</td>
</tr>
<tr>
<td><strong>11. How good is the force at building, developing and looking after its workforce and encouraging an ethical, lawful and inclusive workplace?</strong></td>
<td>11.1 The force promotes an ethical and inclusive culture at all levels.</td>
</tr>
<tr>
<td></td>
<td>11.2 The force understands the wellbeing of its workforce and uses this understanding to develop effective plans for improving workforce.</td>
</tr>
<tr>
<td></td>
<td>11.3 The force maintains and improves the wellbeing of its workforce and understands the effect of the action it is taking.</td>
</tr>
<tr>
<td></td>
<td>11.4 The force is building its workforce for the future.</td>
</tr>
<tr>
<td></td>
<td>11.5 The force is developing its workforce to be fit for the future.</td>
</tr>
<tr>
<td></td>
<td>11.6 Proactive and disruptive action taken by the force and effective vetting management reduce the threat and risk posed by police corruption.</td>
</tr>
</tbody>
</table>
Scores for each force are published in an online dashboard along with written reports from all inspections. Where forces have been assessed to have underperformed in some or all of the core questions, they move into ‘engaged status’, often informally referred to as ‘special measures’. According to an interviewee from HMICFRS, the force’s chief officer and PCC must then prepare an improvement plan to present to an HMICFRS police performance oversight group, which offers feedback on the plan and practical advice and support to implement measures to improve the force’s performance. The force liaison officers continue to monitor performance as the improvement plan is implemented, and this may be reported on in the subsequent inspection report(s).

While data sources such as the ADR are used to inform police performance measurement in all inspections, the type of data and how they are used otherwise varies greatly by inspection.

A policymaker interviewee described the following process. The Home Office and the ONS share data from the ADR with HMICFRS on the force’s efforts in preventing, deterring and investigating crime, and how the force functions as an organisation. Other evidence sources include force management statements, which are self-assessments produced by each force annually setting out expectations on future policing demands and how the force plans to respond to them; data on emergency call answering times provided by British Telecom; and data on wanted persons from the Police National Computer provided by the Home Office. The inspections also use data gathered directly from forces or through fieldwork on-site in the force. These data include interviews and surveys with police officers and staff to understand how the force operates, and

---

208 INT2_EW.
perceptions of how it is performing against the core questions; observations of internal performance meetings, including the type and quality of data that forces use to measure their performance, and the information technology tools (such as Microsoft Power BI and Qlik Sense) used to support this process.\footnote{INT3_EW.} Inspections also include victim service assessments that follow victims’ journeys from reporting a crime through to the final outcome.\footnote{HMICFRS (2022).}

Two interviewees from HMICFRS reported that how these data are used varies by inspection, reflecting how closely tailored the assessments are to the profile of the force, findings from their previous HMICFRS inspections, and the data available around their performance. One interviewee commented that ‘…everything’s being inspected against the same framework, but the action inspection activity could look quite different’\footnote{INT2_EW, INT3_3.}. Another interviewee, however, highlighted the perception among some policing stakeholders that how HMICFRS uses these data in judging performance is ‘opaque’,\footnote{INT1_EW.} and that information on methodology in the publicly available assessment reports is scant.\footnote{See, for example, the report on the PEEL inspection of Avon and Somerset Constabulary 2021/2022, HMICFRS (2023).}

4.2.3. Police and crime commissioners (PCCs)

The third major stakeholder in measuring police performance is each force’s PCC, who has responsibility for, among other things, preparing the force’s police and crime plan, and holding the chief officer accountable for the force’s performance under that plan.

Data from police interviewees and the literature review show that different forces’ approaches to the process of developing and measuring performance against their police and crime plan varies markedly, reflecting the operational independence that is a part of the British model of policing. As such, it is difficult to make generalisations about the design of indicators or data sources used by PCCs. For instance, three interviewees described a range of approaches that PCCs in different forces have adopted to developing the plan for their force.\footnote{INT_6, INT_7EW, INT_EW8.} These include an extensive process of public consultation led by the PCC office to help identify priority areas and associated performance indicators, PCCs deferring to the force’s chief officer in setting priorities and building a performance framework around those priorities, and a collaboration between the PCC and police personnel.

Police and crime plans are typically set out around five priority areas, for example reducing acquisitive crime, with indicators assigned to measure the force’s performance in each area.

Indicators for reducing acquisitive crime may rely on quantitative data sources such as recorded crimes and outcomes of investigations, as well as qualitative assessments of how effective the force has been in its acquisitive crime prevention strategies, the extent to which partnership resources have been aligned to support these strategies and improving public confidence in how the police deal with these crimes.\footnote{INT6_EW.}
International approaches to police performance measurement

Other reported priority areas for measuring force performance include the use of stop and search powers, with indicators covering the rates and types of searches, and perceptions of their use in the community; and violence against women and girls, with indicators around how satisfied victims feel with how the police handled the matter.216

Data sources were reported to vary widely by force, reflecting the different approaches to how performance is measured locally.

Data sources were reported to include recorded crime data from the ADR and crime victimisation data from the CSEW, and qualitative data from crime victims and specialist forums with members of the public about their views on how the police are performing in the local priority areas.217 Overall, there was mixed feedback from interviewees about the use of quantitative data to measure performance. One police interviewee with responsibilities for measuring performance in their force commented:

…the numbers, how we would choose them, some of that we don’t spend a great deal of time and effort on because we’re not as interested in the numbers as we are for the qualitative explanation for the numbers.218

However, this interviewee, and two others, acknowledged that other forces rely much more on quantitative data, but noted the general, if gradual, trend towards using qualitative data to both inform the interpretation of quantitative data and to better understand the experience and perceptions of police personnel and the public about the force’s performance.219

Such discrepancies in the types of data sources used by forces renders comparisons between them particularly challenging.

4.3. Challenges and complexities in measuring police performance

4.3.1. Issues with the reliability, validity and representativeness of data

All interviewees commented on concerns with the quality of crime data provided as part of the ADR and used by all stakeholders to help measure police performance.

While the Home Office sets a national standard for the recording of offences by forces in England and Wales, referred to as the ‘Home Office Counting Rules’, interviewees noted that significant issues around the accuracy and completeness of these data remain in all forces for several reasons. Two interviewees highlighted significant challenges in information technology systems where the data are recorded, noting lags in data availability and limited functionality to input and store aspects of criminal history information and police activities in dealing with criminal offences.220 One interviewee noted that forces across England and Wales use ‘five or six different computer systems’, creating challenges around

216 INT8_EW.
217 INT_6, INT_7EW, INT_EW8.
218 INT6_EW.
219 INT_6, INT_7EW, INT_EW8.
220 INT6_EW.
meaningfully comparing and aggregating crime data across forces. One interviewee also noted that the Counting Rules have become so complex that errors in how crime data are recorded by the police have increased as a result. Available resources for crime recording and the data literacy of staff recording the crime data were also reported to vary across forces.

One interviewee noted that HMICFRS inspections of different forces are reported at different times, making it hard to understand and compare different force’s crime recording standards. Furthermore, as noted above, CSEW data indicate that most crimes in England and Wales are not reported to police, resulting in gaps in the understanding of crime incidence and related aspects of police performance. CSEW data are used by all stakeholders to provide a more complete picture of crime victimisation. One police interviewee also suggested that more routine, independent audits of forces’ crime data would help to ensure greater consistency in data recording and provide greater confidence to forces that they are being assessed and compared fairly against other forces.

4.3.2. Other challenges/complexities

Differences in how aspects of police work are perceived and prioritised between national and local policing stakeholders are reflected in different approaches to measuring police performance.

All interviewees acknowledged the value of a standardised national framework by which police force performance could be measured, albeit one that also provided the flexibility to reflect the differences in local policing priorities, resources and contexts. One police interviewee commented: ‘I like the efficiencies and comparability of a national framework and I think other performance managers in my position would probably say the same – that structure is useful.’

However, having national stakeholders engaged in the process of measuring police performance was also reported to present some challenges. Two interviewees commented on the overlapping roles of the Home Office and HMICFRS in this process, with one interviewee describing them as organisations ‘that would probably benefit from having their responsibilities and roles mapped more clearly’.

Furthermore, four interviewees noted tensions around perceptions of the Home Office agenda for policing and force/PCC local priorities, and the issues generated in different approaches to measuring police performance as a result. Three interviewees commented on concerns around the perceived politicisation of police performance measurement by the Home Office. One interviewee, for example, noted that while burglary rates are included in the National Crime and Policing Measures, perhaps on the potential strength

221 INT2_EW.
222 INT1_EW.
223 INT6_EW.
224 INT3_3, INT6_EW.
225 INT6_EW.
226 INT6_EW, INT7_EW.
227 INT6_EW, INT7_EW.
228 INT_6.
229 INT1_EW, INT4_EW, INT_6, INT_7EW.
230 INT1_EW, INT6_EW, INT7_EW.
of its appeal to voters, including such an indicator requires forces to allocate significant resources to an activity viewed by many forces as of lower priority.\textsuperscript{231} This tension is also reflected to some extent by recent changes in the position of successive home secretaries on the Home Office’s role in measuring police performance, described above, and concerns around whether such a role has made a constructive contribution to ensuring effective policing, or rather generated bureaucratic hurdles and perverse incentives. Three interviewees also highlighted more generally differences in how aspects of police work are perceived and prioritised by the Home Office compared to the forces themselves.\textsuperscript{232} One stated, for example, that efforts to minimise crime harm is higher on the agenda at the local level, with many forces using the Cambridge Crime Harm Index,\textsuperscript{233} a system that takes the extent to which crime harms victims, rather than the volume of reported crimes, as a performance measure. This interviewee also noted that violence against women and girls was more of a priority for many forces than is reflected in the Home Office’s performance measures.\textsuperscript{234} These interviewees commented on the pressure that their forces felt to perform well against these measures, while also addressing and performing well in responding to local priority areas.

Finally, three interviewees highlighted the variation in approaches undertaken at the local level by different PCCs to measuring police performance. They also expressed some criticism about how other forces set their performance indicators and the rigour with which they use data to measure their performance.\textsuperscript{235}

4.4. Perceived innovative and/or promising practices in measuring police performance

Interviewees from HMICFRS and two police forces commented that the relationship between HMICFRS and the forces was generally viewed as constructive on both sides, particularly since the inspection approach changed to ongoing engagement, with dedicated HMICFR staff assigned to individual forces.\textsuperscript{236} One HMICFRS interviewee commented that this new approach enhances the quality of the assessment, as force liaison leads are well informed about the operation of forces allocated, how they are and have been performing under the assessment framework, and the drivers of identified challenges and strengths in their performance:

As well as leading the regular inspection into that force, [liaison leads] would have regular day-to-day contact with that force. So, they have got really good knowledge of what is going on in that force. You’re not starting from a blank slate or piece of paper; from each inspection, you already know quite a lot about it.\textsuperscript{237}

\textsuperscript{231} INT6\_EW.
\textsuperscript{232} INT5\_EW, INT6 INT8\_EW.
\textsuperscript{233} Sherman, P. Neyroud and E. Neyroud (2016).
\textsuperscript{234} INT6\_EW.
\textsuperscript{235} INT6\_EW, INT7\_EW, INT8\_EW.
\textsuperscript{236} INT2\_EW, INT3\_EW, INT4\_EW, INT7\_EW.
\textsuperscript{237} INT2\_EW.
This interviewee also noted that forces have used inspection reports to support appeals to their local PCC on the need for additional or different resource allocation to address areas where they have been assessed to have performed poorly.

4.5. Concluding comments

Stakeholders in police performance measurement in England and Wales have overlapping and at times conflicting roles. Interviewees reported that there is consultation at the national level across policing stakeholders on the development of indicators for measuring performance. However, there are clear differences in approach, particularly between the Home Office and PCCs, resulting in tension and confusion among some forces about how to align national measures with those developed locally. Overall, however, interviewees felt that a standardised, national-level approach to measuring performance, with sufficient flexibility to reflect and capture local priorities and expectations about the role of the police, provides meaningful and comparable insights into their performance.

At the local level, PCC approaches to measuring performance were reported by interviewees to vary widely across forces, reflecting differences in organisational models, local concerns and expectations of the police, as well as levels of data literacy among stakeholders engaged in designing and assessing performance against indicators.

Interviewees also reported concerns around the validity, reliability and representativeness of recorded crime data, which all stakeholders use to some extent in measuring performance. Challenges in data quality are driven by multiple complex issues, such as limitations with information technology systems, which would require substantial time and resource to address. Furthermore, interviewees reported that these data were of limited utility in understanding the effectiveness of police work. However, HMICFRS and at least some force PCCs have moved away from a reliance on quantitative indicators and data sources towards efforts to construct more comprehensive, holistic assessment frameworks that seek to capture the complexities of police work and how it can be measured. These approaches are characterised by continuous, intelligence-led force engagement and assessment using a variety of indicators and data sources, and considering local needs, resources and priorities.
5. Case study 2: Israel

Box 3: Key findings in Israel

How did the methodology for the performance measurements come about? Were stakeholders involved? If so, which ones?
- Key stakeholders in the collection and assessment of police performance data in Israel are the local police units of the Israeli National Police (INP), the National Headquarters and the Ministry of National Security.
- Police performance measurement has become increasingly localised in Israel.
- This localised approach to measuring police performance is exemplified by the framework for performance measurement introduced under a recent police reform known as EMUN.

For what purposes are the performance measurements used?
- Evidence suggests that the main purpose for measuring police performance under EMUN was to increase public trust in the police and to support police stations in addressing local needs.

How is police performance measured in Israel? What aspects of policing do the measurements cover?
- In accordance with EMUN’s framework for measuring police performance, police units assessed their performance by looking at four different measures.
- Local police units could choose what annual targets to attach to each of the four performance measures, and across which geographical area (known as a ‘polygon’) they would apply.
- Performance dashboards were developed alongside the EMUN reforms to allow officers of all ranks to track their unit’s performance in real-time.

What are the (further) advantages and disadvantages of the performance measurements? Are any (side) effects of the performance measurements known? If so, which ones?
- Challenges for Israel include the fact that measuring performance exclusively within certain areas may fail to capture crime displacements.
- Evidence further indicates that officers may have experienced excessive pressure under EMUN, inadvertently encouraging undesirable results.
- Disproportionate emphasis on performance targets under EMUN reforms may have also influenced the quality of data collected from police incident registration systems.
- Interviewees highlighted the importance of considering broader strategic priorities for the police when developing a new framework for performance measurement.
- Interviewees also noted that buy-in from police officers was felt to be crucial for the successful implementation of a performance measurement system, but that, as in other case study jurisdictions, this can be difficult to achieve.

How reliable, valid and representative are the performance measurements?
- Interviewees expressed concern about the longevity of frameworks for police performance measurement in Israel.
5.1. Organisational context, stakeholders and the evolution of performance measurement

Israel has one national police force, the Israeli National Police (INP), which is headed by a chief commissioner appointed by the Ministry of National Security (formerly known as the Ministry of Public Security). The INP is divided into the seven regional districts of Northern, Coastal, Central, Tel Aviv, Jerusalem, Southern, and Judea and Samaria. District commanders oversee a total of between 70 and 80 local police units, each of which is headed by a station commander. The INP employs around 28,000 police officers. The National Headquarters (HQ), located in Jerusalem, oversees the INP. Together with the regional districts, the HQ’s Planning, Management and Information Unit (PMIU) is responsible for approving the strategic plans that each local police unit sets out annually. The Ministry of National Security, which is responsible for Israel’s homeland security, heads the Israeli police (with direct supervision of HQ).

Every three to four years, a new chief commissioner is appointed to lead the INP. Each commissioner tends to develop a new set of police reforms, replacing the existing system. As a result, the INP’s police strategic priorities, and therefore its framework for performance measurement, changes regularly.

In Israel, police performance measurement has become increasingly localized.

In the early 1990s, Israel’s police performance measurement systems were heavily influenced by the development of CompStat in New York. Performance data included high-level, quantitative crime-based data, collected and compared nationally (for a more detailed description of CompStat, see Section 7.1 and A.5). However, between 1999 and 2016, performance measurement became increasingly decentralised, shifting to a more localised approach. Whereas performance objectives and targets were previously applied to all police stations, more recent performance measurement systems have allowed police stations to set local objectives and targets, informed by local needs (see Section 5.2).

This localised approach to measuring police performance is exemplified by the framework for performance measurement introduced under a recent police reform known as EMUN.

---

238 Gofen and Gassner (2020).
239 Gofen and Gassner (2020); additional information provided by INT5_ISR via email.
242 INT7_ISR.
244 Gofen and Gassner (2020).
International approaches to police performance measurement

EMUN (which translates as ‘trust’) was a large-scale police reform introduced under former Israeli Police Chief Commissioner Roni Alsheikh in 2016. The reform included a new framework for performance measurement developed and managed by INP’s Strategy and Research Department. This framework was designed to focus primarily on issues of crime prevention and to allow units to concentrate on a mix of crime and quality of life offences (e.g. noise complaints, littering, disputes between neighbours), depending on local priorities.

While this chapter will focus primarily on police performance measurement under EMUN, as it incorporated practices identified as innovative in both the literature and by interviewees (see Section 5.4), it should be noted that EMUN in its original version is no longer in use. The INP is currently developing a new reform under the new government.

Under EMUN, the main purpose for measuring police performance was to increase public trust in the police and to support police stations in addressing local needs.

Due to the heterogeneity of Israeli society, the INP faces particular challenges regarding increasing public trust in the police. According to Perry and Jonathan-Zamir, two leading experts in the field of police performance measurement in Israel, the ‘police in Israel have a special role due to the unique security situation, ethnic diversity, and significant political, religious, and cultural differences and tensions’. Whilst the Netherlands does not face these problems to the same extent, police performance measurements used in Israel may still be of interest, as looking at how the INP has adapted its framework in light of its issues can increase the adaptability of Dutch stakeholders to changing tensions and challenges at home.

A number of academic studies from the past ten years focus exclusively on the relationship between the police and the Israeli community, noting the lack of trust in the police among citizens. In particular, minorities living in Israel continuously report feeling discriminated against and unfairly targeted by Israeli police officers, and the Israeli police has repeatedly come under scrutiny of both the public and the government. One of EMUN’s goals was therefore to increase public trust in the police.

At the same time, EMUN was intended to support police units in addressing local needs. Under previous performance measurement frameworks, police stations had complained that their performance was being measured unfairly because of a national agenda that did not take into consideration local differences in, for example, crime rates or the availability of resources. EMUN was therefore designed to give individual

---

245 Gofen and Gassner (2020).
246 Gofen and Gassner (2020).
247 INT3_ISR; INT4_ISR.
248 INT1_ISR; INT2_ISR.
249 INT7_ISR.
250 Perry and Jonathan-Zamir (2014).
251 Saunders et al. (2014); Ali and Rosenberg (2022).
253 Saunders et al. (2014, p. 2).
254 INT4_ISR; INT5_ISR; Ministry of National Security (2018); Saunders et al. (2014); Vugalter and Even (2016).
255 INT2_ISR; INT5_ISR.
256 INT2_ISR; INT5_ISR.
police stations more agency and autonomy over their yearly performance targets, while maintaining a certain level of uniformity between the stations (see Section 5.2). In parallel, the reforms also outlined that promotions and demotions of station commanders could no longer be tied to the performance of individual stations, as had been the case under previous frameworks for police performance measurement.

5.2. Current practices in measuring police performance

In accordance with EMUN’s framework for measuring police performance, police units assessed their performance by looking at four different measures.

Under EMUN, each police unit designed an annual workplan to determine its strategic priorities. As part of these workplans, each unit was asked to choose three measures, referred to by interviewees as ‘problems’, by which its performance would be assessed. Once the plans had been drawn up by the unit, they would be submitted to the district command and PMIU for authorisation.

One target was set nationally (the number of traffic accidents), while the other three differed per unit and were selected to reflect the local context. Individual station commanders could determine one of the three measures based on local crime data from previous years, as long as it demonstrated a ‘real and consistent problem’, without which the workplan would not be approved by the district command and PMIU. The final two measures were determined in consultation with residents in the relevant community, whose views were primarily gathered through community surveys administered to a representative sample of the local area. Examples of measures addressed by units under EMUN include property crime, drug-related crime and violent crime. The extent of such measures was subsequently measured using crime numbers and the rate of reported crime.

Under EMUN, local units could choose what annual targets would be attached to each of the four performance measures, and across which geographical area they would apply.

Once a unit had chosen its performance measures, it would then set appropriate annual targets based on what was felt to be reasonable and in line with the unit’s strategic objectives for the area in terms of incident rates. These targets were formulated for particular geographical areas within the units’ wider geographical area of responsibility, chosen by the unit itself and referred to as ‘polygons’, across which the performance of the unit would be tracked using data from police registration systems. The polygons corresponded to

257 Gofen and Gassner (2020).
258 Vugalter and Even (2016); INT4_ISR.
259 INT5_ISR; INT6_ISR.
260 Weisburd et al. (2020).
261 INT5_ISR.
262 INT2_ISR; Weisburd et al. (2020).
263 Weisburd et al. (2020).
264 INT2_ISR.
265 INT5_ISR; INT6_ISR.
266 INT5_ISR.
267 INT6_ISR.
International approaches to police performance measurement

certain subsections of the unit’s jurisdiction. The targets were also set in comparison to a unit’s performance in previous years, not to the performance of other units (e.g. ‘reducing property crime or drug-related crimes or violent crimes in the polygon by 30 per cent compared to the previous year’).\textsuperscript{268}

One interviewee explained that under a previous police reform known as MIFNE, targets were rigidly set based on a unit’s performance in the previous year; however, this approach often caused excessive pressure for units.\textsuperscript{269} For example, if a unit arrested ten murder suspects in one year, it was expected to arrest even more suspects the next year. However, if there were fewer murders committed in the next year, the unit might face negative consequences as it was unable to reach its target of arresting more than ten murder suspects.\textsuperscript{270} As a result, police units were given the opportunity to contextualise and explain their performance results annually under EMUN.\textsuperscript{271}

**Performance dashboards were developed alongside EMUN reforms to allow officers of all ranks to track their unit’s performance in real-time.**

The EMUN framework included live performance-tracking dashboards updated in real-time with data from police registration systems, which were freely accessible by officers of all ranks (Figure 3).\textsuperscript{272} The dashboards were considered particularly useful as they allowed for the collection, integration, analysis and evaluation of different types of data.\textsuperscript{273} At the same time, they presented this information in an intuitive way that could easily be understood by officers with no background in data science.\textsuperscript{274} On the ‘problem screen’, the unit’s targets, workplan and outputs would be presented, with colour-coding indicating how a unit was performing.\textsuperscript{275,276} A second screen included a live Geographic Information System (GIS) that mapped crime or traffic data in the unit’s chosen polygon linked to information on legal outcomes (e.g. information on court proceedings).\textsuperscript{277}

Two interviewees noted that the use of technology in the form of the dashboard was felt to be an especially helpful tool to motivate officers, as they were able to track their station’s performance in real-time. Evidence

\textsuperscript{268} INT5_ISR.
\textsuperscript{269} INT4_ISR.
\textsuperscript{270} INT4_ISR.
\textsuperscript{271} INT4_ISR.
\textsuperscript{272} Gofen and Gassner (2020); INT2_ISR.
\textsuperscript{273} Gofen and Gassner (2020); INT6_ISR.
\textsuperscript{274} Weisburd et al. (2020).
\textsuperscript{275} Weisburd et al. (2020).
\textsuperscript{276} For a more in-depth description of the dashboard and its use, see Weisburd et al. (2020, p. 950): ‘The ‘home page’ of the system is termed the ‘problems page’ (...). On this page the three chosen problems of that year (as well as the fourth generic ‘traffic accidents’ problem) are displayed with clear visual indications of how the station is performing. The screen is divided into four, a section for each problem. In each section, the problem category is named alongside the geographical location (i.e., Domestic Noise and Flowers Neighbourhood) and, below this, the statistics indicating whether crime reductions were being achieved across time. The performance was displayed using both a percentage score and a color-coded indication if the station was meeting its targets of reducing crime in comparison with the previous year. Different colors (blue, black, and red) provided a quick indication of the station’s status regarding crime within the polygon (the chosen area the station focused on) as well as outside it in the entire station area. At the bottom of this problem screen, a different score was displayed, reflecting the implementation of the action plan (i.e., were the outputs listed in the action plan carried out), with the three types of resources (internal, district, and external) each given a score (...).’
\textsuperscript{277} INT5_ISR.
from both literature and interviews indicates that officers at times derived a sense of accomplishment from
the fact that they could directly see when their station met its targets.\textsuperscript{278} One interviewee noted that EMUN
encouraged collaboration between different officers, as they knew their performance would be measured at
a station level and therefore felt a sense of ownership over its performance.\textsuperscript{279} Three interviewees indicated
that this sense of ownership was felt to have a positive impact on the quality of performance data.\textsuperscript{280}
Moreover, one interviewee noted that the collected performance data became more reliable in comparison
to data collected under previous performance measurement systems, as officers had a better understanding
of why performance data were being collected under EMUN and therefore felt a greater sense of
responsibility over the quality of the data.\textsuperscript{281}

\begin{figure}
\centering
\includegraphics[width=\textwidth]{figure3.png}
\caption{The EMUN dashboard}
\end{figure}

\textit{Note: This screenshot shows a translated version of an EMUN dashboard for an anonymised police station in southern
Israel in May 2017. Reproduced with permission from the authors.}

Source: Litmanovitz, Weisburd and Hasisi (2023).

5.3. Challenges and complexities in measuring police performance

5.3.1. Issues with the reliability, validity and representativeness of data

According to interviewees, measuring performance exclusively within polygons may fail to capture
crime displacement.

\textsuperscript{278} Weisburd et al. (2020); INT2_ISR; INT5_ISR.
\textsuperscript{279} INT5_ISR.
\textsuperscript{280} INT2_ISR; INT5_ISR, INT6_ISR.
\textsuperscript{281} INT6_ISR.
The stakeholder interviews highlighted several challenges and risks associated with the approach under EMUN to measuring police performance. Critics warned that focusing on reducing crime exclusively within the station’s chosen polygons might come at the expense of crime prevention and detective work in other areas. One interviewee noted that researchers saw an initial increase of crime outside a number of polygons at the beginning phase of EMUN; however, this was later remedied by including areas outside the polygon in the station’s performance measurement. Moreover, an evaluation study of EMUN found no evidence that the programme increased crime in certain areas. Although performance targets were still primarily linked to the polygons, less ambitious targets were set for areas outside these polygons. For example, this meant that a unit’s performance target might be to reduce property crime by 30 percent within the polygon, and by 10 percent outside of the polygon.

Disproportionate emphasis on performance targets under EMUN reforms may have influenced the quality of data collected from police incident registration systems.

One interviewee noted that pressure to reach targets set under the EMUN performance framework sometimes led to stations manipulating performance data. One such manipulation included wrongly reporting the location of a crime to outside of the polygon to reduce the recorded number of crimes inside the polygon, where a station’s performance was being measured. In another high-profile case, a police chief was demoted after encouraging citizens not to report crimes to the police.

5.3.2. Other challenges/complexities

Interviewees highlighted the importance of considering broader strategic priorities for the police when developing a new framework for performance measurement.

Frameworks for police performance measurement may not provide a meaningful assessment of how the force is performing if they are not tied to the broad strategic priorities and plans of the force. One interviewee proposed that performance measurement should be a driving factor in the development of new policing strategies, rather than being introduced only after organisational changes have been implemented. This is more likely to ensure that how performance is measured aligns with and provides a meaningful assessment of the work of police officers. A second interviewee noted that Israel currently relies primarily on quantitative data to measure police performance. However, this limits the kind of performance information that can be gathered regarding the public perception of police work. As such, to give a representative assessment of how the force is performing against the strategic objective of increasing public satisfaction...
trust, it would be useful if forces also took qualitative data into account. Other interviewees pointed to the complex balance between measuring enough, but not too many, indicators, to avoid losing sight of police performance targets and putting an unnecessary strain on resources. Another interviewee stressed the importance of developing a performance measurement framework that can be adapted to changing local (crime) issues. For example, future performance measurement systems will need to develop indicators that consider cybercrime, as well as the policing of other crimes that do not take place in physical spaces.

**Interviewees noted that buy-in from police officers was felt to be crucial for the successful implementation of a performance measurement system, but that this can be difficult to achieve.**

Although EMUN attempted to improve stakeholder buy-in relative to previous police performance frameworks by making performance data accessible to officers of all ranks, interviewees nonetheless observed that officer buy-in remains a challenge for the INP. Two interviewees proposed that this is especially true where the implementation process was complicated and necessitated a large degree of organisational change. Moreover, they noted that in cases where police officers and commanders do not consider performance evaluations as part of their responsibility, the quality of performance data can suffer.

One interviewee also stressed the importance of buy-in across and clear communication between all stakeholders involved in measuring police performance. In Israel, the police are ultimately overseen by HQ. At the same time, district commanders have a say in how police resources are allocated among the different police units. While HQ started measuring police performance using EMUN metrics and goals when EMUN was implemented in 2016, one interviewee noted that district commanders often continued to measure police stations according to the old measurement system. This left station commanders uncertain about where their resources should be focused. The study team were unable to establish the exact cause for this dissonance between HQ and district police stations regarding performance measurement.

Interviewees expressed concern about the longevity of frameworks for police performance measurement in Israel. Chief commissioners tend to hold their post for three to four years, during which time they normally introduce substantial police reforms. Several interviewees noted that this quick succession of reforms was disruptive to any major, sustained implementation of a performance measurement framework. Interviewees highlighted that the transition to a new performance measurement framework can be a slow process, and that jurisdictions should therefore take time to fully implement and evaluate its impact, rather than introducing new elements in quick succession as strategic priorities change.

---

292 INT4_ISR.
293 INT6_ISR.
294 INT2_ISR; INT7_ISR.
295 INT2_ISR; INT7_ISR.
296 INT1_ISR.
297 INT2_ISR.
298 INT2_ISR.
299 INT1_ISR; INT2_ISR; INT6_ISR; INT7_ISR.
301 INT7_ISR.
5.4. Perceived innovative and/or promising practices

Stakeholders and experts interviewed for this study highlighted two aspects of police performance measurement under EMUN as particularly innovative: 1) its localised approach to performance measures and targets; and 2) the performance dashboards.

Under previous police reforms, experts noted the importance of integrating regional variables into the INP’s approach to performance measurement. In previous systems, there had been a tension between the national police agenda and local needs, with police stations protesting that the national agenda was too rigid for them to address local crime issues. As described in the preceding sections, the EMUN reform was designed to allow for a more localised approach to performance measurement, and interviewees praised the incorporation of both national and local measures and targets in EMUN’s performance measurement system. Several interviewees deemed this emphasis on local performance measures and targets as innovative, but with the caveats of the challenges set out in Section 5.3. They noted that it gave police stations agency in deciding which issues they wanted to address, while simultaneously allowing a more community-centred approach to performance measurement by incorporating input from the public.

Interviewees also praised EMUN’s use of dashboards, noting that they enabled officers of all ranks to track the performance of their unit in an intuitive and simple manner. They proposed that the real-time performance data on the dashboard allowed officers to take greater responsibility over their station’s performance and improve their understanding of how and why performance data were being collected, thereby encouraging officer buy-in.

5.5. Concluding comments

The approach to performance measurement under EMUN combined both national and local measures and targets, allowing for a community approach to performance measurement while maintaining a degree of consistency between different police stations. In this sense, the framework shared similarities with the Netherlands (national, regional and local measures and targets) and, as described in the following chapter, North Rhine-Westphalia (state and regional measures and targets).

However, unlike the forces in these jurisdictions, Israeli police units under EMUN assessed their performance by looking at just four measures and targets and could choose the geographical area across which they would take measurements. According to interviewees, this approach was not without its challenges, as it could cause displacements in crime to go unnoticed.

Although the framework introduced under EMUN no longer exists in its original form, certain aspects of police performance measurement under EMUN were praised as innovative by interviewees, with the

---

303 INT2_ISR; INT7_ISR.
304 INT4_ISR; INT5_ISR; INT7_ISR.
305 INT4_ISR; INT5_ISR; INT7_ISR.
306 INT5_ISR; INT7_ISR.
307 INT6_ISR.
operational dashboards developed to allow officers of all ranks to track their local unit’s live performance highlighted as particularly valuable. The implementation of these dashboards was felt to increase officer buy-in, rendering the performance data collected more reliable as officers had a better understanding of why they were being collected, and therefore felt a greater sense of responsibility for the quality of the data.
6. Case study 3: North Rhine-Westphalia

Box 4: Key findings in North Rhine-Westphalia (NRW)

How did the methodology for the performance measurements come about? Were stakeholders involved? If so, which ones?
- Key stakeholders in police performance measurement are the Ministry of the Interior (Ministerium des Innern), the individual police units (KPBs, Kreispolizeibehörden) and the State Office for Central Police Services (LZPD NRW).
- In recent decades, police performance measurement in NRW has become increasingly localised, as targets are set by regional units.

For what purposes are the performance measurements used?
- Evidence suggests that performance data are used to support the continuous improvement of police work. In particular, police performance is primarily measured with the aim of ensuring cost efficiency and improving public safety and public satisfaction with the police.

How is police performance measured in North Rhine-Westphalia? What aspects of policing do the measurements cover?
- Individual regional units set their own performance targets on an annual basis, following the strategy formulated by the NRW Ministry of the Interior.
- The measures used to evaluate police performance in NRW are primarily quantitative.
- While most performance data are gathered automatically through the force’s internal management information system (FISPol), some data continue to be collected manually by officers.

What are the (further) advantages and disadvantages of the performance measurements? Are any (side) effects of the performance measurements known? If so, which ones?
- NRW uses comparison clusters to make comparisons between regional police units fairer and more meaningful.
- Interviewees identified the large number of regional police units in NRW as a challenge for police performance measurement.
- Evidence further suggests that securing buy-in from officers at all levels can present a challenge to police units across NRW. However, interviewees proposed that raising awareness around the benefits of measuring performance could improve officer engagement.

How reliable, valid and representative are the performance measurements?
- As in other case studies, the need to include more diverse measures of police performance across NRW was identified as a challenge.
- A second challenge is proving causality between police output and desired outcome.

Which (parts of) foreign performance measurements can be labelled as 'qualitatively good', 'innovative' and/or 'promising'?
- The ability of KPBs to set targets that meet their local needs, while simultaneously pursuing the same state-wide policing strategy, was highlighted as an example of good practice in both literature and interviews.
- Promising developments in NRW include the introduction of audits to supplement the state’s current performance measurement framework.
6.1. Organisational context, stakeholders and the evolution of performance measurement

In Germany, each of the 16 states (Bundesländer) is responsible for operating its own police force. The state of North Rhine-Westphalia (NRW) has the biggest police force in Germany, employing around 57,000 people, 40,500 of whom are police officers, across 47 individual police units or Kreispolizeibehörden (KPBs).

In recent decades, police performance measurement in NRW has become increasingly localised, as targets are set by regional units.

Over the last 20 years, police performance measurement in NRW has changed significantly. Whereas the same performance indicators and targets were previously used for all KPBs, performance measurement has become more localised. Currently, KPBs have more agency over NRW’s performance measurement framework than previously, as they are responsible for formulating performance targets and deciding on input measures at a local level (see Section 6.2). At the same time, the performance measures have become more focused: whereas performance of the NRW KPBs was previously measured using 1,605 measures, NRW has been using around 270 different measures since 2005. NRW has also moved away from crime reduction (outcome measures) as the sole target of police operations, towards a broader approach that also considers crime and risk prevention (output measures) in an attempt by the Ministry of the Interior to encourage more proactive policing.

Key stakeholders in police performance measurement are the Ministry of the Interior, the KPBs and the State Office for Central Police Services (LZPD NRW).

Since 1990, the NRW police has experienced several rounds of restructuring. Each restructuring effort has influenced which stakeholders played a role in measuring police performance. The NRW Ministry of the Interior was previously responsible for measuring the performance of all 47 KPBs. However, one interviewee explained that the ministry found that it was unable to conduct in-depth performance evaluations due to the sheer number of KPBs. To address this problem, it reoriented the NRW policing structure in 2019 and 2020, creating a three-tier system. Within the new structure, the three state superior authorities – the State Office for Training, Continuing Education and Personnel Affairs (LAFP NRW), the State Criminal Office (LKA NRW), and the State Office for Central Police Services (LZPD NRW) – took
International approaches to police performance measurement

on functional supervisory capacities. The LZPD NRW, with support from the LAFP NRW and the LKA NRW, became responsible for the performance evaluation of the NRW police, including its performance measurement. The Ministry of the Interior supervises the three state superior authorities and maintains final authority over police performance measurement in NRW.

In 2003, the NRW Ministry of the Interior asked the LKA NRW to create the Central Evaluation Office (ZEVA). ZEVA is unique in Germany, and is responsible for conducting evaluations of specific police programmes for the NRW police and the NRW Ministry of the Interior, such as on the use of police video surveillance or a police initiative to increase safety during football matches. As such, it is not directly responsible for police performance measurement, but rather supports NRW police in identifying evidence-based police practices, tools and methods, ensuring that police resources are used efficiently and using evaluation to make management decisions more transparent. ZEVA relies on mixed-method approaches, using both quantitative and qualitative data such as staff interviews, focus groups, observations, online questionnaires and the triangulation of different types of data sets.

ZEVA currently consists of five police officers who have received additional training before joining the evaluation team. Their work is informed through collaboration with DeGEval – Evaluation Society, as well as other police experts in a process that one interviewee referred to as ‘cooperative evaluation’. Both the Ministry of the Interior and individual KPBs can ask ZEVA to perform an evaluation. Although one interviewee noted that ZEVA initially faced scepticism from some KPBs, they believed that it is slowly receiving a higher level of acceptance. For example, the same interviewee indicated that one KPB recently reached out to ZEVA for support on deciding how to best address a shortage of senior staff. ZEVA has also supported the development and introduction of audits within the NRW police.

In NRW, police performance is measured to ensure cost efficiency and to improve public safety and public satisfaction with the police.

Evidence indicates that police performance measurement has three main purposes in NRW. First, according to the head of ZEVA, Ingo Dungs, performance is measured to ensure that the police budget and police resources, including staff, are allocated and used efficiently. Second, evidence from the literature suggests
that performance is measured to improve public satisfaction with and trust in the police, as a good public image of the police is known to improve civilian cooperation. And third, police performance is measured to ensure and improve public safety.

6.2. Current practices in measuring police performance

Individual regional units set their own performance targets annually, following the strategy formulated by the NRW Ministry of the Interior.

KPBs operate in the context of both state-wide and local policing strategies. The state-wide strategy is set out in functional strategies (Fachstrategien), which outline the policing priorities and plans formulated by the NRW Ministry of the Interior. The functional strategies include objectives centred around three main areas: operational management, road safety and crime reduction.

Based on the state-wide strategy, the KPBs develop an annual security programme (Sicherheitsprogramm), in which they outline local targets aligned with each of the national targets. They further outline specific inputs with which they hope to achieve these targets, such as an increased police presence in crime hotspots to reduce the rate of incidents. Each KPB analyses its local crime and road safety status as a base for its security programme. KPBs may consider the number of incidents for various crime types, as well as the number of road traffic accidents or the availability of personnel when formulating their targets and deciding which inputs are needed to achieve them.

Interviewees noted that KPBs formulate their local performance targets relative to the unit’s previous performance in a specific area. For instance, in line with the state-wide objective to improve road safety, a KPB might set a specific target to reduce the number of bike accidents compared to the previous year, or the previous five-year average.

Unfortunately, the functional strategies are confidential and could not be shared with the study team. As such, it is not possible to provide an overview of precisely which targets are used across KPBs in NRW in any given year.

The measures used to evaluate police performance in NRW are primarily quantitative.

NRW uses around 270 quantitative output and outcome measures to assess police performance. Each of these measures sits under the umbrella of one of the broader strategic priorities defined in the state-wide Fachstrategien. Response time, for example, is one of the output measures under the umbrella of operational

---

332 Hermanutz and Weigle (2017).
333 Additional information provided by INT3_NRW via email.
334 INT3_NRW.
335 Wehe (2017); INT4_NRW.
336 INT5_NRW.
337 Wehe (2017); INT1_NRW; INT4_NRW.
338 INT5_NRW.
339 INT2_NRW; INT4-NRW.
management, as are measures concerning absence days due to sickness and the number of officers who participated in further education.\textsuperscript{341} Under the umbrella of road safety, KPBs may use outcome measures, such as the annual number of traffic accidents, as well as output measures, such as the number of times officers have tested for vehicle speed. Crime reduction measures include input and output measures such as the number of reported crimes and police clearance rates, as well as measures meant to increase the quality of criminal investigations (e.g. how often DNA analysis or fingerprint collection is used).\textsuperscript{342}

During the annual performance review cycle, the Ministry of the Interior and KPBs can propose new measures, or change existing ones when deemed no longer meaningful or representative of performance.\textsuperscript{343} This takes place through the Ministry’s ‘Plan-Do-Check-Act’ (PDCA) cycle, a four-step model that seeks to ensure performance measures and targets are updated annually based on KBP feedback.\textsuperscript{344,345}

**While most performance data are gathered automatically through the force’s internal management information system (FISPoL), some data continue to be collected manually by officers.**

Most performance data are collected automatically through the police’s internal management information system, known as FISPoL NRW. FISPoL collects data such as response time or the number of emergency calls received by the police. Other data are collected manually by police officers, particularly in the area of occupational health and safety.\textsuperscript{346}

**NRW uses comparison clusters to make comparisons between KPBs fairer and more meaningful.**

KPBs collate their performance data and send it to the three state superior authorities, which analyse which targets have been achieved, which targets have not been achieved, and which targets might need to be adjusted in subsequent years.\textsuperscript{347} Following criticism from KPBs that comparing the performance of KPBs on a state-wide level was unfair due to the varying contexts in which they operate,\textsuperscript{348} NRW introduced comparison clusters to make comparisons fairer and more meaningful.\textsuperscript{349} These comparison clusters consist of five to seven KPBs that are similar in terms of organisational structure, personnel growth and crime incidents.\textsuperscript{350} Four interviewees proposed that the clusters make it easier to accommodate local needs, targets and indicators within performance measurement, while also retaining some level of uniformity by means of the state-wide strategy.\textsuperscript{351} The performance measures of each KPB are displayed on FISPoL, which can be accessed by the KPB in question, other members of the comparison cluster, the state superior authorities and the Ministry of the

\textsuperscript{341} Wehe (2017); INT1_NRW.
\textsuperscript{342} Wehe (2017); INT1_NRW; INT5_NRW.
\textsuperscript{343} INT5_NRW.
\textsuperscript{344} INT2_NRW.
\textsuperscript{345} Dungs (2017).
\textsuperscript{346} INT5_NRW; additional information provided by INT3_NRW via email.
\textsuperscript{347} INT1_NRW.
\textsuperscript{348} INT1_NRW.
\textsuperscript{349} INT4_NRW; INT5_NRW.
\textsuperscript{350} Wehe (2017); INT1_NRW; INT2_NRW.
\textsuperscript{351} INT1_NRW; INT2_NRW; INT4_NRW; INT5_NRW.
Key crime figures are made available to the public in annual Police Crime Statistics (*Polizeiliche Kriminalstatistik*, PKS).353

**Performance data are used to support the continuous improvement of police work.**

Interviewees stressed that the aim of police performance measurement in NRW is to provide an opportunity for institutional learning.354 After evaluating performance data, the LZPD provides feedback to each of the KPBs on what they can improve upon in the next year.355 If a KPB is found to be underperforming considerably, the Ministry of the Interior invites its leadership to discuss strategies to improve performance.

### 6.3. Challenges and complexities in measuring police performance

#### 6.3.1. Issues with the reliability, validity and representativeness of data

Interviewees identified a need to include more diverse measures of police performance across NRW.356 As in the Netherlands, England and Wales, and Sweden, the police performance measures used in NRW were not developed through consultation with citizens, as was the case in Israel.357

Furthermore, several stakeholders emphasised the importance of considering both quantitative and qualitative measures.358 They considered collecting qualitative data, such as qualitative feedback from surveys or other forms of citizen consultation, particularly crucial to gaining a better understanding of levels of public trust and public satisfaction with the police.359

**Establishing causality between police outputs and desired outcomes is complex.**

As in other jurisdictions, previous research on the topic of police performance measurement in NRW has also highlighted the complexity of linking large, overarching policing strategies, such as furthering justice and ensuring public safety, to concrete measures.358,359 Determining causality can be challenging when multiple measures are dependent on one another. For example, one interviewee noted that the number of reported crimes is partially dependent on staff resources to record crimes.360 As such, it may be difficult to establish whether an increase in the crime rate for a particular crime type is due to more crimes taking place, a greater willingness among the public to report crimes or an increase in the number of police staff.

---

352 INT4_NRW.
353 Polizei NRW (2023a).
354 Wehe (2017); INT1_NRW; INT2_NRW.
355 INT4_NRW.
356 INT5_NRW.
357 INT2_NRW.
358 Dungs (2017); INT1_NRW; INT5_NRW.
359 INT1_NRW.
360 Dungs (2017); INT5_NRW.
361 Wehe (2017); Dungs (2018); INT1_NRW.
362 INT1_NRW.
Moreover, both crime rates and the number of traffic accidents are influenced by a range of exogenous factors, for example weather conditions or the COVID-19 pandemic, which complicate the task of measuring the impact of any police activities.\textsuperscript{363}

6.3.2. Other challenges/complexities

The number of KPBs in NRW poses a challenge for police performance measurement in NRW. Although interviewees generally positively described the local character of police forces, one interviewee noted that the sheer number of KPBs poses a bureaucratic obstacle to conducting in-depth performance reviews for each of the 47 units.\textsuperscript{364} This issue was partially addressed by introducing audits, an approach examined in greater depth in Section 6.5.

Evidence further suggests that securing buy-in from officers at all levels can present a challenge to police units across NRW. It can be difficult to ensure that police officers at all levels engage with the framework for performance measurement.\textsuperscript{365} One interviewee noted that producing detailed information on incidents to provide accurate data for performance measurement may not be a priority for officers, as it requires time that could be spent on other policing tasks.\textsuperscript{366} However, interviewees identified buy-in from police officers at all levels as crucial for the successful implementation of a performance measurement system.\textsuperscript{367}

6.4. Perceived innovative and/or promising practices

The ability of individual KPBs to set targets that meet their local needs, while simultaneously pursuing the same state-wide policing strategy, has been highlighted as a good-practice example in both literature and interviews. Stakeholders praised NRW’s use of both local and state-wide strategies, noting that this approach allows for the consideration of local needs, while also ensuring a level of uniformity among all 47 KPBs.\textsuperscript{368} Furthermore, comparison clusters were felt to be an innovative approach to addressing the difficult task of measuring the performance of police stations pursuing both state-wide and local targets.\textsuperscript{369}

6.5. Recent developments

NRW recently introduced audits to improve its performance measurement framework.\textsuperscript{370}

\textsuperscript{363} Wehe (2017); INT1_NRW; INT2_NRW; INT4_NRW; INT5_NRW.
\textsuperscript{364} INT4_NRW.
\textsuperscript{365} Dungs (2018); Wehe (2017); INT2_NRW; INT5_NRW.
\textsuperscript{366} INT2_NRW.
\textsuperscript{367} Dungs (2018); Wehe (2017); INT2_NRW; INT5_NRW.
\textsuperscript{368} Wehe (2017); INT4_NRW; INT5_NRW.
\textsuperscript{369} INT1_NRW.
\textsuperscript{370} INT4_NRW.
These audits are conducted by the state superior authorities under a state-wide programme, which is tailored specifically to each KPB. The literature highlighted NRW’s ‘eye-to-eye level’ approach to auditing, in which auditors, experts and KPBs have the chance to exchange views and best practices, while identifying potential areas for improvement. These audits are not used to evaluate the performance of the regional units, but are intended to support KPBs in identifying risks. Furthermore, they are meant to improve KPBs’ understanding of how particular processes can be improved, how resources can best be used and where further training might be beneficial. Areas of interest include processes regarding the prevention of hazards, crime prevention, the prevention of traffic accidents, leadership processes and police administration processes. Unlike the above-described methods for police performance measurement in NRW, these audits do not attempt to capture the breadth of police work, but rather focus on exploring a smaller number of issues in greater depth. To ensure that this is possible, audit measures rely on a combination of quantitative and qualitative data.

All KPBs are audited within a period of three years. After an audit is conducted, the state superior authority suggests follow-up measures, which have varying degrees of obligation. Whereas some measures are compulsory for the KPBs to implement, others constitute recommendations for the improvement of a certain process. KPBs need to report back to the state authority and outline the improvements they implemented, either 3 or 12 months after the initial audit. According to one interviewee, audits promise to be a beneficial addition to NRW’s current performance measurement system.

**Interviewees suggested that raising awareness of the benefits of measuring performance could improve officer buy-in.**

Interviewees proposed several recommendations to help facilitate police officer buy-in. First, interviewees suggested that raising awareness within police units around the purpose of measuring performance and the opportunities for improvement it presents may increase officer engagement with the existing framework for performance measurement. This might improve information recording efforts, as officers become more aware of the benefits of securing meaningful data on performance, which would be especially important in units where some data are collected manually by officers.

Second, several interviewees noted that minimising the effort required from police officers to record performance data would be beneficial to staff buy-in. They proposed avoiding manual data collection and using technology instead to collate different sets of data.

---

571 INT4_NRW; INT5_NRW.
572 Lesmeister, Henkel and Beyer (2022).
573 Lesmeister, Henkel and Beyer (2022); INT4_NRW.
574 Lesmeister, Henkel and Beyer (2022).
575 Lesmeister, Henkel and Beyer (2022); INT4_NRW.
576 INT4_NRW.
577 INT4_NRW.
578 INT4_NRW.
579 INT3_NRW.
580 INT2_NRW; INT4_NRW; INT5_NRW.
Finally, interviewees strongly advised that performance measures should be used as opportunities for learning and not to issue sanctions, as too much pressure on particular performance measures might encourage data manipulation.381

6.6. Concluding comments

Police performance measurement in NRW has become increasingly localised in recent years. As is the case in the Netherlands to some extent, regional units set their own targets but follow a larger government strategy. Where the Dutch police builds on the Security Agenda set by the national-level Ministry of Justice and Security, the KPB follow the strategy formulated by the NRW Ministry of the Interior.

Furthermore, as in the Netherlands, interviewees identified a need to include more diverse measures of police performance across NRW. Both jurisdictions highlighted the lack of measures relating to citizen satisfaction with and confidence in the police. As detailed in Chapter 7, practices recently adopted by the Seattle Police Department may offer actionable insights in this regard.

Other aspects of police performance measurement in NRW were identified as innovative in the literature and stakeholder interviews. This includes the establishment of comparison clusters between KPBs, the evidence base underlying some of the performance measures and indicators, and the inclusion of both state-wide and local performance targets in performance evaluations. Stakeholders also highlighted that the additional support provided through newly established audits and evaluations conducted by ZEVA may benefit performance measurement.

381 INT2_NRW; INT5_NRW.
7. Case study 4: Seattle

Box 5: Key findings in Seattle

How did the methodology for the performance measurements come about? Were stakeholders involved? If so, which ones?
- There is no nationwide approach to police performance measurement in the United States. The Seattle Police Department (SPD) is responsible for collecting and analysing its own performance data. As such, it is the main stakeholder in performance measurement in this jurisdiction.
- In 2012, the city of Seattle entered a Consent Decree that bound the city to reform policing standards and required the SPD to use evidence when making decisions.
- This led to a police reform and the introduction of new initiatives, most notably the Equity, Accountability and Quality (EAQ) programme.

For what purposes are the performance measurements used?
- Evidence suggests that the purpose of measuring police performance within the SPD, and especially EAQ, is twofold: 1) to provide accountability for police action to the government and the public; and 2) to increase public safety by controlling collateral harms that can result from policing.

How is police performance measured in Seattle? What aspects of policing do the measurements cover?
- As in other case studies, the SPD obtains community input through different channels to measure police performance and subsequently prioritise policing in the department.
- Currently, the EAQ programme operates in combination with other well-known initiatives such as CompStat but offers a different focus of measurement.
- The three pillars of EAQ measure different aspects of policing: 1) racial disparities; 2) ‘under’ and ‘over’ policing; and 3) quality of interactions between police and the public.
- To augment the EAQ pillars, new developments are ongoing and further expected, including analysis of community satisfaction through customer feedback.

What are the (further) advantages and disadvantages of the performance measurements? Are any (side) effects of the performance measurements known? If so, which ones?
- Changing the mindset of police officers was seen as the most challenging aspect in the SPD, but necessary to enable the introduction of new performance measurement practices.
- As observed in previous chapters, other challenges for Seattle include the abandonment of initiatives due to a lack of stakeholder engagement, and researchers and evaluators remaining uncertain about what good policing looks like. In addition, the ‘data side’ and the ‘operational side’ of EAQ do not seem sufficiently connected, and communication between the two teams is limited.

How reliable, valid and representative are the performance measurements?
- One interviewee expressed concern that a lack of stakeholder involvement could lead to the discontinuation of initiatives, as new practices may fade due to staff turnover and the consequent loss of institutional memory.
7.1. Organisational context, stakeholders and the evolution of performance measurement

The Seattle Police Department (SPD) is a large municipal law enforcement agency in Washington State that employs approximately 1,200 sworn officers and 631 civilian employees.\(^{382}\) The SPD has more specialty units than other departments in the state, including traffic, major crimes, property crimes, crisis response and collaborative community policing.\(^{383}\)

There is no nationwide approach to police performance measurement in the United States (see Annex A). However, one of the most frequently used tools for measuring police performance is a performance management system called CompStat,\(^{384}\) which primarily focuses on analysing crime statistics with the aim of measuring the performance of US police departments.\(^{385}\) CompStat has been implemented in Seattle,\(^{386}\) although the city renamed it 'SeaStat'.\(^{387}\)

While CompStat is regularly criticised (see Box 6), it should be noted that over the last decade, the SPD underwent a significant reform that introduced various additional performance measurement initiatives. According to five stakeholders interviewed in the context of this study, this new framework represents a significant innovation in police performance measurement in the United States.\(^{388}\)
Box 6: CompStat

CompStat, short for Computer Statistics, is a quantifying performance measurement tool designed by the New York City Police Department in 1994, which has since been adopted in different jurisdictions throughout the United States. The system is designed to reduce crime and achieve the goals set by police forces. It emphasises information sharing and accountability and is best known for its regularly held meetings where departments can discuss and analyse issues related to crime.\(^{389}\) One interviewee described CompStat as an accountability forum that allows leadership to ask questions to districts.\(^{390}\) Despite its prevalence in US jurisdictions, CompStat is often criticised. Eterno et al., for example, warn that CompStat’s ‘performance management can morph into a top-down numbers crunching bureaucracy.’\(^{391}\) They argue that it needs to be updated to align better with contemporary policing needs.\(^{392}\)

The city of Seattle entered a Consent Decree that bound the city to reform policing standards and required the SPD to increase the use of evidence in their decision making.

The reform of the SPD was largely fuelled and shaped by concerns about a pattern of constitutional violations regarding the use of force by the police. The use of force, as well as biased policing, have been common concerns in the United States for many years, visible for instance in the United Nation’s reporting about US police brutality.\(^{393}\) In 2011, the US Department of Justice (DoJ) found that the SPD engaged in a general pattern of ‘unnecessary or excessive force’.\(^{394}\) The DoJ attributed this pattern to a lack of research into the use of force, a lack of adequate policy implementation of the proper use of weapons and a lack of training of officers in the use of force.\(^{395}\)

To improve SPD practices, a Consent Decree was issued in 2012. Such decrees bind the city to certain objectives for reform, specifically regarding the use of excessive force and biased policing, but they are not strict judgements. Instead, Consent Decrees refer to settlements that both parties submit to the court and that are approved by a judge, after which the agreement becomes enforceable for both parties.\(^{396}\) The settlement agreement between the City of Seattle and the DoJ required the SPD to meet certain reform standards that would later form the basis for a new police performance measurement system in Seattle.

Under the Consent Decree, the SPD received particular attention, visible for instance in the 2022 Comprehensive Assessment of the Seattle Police Department, which analyses the progress Seattle has made. According to this review, the decree provided a basis for ‘sustainable progress’ towards lawful policing.\(^{397}\) Interviewees perceived the results as generally positive.\(^{398}\) In 2020, the city of Seattle therefore requested that the Consent Decree be removed.\(^{399}\) However, before the hearing to discuss this could take place, the

\(^{390}\) INT6_US.
\(^{391}\) Eterno, Silverman and Berlin (2021).
\(^{392}\) Eterno, Silverman and Berlin (2021).
\(^{394}\) Seattle Police Monitor (2022, p.26).
\(^{395}\) Seattle Police Monitor (2022, p.26).
\(^{396}\) Cornell Law School (2022).
\(^{397}\) Cornell Law School (2022).
\(^{398}\) INT5_US.
\(^{399}\) INT5_US.
International approaches to police performance measurement

death of George Floyd\textsuperscript{400} and the ignited protests across the country (including in Seattle) led to historic levels of misconduct complaints.\textsuperscript{401} The city of Seattle subsequently withdrew its request to end the Consent Decree in 2020 and has continued its review of the SPD.\textsuperscript{402} In March 2023, the City of Seattle and the DoJ filed again a joint proposal for termination.\textsuperscript{403}

The Consent Decree led to police reform and the introduction of new initiatives, most notably the Equity, Accountability and Quality (EAQ) programme.

Despite the withdrawal of the Consent Decree request, the changes that the SPD has implemented since 2012 were highlighted by interviewees as examples of effective reform for other agencies and police forces to emulate.\textsuperscript{404} One interviewee explained that the SPD was required to reassess its policies and practices in relation to the use of force and assess the quality of police interactions with the public.\textsuperscript{405} More specifically, the SPD has implemented an Equity, Accountability and Quality (EAQ) programme and training, explored disparity concerns and collateral harms in society, and improved data processes.\textsuperscript{406} According to an interviewee in the SPD leadership, the most immediate result of the reforms was an improvement in the process of documenting the use of force, and in the structured process for reviewing these incidents.\textsuperscript{407} The improved data processes would also benefit measures taken to assess police performance (see Section 7.4).

The SPD set up a new Performance Analytics and Research team, which led to the creation of the EAQ. The SPD pursues the three components of the EAQ programme, equity, accountability and quality, by assessing disparate outcomes of policing, over- and under-policing of communities,\textsuperscript{408} and the quality of service.\textsuperscript{409} The programme covers these three aspects in great depth, and therefore presents an interesting and unique part of the larger police performance measurement strategy in Seattle.

The remainder of this chapter will explore the specifics regarding EAQ, unless otherwise stated.

There are several key stakeholders in Seattle’s current police performance measurement system.

First, the settlement agreement bound, and continues to bind, the US Department of Justice and the Seattle Police Department to reforming its policing standards. Both departments are among the largest stakeholders of police performance measurement in Seattle. The DoJ has a stake in the improvement of the departments’ performance to increase overall performance and public satisfaction. The Seattle Police

\textsuperscript{400} George Floyd, a black man, was killed while being arrested by the police in Minneapolis. The incident was filmed, and protests over police brutality and racial discrimination quickly spread throughout the country, as well as internationally. See, for instance, BBC News (2020).

\textsuperscript{401} Seattle Police Monitor (2022, p.26).

\textsuperscript{402} INT5_US.

\textsuperscript{403} Sforza (2023).

\textsuperscript{404} INT4_US; INT5_US; INT7_US.

\textsuperscript{405} INT5_US.

\textsuperscript{406} RTI International (2023).

\textsuperscript{407} INT3_US.

\textsuperscript{408} A lesser, or greater level of police presence than generally expected, based on public demand, current policing strategies for the given location and community preference. See for instance Atherley, Hickman, Parkin and Helfgott (2022).

\textsuperscript{409} RTI International (2023).
Department has simultaneously sought to generate greater insights into the performance of its officers to facilitate change and reform in line with the stipulations of the Consent Decree.

One of the main objectives of the settlement agreement between the DoJ and the SPD is to better serve the community, which is the recipient of policing services. The public is also directly involved in measuring (and guiding the improvement of) police performance through various channels. Moreover, within EAQ there are plans to start incorporating feedback from the public.

The purpose of measuring police performance within the SPD, and especially EAQ, is twofold: 1) accountability for police action to the government and the public; and 2) to provide public safety while controlling collateral harms that can result from policing.

First, in line with the Settlement Agreement, the main purpose of Seattle’s new police performance measurement framework is to ensure greater accountability for the use of force by SPD officers to the federal government and the public.

Second, interviewees believed that EAQ allows officers to serve the public better. One interviewee, for instance, highlighted that the power given to law enforcement is directly dependent on the public: if the public does not agree with current policing approaches, different measures will have to be found. The SPD’s focus on the public is also visible through the many initiatives for obtaining public input in SPD practices (see Figure 4).

To serve the public better, the SPD felt that further insights into the quality of interactions between police officers and the public were required. As such, EAQ has sought to establish a baseline of what ‘good policing’ in this regard may look like. One interviewee noted that EAQ proposes a baseline method for determining if the behaviour of an officer towards the public is deteriorating, and whether early intervention would be appropriate:

Early intervention systems in the United States and specifically in Seattle tend to rely on data that has already been created in negative encounters, for example, complaints […] Those are the bad outcomes that we are trying to avoid with early intervention. […] If you have got an officer whose baseline is [up] here […] and you see them trending down, they have not received a complaint yet, they have not got into a bad use of force yet, they have not gotten in any collisions, but you are just beginning to notice that their interactions with the public are a little bit more abrasive than they were before, for example […] it gives us the opportunity to evaluate that officer, what is changed and whether we can intervene with some training or some time off, some restorative time off or some counselling, or whatever it might be to bring their trend line back up.

However, it is important to note that EAQ is not intended to measure the performance of individual officials, nor is it intended to act as a punitive measure. The programme is strategic in nature and not meant

410 INT5_US.
411 INT4_US; INT5_US.
412 INT5_US.
to replace good supervision. Rather, EAQ measurements and the resulting findings are meant to review practices and spark conversations about what officers and the SPD can do better as a team, and for the team.

7.2. Current practices in measuring police performance

The **EAQ programme operates in combination with other initiatives such as CompStat, but offers a different focus of measurement.**

The EAQ framework provides a new focus for police performance measurement in Seattle and complements rather than replaces CompStat (or SeaStat). According to one interviewee, EAQ borrows aspects from CompStat, for instance in terms of holding regular discussion meetings. However, the focus of performance measurement has shifted significantly, as interviewees noted that the SPD is now able to examine not only crime statistics, but also the entire organisation and corresponding behavioural changes.

The reasoning behind this approach is that the police, while only being able to influence crime statistics to a certain extent, can generally exert more influence on the three pillars of EAQ, namely the fairness of policing outcomes, the accountability of resources in the field and the quality of interactions with the public. The three components of EAQ, referred to as ‘pillars’ by interviewees and in the literature, and their purpose, as well as the measures and data used within each component, are detailed below.

**The three pillars of EAQ measure different aspects of policing, namely racial disparities, ‘under’ and ‘over’ policing, and the quality of interactions.**

**Equity.** The first component of EAQ focuses on the racial component of an investigative stop – also referred to as a ‘Terry Stop’. The quality assurance reviewers of EAQ explain that this does not concern the reasoning for choosing who to stop, but refers to the post-stop (and frisk) outcomes. SPD uses propensity score matching (see Box 7 for a brief overview of the relevant reasoning and methodology used) to measure decision-making disparities in police stop outcomes. The data around this measure of investigative stops is drawn from the filed police reports following such a stop, which are gathered from the SPD’s records management system.
Box 7: Assessing disproportionality in ‘Terry Stops’ using the equity pillar

According to Terry v Ohio, the decision to frisk someone requires an officer to have reasonable suspicion that the person is armed and presently dangerous. The creator of EAQ explained that with a propensity score matching model, the SPD’s research and analytics team’s aim under the equity pillar is to find as many variables as possible that ‘might represent the reasonable officer’s assessment of what dangerousness might look like for that person’. The propensity score is used to weight the observation.418

Propensity score matching is a quasi-experimental method and therefore makes use of a treatment and a control group. The method balances descriptive situational factors (e.g. time and day, location and/or what the original call was for), as well as personal factors (e.g. the officer’s age, experience, race and/or gender) with local factors (e.g. the number of times that the person is being stopped, frisked and a weapon has been found).419 These types of variables aim to approximate certain local knowledge. This is then abstracted over three-month periods, which helps to measure a broader picture – not just representing what the officer might think about a person, but also what the officer might already know about that person.420

In short, by controlling for other, non-race factors present during the stop, evaluators aim to identify and compare equity issues.421 The process helps measure the disparate impact of police services and facilitates better understanding of police performance in Seattle.422

1. Accountability. This component relates to measures of accountability in investigations through ‘under’ policing (neglect) or ‘over’ policing (harassment). For this, the SPD uses calls for service (from the community) and automatic vehicle locator (AVL) data, which track an officer’s location every few seconds. The calls for service are used to create a geostatistical model approximating what an officer might know about where calls will come from in their area. The SPD uses that model to explain police presence, measured through AVL, and visualise only the points not explained, or the residuals.423 Once these data have been collected,424 the data analytics team look for disparities in police officers’ locations against where police are needed throughout the city based on the calls for service coming into the police department.425 According to one interviewee, this analysis might show whether demand for police services in a particular area is being met. This serves as an indication that changes need to be made in policing strategies.426 This is not seen by the SPD as

418 INT3_US; INT4_US.
419 The SPD refer to these as ‘context’ and ‘hyper context’, with the latter referring to the abstracted calculations used to approximate what the officer might know about the subject and/or the situation that might contribute to their decision to conduct a frisk, but that is not explicitly in the fielded data (INT4_US).
420 INT3_US; INT4_US.
421 RTI International (2023).
422 INT4_US.
423 INT4_US.
424 The SPD receives information about 1) emerging demands; 2) situational awareness information to convey to the field (if an attacker can determine a predictable pattern of officer presence from AVL, they could anticipate police patrol patterns or presence for an ambush); and 3) over policing or false positive assumptions on the part of the officer about where demand for their services might be concentrated. INT4_US.
425 RTI International (2023).
426 INT4_US.
predictive policing or crime prediction. Rather, the SPD are modelling requests for police service to explain discretionary patrol activity.\textsuperscript{427}

2. Quality. The final component relates to quality measures of police services and is measured through the automated transcription of recordings from body-worn surveillance cameras. While this is a functional system, interviewees have noted that it is still in its testing phase.\textsuperscript{428} The transcription is conducted by a partner agency, which runs it through a machine learning program that assigns value to different phrases within the recorded conversation, leading to scores on the quality of the conversations. This helps assess a baseline for basic police services.\textsuperscript{429} However, the eventual aim of this measurement is not to show individual ratings, but to obtain overall performance numbers to establish departmental trends. This information can serve as input for further action such as training.\textsuperscript{430}

Figure 4: The three EAQ components

Source: Graph created based on information collected from interviews and RTI International (2023).

Community input, collected in various ways, is used to measure police performance under EAQ and to determine police priorities in the SPD.

One interviewee noted that community input, gathered through multiple channels under EAQ, is highly valued in the SPD as a whole.\textsuperscript{431} Some of the channels used include annual surveys, micro-community

\textsuperscript{427} INT4_US.
\textsuperscript{428} INT6_US.
\textsuperscript{429} See for instance Cision PR Newswire (2021).
\textsuperscript{430} INT3_US; INT4_US; INT5_US.
\textsuperscript{431} INT5_US.
policing plans where different groups can address their biggest safety concerns, advisory councils within each police precinct, demographic advisory councils for specific communities (e.g. LGBTQ, East Africa, Philippines) and different committees that meet to discuss their community’s concerns about policing. Qualitative and quantitative data from each of these instruments feeds into a collaborative policing bureau that analyses whether the police are working with the community to provide appropriate police services as intended. One interviewee noted that obtaining community input has become even more important to meaningful police performance measurement after the events of 2020.

To improve the pillars of EAQ specifically, further developments regarding community input are expected, such as analysis of community satisfaction.

According to an interviewee, all pillars of EAQ will likely be enhanced in coming years. One example regarding community input relates to a new initiative to collect customer satisfaction survey feedback. This initiative was recently added to the quality pillar to complement the automated transcription of body-worn surveillance. To collect such data, the SPD will be using ‘SPIDR Tech’, a software platform that sends customised text or email messages to people who call 911 (including crime victims or those who have reported a crime). It is used to update the status and set and maintain expectations of the police, as well as measure community satisfaction at various points of contact along the call processing workflow (from beginning to end).

The process of interacting with the community will soon be accelerated through EAQ by collecting what an interviewee refers to as ‘near real-time feedback’. This means that once a police officer closes a call, the person who called for help will receive a text message requesting information about their experience with the SPD. The information will be collected based on the Net Promoter Score. This procedure will be used to set expectations (e.g. providing real-time updates on the status of people’s cases) and to measure broader performance management of the SPD. Based on the scores and certain descriptive statistics, such as response time and time spent on a call, the officer will receive an average satisfaction score intended to give the supervisor information about the team’s performance. However, the interviewee stressed that such early-intervention systems are not meant to be punitive and should always be supportive in promoting good behaviour.

432 INT5_US.
433 INT5_US.
434 INT5_US.
435 Written correspondence in follow up to INT4_US.
436 Written correspondence in follow up to INT4_US.
437 INT4_US.
438 INT4_US.
439 INT4_US.
440 INT4_US.
441 The Net Promoter Score is a management tool measuring customer satisfaction and asking how likely customers are to recommend the product or service. As of 14 November 2023: https://www.netpromotersystem.com/about/measuring-your-net-promotor-score/?#:~:text=Your%20Net%20Promoter%20Score%20is,who%20scored%200%20to%206
442 INT3_US.

82
One interviewee perceived two potential challenges with implementing this near-real-time feedback plan: avoiding selection bias and subjective responses and ensuring that the public understands the importance of the project. They noted that it is therefore important to first ensure awareness of these plans among SPD staff, and that police officers explain to the public why they are doing what they are doing in a specific case (e.g. arresting someone, which can lead to negative feedback). 443

7.3. Challenges and complexities in measuring police performance

7.3.1. Issues with the reliability, validity and representativeness of data

One respondent expressed concern that a lack of stakeholder involvement could lead to the discontinuation of initiatives.

One interviewee working in the field of learning and development pointed out that the new practices introduced by the team still rely heavily on individuals to implement them, with the risk that they may fade due to staff turnover and the consequent loss of institutional memory. 444 As such, the interviewee emphasised the need to consult management and share with them the value of the practices in question for police performance, so that they may be integrated more fully by the organisation. 445

7.3.2. Other challenges/complexities

Changing the mindset of officers was seen as the most challenging, but essential, aspect of implementing new performance measurement practices.

One stakeholder from the SPD leadership interviewed for this study noted that it can be challenging to ensure that officers are open to new inputs, such as EAQ and associated leadership tactics. As such, this interviewee also envisioned difficulties with the implementation of action following the results of EAQ analyses. They explained that should the data begin to show certain trends about a particular squad’s performance, there may be some resistance in accepting the results. 446 A separate interviewee from the Performance Analytics and Research team provided a possible explanation for the resistance, namely that assuming responsibility may lead to liability. Therefore, departments may refrain from directly reporting poor performance. 447

According to the SPD leadership stakeholder, such a change in police culture, for example to introduce new and complex police performance measurement systems, needs to receive both top-down and bottom-up support. This may be achieved by appropriately training officers and driving the department’s mission with leadership. 448 Another interviewee working in learning and development spoke about a change in mindset regarding new human resource (HR) initiatives, such as increased coaching opportunities for leaders in the SPD and better employee engagement. Although this interviewee was hired as part of SPD’s response to

443 INT5_US.
444 INT7_US.
445 INT7_US.
446 INT5_US.
447 INT4_US.
448 INT5_US.
the need for police reform, they noted that the initiatives still required ‘pushing with command staff’. The interviewee nevertheless suggests that the reform within the SPD has been impressive so far in achieving an overall change of the philosophy of performance ‘management’ to performance ‘development’. They highlighted that this is crucial, because change cannot be achieved without focusing on the underlying mindset first.

Interviewees indicated that the ‘data side’ and the ‘operational side’ of EAQ do not seem sufficiently connected, and that communication between the two teams tends to be limited.

One interviewee mentioned that their preference is not to share detailed information about EAQ measurements and calculations with police commanders because of the complexity of the programme. They noted that there is a fear that the police leadership in question may make mistakes when talking about it to the public, which in turn may create issues for the department. This is in direct contradiction to one of the above-mentioned preferred practices of another interviewee, namely that management should be consulted to ensure the continuity of projects. This interviewee, however, also noted that the SPD has deliberately introduced EAQ to other team members slowly, starting with a descriptive product before moving to a more enhanced product that can be understood and implemented. Improvements on the connectedness are therefore said to be expected.

7.4. Perceived innovative and/or promising practices

The detection of anomalies in the performance trends of officers under the equity pillar can be used to promote conversations around early intervention and the needs of police departments.

The equity pillar of EAQ seeks to achieve completely equal interactions between police officers and the public (i.e. with no racial disparities), as well as high levels of public satisfaction with policing services. According to EAQ’s creator, the equity pillar is an innovative ‘process of identifying and isolating anomalies in the patterns’. By analysing the quality of interactions between the police and the public, anomalies in patterns or trends may be detected that indicate the performance of individual police officers is deteriorating. Rather than waiting for further deterioration or even incidents, these findings can steer the conversations that need to be held within police departments’ leadership about the performance of the squad, for instance to intervene in a timely manner or address any particular needs of the department (e.g. through providing relevant training).

The use of automated transcriptions for body-worn surveillance cameras under the quality pillar may offer a best practice example to establish clearer baselines and obtain buy-in.

According to one interviewee, the analysis of thousands of automated transcriptions for body-worn surveillance cameras under the quality pillar of EAQ allows the SPD to establish a clearer picture of police–
public interactions, and thus a clearer baseline for equitable treatment. \(^{454}\) EAQ’s quality assurance partners consider it likely that the technology used will be more widely implemented in the next decade. \(^{455}\) Although using artificial intelligence to establish baselines has its own limitations (e.g. the loss of sound in conversation, or failure to pick up on intonation or sarcasm), interviewees report that the transcription service uses a robust methodology. \(^{456}\) They note that the automated transcription of body-worn surveillance cameras is a valuable and efficient technology, especially as it significantly reduces man-hours.\(^ {457}\)

**Overall, EAQ is a promising initiative to address existing challenges and formulate answers to questions that other departments may not yet be asking.**

One of the interviewees highlighted that EAQ is one of the few initiatives that has been able to begin to address existing challenges, such as those identified in Section 7.3.\(^ {458}\) They explained that EAQ has already enabled the SPD to start to formulate answers to questions that other departments are ‘still too afraid to ask’, such as those around equity and disproportionality.\(^ {459}\)

The same interviewee also emphasised that it is important not to be daunted by complexity when measuring police performance, noting that this is an area in which EAQ is particularly helpful as it ‘looks for pieces of data that can help drive police performance or changes within an environment without having to make drastic decisions’. \(^ {460}\) That is, the results from EAQ analyses do not generally steer the agency towards large-scale restructuring of their strategy or environment. As mentioned, trends detected in the EAQ results (e.g. anomalies in equitable behaviour) are more likely to guide discussions between police forces and management, or serve as input for requesting more training, for example.\(^ {461}\)

### 7.5. Recent developments

**Besides ongoing developments within the three EAQ pillars, EAQ results will likely be incorporated into the SPD’s HR and learning and development structures.**

The measures implemented under EAQ are under continuous development to improve current practices. At the same time, EAQ results are also likely to be integrated into learning and development structures within the SPD in the near future.\(^ {462}\) Certain trends that emerge could trigger further training on certain topics or in certain departments, or even provide opportunities for cross-training (e.g. from one department to another), encouraging collaboration and knowledge exchange within the whole SPD. These initiatives

---

\(^{454}\) INT4_US.  
\(^{455}\) INT4_US.  
\(^{456}\) INT6_US.  
\(^{457}\) INT6_US.  
\(^{458}\) INT1_US.  
\(^{459}\) INT5_US.  
\(^{460}\) INT3_US.  
\(^{461}\) INT7_US.  
\(^{462}\) INT7_US.
are intended to improve SPD performance, including through better employee engagement and training opportunities.\textsuperscript{463}

**EAQ is also expected to be transferred to other jurisdictions.**

According to two interviewees, several other police departments have shown interest in the approach to police performance measurement taken by the SPD, and talks are currently ongoing between the creator of EAQ and police departments in the United States (for example those under a Consent Decree) and abroad about transferring the programme. Both interviewees proposed that EAQ, or certain aspects of the programme, are transferable to other departments or countries.\textsuperscript{464}

### 7.6. Concluding comments

EAQ presents an innovative approach to police performance measurement at a local level that offers new ways of using existing data. The development of this framework was driven by a critical need to reform policing standards and develop evidence-based practices.

As in other jurisdictions, such as the Netherlands (at a local level in some regions), Israel and Sweden, community input is used to measure the performance of the SPD under EAQ. This input is gathered through various channels, including new initiatives such as SPIDR Tech, a software platform that sends a customised survey text or email message to crime victims or those who have reported a crime, and the automated transcription of recordings from body-worn surveillance cameras.

However, as has also been highlighted in the chapters on Israel and North Rhine-Westphalia, such novel approaches to performance measurement rely on securing buy-in from police officers, which interviewees reported as a challenge. EAQ has sought to mitigate this by using results, such as the detection of anomalies in performance trends, to start discussions about the needs of police forces.

Overall, EAQ presents as a promising initiative that may address existing challenges and formulate answers to questions around racial disparities, ‘under’ and ‘over’ policing, and quality of interactions that other police jurisdictions have yet to ask.
8. Case study 5: Sweden

Box 8: Key findings in Sweden

<table>
<thead>
<tr>
<th>How did the methodology for the performance measurements come about? Were stakeholders involved? If so, which ones?</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Key stakeholders in police performance measurement are the Swedish Police Authority, the Ministry of Justice and the National Council for Crime Prevention.</td>
</tr>
<tr>
<td>• Before the 2015 move to a national police force, the Swedish police were criticised for a lack of effectiveness and a dysfunctional division of tasks and governance across authorities. The reform subsequently introduced a standardised approach to police performance measurement.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>For what purposes are the performance measurements used?</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Evidence suggests that the main purpose of measuring the performance of the Swedish Police Authority is to improve its practice, to show internal and external accountability, and to increase public confidence in the police.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>How is police performance measured in Sweden? What aspects of policing do the measurements cover?</th>
</tr>
</thead>
<tbody>
<tr>
<td>• The Swedish Ministry of Justice primarily uses quantitative primary and secondary indicators to measure the performance of the Swedish Police Authority in relation to its strategic priorities.</td>
</tr>
<tr>
<td>• Quantitative crime statistics are analysed and published in monthly and annual reports.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>What are the (further) advantages and disadvantages of the performance measurements? Are any (side) effects of the performance measurements known? If so, which ones?</th>
</tr>
</thead>
<tbody>
<tr>
<td>• A challenge for Sweden is that setting quantitative performance targets for the police has previously led to perverse incentives.</td>
</tr>
<tr>
<td>• It also remains difficult for stakeholders to establish what good police performance entails, and criteria for success are not established at a national level. Other challenges include budget and capacity constraints, and difficulties in obtaining stakeholder approval.</td>
</tr>
<tr>
<td>• Among recent developments, Sweden is seeking to mitigate some of the challenges through an increasing focus on research and by facilitating technical development.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>How reliable, valid and representative are the performance measurements?</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Interviewees praised the quality and availability of data on police performance in Sweden.</td>
</tr>
<tr>
<td>• However, as in other jurisdictions, interviewees perceived the indicators as not being entirely representative of police performance.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Which (parts of) foreign performance measurements can be labelled as 'qualitatively good', 'innovative' and/or 'promising'?</th>
</tr>
</thead>
<tbody>
<tr>
<td>• There are several promising initiatives aimed at incorporating various measurements into current policing practices to better capture police performance, including the harm index and the complexity index.</td>
</tr>
<tr>
<td>• The police are also working on strategies to improve current measurement practices, including the continuous development of indicators and new surveys.</td>
</tr>
</tbody>
</table>
8.1. Organisational context, stakeholders and the evolution of performance measurement

The Swedish Police Authority is the national police force of Sweden and is led by a National Police Commissioner. The force is comprised of the Office of the National Police Commissioner, seven police regions, the Department of National Operations, the National Forensic Centre, six national departments, and the Department of Special Investigations (see Figure 5). It is further divided into 95 local police districts and 25 police districts, although these numbers are approximate as organisational changes are ongoing. In total, the Swedish Police Authority employs 25,000 officers.

Each of the seven police regions is responsible for policing in its geographical area. The regions may also be tasked with police development activities or, for example, supporting the national police whenever required. Decision-making authority is also given to police regions, although it should be noted that some tasks are coordinated at the regional level.

The current police system came into force in 2015, after an extensive reform of the Swedish police. During this reform, the police changed from 21 separate police authorities and a National Police Board into the single national Swedish Police Authority it is organised as today.

Figure 5: Organisation of the Swedish Police Authority

Source: Adapted from Polisen (2023).

---

465 Polisen (2023).
466 OSCE POLIS (2023b).
467 OSCE POLIS (2023b).
468 Sveriges Riksdag (2014).
Before the 2015 reform, the Swedish police were criticised for, among other things, a lack of effectiveness and a dysfunctional division of tasks and governance across authorities. Interviewees noted that the organisational reform has generally led to a standardised approach to policing in Sweden, including regarding police performance measurement. Most notably, the authority is expected to deliver and report on results in annual reports. One interviewee explained that the strategic aims for the police are now generally the same at both national and regional levels. These strategic aims are guided by the two main objectives for law enforcement set by the Ministry of Justice: 1) to reduce crime and increase safety/security under criminal law; and 2) to preserve the rule of law. However, the same interviewee added that individual police regions are also able to develop their own performance measurements to sit alongside the national framework.

A second interviewee stressed that the process towards a national approach to police performance measurement, initiated by the 2015 reforms, has yet to be completed. They noted that performance measurement has not been consolidated nationally in its entirety as there are ongoing initiatives not being implemented at the national level. Furthermore, the reform has been criticised for being overly instrumental and neglecting human elements such as organisational cultures and social relations, as well as losing the benefits of local expertise through centralisation.

Key stakeholders in police performance measurement are the Swedish Police Authority, the Ministry of Justice and the National Council for Crime Prevention.

The Police Authority measures performance in the regions (regional measurement) and nationally (e.g. for the annual reports). It may also request evaluations of specific interventions by regions or independent researchers. Police regions may measure (regional) performance in parallel to the authority’s assessment.

The Ministry of Justice makes decisions based on the results presented by the national police. However, they have limited capacity to use or create data beyond what they are provided with by the police and other government authorities.

The National Council for Crime Prevention (Brottsförebyggande rådet or Brå) produces official crime statistics, carries out research and supports local work on crime prevention. Research carried out includes analysis and evaluation of working methods of the police and police investigations. It also publishes public

---

469 Holmberg (2014).
470 The annual report highlights the policing activities and outcomes of a given year, as well as resources, future priorities and any other special topic areas that may have been assigned by the government [INT5_SWE].
471 Written responses by interviewee in follow up to INT2_SWE.
472 Written responses by interviewee in follow up to INT2_SWE.
473 INT3_SWE.
474 Ivarsson Westerber (2020).
475 Polisen (2022).
476 INT1_SWE.
477 Written responses by interviewee in follow up to INT2_SWE.
478 Written responses by interviewee in follow up to INT2_SWE.

89
Other authorities that conduct research on topics related to policing in Sweden are the Swedish Agency for Public Management and the Swedish National Audit Office.\(^{480}\)

The National Police Commissioner is responsible for decisions around actions and changes within the police,\(^{481}\) and other government bodies use the evidence gained from police performance measurements to inform broader national priorities and policies.\(^{482}\)

The Police Authority measures its performance to improve its practice, to show internal and external accountability, and to boost public confidence in the police.

One of the main purposes for measuring police performance in Sweden, according to four interviewees, is to facilitate organisational improvement.\(^{483}\) According to one of those interviewees, the police use performance measurement results for internal assessments on how the police are progressing towards achieving the strategic aims set by the Ministry of Justice.\(^{484}\)

A second reason for measuring the performance of the Swedish police is to show accountability. One interviewee highlighted the importance of external accountability – to be able show the public ‘what is being done’.\(^{485}\) A second interviewee noted the need to be able to demonstrate to management or police leadership what the police has achieved.\(^{486}\) Two interviewees within the police authority also stated that internal accountability is especially important, because in their experience the highest layer of management is not always sufficiently informed of what is happening ‘on the ground’.\(^{487}\)

Interviewees also noted that measuring police performance may serve to encourage public confidence in the police. One interviewee suggested that the media can affect public attitudes to the police by focusing on increasing crime rates or reporting on perceived negative results. They noted that demonstrating and explaining certain aspects of performance results can help mitigate the impact on public trust of such reporting, for instance by explaining that cases may take longer to resolve because they have become more complex (using the so-called ‘complexity index’, see Section 8.4).\(^{488}\)

8.2. Current practices in measuring police performance

Police performance in Sweden is currently measured primarily through quantitative crime indicators. Stakeholders at all levels predominantly measure police performance by analysing crime statistics (e.g. reported cases, suspects, prosecutions) and by administering and analysing surveys.\(^{489}\) The

\(^{479}\) Brötsförebyggande rådet (Brå) (2023a).
\(^{480}\) Bogestam and Bergnor (2018).
\(^{481}\) Polisen (2021).
\(^{482}\) Brötsförebyggande rådet (Brå) (2023b).
\(^{483}\) INT1_SWE; INT2_SWE; INT3_SWE.
\(^{484}\) INT3_SWE.
\(^{485}\) INT5_SWE.
\(^{486}\) INT5_SWE; INT7_SWE.
\(^{487}\) INT7_SWE.
\(^{488}\) INT7_SWE.
\(^{489}\) INT1_SWE; INT2_SWE; INT5_SWE; INT6_SWE.
International approaches to police performance measurement

corresponding indicators used are primarily quantitative.490 According to interviewees, this approach has been chosen because measuring qualitative aspects using data from predominantly quantitative case management systems was deemed challenging,491 and quantitative data are generally easier to produce.492

The Ministry of Justice uses primary and secondary indicators formulated to measure the performance of the Swedish Police Authority.

As mentioned, police performance measurement in Sweden has not yet been completely consolidated at a national level. However, the Ministry of Justice has formulated four primary performance indicators, aligned with the government’s strategic priorities (reduce crime, increase safety, preserve the rule of law), whereby the performance of the national police is measured annually:

1. Number of reported crimes for different crime types
2. Perceived safety in the population
3. Victimisation rates
4. The level of trust in the police.

The Police Authority is expected to report on the results for each of these indicators in its annual report.493 Interviewees noted that there are no specific targets attached to these indicators, but instead an understanding of in which direction the numbers should be going.

There are also several secondary indicators (Figure 6) that sit under the strategic priorities.494 According to one interviewee, these indicators were formulated by the Ministry of Justice to allow the Police Authority to report on its performance at a more detailed level. The sub-indicators also leave room for further aggregation for each authority.495

---

490 INT3_SWE.
491 Written responses by interviewee in follow up to INT2_SWE; INT4_SWE.
492 INT3_SWE.
493 INT2_SWE.
494 Written responses by interviewee in follow up to INT2_SWE.
495 Written responses by interviewee in follow up to INT4_SWE.
Figure 6: Sub-indicators designed by the Ministry of Justice

Overall, as both the indicators and the sub-indicators demonstrate, the main areas of performance measurement concern victimisation rates, fear of crime and confidence in the criminal justice system.\(^{496}\) This also aligns with the perceived focus of the public when it comes to policing in Sweden. According to one interviewee, public debate has primarily concentrated on traditional indicators (most notably clearance rates and levels of public trust).\(^{497}\)

The Swedish government has expressed a desire to move away from a focus on cost-effectiveness and value for money. Instead, it has recently opted for what one interviewee termed a more ‘trust-based’ form of governance, which is built on the assumption that with clear objectives, employees can and should be trusted to find ways to achieve these objectives, while needing fewer performance indicators.\(^{498}\)

\(^{496}\) Brottsförebyggande rådet (2023c).

\(^{497}\) INT1_SWE.

\(^{498}\) Written responses by interviewee in follow up to INT2_SWE.
Quantitative crime statistics are analysed and published in monthly and annual reports.

Police performance in Sweden is most often measured through the analysis of both crime statistics and public survey results.

For the four main indicators set by the Ministry of Justice, data are supplied by Brå. The first indicator (number of reported crimes by different crime types) is measured through data from Brå’s official statistics, which are collected from police registration systems. The analysis of these crime statistics happens both monthly and ad hoc. To conduct such analyses, the financial department of the Police Authority works with a case management system that communicates the relevant crime statistics to data analysts. According to interviewees, the main interest is the number of cases investigated versus whether they were ‘successfully’ closed.\footnote{INT3\_SWE.}

For the remaining three indicators, data stem from Brå’s National Crime Survey,\footnote{Written responses by interviewee in follow up to INT2\_SWE.} which was highlighted by several interviewees as an important part of measuring police performance in Sweden.\footnote{See for instance INT2\_SWE; INT3\_SWE} This is explained in more detail in Box 9.
Box 9: The Swedish National Crime Survey

The Swedish National Crime Survey, produced by Brå, is an annual survey that collects data on the opinions and experiences of the Swedish population in relation to victimisation, fear of crime, confidence in the criminal justice system and crime victims’ contacts with the criminal justice system. It has been conducted since 2006.502 In the most recent crime survey (2022), almost 65,000 (out of a sample of 200,000 people) participated. The sample is drawn from the population register from Statistics Sweden.503 It is noted that younger people are overrepresented in the survey sample.504 During data analysis, cases are weighted to account for both overrepresentation and for differences in non-response among groups.505 Over time, the survey has changed. The method was revised in 2017, changing from telephone interviews to Internet and postal surveys so that more people would be able to participate.506 Furthermore, since its first iteration in 2006, some questions have been revised and new questions have been added. For example, questions concerning certain relevant topics for today’s society such as pickpocketing, credit card fraud and online harassment have been added.507

The survey is one of the most important measures of police performance in Sweden. However, Brå points out that the sample is not representative of all age groups (only those aged 16–84 are sampled), nor does it include people who reside in institutions (e.g., prison). It is also assumed that the most marginalised people in society are not well represented.508 One interviewee highlighted that with such surveys it is generally more difficult to reach certain groups either because of a lack of trust in the police or language barriers.509

Interviewees praised the quality and availability of data on police performance in Sweden.

Four interviewees noted that the data that stakeholders collect and produce to measure police performance are generally of good quality and readily available. Two interviewees mentioned that Sweden is very rich in data – and ‘more or less everything is documented’, indicating that data covering many policing aspects are available.510 Two other interviewees confirmed that the data coming from the case management systems are ‘generally very good’.511

8.3. Challenges and complexities in measuring police performance

8.3.1. Issues with the reliability, validity and representativeness of data

Some of the indicators formulated by the Ministry of Justice have caused undesirable results. For example, one interviewee noted that the police were previously given specific targets for the number of breathalyser tests to be taken per region. However, it was subsequently found that officers were purposefully conducting tests during periods when fewer drunk drivers were expected to be on the road, meaning that fewer drivers

502 Brottsförebyggande rådet (2023c).
503 Statistics Sweden (2023).
504 Brottsförebyggande rådet (2023c).
505 Brottsförebyggande rådet (2023c).
506 INT2_SWE.
507 Brottsförebyggande rådet (2023c).
508 Brottsförebyggande rådet (2023c).
509 INT8_SWE.
510 INT6_SWE.
511 Written responses by interviewee in follow up to INT2_SWE.
would test positive for alcohol. Given that the processing of drivers who test positive can take a substantial amount of time, officers were able to carry out more tests in this way.\textsuperscript{512} The same interviewee highlighted that cases involving repeat offenders are more likely to be brought to court in Sweden as they are less complex than cases involving ‘new’ offenders, which allows the police to reach targets related to prosecution rates more easily.\textsuperscript{513}

8.3.2. Other challenges/complexities

Interviewees felt that existing indicators are not entirely representative of police performance. A common criticism among interviewees was that the set of indicators is not fully representative of the nature of police work.\textsuperscript{514} Some interviewees noted that the indicators are too narrow,\textsuperscript{515} while others highlighted that they do not necessarily provide a better understanding of how the police are contributing to the objectives set by the Ministry of Justice.\textsuperscript{516} It has also been proposed that the indicators do not show the complexity of cases that the police handle, and that they are too simplified to meaningfully measure police performance.\textsuperscript{517}

Two other interviewees mentioned that the indicators do not touch upon services provided by the police that are not directly crime-related, such as traffic control or social services. They noted that while they believe it should be possible to capture more of this work, a stronger evaluation culture within the law enforcement context (and relevant government authorities) would need to be developed.\textsuperscript{518}

One interviewee similarly re-emphasised that while there is generally a willingness to improve current practices, evaluations are often only carried out in relation to a specific intervention, such as gang violence, which leads to an inability to detect any overall effects of policing.\textsuperscript{519}

Evidence from interviews also suggests that there is no consensus on what ‘good’ police performance entails, especially as criteria for success are not standardised at the national level. As in other jurisdictions, such as Seattle (see Chapter 7), two interviewees mentioned a lack of consensus on how ‘good’ policing should be defined. This complicates the task of measuring police performance,\textsuperscript{520} which is made even more difficult by the fact that criteria for success differ across regions.\textsuperscript{521}

Another interviewee also mentioned that there is a discrepancy between the wants and needs of the police, and what can feasibly be measured with existing data.\textsuperscript{522}

\textsuperscript{512} INT2\_SWE.
\textsuperscript{513} INT2\_SWE.
\textsuperscript{514} For instance, written responses by interviewee in follow up to INT2\_SWE; INT3\_SWE; INT6\_SWE; INT8\_SWE.
\textsuperscript{515} INT1\_SWE.
\textsuperscript{516} Written responses by interviewee in follow up to INT2\_SWE.
\textsuperscript{517} INT3\_SWE.
\textsuperscript{518} INT6\_SWE.
\textsuperscript{519} INT1\_SWE.
\textsuperscript{520} INT2\_SWE.
\textsuperscript{521} INT3\_SWE.
\textsuperscript{522} INT1\_SWE.
Other challenges include budget and capacity constraints.

Several interviewees noted that resource constraints, especially in terms of budget and capacity, had imposed limits on police performance measurement. One interviewee noted that there have been issues obtaining funds for evaluation research. A second interviewee talked about how the capacity to analyse results in relation to resources is still underdeveloped. According to this same interviewee, resource constraints became an issue after the 2015 reform, when more money was spent on hiring new police officers. The constraints have led to uncertainty about the extent to which the police can now measure and demonstrate their performance results, and thus how well they will be able to demonstrate any returns on investment.

For two other interviewees who work on a separate project for which they have developed a new survey (see section 8.5), budget was also an area of concern, and the main reason why the new survey has not been conducted in more areas yet, and why the results have not been analysed to the extent that the interviewees would have liked.

Interviewees highlighted challenges with obtaining stakeholder approval and engagement for new measures or initiatives.

Interviewees noted that securing buy-in from stakeholders for new initiatives to improve police performance measurement often proved challenging. This was the case regarding a new initiative to start designing and applying a ‘complexity index’ (see Section 8.4), with the creator explaining that it takes time to create enthusiasm for new ways of thinking. Two other interviewees working on the above-mentioned survey initiative also noted that to improve current methods, all relevant parties – in this case municipalities – must be ‘on the same page’ and agree and support the changes to be implemented. These interviewees noted that they had found it challenging to involve stakeholders.

Partly related is a comment from an interviewee who experienced issues with communication and the lengthy bureaucratic processes of the Swedish government and Police Authority, which resulted in delays in obtaining data to be used to evaluate the police.

8.4. Perceived innovative and/or promising practices

There are several ongoing promising strategies and initiatives in Sweden that contain elements designed to counter some of the above challenges.

To mitigate some of the challenges involved with police performance measurement, stakeholders in Sweden have recently demonstrated an increased focus on research and technical development.

Evidence from the literature and interviews suggests an increased focus on conducting research and implementing research findings related to police performance measurement in Sweden. One interviewee

523 INT1_SWE.
524 Written responses by interviewee in follow up to INT2_SWE.
525 INT8_SWE.
526 INT7_SWE.
527 INT8_SWE.
528 INT1_SWE.
International approaches to police performance measurement

highlighted increased cooperation with non-police researchers. \(^{529}\) Similarly, the Swedish government recently commissioned a report published by Brå (2022) that proposes a new model for measuring police performance, and supports the development of new indicators for performance measurement in Sweden. \(^{530}\) Interviewees emphasised the need for cooperation between all authorities to design the indicators, and to design and implement a strategy that is holistic and comprises the whole judicial system (i.e. not just the police, even though some indicators do purely measure police performance). They refer to this as the ‘umbrella perspective’, and note that it is important for multiple reasons: 1) to see how the police has actually performed (e.g. after prosecution of a case); 2) to allocate resources effectively; and 3) to adapt parts earlier in the process:

I view the workload of the criminal justice system as a whole is like a balloon and if you press a part of the balloon, it squeezes, it extends somewhere else. […] so, in order to have the whole balloon in front of you, you have to watch the whole system. \(^{531}\)

Among others, Brå suggested a set of indicators (18 in total) that are a combination of existing, new and updated indicators. Two important considerations for the development of this set of indicators were highlighted by interviewees: 1) the indicators have been designed in dialogue with all relevant authorities in the criminal justice system, because they themselves have to feel that the indicators are useful and operational for their work; and 2) the indicators have been designed to be easily usable and the framework includes indicators that are more ‘statistically adequate’ and measurable. \(^{532}\) The proposed indicators are shown in Table 18. According to the interviewees, most leave room for the implementation of specific measurements for each authority. \(^{533}\)

Table 18: Suggested new indicators

<table>
<thead>
<tr>
<th>Themes</th>
<th>Performance indicator</th>
<th>Description of indicator</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Volume management (^{534})</td>
<td>Percentage of offences linked to a suspect that result in a sanction.</td>
<td>The ratio between the offences linked to a suspect and the number of offences that result in a sanction.</td>
<td>Monitoring the proportion of all offences linked to registered suspects that result in a sanction.</td>
</tr>
<tr>
<td></td>
<td>Backlog size.</td>
<td>The ratio between the number of offences linked to a registered suspect that remain in the backlog at some point in the criminal justice process at the end of the current year, and the total number of offences linked to a registered suspect that are processed by the investigative, prosecutorial and</td>
<td>Monitoring the ability of the investigative and judicial authorities to manage their case flow.</td>
</tr>
</tbody>
</table>

\(^{529}\) INT1_SWE.

\(^{530}\) Written responses by interviewee in follow up to INT2_SWE.

\(^{531}\) INT4_SWE.

\(^{532}\) INT4_SWE.

\(^{533}\) Email correspondence in follow up to INT4_SWE.

\(^{534}\) Volume management refers to case flows; the process from offence report until (possible) sanction.
<table>
<thead>
<tr>
<th>Themes</th>
<th>Performance indicator</th>
<th>Description of indicator</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Regional consistency</strong></td>
<td>Backlog age.</td>
<td>The time (median number of days) that has passed as of 31 December since the offences comprising the backlog were registered in the criminal justice system.</td>
<td>Monitoring the length of the processing times associated with the cases that remain open at the end of the year.</td>
</tr>
<tr>
<td></td>
<td>Duration of the criminal justice process.</td>
<td>The time (median number of days) from the reporting of an offence to the publication of a verdict in a district court.</td>
<td>Monitoring the timeframe regarding a decision on the issue of guilt for complainants and offenders.</td>
</tr>
<tr>
<td></td>
<td>Investigations requiring additional investigative measures following final submission to prosecutor.</td>
<td>The number of decisions made by prosecutors requiring that additional investigative measures are conducted in police investigations submitted as having been completed.</td>
<td>Monitoring the extent to which the decisions made by prosecutors concerning completed police investigations are made in a uniform manner.</td>
</tr>
<tr>
<td></td>
<td>Charged offences dismissed by district courts as unproven.</td>
<td>The number of charged offences dismissed as unproven in district court judgements during the year, in relation to the total number of charged offences that have led to a conviction in the same year.</td>
<td>Monitoring the extent to which the charges presented by prosecutors to the courts are of equal quality across the country.</td>
</tr>
<tr>
<td></td>
<td>Rulings changed by the courts of appeal.</td>
<td>The number of charged offences for which the verdict of a court of appeal differs from that of the district court, in relation to the total number of charged offences reviewed by the courts of appeal during the year.</td>
<td>Monitoring the extent to which the investigations that result in an appeal are of equal quality across the country.</td>
</tr>
<tr>
<td><strong>Disruptive measures</strong></td>
<td>Number of resource-intensive criminal investigations and cases.</td>
<td>The number of criminal investigations in which at least four out of seven conditions indicating a resource-intensive case are met. The seven conditions are: 1) more than (x) registered suspects; 2) more than (x) complainants; 3) at least (x) months of pre-trial detention?; 4) the use of at least (x) covert coercive investigative measures; 5) at least (x) applications for international legal assistance; 6) investigation results in a main court hearing lasting at least (x) hours; and 7) investigation results in prison</td>
<td>Monitoring the ability of the criminal justice system to investigate, prosecute and try particularly resource-intensive cases.</td>
</tr>
</tbody>
</table>

---

535 Regional consistency refers to measuring the legal principles of uniformity, predictability and consistency.

536 Disruptive measures target investigative and prosecutorial operations that aim to reduce criminal activity.
## International approaches to police performance measurement

<table>
<thead>
<tr>
<th>Themes</th>
<th>Performance indicator</th>
<th>Description of indicator</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>sentences amounting to at least (x) years.</td>
<td>Confiscation of criminal proceeds.</td>
<td>Monitoring the extent to which the criminal justice system is able to confiscate criminal proceeds from offenders.</td>
</tr>
<tr>
<td></td>
<td>The number of requests to confiscate criminal proceeds approved each year.</td>
<td>Company fines.</td>
<td>Monitoring the extent to which the criminal justice system acts to prevent companies being used to commit offences.</td>
</tr>
<tr>
<td></td>
<td>The number of such disqualifications imposed each year.</td>
<td>Disqualifications from commercial activity.</td>
<td>Monitoring the extent to which the criminal justice system investigates and prosecutes individuals who have committed sophisticated offences using businesses as a vehicle in the commission of these crimes.</td>
</tr>
<tr>
<td></td>
<td>The number of persons per 100,000 of the population who appear as complainants in initiated police investigations.</td>
<td>Complainants in initiated police investigations.</td>
<td>Monitoring the ability of the criminal justice process to safeguard the rights of victims of crime in cases referred to the criminal justice system.</td>
</tr>
<tr>
<td></td>
<td>The number of complainants who have had their case heard by a district court in relation to the total number of complainants registered in connection with investigated offences that have been linked to a suspect.</td>
<td>Complainants whose cases have been heard in court.</td>
<td>Monitoring the proportion of complainants in initiated police investigations whose cases are heard by a district court.</td>
</tr>
<tr>
<td></td>
<td>The proportion of people who report very or relatively positive experiences of the justice system authorities minus the proportion who report having very or relatively negative experiences.</td>
<td>Experiences of the criminal justice process.</td>
<td>Monitoring the perceptions of crime victims and others involved in the criminal justice process regarding their contacts with the criminal justice system.</td>
</tr>
<tr>
<td></td>
<td>The proportion of people with a very or relatively high level of confidence in the criminal justice system, minus the proportion with a very or relatively low level of confidence in the criminal justice system.</td>
<td>Confidence in the criminal justice system.</td>
<td>Monitoring the proportion of Swedish residents who have confidence in the authorities involved in the criminal justice process.</td>
</tr>
</tbody>
</table>

---

537 This area focuses on the criminal justice process, legal security and recidivism.
There are also some ongoing digital developmental innovations, for instance a large-scale project to ‘develop a digital, case-level system connecting all the authorities in the criminal justice system’ that includes the Police Authority, the Public Prosecutor and the courts.\(^{538}\) This aligns with the suggestion from one interviewee, quoted above, that law enforcement should be viewed as just one element of the judicial system, which requires cooperation and information sharing to function properly.

**Some initiatives aim to incorporate various measurements into current policing practices to better capture police performance, including the crime harm index and the complexity index.**

In Sweden, there is a shift towards the inclusion of different measurements, most notably two indexes. The **harm index** is a system that takes the extent to which crime harms victims, rather than the volume of reported crimes as a measure. While the index originated in Cambridge, UK, it has been adapted for use in several countries (see also the New Zealand Crime Harm Index in Annex A).\(^{539}\) In response to the need to better understand data, and to explain to the public what the police does, the Swedish version of the harm index is currently being implemented in Swedish performance measurement. Interviewees noted that they expect this index to improve accountability, legitimacy and communication about the effects of police work. They likewise noted that the index is meant to counter some of the above-mentioned criticisms, including calls for the development of more representative and meaningful indicators.\(^{540}\) One interviewee highlighted

<table>
<thead>
<tr>
<th>Themes</th>
<th>Performance indicator</th>
<th>Description of indicator</th>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recidivism</td>
<td>Recidivism level.</td>
<td>The proportion of convicted offenders released from prison, or receiving a non-custodial sanction, who subsequently receive a new conviction within a year.</td>
<td>Monitoring the proportion of convicted offenders who reoffend.</td>
</tr>
<tr>
<td></td>
<td>Correctional system clients who have completed risk-reduction interventions.</td>
<td>The proportion of correctional system clients with identified risk-reduction needs in relation to recidivism who have completed at least one risk-reduction intervention focused on recidivism during the course of their sanction.</td>
<td>Monitoring the extent to which correctional system clients complete needs-based interventions against recidivism.</td>
</tr>
<tr>
<td></td>
<td>Correctional system clients in occupational/treatment activities at correctional facilities.</td>
<td>The proportion of the total time available that is spent in such activities.</td>
<td>Monitoring the extent to which inmates have access to activities that can improve their chances of not reoffending (such as work, education or treatment) at their correctional facility.</td>
</tr>
</tbody>
</table>

Source: Brottsförebyggande rådet (2022). ‘En modell för uppföljning av rättskedjans myndigheter.’

---

\(^{538}\) INT3\_SWE.

\(^{539}\) For more info see for instance: University of Cambridge (2023).

\(^{540}\) INT5\_SWE; INT7\_SWE.
that it may sometimes be more important to establish where the most harm is done, rather than considering
the volume of cases, so that police priorities can be adjusted accordingly.\textsuperscript{541}

The \textbf{complexity index} is an initiative of an employee within one of Sweden’s police departments to improve
the current (according to one interviewee) simplistic framework for police performance measurement.\textsuperscript{542}
The programme has been praised as promising and necessary by various interviewees.\textsuperscript{543}

The index is designed to give a complexity rating to every case handled by the police. It considers the type
of crime and the number of resources required, most notably documents and interviews, to close a case.
According to the creator, the ratings therefore balance each other out: a murder, for instance, will receive a
higher complexity score, but a straightforward investigation will involve fewer interviews and documents,
and thus receive a lower complexity score.\textsuperscript{544}

The index has been designed to demonstrate to the public, media and internal management the effort
required from the police for different cases. The idea is that cases have become more complex over time, for
instance in relation to organised crime, and thus require more resources to be resolved.

The index is part of a software application, the results of which will begin to complement the current picture
of police performance of the department in question.\textsuperscript{545}

\textbf{8.5. Recent developments}

\textbf{The police authority in Sweden is working on strategies to improve current measurement practices,}
\textit{including the continuous development of indicators and new surveys.}

Within the police, efforts are ongoing to create more variety in the performance indicators used, for instance
by trying to design more qualitative indicators.\textsuperscript{546}

The Police Authority and the Swedish University Linnaeus have also recently begun to conduct ‘safety
surveys’. The coordinators of this project emphasise that this survey is different to the National Crime
Surveys,\textsuperscript{547} with topics including (people’s experiences with) criminal networks or gangs, and domestic
violence. The interviewees explained that the reason for changing the topics and the types of question was
that the National Crime Survey topics were felt to be outdated and no longer fully relevant to police work.

Interviewees argued that the main benefit of this new survey is that data can start to be collected at a local
level, such as for specific areas or streets. Such knowledge can support police activities and prioritisation,
and according to interviewees cannot be gathered from the National Crime Survey.\textsuperscript{548}
8.6. Concluding comments

Sweden has a clear set of limited and well-communicated indicators that, unlike in the Netherlands, are primarily consolidated at the national level. Nevertheless, these indicators have been criticised in both literature and the interviews conducted for this study. This critique primarily focuses on their quantitative nature, their over-simplification and the fact that they may not be representative of the entire spectrum of police activity. At the same time, the reform of the Swedish police in 2015, and the associated changes in policing and police performance measurement, have also presented some challenges.

As this chapter has demonstrated, various initiatives and ideas have been suggested to improve how the performance of the Swedish police is assessed. Examples include an increased use of research in police performance measurement, and the inclusion of the harm and complexity indexes in some departments. According to interviewees, these approaches have the potential to complement traditional crime statistics.

Several interviewees emphasised the need for better quality evaluation of police performance in Sweden. However, they noted that the Swedish authorities would first need to cultivate an ‘evaluation culture’ that facilitates and promotes such activities. Moreover, data sharing processes would need to be improved, and additional resources should be made available.
9. Cross-cutting themes in police performance measurement

The case study police jurisdictions analysed in this study have significant differences in terms of their size, organisational structure and resources, as well as their approaches to measuring performance. Nevertheless, it is possible to identify recurring themes in terms of how these organisations and their stakeholders have developed and implemented performance measurement frameworks. Similarities can also be found regarding the types of indicators used to measure performance and their associated data sources, the aspects of policing that the indicators seek to measure, and the risks and challenges of using these indicators. There are also commonalities in terms of promising and/or innovative practices in measuring performance. This chapter reflects on these themes.

9.1. Developing, implementing and maintaining performance measurement frameworks

A key initial consideration in developing a police performance measurement framework is to establish a shared understanding of its purpose.

The study team identified a variety of goals for measuring performance across and even within jurisdictions from the literature and interviews:

- Increasing and maintaining public trust and confidence in the police (noted by interviewees in England and Wales, Israel, NRW, Seattle, Sweden).
- Supporting learning and improvement within police forces (England and Wales, Israel, NRW, Sweden).
- Demonstrating the return on investment for taxpayers for policing services (England and Wales, the Netherlands, NRW).
- Demonstrating police efficiency and effectiveness (England and Wales, Israel, the Netherlands).
- Supporting crime reduction efforts (England and Wales, Sweden).

Establishing a performance measurement framework also requires broader consideration of how the role of the police should be understood by its stakeholders. In England and Wales, for example, interviewees commented on the view of some national-level policymaking stakeholders that police performance should be measured in terms of effectiveness in crime reduction. However, local-level practitioner stakeholders may
consider that the police’s work in safeguarding vulnerable people, increasing public perception of safety and targeting crimes which create the most harm should be where performance is measured.

**This links to the need for a coordinated approach to police performance measurement in which the functions and tasks of different organisations are clearly defined and delineated.**

Across all jurisdictions, a range of national and local police performance stakeholders were involved in the process of developing, implementing and maintaining measurement frameworks. In Sweden, there is a move towards centralising this process, albeit with local measurements in place, whereas the approach in Israel, NRW and Seattle reflects increasing localisation in policing. For example, in NRW even regional units within the state develop their own measures. In England and Wales, a range of national and local stakeholders have overlapping roles and have developed various performance measures. These can be misaligned and pose challenges for forces to implement. Similar misalignments between national and local stakeholders were reported by interviewees from Israel, who noted that when the National HQ started measuring police performance using EMUN metrics and targets in 2016, district commanders continued to measure police units according to the old measurement system. This meant that it was unclear for station commanders where they should focus their resources.

While Sweden moved to a national police force in 2015, this transition was accompanied by a strategy that sets clear and agreed-upon objectives, goals and indicators for measuring police performance. Interviewees furthermore indicated that within the new system, the relationship between different stakeholders, such as the Police Authority and the Swedish National Council for Crime Prevention (Brå), is clear, as are their respective responsibilities. This is shown, among others, by the execution of their respective tasks concerning data collection and data sharing. However, this system is not fully nationalised, and success criteria are not set or implemented, resulting in variations in local approaches.

**Securing and maintaining engagement from officers in how performance is measured has proved important.**

Evidence from Israel indicates that securing and maintaining officer engagement can be a significant challenge, particularly when the implementation process is complicated and necessitates a large degree of organisational change. The EMUN police reforms sought to tackle this issue by involving officers of all ranks in organisational performance measurement, rather than adopting a strictly top-down approach. To support these efforts, the EMUN framework included a performance-tracking dashboard that was updated in real-time and that officers of all ranks could access freely.

Interviewees from Seattle and Sweden likewise mentioned the importance of ensuring buy-in among force stakeholders. In particular, interviewees from Seattle noted that to bring about change to police performance measurement within police departments, efforts must be made both top-down (by police leadership) and bottom-up (for instance by training officers early on). It was also noted that efforts should be made to meaningfully engage all levels of officers in the process from the very start so that data collection (and analysis) can be operationalised in a focused and systematic way.
9.2. Indicators and data sources used to measure police performance

Table 19 provides an overview of the types of indicators used in the case study jurisdictions to measure police performance. In this typology, indicators can be loosely grouped as pertaining to productivity on preventing, reducing and responding to crime; experiences and perceptions of public safety, particularly relating to traffic incidents; citizen trust, and satisfaction and confidence in the police; and equity and procedural justice, such as proportionality in police checks.

The data used for these indicators are predominantly (even exclusively for some jurisdictions) quantitative.

In particular, stakeholders in all jurisdictions rely strongly on reported crime data gathered by the forces themselves to inform part or all of their performance measurement. While interviewees reported that these data were useful in measuring their performance in relation to crime-related outcomes as they are relatively accessible and perceived by some stakeholders to be easy to interpret, using these data was also reported to pose significant risks and challenges, described below. Seattle also reported using more innovative data sources such as location data from police vehicle location tracking systems and transcriptions from interactions between police and members of the public recorded on body-worn cameras.

Survey data were also used in many jurisdictions to support police performance measurement.

Survey participants include police officers, for example in England and Wales, where they may be invited by inspectors from HMICFRS to share their views on how the organisation looks after them and how the force is organised and managed. Members of the public may also input into the measurement of police performance, typically through participating in surveys about their own experiences with the police and of crime victimisation. Data from crime victimisation surveys can be particularly valuable in providing a more accurate picture in actual crime rates, given widespread concerns about the accuracy and completeness of police recorded crime data. Interviewees from Seattle and England and Wales also highlighted community consultation processes such as advisory councils, citizen forums and extensive qualitative feedback collection from victims about their experiences with the police.

There are challenges associated with using any type of data in measuring police performance, but interviewees identified police recorded crime data as carrying the most risk in terms of validity, reliability and representativeness.

Across all case study jurisdictions, interviewees expressed concerns about the quality of crime data from police management systems. The drivers of poor-quality data were also relatively consistent across jurisdictions and tended to spring from the reliance on police staff to manually input crime data. They included:

- Complex data inputting processes increasing human error.
- Information systems with limited functionality and linkage with related data sources.
- Lack of resources to ensure that data inputters are appropriately trained and have sufficient time to enter correct and complete information.
- Under-reporting by members of the public about their crime victimisation.
Using police recorded crime data was also reported in some jurisdictions to have generated perverse incentives for some officers to manipulate the data.

This manipulation can include under-recording of reported crime, excessive charging of suspects to improve police clearance rates, conducting sobriety tests when the rates of positive results are likely to be lower, or focusing resources on pursuing offenders with higher numbers of charges against them. As some interviewees noted, such activities can be damaging to public confidence and trust in the police, and thus impact their capacity to promote public safety effectively. Preventing recorded crime data manipulation, particularly in areas where quantitative crime targets are used to measure police performance, requires effective organisational leadership and a culture of transparency and accountability.

The relationship between policing input and outputs and their outcomes may be difficult to establish.

Interviewees from NRW highlighted that many desired outcomes, particularly those reflected in police recorded crime data such as reduced crime rates, are affected by exogenous factors over which the police have no control, such as the influence of weather conditions on traffic accidents, or the COVID-19 pandemic on crime incidents. In the Netherlands, interviewees reported that while inputs and outputs can be easy to measure, stakeholders find it a challenge to meaningfully interpret outcomes. In England and Wales, stakeholders from HMICFRS and some forces reported using these kinds of quantitative data to provide a starting point for their performance analysis, with qualitative insights then sought to properly contextualise and give greater depth of meaning to these data.
<table>
<thead>
<tr>
<th>Type of indicator</th>
<th>Examples of indicators</th>
<th>Case study jurisdictions</th>
<th>Typical data source</th>
<th>Reported challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Law enforcement productivity</td>
<td>Recorded (and unreported) crime rates</td>
<td>The Netherlands, England and Wales, Israel, North Rhine-Westphalia, Seattle, Sweden</td>
<td>Police management systems; management systems of the public prosecutor; crime victimisation surveys</td>
<td>Variable data quality; perverse incentives; dark figure of crime; causality</td>
</tr>
<tr>
<td></td>
<td>Clearance rates</td>
<td>The Netherlands, England and Wales, North Rhine-Westphalia, Seattle, Sweden</td>
<td>Police management systems</td>
<td>Variable data quality; perverse incentives</td>
</tr>
<tr>
<td></td>
<td>The number of arrests</td>
<td>The Netherlands, England and Wales, Israel</td>
<td>Police management systems</td>
<td>Perverse incentives; causality</td>
</tr>
<tr>
<td></td>
<td>Response times</td>
<td>England and Wales, North Rhine-Westphalia</td>
<td>Police management systems</td>
<td>Variable data quality</td>
</tr>
<tr>
<td></td>
<td>The value of confiscated assets</td>
<td>The Netherlands</td>
<td>Police management systems</td>
<td>Variable data quality</td>
</tr>
<tr>
<td></td>
<td>Time spent on investigations (from report to closure)</td>
<td>Sweden</td>
<td>Police management systems</td>
<td>Perverse incentives</td>
</tr>
<tr>
<td></td>
<td>The number of open cases</td>
<td>Sweden</td>
<td>Police management systems</td>
<td>Perverse incentives</td>
</tr>
<tr>
<td>Safety</td>
<td>The number of traffic accidents</td>
<td>North Rhine-Westphalia, Sweden, Israel</td>
<td>Police management systems</td>
<td>Variable data quality; causality</td>
</tr>
<tr>
<td>Type of indicator</td>
<td>Examples of indicators</td>
<td>Case study jurisdictions</td>
<td>Typical data source</td>
<td>Reported challenges</td>
</tr>
<tr>
<td>-------------------</td>
<td>------------------------</td>
<td>--------------------------</td>
<td>---------------------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Citizen trust and satisfaction</td>
<td>The number of people, including crime victims, who indicate that they are satisfied with police services</td>
<td>England and Wales, Seattle, Sweden</td>
<td>Surveys; interviews; public forums</td>
<td>Ensuring a representative sample</td>
</tr>
<tr>
<td></td>
<td>The number of people who indicate that they have trust in police services</td>
<td>England and Wales, Sweden</td>
<td>Surveys; interviews; public forums</td>
<td>Ensuring a representative sample</td>
</tr>
<tr>
<td></td>
<td>The number of police complaints</td>
<td>England and Wales</td>
<td>Police management systems</td>
<td>Under-reporting</td>
</tr>
<tr>
<td>Equity and procedural justice</td>
<td>Perceived disproportionality in policing checks</td>
<td>England and Wales, The Netherlands</td>
<td>Survey data</td>
<td>Ensuring a representative sample; consensus on a definition of ‘reasonable suspicion’</td>
</tr>
<tr>
<td></td>
<td>Racial disparities in post-check outcomes</td>
<td>Seattle</td>
<td>Police management systems</td>
<td>Consensus on a definition of ‘reasonable suspicion’</td>
</tr>
</tbody>
</table>

549 The pilot programme described in Section 3.2 sought to explore whether this indicator could be included in future performance measurement in Rotterdam, Utrecht and The Hague.
International approaches to police performance measurement

<table>
<thead>
<tr>
<th>Type of indicator</th>
<th>Examples of indicators</th>
<th>Case study jurisdictions</th>
<th>Typical data source</th>
<th>Reported challenges</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workplace culture and organisation</td>
<td>The extent of ‘over’ and ‘under’ policing</td>
<td>Seattle</td>
<td>Police management systems; police vehicle location tracking systems</td>
<td>Variable data quality</td>
</tr>
<tr>
<td></td>
<td>How the force looks after its workforce</td>
<td>England and Wales</td>
<td>Surveys; interviews; police management systems</td>
<td>Ensuring a representative sample; opaque assessment methodologies</td>
</tr>
<tr>
<td></td>
<td>How the force plans and manages organisationally</td>
<td>England and Wales</td>
<td>Surveys; interviews; police management systems</td>
<td>Ensuring a representative sample; opaque assessment methodologies</td>
</tr>
</tbody>
</table>

Source: Data collected over the course of this study.
9.3. Innovative and promising practices in police performance measurement

This study has identified several innovative and promising practices in measuring police performance, although there is limited evidence to assess their effectiveness.

In England and Wales, interviewees positively highlighted the change in performance assessment by the HMICFRS from an annual inspection to a process of continuous engagement, creating a feedback loop that was reported to enhance the inspectorate’s understanding of the force and its performance. Interviewees also commented on the value of a national framework for performance measurement that facilitated meaningful comparisons between individual forces’ performance, while ensuring sufficient flexibility to capture and reflect local variations in policing priorities, resources and community expectations.

Performance measurement also has a role in supporting learning and improvement within forces. In most jurisdictions, assessments are not used, at least solely, to punish forces for poor performance, but rather to initiate reflection and action by the force on addressing performance issues and the provision of support from local and national stakeholders in the delivery of improvements. In Seattle, new measures to assess police performance have been formally incorporated into the force’s professional learning and development structures, demonstrating one potential approach to the integration of the framework’s strategic aims at all levels of and in all processes within the force. Interviewees from Seattle also highlighted the use of more innovative data sources to inform performance measurement, including data from body-worn cameras.

Findings from these jurisdictions indicate that at least some stakeholders are incorporating more evidence-based approaches to performance measurement. In Sweden and England and Wales, for example, the crime harm index is used to understand how the police are performing in relation to the types of crimes that create the most harm for individuals and communities, rather than looking strictly at crime volume. Sweden also reported the use of the complexity index, a simple framework for police performance measurement.
This chapter draws on the findings from the study to recommend several starting points for improving performance measurement in the Netherlands. The aim is to show that different practices may provide valuable insights in certain contexts. However, which insights will be applicable in the future depends on the choices made by key stakeholders in the Netherlands, such as the Ministry of Justice and Security and the Dutch police force. The study team proposes that Dutch police and policymakers should first consider several fundamental questions about the kind of performance measurement framework they want to implement.

10.1. Key decisions that determine how performance is measured

The findings from the case studies demonstrate that the development of a framework for performance measurement first and foremost requires key stakeholders, such as police forces and policymakers, to make decisions on several fronts before new practices can be put in place in the Netherlands. In particular, there are two key areas on which agreement should first be reached before a new framework for police performance measurement can be developed: 1) careful consideration needs to be given to the purpose for measuring the performance of the police; and 2) consensus needs to be reached on the level (national, regional and/or local) at which parameters for performance should be set and by whom before any of the practices discussed in the case study chapters can be meaningfully adopted in the Netherlands. Practices are not always transferable due to how performance measurement frameworks have been designed in other jurisdictions, which may differ from the Dutch approach.

10.1.1. What is the purpose of measuring the performance of the police?

Before the parameters by which performance will be measured are set out, it is crucial that stakeholders agree on why performance should be measured. An organisation’s objectives, and the associated indicators and targets, will be determined by whether it is seeking to improve the quality of its services, provide strategic direction for its work, show accountability to the public, or demonstrate a return on investment (or a combination of several or all of these factors). However, the study team found that there is currently no agreement among Dutch stakeholders on the purpose of performance measurement.

The study findings suggest that there is more clarity on this in some of the case study countries. For example, the Seattle Police Department has chosen to make accountability and the improvement of the quality of services a central pillar of its work. As such, the inclusion of equity indicators provides a meaningful addition to its performance measurement framework. The police department also uses the automated transcription
of recordings from body-worn surveillance cameras to gain a better understanding of the quality of interactions between the police and the public (see Chapter 7).

Other jurisdictions that have likewise aligned their performance measurement frameworks with a desire to improve service quality and demonstrate accountability may use surveys to capture the level of citizen confidence in and satisfaction with the police. This is the case, for example, for forces in England and Wales (see Chapter 4) and Sweden (see Chapter 8).

However, the inclusion of equity indicators and targets related to citizen trust and confidence may not be appropriate for all units, as the extent to which service improvement and accountability are seen as the primary objectives of performance measurement appears to vary significantly across the country. Whether these instruments can provide useful tools for performance measurement in the Netherlands will depend on whether the Dutch police consider service improvement and accountability to the public to be a key purpose of performance measurement. The findings from this study suggest that this is not currently the case across the different national, regional and local police units in the Netherlands.

10.1.2. Who should set performance objectives, indicators, measures and targets, and at what level (national, regional and/or local)?

Tied to these decisions about why performance should be measured are decisions around who should be involved in setting performance objectives, indicators, measures and/or targets, and whether this should be done at a national, regional or local level.

As described in Chapter 3, in the Netherlands the strategic direction from the Ministry of Justice and Security, laid out in the Security Agenda, is supplemented by both regional and local agreements on what should be measured and against which targets. The study found that key national stakeholders in the Netherlands consequently do not have a clear picture of the performance indicators and targets currently used across the country, as the design of these metrics is partly left to regional and local stakeholders.

Other jurisdictions have opted for different approaches when deciding at what level the parameters for performance should be set and by whom. For example, as Chapter 5 highlighted, a more local approach was introduced in Israel under EMUN reforms to address local needs and encourage trust in the police. Under this framework, each police unit designed an annual workplan to determine its strategic priorities. As part of these workplans, units were asked to choose four measures, referred to by interviewees as ‘problems’. One measure was set nationally (the number of traffic accidents), while the other three differed per unit and were selected to reflect the local context. In contrast, in Sweden most of the strategic priorities for the national police authority are set by the Ministry of Justice and apply to all police units (see Chapter 8).

These choices have also influenced the instruments for measuring performance used in these countries. For example, the decision to employ live operational dashboards at a local level in Israel allowed each unit to develop an understanding of its performance in relation to the measures it chose to track. However, live operational dashboards for local units with individual targets would be a less obvious choice in Sweden because of the desire to consolidate police performance indicators and targets at a national level. If police objectives, indicators and targets continue to be set at both national and regional levels in the Netherlands, any decision to introduce similar instruments would have to take this into account.
10.2. Concluding insights for improving police performance measurement in the Netherlands

Decisions on the two areas for agreement, the purpose of performance measurement and the level(s) at which the framework should be set and by whom, will give greater coherence to the Dutch strategy for police performance measurement and will determine what practices for measuring performance from the case study jurisdictions might be transferable to the Netherlands.

Developing a more coherent strategy for police performance may also increase officer engagement with police performance measurement, as they would have a clearer sense of how the process may bring about improvements across the force. It should be noted that consistent reporting on the performance indicators and targets considered, the data collection methods used, and the lessons provided by the results could provide a further positive development in this regard. Greater officer buy-in is also likely to benefit the quality of performance data by encouraging officers to consider how the data they record using police registration or management systems will be used to analyse how the force is performing against certain indicators.

In conclusion, this study offers several concrete leads for improving performance measurement in the Netherlands. These include the practices adopted by other jurisdictions, but above all the need to consider several more fundamental questions about the type of performance measurement framework that the Dutch police and policymakers would like to see.
References

Aiello, T. 2022. ‘Mamaroneck police using QR codes to survey public about interactions with officers.’ CBS New York, 16 September. As of 18 October 2023:

Akkermans, M., R. Kloosterman, E. Moons, C. Reep and M. Tummers-van der Aa. 2022. ‘Veiligheidsmonitor 2021.’ As of 18 October 2023:
https://www.cbs.nl/nl-nl/longread/rapportages/2022/veiligheidsmonitor-2021

Algemene Rekenkamer. 2016. Landelijke prioriteiten politie. Rapport behorend bij verantwoordingsonderzoek naar begrotingshoofdstuk VI. As of 17 October 2023:
https://www.rekenkamer.nl/publicaties/rapporten/2016/05/18/landelijke-prioriteiten-politie


Ali, N. and D. Rosenberg. 2022. ‘Non-police Intervention Consideration by Ethnic Minorities: The Case of Arab Citizens of Israel.’ Victims & Offenders 6. As of 19 October:
https://doi.org/10.1080/15564886.2022.2100541

American Police Officers Alliance. 2019. ‘What is Police Oversight?’ 9 May. As of 18 October 2023:
https://americanpoliceofficersalliance.com/what-is-police-oversight/

https://doi.org/10.1093/police/paac025

Australia New Zealand Policing Advisory Agency. 2023. ‘About.’ As of 18 October 2023:
https://www.anzpaa.org.au/about


Australian Bureau of Statistics. 2021. General Social Survey: Summary Results, Australia. As of 18 October 2023:

Australian Bureau of Statistics. 2023. ‘Crime and Justice.’ As of 18 October 2023:
International approaches to police performance measurement


https://www.bbc.co.uk/news/world-us-canada-52921418

https://assets.gov.ie/94892/8a69a765-8e39-4b01-88d6-363c8d70d.pdf


Brottsförebyggande rådet (Brå). 2022. *En modell för uppföljning av rättskedjans myndigheter.* As of 19 October 2023:
https://bra.se/download/18.8ee8b9d18418cd0c7f4651/1657229645420/2022%20modell%20for%20uppfollning%20av%20rattskedjans%20myndigheter%20modell%20.pdf

Brottsförebyggande rådet (Brå). 2023a. ‘Forskning och Analys: Polisen.’ As of 19 October 2023:
https://bra.se/forskning-och-analys/polisen.html

Brottsförebyggande rådet (Brå). 2023b. ‘Om Brottsförebyggande rådet.’ As of 19 October 2023:
https://bra.se/om-bra.html

Brottsförebyggande rådet. (Brå). 2023c. ‘The Swedish Crime Survey.’ As of 19 October 2023:

Brown, J. 2021. *Policing in the UK.* House of Commons Library. As of 19 October 2023:
https://commonslibrary.parliament.uk/research-briefings/cbp-8582/

Bureau of Crime Statistics and Research. 2023. ‘About us.’ As of 18 October 2023:


Cision PR Newswire. 2021. 'Truleo Launches First Body Camera Audio Analytics Platform for Law Enforcement.' 6 October. As of 19 October 2023:


City of Seattle. 2023a. 'Department Fact Sheet.' As of 19 October 2023:

https://www.seattle.gov/police/about-us/about-the-department/department-fact-sheet

City of Seattle. 2023b. 'SeaStat.' As of 19 October 2023:

https://www.seattle.gov/police/information-and-data/data/seastat


Cornell Law School. 2022. ‘Consent decree.’ Legal Information Institute. As of 18 October 2023:

https://www.law.cornell.edu/wex/consent_decree

Cruise, C. and J. Moore. 2022. 'Virginia town rolls out Uber-style ratings for police officers.' *WTOP News,* 21 April. As of 18 October 2023:


https://www.rand.org/pubs/technical_reports/TR1153.html
International approaches to police performance measurement


Guardian Score. 2023b. ‘Homepage.’ As of 18 October 2023: https://guardianscores.com/


Home Office. 2010. ‘Police Reform: Theresa May’s speech to the National Policing Conference.’ As of 19 October 2023:

Home Office. 2012. ‘Definition of policing by consent.’ As of 19 October 2023:


Home Office. 2022. ‘Home Office Annual Data Requirement (ADR) data.’ As of 19 October 2023:


https://www.inspectie-jenv.nl/Publicaties/rapporten/2019/12/16/periodiek-beeld-opsporing

https://www.inspectie-jenv.nl/Publicaties/rapporten/2020/09/01/periodiek-beeld-handhaving

Inspectorate of Justice and Security. 2021a. *Noodhulp. Incidentafhandeling door de politie*. As of 17 October 2023:
https://www.inspectie-jenv.nl/Publicaties/rapporten/2021/09/09/rapport-noodhulp

https://www.inspectie-jenv.nl/Publicaties/rapporten/2021/01/28/rapport-onderzoek-le-deelrapport-dlio

https://www.inspectie-jenv.nl/Publicaties/rapporten/2022/01/25/rapport-onderzoek-naar-de-taakuitoering-bij-de-landelijke-eenheid

https://www.inspectie-jenv.nl/Publicaties/plannen-van-aanpak/2022/03/23/plan-van-aanpak-kwaliteitszorg-politie

Inspectorate of Justice and Security. 2023a. ‘Over de Inspectie JenV.’ As of 17 October 2023:
https://www.inspectie-jenv.nl/organisatie

Inspectorate of Justice and Security. 2023b. ‘Toezichtgebieden: Politie.’ As of 17 October 2023:
https://www.inspectie-jenv.nl/toezichtgebieden/politie
Ivarsson Westerberg, A. 2020. ‘Reform i uniform – därför gick polisens stora omorganisation fel.’ Forskning.se, 30 November. As of 19 October 2023:


International approaches to police performance measurement


New South Wales Police Force. 2023. ‘About us.’ As of 18 October 2023:

https://www.justice.govt.nz/assets/Cycle-4-Core-Report-v0.20-20220628.pdf


New Zealand Police. 2023a. ‘About EBPC.’ As of 18 October 2023:

New Zealand Police. 2023b. ‘Find Police stations by A–Z.’ As of 16 October 2023:
https://www.police.govt.nz/contact-us/stations/a2z?nondesktop

New Zealand Police. 2023c. ‘Police districts.’ As of 18 October 2023:

https://doi.org/10.1093/police/pan049

Northern Territory Government. 2022. *Budget 2022–23*. As of 18 October 2023:

New York Police Department. 2023. ‘CompStat 2.0.’ As of 18 October 2023:
https://www1.nyc.gov/site/nypd/stats/crime-statistics/compstat.page


OSCE POLIS. 2023a. ‘Country Profile: Canada – Municipal Level.’ As of 17 October 2023:
https://polis.osce.org/country-profiles/canada#municipal-level-10014

OSCE POLIS. 2023b. ‘Country Profile: Sweden.’ As of 17 October 2023:
https://polis.osce.org/country-profiles/sweden

Parsons, J., C. Gokey and M. Thornton. 2013. *Indicators of Inputs, Activities, Outputs, Outcomes and Impacts in Security and Justice Programming*. UK Department for International Development. As of 16 October 2023:

International approaches to police performance measurement

Phillips, T. 2022. ‘US police officers pilot contactless badges that citizens can tap with their NFC phone.’ NFCW, 16 June. As of 18 October 2023:


Police of Finland. 2023a. Poliisihallituksen tilinpäätös ja toimintakertomus vuodelta 2022. As of 18 October 2023:
https://poliisi.fi/documents/25235045/27075255/Poliisihallituksen+tilinp%C3%A4%C3%A4%C3%A4%C3%B6s+ja+toimintakertomus+2022.pdf/74e0f645-8ff6-e735-51c5-0887f15f950d/Poliisihallituksen+tilinp%C3%A4%C3%A4%C3%A4%C3%B6s+ja+toimintakertomus+2022.pdf?t=1678786478419

Police of Finland. 2023b. Twitter post: ‘Näillä kahdella indeksillä voidaan seurata poliisin valvontatoiminnan vaikuttavuutta.’ As of 18 October 2023:
https://twitter.com/SuomenPoliisi/status/1623667767051186176?s=20

Police of Finland. 2023c. Twitter post: ‘Poliisin 90-luvulta asti tulosmittarina käytämä katuturvallisuusindeksi on huonontunut, mutta Poliisihallitus haluaa täsmentää, että konkreettinen katujen ja muiden yleisten paikkojen turvallisuus ei kuitenkaan ole heikentynyt Suomessa. Tämä johtuu siitä, että katuturvallisuusindeksi ei kovin hyvin mittaa katujen ja muiden yleisten paikkojen turvallisuutta, koska indeksiin sisältyvät myös yksityisillä paikoilla kuten kodeissa tapahtuneet rikokset.’ As of 18 October 2023:
https://twitter.com/SuomenPoliisi/status/1623667760013123584?s=20

Police of Finland. 2023d. Twitter post: ‘Tästä syystä poliisihallitoksessa on kehitetty uusi indeksi – häiriöindeksi. Se kuvaa paremmin katujen turvallisuutta ja tulee korvaamaan vanhan katuturvallisuusindeksin. Häiriöindeksin mukaan katujen turvallisuus on pääin astuvan kehittynyt parempaan suuntaan.’ As of 18 October 2023:
https://twitter.com/SuomenPoliisi/status/1623667763297255424?s=20

Police University College. 2023a. ‘Police Barometer.’ As of 18 October 2023:
https://polamk.fi/en/police-barometer/

Police University College. 2023b. ‘Research, development and innovation activities (RDI) in brief.’ As of 18 October 2023:
https://polamk.fi/en/rdi-in-brief

Polisen. 2021. 'Tillsynsenheten.' As of 19 October 2023:
https://polisen.se/om-polisen/organisation/tillsynsfunktionen/

Polisen. 2022. ‘Polisens organisation 123ret tt haveri utan dess like’. Expressen, 19 May. As of 19 October 2023:
https://www.expressen.se/debatt/polisens-organisation-ar-ett-haveri-utan-dess-like/

Polisen. 2023. ‘The Swedish Police Authority.’ As of 19 October 2023:

Politie Eenheid Rotterdam. 2022. Jaarverslag 2022. As of 18 October 2023:
https://www.regioburgemeesters.nl/save823/

Politie Nederland. 2017. Aanvaksprogramma Informatievoorziening Politie: Eindrapport. As of 18 October 2023:
https://zoek.officielebekendmakingen.nl/blg-828214.pdf


Polizei NRW. 2023a. ‘Polizeiliche Kriminalstatistik.’ As of 19 October 2023: https://lka.polizei.nrw/artikel/polizeiliche-kriminalstatistik


Polizei NRW. 2023c. ‘Die Zentralstelle Evaluation (ZEVA).’ As of 19 October 2023: https://lka.polizei.nrw/zeva


Rijksoverheid. 2022. *Veiligheidsagenda 2023-2026.* As of 18 October 2023: https://open.overheid.nl/documenten/ronl-fefff06abdd2c99d8ad9be531c68a1df6deae72d/pdf
International approaches to police performance measurement


Stals, L. And S. Van der Walle. 2020. *Het monitoren van publiek vertrouwen in de politie.* As of 18 October 2023: https://repository.wodc.nl/handle/20.500.12832/3006


University of Cambridge. 2023. ‘Cambridge Crime Harm Index (CCHI).’ Institute of Criminology. As of 19 October 2023: https://www.crim.cam.ac.uk/research/thecambridgecrimeharmindex


International approaches to police performance measurement


Veiligheidsmonitor. 2022. ‘Databank.’ As of 18 October 2023: https://veiligheidsmonitor.databank.nl/viewer


Zouridis, S. 2016. ‘Doe de goede dingen, maar doe ze wel steeds beter.’ *Tijdschrift voor de politie* jg.77/nr.7/15.

As of 18 October 2023:
https://www.politieacademie.nl/kennisonderzoek/kennis/mediatheek/pdf/91449.pdf

https://repository.wodc.nl/bitstream/handle/20.500.12832/2056/2341-volledige-tekst_tcm28-73068.pdf?sequence=2&isAllowed=y
Annex A. Non-case study countries

This annex sets out the findings from Phase 1 for those countries that, using the selection criteria set out in Chapter 2, were not chosen as case studies to be researched in greater depth in Phase 2. These were Australia, Canada, Finland, New Zealand and the United States (looked at as a whole). For each country, a brief introduction is followed by a high-level overview of current practices, of challenges and complexities in measuring police performance meaningfully, and of recent developments. It should be noted that these findings are exclusively based on the data collected during Phase 1 of the study. As such, they are not meant to be comprehensive, but rather seek to provide a snapshot of approaches to police performance measurement.
A.1. Australia

A.1.1. Background and context

There are two main law-enforcement components in Australia’s judicial system: the Australian Federal Police (AFP) and state- and territory-specific police forces. It should be noted that the AFP is responsible for policing in the Australian Capital Territory (ACT).

The Australian Federal Police was founded in response to a 1987 terrorist attack and is responsible for the national security of Australia. It employs over 7,000 people and its agents primarily focus on preventing, investigating and disrupting transnational, serious, complex and organised crime. Its mission is to ‘reduce criminal and security threats to Australia’s collective economic and societal interests through cooperative policing services’, and to provide ‘a safe and secure environment through policing activities on behalf of the Australian Capital Territory Government’. The AFP is accountable to the Attorney General.

State and territory police agencies are responsible for policing at state and local levels. Their responsibilities include maintaining social order; promoting safe road-user behaviour; preventing, detecting and investigating crime; responding to calls for assistance; and performing and coordinating emergency and rescue operations. Each police agency is made up of a number of police stations that operate locally and are organised into regions and/or districts; they are administered by the relevant state government. The police commissioner for each state is appointed by the relevant territory government and is accountable to the state parliament through an elected police minister.

Key stakeholders in police performance measurement in Australia are the Australian Productivity Commission, the Australian Bureau of Statistics, state crime statistics agencies and the Australia and New Zealand Policing Advisory Agency.

The Australian Productivity Commission reviews the performance of the state and territory police agencies, including the AFP’s work on behalf of the Australian Capital Territory Government, and publishes annual reports on the work of these agencies. Other agencies also contribute to police performance measurement in Australia. The Australian Bureau of Statistics, Australia’s national statistics agency, collects, develops and analyses data on a wide variety of topics, including crime and criminal justice.

---

553 INT3_AUS.
555 Victoria Police Force (2023a).
556 Western Australia Police Force (2022).
557 Victoria Police Force (2022).
558 Queensland Police Service (2022a).
560 Australian Government (2023a).
561 INT1_AUS.
International approaches to police performance measurement

also responsible for conducting the annual General Social Survey, which includes questions about community trust in the police.\textsuperscript{563} At the state and territory level, there are crime statistics and research agencies that contribute to police performance measurement. Examples are the \textbf{New South Wales Bureau of Crime Statistics and Research (BOCSAR)}, which has as one of its objectives to ‘identify factors that affect the effectiveness, efficiency or equity of the NSW criminal justice system’, and \textbf{Crime Statistics Agency Victoria}, which is responsible for ‘processing, analysing and publishing Victorian crime statistics’.\textsuperscript{564,565,566}

The \textbf{Australia and New Zealand Policing Advisory Agency (ANZPAA)} is also an important stakeholder in police performance measurement in Australia. Its members are the police commissioners from each jurisdiction, as well as ACT’s Chief Police Officer and the Commissioner of Police of the New Zealand Police.\textsuperscript{567} As an organisation, its purpose is to provide advice on ‘current and emerging policing priorities’, and to promote good practice, including in the area of police performance measurement.\textsuperscript{568} ANZPAA runs the annual National Survey of Community Satisfaction with Policing on behalf of all Australian state and territory police forces.\textsuperscript{569}

\textbf{A.1.2. Current practices in performance measurement}

\textit{Australian Federal Police}

\textbf{AFP’s framework for performance measurement is produced annually in accordance with national legislation.}

The Public Governance Performance and Accountability Act 2013 governs financial management, governance, performance and accountability for Commonwealth government departments, including the Federal Police.\textsuperscript{570} In line with this legislation, the AFP is required to produce three documents that relate to its performance: the Portfolio Budget Statement, the Corporate Plan and the Annual Performance Statement.\textsuperscript{571-572} The Portfolio Budget Statement asks the organisation to set performance criteria and targets.\textsuperscript{573} While the Attorney General provides ministerial direction to the AFP and sets its strategic priorities, it is the agency itself that is responsible for formulating the indicators by which it measures its performance.\textsuperscript{574} The Corporate Plan defines the activities that the AFP will undertake to meet these targets.

\begin{itemize}
\item \textsuperscript{563} Australian Bureau of Statistics (2021).
\item \textsuperscript{564} Bureau of Crime Statistics and Research (2023).
\item \textsuperscript{565} Victoria Police Force (2023b).
\item \textsuperscript{566} INT3_AUS.
\item \textsuperscript{567} Australia New Zealand Policing Advisory Agency (2023).
\item \textsuperscript{568} Australia New Zealand Policing Advisory Agency (2023).
\item \textsuperscript{569} INT3_AUS.
\item \textsuperscript{570} INT3_AUS.
\item \textsuperscript{571} INT3_AUS.
\item \textsuperscript{572} Australian Federal Police (2022).
\item \textsuperscript{573} INT3_AUS.
\item \textsuperscript{574} INT3_AUS.
\end{itemize}
and talks about the broader context in which the agency operates, while the Annual Performance Statement reports on whether the various targets have been met.\textsuperscript{575}

**Figure 7: Australian Federal Police: Performance criteria, targets and outcomes 2021-2022.**

The AFP analyses a combination of qualitative and quantitative data to assess how it is performing against targets outlined in its annual Portfolio Budget Statement. Qualitative and quantitative data are used to report on the performance criteria outlined in Figure 7. For example, in 2021 and 2022, a combination of data from surveys, statistical data and case studies was used to measure the agency's performance.\textsuperscript{576}

\textsuperscript{575} INT3_AUS.

\textsuperscript{576} Australian Federal Police (2022).
As highlighted in Figure 7, ‘high community confidence’ was set as one of the performance criteria in 2021–22. The relevant section in the annual report elaborates that this specifically relates to ‘Australian community confidence in the AFP’s contribution to law enforcement and national security’. To measure the AFP’s performance in this area, the agency commissions an annual telephone survey that uses a random sample (n=1,044 in 2021–22) of the Australian population stratified by age, gender and state/territory. Participants are asked to give a confidence rating of one to ten, with eight out of ten or higher being considered a ‘high confidence rating’.

The ‘return on investment’ of AFP operations is used to assess their value.

The performance criteria ‘return on investment’ (ROI) is used to assess the value of some AFP operations, including its transnational (meaning ‘domestic’ in this context) operations. Here, the ROI is calculated by means of a cost-benefit analysis that uses data from the AFP’s SAS Firefly software, a visual analytics tool, on three different measures: the cost of relevant operations, the drug harm index (DHI) and the estimated financial return (EFR) (Figure 8). The AFP’s target in this area is for the ROI to be greater than one.

Figure 8: Return on investment calculation: Transnational crime efforts of the AFP

The AFP also uses case studies to evaluate its activities in the areas of prevention, disruption and enforcement, as well as how officers respond to incidents. These provide detailed examples of the agency’s activities and highlight how the AFP delivers ‘public value’. For instance, the agency’s annual report for 2021–22 includes a case study that highlights Operation Phoenix and Operation Wilmot, which had the aim of ‘protecting Australian high office holders from harm during the 2022 federal election’. It details why these operations were established, what actions were undertaken and how they helped prevent incidents from occurring. One interviewee noted that this case study approach allows the AFP to assess its performance in a more nuanced manner than if only statistical and survey data were employed.
State and territory police

State and territory legislation governs how individual state and territory police forces measure their performance, but there appear to be similarities in terms of the indicators and data sources used.

State and territory police forces manage their performance separately from the AFP and are governed by state legislation, such as the Annual Reports (Departments) Regulation 2015 in New South Wales, the Financial and Performance Management Standard 2019 in Queensland and the Financial Management Act 1994 in Victoria.\(^{589}\) All state and territory forces report on their performance in an annual report.\(^{590}\) Indicators and targets may be defined within a budget statement at the beginning of the year, as in Victoria and the Northern Territory, in operational plans\(^{591}\) or annual reports.\(^{592}, 593, 594\)

State and territory forces generally use indicators that relate to the prevalence of crime, community safety and road safety to measure their performance.\(^{595}, 596, 597, 598\) These may include:

- The number of incidents reported per crime type\(^{599}, 600, 601, 602\)
- Community confidence in the police\(^{603}, 604, 605, 606\)
- Speeding infringement notices issued\(^{607}\)
- Average response times.\(^{608}\)

To assess how state and territory forces are performing against these indicators, quantitative data sources such as records from police systems and statistics agencies, and survey results are used.

To report on the above-mentioned indicators, state and territory police forces analyse a range of primarily quantitative data sets.\(^{609}\) These may include records from police dispatch systems and incident registration

\(^{589}\) INT1_AUS; INT2_AUS; INT3_AUS.
\(^{590}\) INT3_AUS; INT2_AUS.
\(^{591}\) Queensland Police Service (2022b).
\(^{592}\) Victoria Police Force (2022).
\(^{593}\) Northern Territory Government (2022).
\(^{594}\) South Australia Police Force (2022).
\(^{595}\) South Australia Police Force (2022).
\(^{596}\) New South Wales Police Force (2022).
\(^{597}\) Western Australia Police Force (2022).
\(^{598}\) Victoria Police Force (2022).
\(^{599}\) New South Wales Police Force (2022).
\(^{600}\) South Australia Police Force (2022).
\(^{601}\) Tasmania Police Force (2022).
\(^{602}\) Victoria Police Force (2022).
\(^{603}\) New South Wales Police Force (2022).
\(^{604}\) South Australia Police Force (2022).
\(^{605}\) Victoria Police Force (2022).
\(^{606}\) Queensland Police Service (2022b).
\(^{607}\) Tasmania Police Force (2022).
\(^{608}\) Tasmania Police Force (2022).
\(^{609}\) INT2_AUS.
International approaches to police performance measurement

systems, the Australian Bureau of Statistics, and state-specific crime statistics agencies, such as the New South Wales Bureau of Crime Statistics and Research (BOSCAR) and the Crime Statistics Agency Victoria.

ANZPAA runs the annual National Survey of Community Satisfaction with Policing on behalf of all Australian state and territory police forces. This survey collects data on levels of public satisfaction with police contact and policing outcomes. As the survey uses a relatively large sample (n=27,926 in 2021–22), it allows police stations to use the resulting data to assess their local situation. State and territory forces may also use additional surveys to gather data on how specific aspects of policing are perceived. For instance, the Western Australia Police Force commissioned an additional quantitative online survey to measure the effectiveness of a series of road safety awareness campaigns between 2017 and 2022.

The Australian Productivity Commission uses annual reports produced by state and territory forces to assess police performance.

The performance of state and territory police forces is also reported on at a federal level by the Australian Productivity Commission, which produces an annual report that includes data from all states and territories, partly taken from their individual annual reports and the National Survey of Community Satisfaction with Policing. The commission uses a performance indicator framework that includes indicators on equity, effectiveness and efficiency, separated into those that relate to outputs and those that relate to outcomes. It also reports separately on data for Aboriginal and Torres Strait Islander people in relation to the established indicators.

A.1.3. Challenges and complexities to measuring police performance

Inconsistencies in data from police incident registration systems present a challenge to measuring police performance in Australia.

Data from the interviews conducted for this study suggest that the quality of the administrative data captured in the registration systems of both the AFP and the state and territory police forces is variable as it is largely dependent on the entry created by responding officers. One interviewee noted that the priority of these officers is to produce reports that can be used at the operational level, not to create data sets that

610 Western Australia Police Force (2022).
612 South Australia Police Force (2022).
615 INT3_AUS.
616 Western Australia Police Force (2022).
617 INT3_AUS.
618 Western Australia Police Force (2022).
620 Australian Government (2023b).
621 Australian Government (2023b).
622 INT2_AUS; INT3_AUS.
may be used to produce insights into how the force in question is performing.\textsuperscript{623} They highlighted that data are therefore not always recorded in a consistent manner that allows data analysts to produce official statistics.\textsuperscript{624} A second interviewee noted that this issue has been addressed to some extent by the introduction of the National Standard Crime Recording Standard (NCRS) in 2009, which has sought to bring more uniformity to police recording processes.\textsuperscript{625,626} However, according to the first interviewee, inconsistencies in initial police records continue to pose a problem.\textsuperscript{627}

A.1.4. Recent developments

The AFP is developing a community survey to better capture levels of satisfaction with and confidence in the federal police.

The AFP is in the process of improving its ability to measure community satisfaction with and confidence in the federal police so that this factor may be integrated more fully into organisational performance measurement. In 2021, the AFP ran its community survey twice a year, rather than just once, as it had carried out a major operation and wanted to see whether it had resulted in a change in public perceptions of the force.\textsuperscript{628} The AFP is currently in the process of developing measures that consider community confidence in the force not only at an aggregate level, but also in specific contexts, for instance regarding cybercrime and counterterrorism. One interviewee noted that in this way the force hopes to be able to construct a fuller picture of how it is performing against its strategic priorities.\textsuperscript{629}

\textsuperscript{623} INT3_AUS.
\textsuperscript{624} INT3_AUS.
\textsuperscript{625} INT2_AUS.
\textsuperscript{626} Australian Bureau of Statistics (2013).
\textsuperscript{627} INT3_AUS.
\textsuperscript{628} INT3_AUS.
\textsuperscript{629} INT3_AUS.
A.2. Canada

A.2.1. Background and context

The police system in Canada is generally distinguished by three levels of policing service overseen by two bodies.

The Royal Canadian Mounted Police (RCMP) is Canada’s national police force, alongside the Canadian Forces Military Police (CFMP). The RCMP is rather unique as it combines international, federal, provincial and municipal policing.630 It is also the largest force in the country, and enforces federal law, investigates (financial and organised) crime, and protects national security.631

Under the Canadian Constitution, provinces are usually responsible for public policing. In practice, however, only Ontario, Quebec, Newfoundland and Labrador maintain their own police force. The RCMP provides policing services to all other provinces.632

The provinces further delegate the responsibility of local police services to municipalities through police acts. These municipal police forces are usually governed directly by municipal councils or communities.633 Larger cities commonly maintain their own force, while smaller cities may contract police services from the RCMP, which is the case in 197 municipalities.634 Municipal-level police are responsible for tackling overall criminality within their jurisdiction.

The RCMP is overseen by two bodies: 1) the Royal Canadian Mounted Police External Review Committee, which promotes fair and equitable labour relations and conducts independent reviews of appeals; and 2) the Civilian Review and Complaints Commission (previously Commission for Public Complaints Against the RCMP), which receives complaints from the public about RCMP members and investigates them objectively. The Civilian Review and Complaints Commission also conducts public hearings and research.635

A.2.2. Current practices in performance measurement

Canada currently does not have a streamlined approach to measuring police performance.

Literature and interviewees highlight the lack of a national police performance measurement system. The literature suggests that due to the lack of a streamlined approach, police performance is often still measured by local police organisations,636 which also select performance indicators themselves.637 Furthermore, there are no legislative requirements in Canada that dictate the use of specific measures or indicators for measuring

630 OSCE POLIS (2023a).
631 Shearing, Stenning and de Bruin (2013).
632 Canada has 10 provinces: Alberta, British Columbia, Manitoba, New Brunswick, Newfoundland and Labrador, Nova Scotia, Ontario, Prince Edward Island, Quebec, and Saskatchewan.
633 Shearing, Stenning and de Bruin (2013).
634 OSCE POLIS (2023a).
635 OSCE POLIS (2023a).
636 Hodgkinson, Caputo and McIntyre (2019).
637 Mazowita and Rotenberg (2019).
police performance, which demonstrates the lack of a national framework. There are some regional guidelines. In Ontario, for instance, police must report on specific indicators of performance under municipal reporting regulations. Nonetheless, according to Public Safety Canada, there are no evidence-based performance standards that Canada can build on.638

The most common way of measuring police performance is by analysing quantitative indicators and public opinion.

One interviewee noted that currently, the most common way of assessing performance is by using statistical counts.639 In 2019, it was reported that police mostly relied on crime rates and the crime severity index, clearance rates, police strength (number of officers per 100,000 population), and expenditure.640 Canadian policing services therefore often rely on traditional metrics in the absence of national indicators, or a framework.641 Statistics Canada conducts a national survey on police performance every five years. According to Maslov, this survey is the only source of performance metric that allows for comparison in Canada.642

Efforts have been made to streamline the measurement of police performance across the country, resulting in a new performance measurement framework.

In recent years, attention has been paid to streamlining performance measurement in the country. The most promising attempt, according to literature and interviewees, is the Canadian Police Performance Metrics Framework, which was introduced as a response to the challenges and criticism that existing indicators do not reflect the complexity of contemporary policing. This framework, approved in 2017, is broadly organised around four pillars:

1. **Crime and victimisation.** This pillar comprises the incidences of crime, victimisation and offenders, and includes traditional performance metrics (Table 20). Measures used include annual data statistics and victimisation rates collected by the General Social Survey (GSS) conducted every five years.643

---

638 Public Safety Canada (2022).
639 INT2_CAN.
640 Mazowita and Rotenberg (2019).
641 Mazowita and Rotenberg (2019).
642 Maslov (2016).
643 Mazowita and Rotenberg (2019).
International approaches to police performance measurement

Table 20: Canadian Police Performance Metrics Framework: Crime and victimisation metrics

<table>
<thead>
<tr>
<th>Crime</th>
<th>Victimisation</th>
<th>Offenders</th>
<th>Re-contact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crime rates</td>
<td>Reporting rates</td>
<td>Clearance rates</td>
<td>Repeat victimisations</td>
</tr>
<tr>
<td>Youth crime rates</td>
<td>Perceptions of safety</td>
<td>Court outcomes</td>
<td>Chronic offending rates</td>
</tr>
<tr>
<td>Crime severity</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Traffic violations</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Source: Mazowita and Rotenberg (2019).

2. **Police activities and deployment.** This pillar covers measures of policing activities, including calls for service, activity reporting and targeted policing (Table 21). A Uniform Calls for Service system is under development, but there are no other measures to effectively measure this pillar of policing.644

Table 21: Canadian Police Performance Metrics Framework: Police activities and deployment metrics

<table>
<thead>
<tr>
<th>Calls for service</th>
<th>Activity reporting</th>
<th>Targeted policing initiatives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calls for service by population, or source</td>
<td>Time allocation of officers logged into the activity reporting system</td>
<td>Existence of such initiatives</td>
</tr>
<tr>
<td>Proportion of dispatched calls for service (plus those not resulting in a founded criminal incident)</td>
<td>N/A</td>
<td>The amount of human resources allocated to such initiatives</td>
</tr>
<tr>
<td>Number of units or officers allocated by call type</td>
<td>N/A</td>
<td>The amount of financial resources allocated to such initiatives</td>
</tr>
<tr>
<td>Total service time by call type and priority and median response time</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Re-contact in criminal and non-criminal occurrences</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

---

644 Mazowita and Rotenberg (2019).
### Table 22: Canadian Police Performance Metrics Framework: Police resources metrics

<table>
<thead>
<tr>
<th>Police personnel</th>
<th>Workforce health and motivation</th>
<th>Training and development</th>
<th>Operating expenditure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Police officer strength per 100,000 population</td>
<td>Number of departures of officers for reasons other than retirement</td>
<td>Number of training hours per officer</td>
<td>Expenditure relative to measures of policing outcomes</td>
</tr>
<tr>
<td>Proportion of civilian staff, women and minorities</td>
<td>Number of departures of officers with less than 10 years of experience</td>
<td>Expenditure on training as a proportion of overall expenditure</td>
<td>Overtime pay as a proportion of overall salary expenditure</td>
</tr>
<tr>
<td>Calls for service requiring a police presence, per officer</td>
<td>Proportion of officers eligible to retire</td>
<td>Investment in technological resources</td>
<td>Average salaries between comparable sized police services</td>
</tr>
<tr>
<td>Reported crimes per officer</td>
<td>Number of personnel on long-term leave</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Source: Mazowita and Rotenberg (2019).
International approaches to police performance measurement

Table 23: Canadian Police Performance Metrics Framework: Trust and confidence metrics

<table>
<thead>
<tr>
<th>Trust and confidence in the police</th>
<th>Public engagement</th>
<th>Use of force and authority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceptions of ensuring safety of citizens</td>
<td>Police website total page views</td>
<td>Person stops, street checks and searches</td>
</tr>
<tr>
<td>Perceptions of enforcing the law</td>
<td>Police Facebook page total page views</td>
<td>Use of physical force or weapons</td>
</tr>
<tr>
<td>Perceptions of promptly responding to calls</td>
<td>Police Twitter (now X) page total followers</td>
<td>Number of reviews of police misconduct</td>
</tr>
<tr>
<td>Perceptions of providing information on crime prevention</td>
<td>N/A</td>
<td>Number of citizen complaints (substantiated vs. unsubstantiated)</td>
</tr>
<tr>
<td>Perceptions of equitable treatment</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Perceptions of approachability</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Source: Mazowita and Rotenberg (2019).

Statistics Canada notes that each dimension under the pillars should be interpreted together. One interviewee mentioned, however, that the framework has not yet been implemented universally, despite the promising outlook of the programme. The reasons for this are unclear, although the interviewee mentions that funding could play a role. Currently, as aforementioned, local police services choose their own performance indicators and can subsequently ‘pick and choose’ from the suggested indicators, if desired.

A.2.3. Challenges and complexities to measuring police performance

Literature and interviews suggest that police performance measurements are not streamlined nor universally implemented, and the tools to do so are lacking.

The literature shows that there is currently no nationally implemented performance measurement system to measure the performance of the Canadian police, which leads to disparities between local regions. Mazowita and Rotenberg note that because local police agencies select performance indicators themselves, there is a tendency to measure ‘everything and anything’.649

The traditional metrics currently used do not reflect performance outcomes nor the complexities of contemporary Canadian policing.

The literature suggests that within the Canadian system there has been little attention to developing measurement frameworks that emphasise longer term police performance outcomes in society and/or better

---

647 Mazowita and Rotenberg (2019).
648 INT1_CAN.
649 Mazowita and Rotenberg (2019).
quality of policing services to the public. 650 This statement aligns with the views of one interviewee, who mentioned that there is little attention for improving quality measurement efforts in Canada. 651

The literature also suggests that besides the traditional use of metrics, few metrics or measures are provided that allow for the aggregation or analysis of types of crime, or the context or complexity of crime. For instance, not all assaults, despite being classified as an assault, require the same police response and case investigation. Traditional metrics such as clearance rates do not necessarily indicate the quality of police performance. 652

A third identified challenge in the interviews is that the data are currently considered insufficient for the purposes of measuring performance.

One interviewee explained that although the police do record data based on their cases, it is mainly for case management purposes, with the primary aim of entering data to close cases or submit them to the court, which can encourage perverse incentives. In particular, it was pointed out that when officials did not have to file a report to close a case if they listed the incident as ‘unfounded’, an increase in unfounded cases was seen. 653 A second interviewee reported a similar occurrence: police officers started registering the police station as the crime scene to speed up data entry, which led to the police station being identified as a ‘crime hotspot’ when evaluating the data. 654

A.2.4. Recent developments

Despite the challenges, the use of police performance measurement in Canada has been slowly developing, and some local initiatives seem promising.

Some local regions are increasingly implementing performance measurement frameworks, with some already making good progress in developing appropriate measurement standards. 655

One interviewee also noted that certain regions are developing police performance measurement policies. For example, Toronto, where crime levels have recently been included in assessments. 656

650 Mazowita and Rotenberg (2019).
651 INT1_CAN.
652 Hodgkinson, Caputo and McIntyre (2019).
653 INT1_CAN.
654 INT2_CAN.
655 Public Safety Canada (2022).
656 INT1_CAN.
A.3. Finland

A.3.1. Background and context

Finland has a national police force that consists of the National Police Board and 11 local police departments; it is overseen by the Ministry of the Interior. Local police departments are not organised under local government organisations, but sit directly under the National Police Board. There are also two national police agencies: the National Bureau of Investigation and the Police University College of Finland. As of 2022, the Finnish national police employs approximately 7,500 officers.

The National Police Board is the central administrative authority of the police force. It is headed by the National Police Commissioner and directs, manages and supervises the activities of the various police departments and agencies in accordance with the Act on Police Administration. The National Police Board is also responsible for measuring the performance of the Finnish police.

The Police University College conducts research to inform policing activities and strategic plans. It also produces reports on police performance and analyses the results from the biennial national community survey on perceptions of police operations and internal security. Furthermore, while police statistics are produced by the National Police Board, the Police University College hosts statistics at the national level.

A.3.2. Current practices in performance measurement

The performance measurement framework of the Finnish police is agreed upon annually by the Ministry of the Interior and the National Police Board.

The indicators and targets used to measure the performance of the Finnish police are agreed upon annually by the Ministry of the Interior and the National Police Board in so-called ‘performance agreements’. As part of this framework, performance indicators are linked to wider police objectives, which are in turn linked to the wider strategic themes of the government and tracked over the course of one year. As Table 24 below shows, the performance of the Finnish police force is primarily measured using quantitative measures...
and qualitative constructs that are measured in a quantitative manner, such as surveys measuring ‘citizen trust in the police’. Moreover, there is a relatively large thematic spread, with indicators relating to the carbon footprint of the police and staff training also included in the performance agreement.

### Table 24: Examples of indicators per strategic theme of the Finnish government

<table>
<thead>
<tr>
<th>Corresponding strategic theme</th>
<th>Police objective</th>
<th>Police performance indicator</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A safe and secure Finland</strong></td>
<td>The police ensure quality police services on a national level by allocating resources according to the operating context and tasks.</td>
<td>Annual clearance rate of criminal offences.</td>
</tr>
<tr>
<td></td>
<td>To increase people’s sense of security and confidence by providing reliable emergency services and by strengthening the police’s role in crime prevention.</td>
<td>Investigation time for criminal offences.</td>
</tr>
<tr>
<td></td>
<td>To ensure and promote equal and fair treatment of staff and customers in all situations.</td>
<td>Citizens’ trust in the police according to the Police Barometer.</td>
</tr>
<tr>
<td><strong>A carbon neutral and biodiversity friendly Finland</strong></td>
<td>To reduce the carbon footprint by 50% by 2027 (from 2018 levels).</td>
<td>Percentage of all-electric and plug-in hybrid cars in civil general-purpose vehicles.</td>
</tr>
<tr>
<td><strong>A fair, equal and inclusive Finland</strong></td>
<td>To ensure that the skills of well-performing personnel meet the requirements of the changing operating environment.</td>
<td>Administrative complaints and Ethics Review Board reports.</td>
</tr>
<tr>
<td><strong>Best public administration in the world</strong></td>
<td>To ensure that the skills of well-performing personnel meet the requirements of the changing operating environment.</td>
<td>Working time spent on training as a percentage of total working time.</td>
</tr>
</tbody>
</table>


After the performance agreement between the Ministry of the Interior and the National Police Board has been agreed upon, the National Police Board drafts performance frameworks and sets objectives for the two national police agencies and the 11 departments. The strategies for the departments contain the same performance indicators used at the national level, but they may be prioritised differently in accordance with regional and local needs. Figures on how each of the departments is performing against the indicators set out in their performance agreements are subsequently published by the National Police Board in an annual report, alongside its financial statements.

To measure how the police force as a whole and individual departments are performing against these indicators, quantitative data from police systems and public surveys are analysed.

To assess police performance in relation to the indicators set out in the performance agreements, various quantitative data sources are used. Information relating to crime statistics, clearance rates, operational readiness times and the number of administrative complaints is gathered from police systems such as the

---

672 Finland Ministry of the Interior (2019).
673 Finland Ministry of the Interior (2019).
674 Police of Finland (2023a).
International approaches to police performance measurement

Police Performance Information System. At the same time, citizen trust in the police is measured by means of the Police Barometer, a biennial survey that has been conducted by Statistics Finland in cooperation with the Police University College since 1999. A quota sample of around 1,000 people (n=1,082 in 2020) aged 15 to 79 living in mainland Finland are interviewed. Survey interviewees are asked several questions on different aspects of trust in the police and police legitimacy, such as whether the police make fair and impartial decisions; treat citizens with respect and justify their decisions; whether the actual impact of policing is in line with the intended benefits; and whether citizens’ own values and perceptions of justice are consistent with the values and practices embodied in police actions. For each of these questions, interviewees are asked to give a score between one and four.

A.3.3. Challenges and complexities in measuring police performance

The Finnish Ministry of the Interior recognises that measuring the performance of the Finnish police in crime prevention is especially complex.

The performance of the Finnish police is primarily assessed using quantitative indicators that are tracked over one-year periods. A publication by the Finnish Ministry of the Interior recognises that this framework for performance measurement does not allow the force to measure the impact of its preventative work in a meaningful manner. The authors highlight that demonstrating the results of such efforts using police statistics can be challenging, as it is often impossible to verify causal links between a decrease in offences and preventative work given that lower crime figures may be partly due to other factors such as a decline in public willingness to report incidents.

The Police Barometer may under-represent or exclude the views of disadvantaged sections of Finnish society.

Furthermore, the authors of the report that accompanies the most recent iteration of the Police Barometer qualify the results of the survey by noting that it may under-represent or completely exclude disadvantaged people living on the margins of Finnish society, including those who are homeless, use drugs and have mental ill health. Moreover, they recognise that various factors, such as personality and circumstances in life, may systematically influence the response propensity of those in the survey population, which may introduce bias into the data and results.

675 Police of Finland (2023a).
676 Vuorensyrjä and Rauta (2020).
677 Vuorensyrjä and Rauta (2020).
678 Statistics Finland (2023).
679 Vuorensyrjä and Rauta (2020).
680 Vuorensyrjä and Rauta (2020).
681 Vuorensyrjä and Rauta (2020).
682 Finland Ministry of the Interior (2019).
684 Vuorensyrjä and Rauta (2020).
685 Vuorensyrjä and Rauta (2020).
A.3.4. Recent developments

The Street Safety Index, a value used to track the changes in safety across time, has recently been replaced by the Disturbance Index.

One indicator that was consistently included in the annual performance agreements of the Finnish police force was the Street Safety Index, a relative value that can be compared to the reference value (100 in 1999) to track changes in safety across time. It was calculated by dividing the population in a particular area by the combined weighted sum of reported crimes for certain crime types, such as robberies, assaults, vandalism and drunk driving incidents, in any one year.686 In February 2023, the Finnish police announced that it would no longer be using the Street Safety Index, citing that it did not produce accurate and meaningful results because it also includes crimes that did not occur in public places.687 At the same time, it was announced that the Street Safety Index would be replaced with the Disturbance Index, 688 which is calculated in a similar manner, dividing the population by the weighted sum of reported crimes for certain crime types; however, it only includes crimes that have occurred in public places.689 According to the Finnish police, this index can be used to monitor the results of police work more effectively.690

686 Police of Finland (2023a).
687 Police of Finland (2023c).
688 Police of Finland (2023d).
689 Police of Finland (2023a).
690 Police of Finland (2023b).
A.4. New Zealand

A.4.1. Background and context

New Zealand Police is New Zealand’s national police force and is overseen by the Commissioner of Police. Within the organisation there are 12 policing districts administered from Police National Headquarters in Wellington (Figure 9). Each district is headed by a district commander and has a central station that oversees the functioning of several subsidiary stations. As of 2022, there are 325 police stations in New Zealand and the force employs 14,600 in total.

The New Zealand Evidence-Based Policing Centre (EBPC) is a research organisation that aims to inform police practice and improve the allocation of resources across the organisation by using practitioner-based research, police data and crime-science to produce insights on a variety of issues, including police performance measurement. The EBPC is a partnership between the New Zealand Police, the University of Waikato, the Institute of Environmental Science and Research, and New Zealand Police’s strategic partner, Vodafone New Zealand. Since 2017, it has been responsible for developing a new framework for performance measurement.

As in Australia, the Australia and New Zealand Policing Advisory Agency (ANZPAA) is an important stakeholder in police performance measurement in New Zealand. Its members are the police commissioners from each jurisdiction, as well as the Australian Capital Territory’s Chief Police Officer and the Commissioner of Police of the New Zealand Police. As an organisation, its purpose is to provide advice on ‘current and emerging policing priorities’, as well as to promote good practice, including in police performance measurement. However, it should be noted that in New Zealand, unlike in Australia, the annual survey that the police use to measure levels of citizen satisfaction with the police is run by the Ministry of Justice, not the ANZPAA.

Unlike its Australian counterpart, the New Zealand Productivity Commission does not appear to measure the performance of the police annually.

---

691 New Zealand Police (2022).
692 New Zealand Police (2023b).
693 New Zealand Police (2023a).
694 INT1_NZ; INT2_NZ.
695 Australia New Zealand Policing Advisory Agency (2023).
696 Australia New Zealand Policing Advisory Agency (2023).
697 New Zealand Ministry of Justice (2022).
698 The most recent assessment of police performance by the New Zealand Productivity Commission appears to be from 2017. See Genet and Hayward (2017).
A.4.2. Current practices in performance measurement

A limited set of measures to assess how the police in New Zealand are performing in relation to their strategic aims was developed in consultation with police stakeholders across the country.

Since 2020, the strategic objectives of New Zealand Police have been threefold: safe homes, safe roads and safe communities.\footnote{INT1\_NZ; INT2\_NZ.}\footnote{INT2\_NZ.} To assess how the police are performing in relation to each of these broad aims, the EBPC developed a new framework for police performance measurement in 2021, which has since become known as the Operational Performance Framework (OPF).\footnote{INT3\_NZ.} The OPF includes a set of 120 measures to which targets are attached, and which are reported on in the organisation’s Annual Account.\footnote{INT1\_NZ; INT2\_NZ.} Each measure sits under one of the strategic objectives and is classed as relating to ‘outcome’, ‘prevention’, ‘response’ or ‘investigation and resolution’ (Table 25).\footnote{New Zealand Police (2022).}

Two interviewees indicated that this framework was developed in consultation with police stakeholders across New Zealand.\footnote{INT1\_NZ; INT2\_NZ.} They noted that the EBPC ran focus groups and workshops throughout the country.

---

\footnote{INT1\_NZ; INT2\_NZ.}
\footnote{New Zealand Police (2022).}
\footnote{INT3\_NZ.}
\footnote{INT1\_NZ; INT2\_NZ.}
\footnote{New Zealand Police (2022).}
\footnote{INT1\_NZ; INT2\_NZ.}
to establish what local police leadership felt were important indicators of the force’s performance in relation to each of the three strategic objectives.  

### Table 25: Examples of measures from the Operational Performance Framework per strategic objective

<table>
<thead>
<tr>
<th>Safe homes</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Percentage of family violence reoffending at same or greater level of seriousness (outcome).</td>
<td></td>
</tr>
<tr>
<td>Aggravated sexual assault victimisation clearance rate (investigation and resolution).</td>
<td></td>
</tr>
<tr>
<td>Compliance with Police Safety Orders (prevention).</td>
<td></td>
</tr>
<tr>
<td>Safe roads</td>
<td></td>
</tr>
<tr>
<td>Number of emergency traffic events responded to (response).</td>
<td></td>
</tr>
<tr>
<td>Number of fatal and serious crashes per 10,000 licensed vehicles (prevention).</td>
<td></td>
</tr>
<tr>
<td>Percentage of survey respondents who had contact with police at the roadside that agree they were treated fairly (investigation and resolution).</td>
<td></td>
</tr>
<tr>
<td>Safe communities</td>
<td></td>
</tr>
<tr>
<td>Percentage of people who agree or strongly agree that police deal effectively with serious crimes (outcome).</td>
<td></td>
</tr>
<tr>
<td>Percentage of youth who reoffend following referral to a Youth Aid intervention (prevention).</td>
<td></td>
</tr>
<tr>
<td>Percentage of charges laid by police that result in conviction (investigation and resolution).</td>
<td></td>
</tr>
</tbody>
</table>


To assess how the police are performing in relation to these measures, data are collected from police registration systems and citizen surveys.

The primary source of data to assess how New Zealand Police is performing against the measures set out in the OPF is police registration and administrative systems, such as the National Intelligence Application. This includes crime rates, reoffending rates and clearance rates.

Data on police performance are also collected through the New Zealand Crime and Victims Survey (NZCVS), an annual random-sample citizen survey run by the Ministry of Justice. Apart from the main sample for this survey (n=4,246 in 2020–21), there is also a Māori booster sample (n=1,998 in 2020–21) aimed at increasing the sample size for Māori New Zealanders. The survey includes 16 key questions around police legitimacy and trust in the national police, including questions on the extent to which respondents feel that police deal effectively with serious crimes and the extent to which respondents are

---

705 INT1_NZ; INT2_NZ; INT3_NZ.
706 INT1_NZ; INT2_NZ; INT3_NZ.
707 INT1_NZ.
709 New Zealand Ministry of Justice (2022, Appendix 2).
satisfied with police service delivery among victims of crime.\textsuperscript{710,711} Results are then analysed to determine how the force is performing in relation to the strategic measures relating to citizen satisfaction, and are subsequently reported on to police leadership.\textsuperscript{712}

**Dashboards have been developed to inform stakeholders across the force of how the organisation is performing in different areas at both national and regional levels.**

Using SAS Visual Analytics and Power BI, police administrative data are brought together in a set of operational and strategic dashboards\textsuperscript{713} that allow stakeholders at both national and regional levels to be kept informed of how they are performing in relation to different crime types.\textsuperscript{714} In designing these dashboards, particular attention was paid to ensuring that the data are displayed in an intuitive manner that allows staff who do not regularly work with data to draw insights from it.\textsuperscript{715}

### A.4.3. Challenges and complexities in measuring police performance

**In the past, New Zealand police have struggled with a lack of consistency in how data are recorded and analysed.**

Data from two interviews indicate that in recent years, police in New Zealand have found it challenging to ensure that information related to police performance is recorded and analysed in a consistent manner.\textsuperscript{716}

As in many other jurisdictions, the quality of data collected from police incident registration systems has been largely dependent on the accuracy of the information entered by the responding officer.\textsuperscript{717} In recent years, this has led to inconsistencies and inaccuracies in the data recorded as there has been no systematic approach to what should be recorded and how.\textsuperscript{718}

Both interviewees expressed that the meaningful measurement of police performance has been hampered by a lack of consensus on the definitions of certain crime types, such as ‘family harm’ and ‘burglary’.\textsuperscript{719} This has resulted in different parts of the police organisation interpreting administrative data from police registration systems, and thus the performance of the force, differently.\textsuperscript{720}

### A.4.4. Recent developments

**In 2023, police performance measurement will for the first time use the New Zealand Crime Harm Index (NZ CHI).**

\textsuperscript{710} INT1\_NZ.
\textsuperscript{711} New Zealand Police (2022).
\textsuperscript{712} INT1\_NZ.
\textsuperscript{713} INT1\_NZ.
\textsuperscript{714} INT2\_NZ.
\textsuperscript{715} INT2\_NZ.
\textsuperscript{716} INT1\_NZ; INT2\_NZ.
\textsuperscript{717} INT1\_NZ.
\textsuperscript{718} INT1\_NZ; INT2\_NZ.
\textsuperscript{719} INT1\_NZ; INT2\_NZ.
\textsuperscript{720} INT2\_NZ.
In 2017, researchers in New Zealand developed the New Zealand Crime Harm Index (NZ CHI), largely replicating the methodology used to build the Cambridge Crime Harm Index (CCHI). However, where the CCHI made use of sentencing guidelines, the NZ CHI was developed using 15 years of sentencing data to develop proxy measures of harm caused. From 2023 onwards, the NZ CHI will be applied to each of the 120 measures set out in the OPF, meaning that police leadership at national and regional levels will be able to see how they are performing against each measure in terms of harm as well as crime volume.

New Zealand Police have recently developed a data dictionary to mitigate some of the above-mentioned challenges and complexities.

In response to the above-described issues with the accuracy and consistency of data relating to police performance, the EBPC has developed an expansive data dictionary that describes each crime type included in the recently developed framework of performance measures. Additionally, the caveats and limitations of the different data gathered to assess the measures are set out in detail. It will also explain that the NZ CHI has been applied to the measure and detail the harm of the crime type in question in terms of equivalent prison days. This ensures consensus among police stakeholders on what is being measured, and that analyses can be replicated in the future.

Stakeholders within New Zealand Police are developing tools to enable police leadership to understand and interpret data related to police performance.

The Commissioner’s Office and the EBPC are developing ways to provide ongoing support to police leadership in the area of performance data and measurement to enable them to understand and draw insights from this component of management information. Stakeholders within the police who are responsible for developing the strategic dashboards are not only focusing on drawing out high-level and strategic trends in an intuitive manner, they are also looking to develop master classes on using performance data to ensure that police leaders can make informed decisions around the deployment of resources.
A.5. The United States

A.5.1. Background and context

United States’ law enforcement organisations include federal law enforcement, state police and county police. Federal law enforcement is divided into eight major divisions, including the US Secret Service, the US Marshall Service, the FBI, the CIA, Immigration and Customs Enforcement (ICE), the Department of Homeland Security, the US Drug Enforcement Administration, and the Bureau of Alcohol, Tobacco, Firearms and Explosives. The state police (sometimes known as state troopers) are responsible for tasks such as the enforcement of traffic laws on state highways and the coordination of cross-jurisdictional police work. County police units are also known as sheriff’s offices, as they are headed by a sheriff who is usually elected by the county. County police officers are responsible for ‘traditional’ police work, including crime fighting and traffic control. Overall, the United States employs over 500,000 police officers across more than 15,000 local police departments. Many police forces around the country are small, with about half employing ten or fewer police officers per station. The police departments are designed to be decentralised, rendering any kind of oversight and performance measurement a complex task.

A.5.2. Current practices in performance measurement

There is no uniform approach to measuring police performance across the United States, and information about data usage is published only sporadically.

In the United States, each jurisdiction is responsible for measuring its own police performance. Often, no information is made public on the methods employed or the data used. However, in some instances, jurisdictions do make their data available to the public. For example, the NYPD publishes limited data online on the number of crime incidents and the development of crime over time. At the same time, aggregate crime data are made available annually through the Uniform Crime Report (UCR), the National Incident-Based Reporting System (NIBRS) and the FBI, which manages the national crime and arrest reporting system.

There are a small number of trends in police performance measurement that can be identified across jurisdictions, revealing some more commonly used indicators.

Because there is no national approach to measuring performance, there are no agreed-upon indicators to list. However, when looking at different jurisdictions, a few trends can be identified. For example, across multiple jurisdictions, police performance measurement focuses on similar measures and indicators, such as crime and crime rates, and traffic incidents. The four main indicators by which police performance appears to be measured across jurisdictions are: 1) reductions in the number of (serious) crime; 2) clearance

---

733 INT1_US.
735 INT1_US; INT2_US.
rate; 3) response time; and 4) enforcement activity. However, Sparrow points to several issues that arise from this narrow focus on police performance (see Section 1.3). For a more in-depth review of his criticism.

The literature shows that multiple jurisdictions use CompStat to measure police performance. Despite the lack of a uniform police performance measurement approach, CompStat (also see Chapter 7) is used by multiple police departments. CompStat was developed by the NYPD in 1994, and has since spread around the world. CompStat collects and analyses quantitative data and focuses primarily on crime reduction measures, with an emphasis on ‘accurate and timely intelligence, effective tactics, rapid deployment and relentless follow-up and assessment’. CompStat is meant to provide a structured approach to police performance, while being adaptable to the needs of different police departments by increasing ‘a police organisation’s capacity to identify, understand and respond to crime problems as they emerge’. Common police performance measures and indicators include measuring crime rates, arrests, clearance rates, response time, traffic accidents, police officer misconduct and use of force. However, it should be stressed that not all US police departments use CompStat.

The purpose of measuring police performance in the United States differs between jurisdictions, but commonly relates to justifying budget allocations, increasing accountability and monitoring misconduct.

Key stakeholders in police performance vary between states and jurisdictions. Stakeholders usually include the police leadership of individual police departments, police unions and police officers, police and fire commissioners, police oversight boards (also known as civilian oversight boards), and the general community. Consequently, the main purposes of performance measurement can vary considerably between jurisdictions and departments. In recent years, many US police forces have come under heavy scrutiny for the use of force and the unequal treatment of citizens – including discrimination and racial profiling. As such, reasons to collect performance data identified in the literature and by stakeholder interviewees included considerations related to justifying the use of police budgets and resources, as well as increasing accountability and monitoring potential personnel misconduct.

---

739 Eterno, Silverman and Berlin (2021, p. 898).
741 INT2_US.
742 INT1_US; American Police Officers Alliance (2019).
743 INT3_US.
744 INT2_US.
746 INT1_US; INT2_US.
A.5.3. Challenges and complexities

The fragmentation of the US policing system makes performance measurement a complex task.

There are a large number of police agencies, jurisdictions, officers and measurement systems in place in the United States, and due to the localisation of police units, there is little uniformity in police performance measurements.\textsuperscript{747} One interviewee noted that the quality of performance data can vary significantly between jurisdictions, as larger police departments will be better equipped to track performance measures compared to smaller departments with less capacity and fewer resources.\textsuperscript{748}

Low budgets pose a risk to effective performance measurement.

As one interviewee explained, a software overhaul that might be needed to implement a new performance measurement system could use up all the budget allocated to a police department for a given year, making it less likely that an organisational change will be implemented.\textsuperscript{749} The interviewee gave an example from the FBI, which recently changed the way it asked police departments to report crime statistics (from the UCR to the NIBRS), leading to a number of police units stopping reporting their crime data altogether, and rendering current national crime statistics less representative than before. According to the interviewee, the decrease in reporting was partially because the change to NIBRS would have been expensive, as departments would have needed to retool their computer records management system.\textsuperscript{750}

However, even when there is sufficient budget available, lack of awareness around police performance can pose a risk to effective performance measurement.

As one interviewee explained, informing officers about the use and purpose of police performance measurement is crucial to ensuring officer buy-in, which helps ensure that the collection of performance data is reliable.\textsuperscript{751} Training for commanding officers and personnel on how to interpret and use performance indicators can mitigate this risk.\textsuperscript{752}

Despite its prevalence in US jurisdictions, CompStat and its focus on crime indicators remain criticised.

Eterno et al., for instance, warn that CompStat’s ‘performance management can morph into a top-down numbers crunching bureaucracy’.\textsuperscript{753} They argue that CompStat requires updating to align better with contemporary policing needs by incorporating, above all, measures that consider community-oriented policing.\textsuperscript{754} Two interviewees furthermore mentioned that public experiences and views of the police are rarely covered in contemporary police performance measurement systems in the United States.\textsuperscript{755}

\textsuperscript{747} INT1_US.
\textsuperscript{748} INT2_US.
\textsuperscript{749} INT1_US.
\textsuperscript{750} INT1_US.
\textsuperscript{751} INT1_US.
\textsuperscript{752} INT1_US.
\textsuperscript{753} Eterno, Silverman and Berlin (2021, p. 886).
\textsuperscript{754} Eterno, Silverman and Berlin (2021, p. 898).
\textsuperscript{755} INT1_US; INT2_US.
International approaches to police performance measurement

In line with the common critique regarding output targets, one interviewee also warned that current performance indicators lead to unintended consequences. For example, an overreliance on arrest rates could lead to higher numbers of arrests than warranted.756

A.5.4. Recent developments

One new innovative practice identified from the interviews is a focus on community satisfaction.

One interviewee identified measuring community satisfaction using QR codes as an innovative measure.757 While no academic literature has been identified in this study’s literature searches about such measures in the United States, there are a number of news stories about the recent introduction of QR codes as a tool for citizens to rate their interactions with police officers anonymously.758 The tool is an initiative of Guardian Score, which advertises the use of QR codes with the slogan: “Why can we rate Uber® drivers but not police officers?”759 The police uses the data results itself.

According to Guardian Score, the process is as follows: following an interaction with a civilian, police officers provide the person with a QR code, leading them to a quick survey.760 The intent of the survey is not for the person to share personal opinions about the police, but to measure procedural justice by asking a person how well the officer explained any reasoning behind the interaction in question, the officer’s listening skills, perceptions of fairness and professionalism, and whether the officer explained potential next steps of the process.761 In the survey, civilians can rate interactions with the police from one to five stars. It also asks for basic demographic information (age, gender, race). The survey scores can subsequently be analysed by police supervisors.762

There is currently no comprehensive list available of all police departments that use the surveys with QR codes. News articles show that they are implemented by several police departments, most notably by police units in the state of New York,763 Virginia,764 Pennsylvania765 and Georgia.766

However, two issues concerning the use of QR codes have already been identified: low response rates and a lack of transparency. One news article found that response rates to the survey ranged from only 10–20 percent.767 Moreover, data from the surveys are not publicly available.768

756 INT2_US.
757 INT2_US.
758 Aiello (2022).
759 Guardian Score (2023b).
760 Guardian Score (2023b).
761 Guardian Score (2023c).
762 Guardian Score (2023c).
763 Aiello (2022).
764 Cruise and Moore (2022).
765 Taylor (2022).
767 Taylor (2022).
768 Guardian Score (2023a).
### Table 26: HMICFRS PEEL Assessment Framework

<table>
<thead>
<tr>
<th>Core question</th>
<th>Indicators</th>
<th>Associated metrics</th>
</tr>
</thead>
</table>
| 1. How good is the force’s service for victims of crime?                      | 1.1 The force manages incoming calls, assesses risk and prioritises the police response well. | 1.1.1 Answers 999 and 101 calls within the force’s agreed time frames a substantial proportion of the time, and has a low number of callers disengaging while waiting for the call to be answered.  
1.1.2 The call handler correctly records the details of the call; the call handler uses and correctly records a structured initial triage and risk assessment to inform the prioritisation given to the call to provide the most appropriate response.  
1.1.3 Call handlers act politely, appropriately and ethically, using clear unambiguous language and give appropriate safeguarding and evidence preservation advice. |
|                                                                               | 1.2 The force deploys its resources to respond to victims and incidents in an appropriate manner. | 1.2.1 The force responds to calls for service within its published time frames based on the prioritisation given to the call and does not inappropriately change the prioritisation of a call (i.e. on the basis of managing down demand when short of resources).  
1.2.2 The force provides an appropriate response, taking into consideration risk and victim vulnerability, including information obtained subsequent to the initial call (i.e. from the public/officers/systems checks). |
|                                                                               | 1.3 The force’s crime recording can be trusted. | 1.3.1 The force is effective at recording reported crime.  
1.3.2 The systems and processes within the force support accurate crime recording.  
1.3.3. The force demonstrates the necessary leadership and culture to meet the national standards for crime recording. |
|                                                                               | 1.4 The force has effective arrangements for the screening and allocation of crimes for further investigation and these take into account vulnerability. | 1.4.1 The force has a suitable crime screening/allocation policy which it adheres to and applies in a consistent way.  
1.4.2 The victim is informed promptly if a crime is screened out. |
|                                                                               | 1.5 The force carries out a proportionate, thorough and timely investigation into reported crimes, with senior level governance providing robust scrutiny. | 1.5.1 All investigative opportunities are considered and those which are proportionate are carried out in a timely manner.  
1.5.2 The Code of Practice for Victims of Crime is adhered to.  
1.5.3 Investigations are appropriately supervised and reviewed.  
1.5.4 The force has an effective investigations policy. |
<table>
<thead>
<tr>
<th>Core question</th>
<th>Indicators</th>
<th>Associated metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.5.5 The force has effective governance arrangements for investigative standards.</td>
<td>1.6 The force makes sure that it follows national guidance/rules for deciding the outcome it gives for each report of crime.</td>
<td>1.6.1 The force has an outcomes policy which aligns with national guidance/rules and which is complied with and applied consistently. 1.6.2 When making an outcome decision, the force’s systems and processes make sure that appropriate consideration is given to the nature of the crime, the offender and the victim. 1.6.3 The force demonstrates the necessary leadership and culture to ensure the use of outcomes is appropriate; standards of compliance with force and national policies are high.</td>
</tr>
<tr>
<td>2. How good is the force at engaging with the people it serves and treating them fairly, appropriately and respectfully?</td>
<td>2.1 The force engages with all its diverse communities to understand and respond to what matters to them.</td>
<td>2.1.1 In liaison with third sector organisations, the force actively identifies and includes all of its varied communities. 2.1.2 The force actively seek views from and engages regularly with local communities to identify local problems and gather intelligence. 2.1.3. The force’s short-term and long-term local policing and partnership activity are influenced by information and intelligence gathered from community engagement. 2.1.4. The force empowers local people to get involved in local policing activity.</td>
</tr>
<tr>
<td>2.2 The workforce understands why and how to treat the public with fairness and respect.</td>
<td>2.3 The workforce understands how to use stop and search powers fairly and respectfully.</td>
<td>2.2.1 The workforce has a sound understanding of unfair behaviour and how to combat it; this knowledge is applied during interactions with the public. 2.2.2 Officers are sufficiently trained in effective communication skills; this knowledge is applied during interactions with the public. 2.3.1 Officers are sufficiently trained in how to use stop and search fairly and appropriately. This knowledge is applied during their interactions with the public. 2.3.2 The overwhelming majority of recorded grounds for stop and search are reasonable. 2.3.3 Body-worn video is used in all stop and search encounters. Interactions between officers and the public are improved as a result.</td>
</tr>
<tr>
<td>2.4 The force understands and improves the way it uses stop and search powers.</td>
<td>2.5 The workforce understands how to use force fairly and appropriately.</td>
<td>2.4.1 The force can demonstrate that its use of stop and search conducted under section 1 PACE [Police and Criminal Evidence Act 1984] and associated legislation is fair and effective. 2.4.2 The force can demonstrate that its use of stop and search under section 60 Criminal Justice and Public Order Act is fair and effective. 2.4.3. The force acts upon scrutiny and challenge received from an external independent forum to improve officers’ use of stop and search powers. 2.5.1 Officers and staff are trained in how to use force fairly and appropriately. This knowledge is applied during their interactions with the public.</td>
</tr>
</tbody>
</table>
| 2.6 The force understands and improves the way in which it uses force.          |                                                                                                                                   | 2.6.1 The force understands how, and with what impact, its officers use force and it uses this knowledge to make improvements. 2.6.2 The force acts upon scrutiny and challenge received
<table>
<thead>
<tr>
<th>Core question</th>
<th>Indicators</th>
<th>Associated metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>3. How good is the force at preventing and deterring crime, anti-social behaviour and vulnerability?</td>
<td>3.1 The force prioritises the prevention of crime, anti-social behaviour and vulnerability.</td>
<td>3.1.1 The force tackles crime, incidents and vulnerability through a focus on prevention activity in force/local performance meetings and tactical tasking and co-ordinating group meetings against the priorities it has set. 3.1.2 The force analyses its own and partner data to establish high-demand and vulnerable locations, people and suspects, including repeat victims, and uses this to support its prevention approach. 3.1.3 The force understands the strengths and needs of local communities and is helping to build resilience and cohesion. 3.1.4 The force acts on results from its use of evidence-based policing methodology.</td>
</tr>
<tr>
<td></td>
<td>3.2 The force uses problem solving and works in partnership to prevent crime, anti-social behaviour and vulnerability.</td>
<td>3.2.1 The force achieves sustainable results through the use of a structured problem-solving model focused on understanding the root cause of crime and vulnerability. 3.2.2 The force understands threat and risk through effective and informative analysis that directs activity. 3.2.3 The force effectively evaluates problem-solving activity, and shares it to inform future activity. 3.2.4 The force works in partnership with a wide range of other organisations in problem solving, crime prevention and early intervention activity, which is effective and achieves positive outcomes and reductions in demand. 3.2.5 The force is undertaking early intervention approaches with a focus on positive outcomes.</td>
</tr>
<tr>
<td></td>
<td>3.3 The force understands demands facing neighbourhood policing teams and manages resources in line with that demand.</td>
<td>3.3.1 The force has a good understanding of neighbourhood demand, and this informs its decision making on resource deployment. 3.3.2 The force is professionalising neighbourhood policing through training, accreditation and CPD (continuing professional development), all focusing on prevention. 3.3.3 The force values successful neighbourhood policing and rewards officers, staff and volunteers who make effective contributions.</td>
</tr>
<tr>
<td>4. How good is the force at responding to the public?</td>
<td>4.1 The force identifies and understands risk effectively at initial contact.</td>
<td>4.1.1 Call handlers answer calls quickly and use a structured approach for risk assessment; they record this on force systems, allowing effective safeguarding of victims and better deployment decisions to be made. 4.1.2 The force understands and promptly identifies vulnerability at first point of contact. 4.1.3 The force seeks advice from experts to inform and help better decision making and risk assessments. 4.1.4 The public are able to contact the force through a range of channels to report that a crime has occurred.</td>
</tr>
<tr>
<td></td>
<td>4.2 The force provides an appropriate response to incidents, including those involving vulnerable people.</td>
<td>4.2.1 The force attends incidents quickly enough to secure scenes, safeguard and protect victims, and provide the quality of service expected. 4.2.2 The force thoroughly assesses a victim’s vulnerability and risk at the initial response.</td>
</tr>
</tbody>
</table>
### Core question | Indicators | Associated metrics
--- | --- | ---
4.3 The force understands the demand faced by officers responding to calls for service and manages its resources to cope with that demand. | 4.3.1 The force has a good understanding of initial emergency response demand; this understanding informs its decision making on resource deployment and the number of officers it needs. 4.3.2 The force can effectively view its demand across all areas, allowing it to consider current demand against current resource. 4.3.3 The force supports and develops the supervisors of initial responders; they provide effective leadership at actual incidents, not just observing and directing remotely. | 4.2.3 The force is effective at managing crime scenes and making the most of early evidence opportunities. 4.3.3 The force supports and develops the supervisors of initial responders; they provide effective leadership at actual incidents, not just observing and directing remotely. |
4.4 The force has a good understanding of the wellbeing needs of its contact management staff and officers initially responding to emergency calls. | 4.4.1 Workloads and working hours in the control room and for response officers are manageable and fair, as the force achieves the right balance between meeting demands and looking after the individual. 4.4.2 Line managers in both the control room and response teams identify wellbeing problems early and, as a result, problems are prevented or don’t get any worse. 4.4.3 The force is professionalising initial contact and emergency response policing through training, accreditation and CPD. | 4.4.3 The force is professionalising initial contact and emergency response policing through training, accreditation and CPD. |
5.1 The force understands how to carry out quality investigations on behalf of victims and their families. | 5.1.1 The force has effective governance in place (strategy, policy and accountability) to make sure that it can address the capacity, capability and standards it needs to achieve quality investigations. | 5.1.1 The force has effective governance in place (strategy, policy and accountability) to make sure that it can address the capacity, capability and standards it needs to achieve quality investigations. |
5.2 The force understands the crime demand it faces and what resources it needs to meet it effectively. | 5.2.1 The force has a comprehensive picture of its crime demand. 5.2.2 Leaders have a good understanding of skills and capabilities required to meet both current and future demand, and feed this into central resourcing and training departments. 5.2.3 Investigations are allocated using effective processes based on a risk assessment of threat, risk and harm that allow the force to meet a wide range of crime demand. 5.2.4 The force understands the capacity and capabilities required to meet digital, cyber and forensic demands (support for investigations). | 5.2.3 Investigations are allocated using effective processes based on a risk assessment of threat, risk and harm that allow the force to meet a wide range of crime demand. 5.2.4 The force understands the capacity and capabilities required to meet digital, cyber and forensic demands (support for investigations). |
5.3 The force provides a quality service to victims of crime. | 5.3.1 The force consistently conducts thorough investigations, leading to satisfactory results for victims. 5.3.2 The force achieves good results for victims by pursuing evidence-led investigations, when appropriate to do so, and actively pursues prosecution on behalf of the victim. 5.3.3 The force maintains victim and witness confidence through adherence to the Code of Practice for Victims of Crime. 5.3.4 Victims and witnesses are provided with adequate support, which encourages them to see criminal justice proceedings through to completion. 5.3.5 Effective and appropriate supervision and review of investigations has a positive effect. | 5.3.1 The force consistently conducts thorough investigations, leading to satisfactory results for victims. 5.3.2 The force achieves good results for victims by pursuing evidence-led investigations, when appropriate to do so, and actively pursues prosecution on behalf of the victim. 5.3.3 The force maintains victim and witness confidence through adherence to the Code of Practice for Victims of Crime. 5.3.4 Victims and witnesses are provided with adequate support, which encourages them to see criminal justice proceedings through to completion. 5.3.5 Effective and appropriate supervision and review of investigations has a positive effect. |
<table>
<thead>
<tr>
<th>Core question</th>
<th>Indicators</th>
<th>Associated metrics</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. How good is the force at protecting vulnerable people?</td>
<td>5.4 The force manages the wellbeing of staff involved in investigations.</td>
<td>5.4.1 Supervisors promote a healthy work-life balance. 5.4.2 Workloads and working hours are manageable and fair as the force achieves the right balance between meeting demands and looking after the individual.</td>
</tr>
<tr>
<td>6.1 The force understands the nature and scale of vulnerability. (This includes all types of vulnerability e.g. older people, hate crime, mental ill health, domestic abuse, children, child sexual exploitation, modern day slavery and human trafficking)</td>
<td>6.1.1 The force has effective governance for vulnerability (strategy, policy and accountability) to make sure it has the capacity and capability to protect vulnerable people. 6.1.2 The force collects victim feedback on a regular basis, including through partner agencies, and uses this to improve services both at an organisational and individual level.</td>
<td></td>
</tr>
<tr>
<td>6.2 The force provides ongoing safeguarding and support for vulnerable people, including those at risk of criminal exploitation.</td>
<td>6.2.1 The force recognises the need for effective and ongoing safeguarding of vulnerable victims. 6.2.2 The force is good at recognising and dealing with harm, including hidden harm (evidenced through the number of victims identified and safeguarded by police), sharing information about vulnerable victims/groups with partner agencies and this prompts appropriate action/support. 6.2.3 The force understands how it uses the powers available to best protect and safeguard vulnerable people and victims, and makes sure they are used when appropriate.</td>
<td></td>
</tr>
<tr>
<td>6.3 The force works effectively and proactively with partners to reduce vulnerability and repeat victimisation.</td>
<td>6.3.1 The force contributes to the effectiveness of multi-agency arrangements.</td>
<td></td>
</tr>
<tr>
<td>6.4 The force has a good understanding of demand and resources, including when working with other agencies.</td>
<td>6.4.1 Resourcing and investment in vulnerability is made using a full understanding of current demand. 6.4.2 The force understands its own likely future demand, and also understands the future investment and resourcing decisions of partner agencies and how this may affect policing and the community.</td>
<td></td>
</tr>
<tr>
<td>6.5 The force maintains and improves the wellbeing of staff involved in protecting vulnerable people and understands the effect of the action it is taking.</td>
<td>6.5.1 The force is clear that many protecting vulnerable people (PVP) roles pose a high risk to wellbeing and it provides an enhanced wellbeing service to the people in these high-risk roles.</td>
<td></td>
</tr>
<tr>
<td>Core question</td>
<td>Indicators</td>
<td>Associated metrics</td>
</tr>
<tr>
<td>---------------</td>
<td>------------</td>
<td>--------------------</td>
</tr>
</tbody>
</table>
| 7. How good is the force at managing offenders and suspects? | 7.1 The force is effective in apprehending and managing suspects and offenders to protect the public from harm. | 7.1.1 Outstanding suspects (not yet apprehended) are prioritised and monitored, and force leaders held to account for apprehending them.  
7.1.2 Processes are in place to promptly circulate wanted persons on the police national computer (PNC) and action is taken to locate these persons.  
7.1.3 Pre-charge bail is always used when appropriate; this is monitored and there is clear accountability for ensuring this happens.  
7.1.4 The force uses ‘released under investigation’ or ‘subject to voluntary attendance’ when appropriate; their use is monitored for both risk and timeliness, and there is clear accountability for making sure this happens.  
7.1.5 The force works with Immigration Enforcement to identify and manage arrested foreign nationals effectively; it makes referrals to ACRO (criminal records office) to check previous overseas convictions for foreign nationals. |
| | 7.2 The force effectively manages the risk posed to the public by the most dangerous offenders | 7.2.1 The force uses nationally recognised risk assessment tools e.g. ARMS (active risk management system), OSP in line with APP [authorised professional practice], and these are completed in a timely manner, including reviews.  
7.2.2 Management of offenders, including reactive management of offenders, is in line with APP.  
7.2.3 The force routinely considers preventive or ancillary orders to protect the public from the most dangerous offenders; breaches are monitored, and action taken.  
7.2.4 Neighbourhood and response teams are aware of registered sex offenders in their area; the awareness is sufficient that they recognise opportunities, take enforcement action, submit intelligence and safeguard victims.  
7.2.5 The force has systems in place to proactively identify from all sources the sharing of indecent images of children. It understands the benefits of, and how best to use, specialist software to proactively identify people sharing indecent images of children. It takes appropriate action, in a timely manner and based on an assessment of risk. Any risk in any backlogs are then visible to senior leaders. |
| | 7.3 The force has an effective Integrated Offender Management (IOM) programme. | 7.3.1 The force understands who its repeat offenders are and takes effective measures to reduce re-offending or change behaviour.  
7.3.2 The force IOM programme is in line with the latest IOM strategy in terms of evidence-based need and is effective in supporting desistance. |
| | 7.4 The force understands the demand and has the resources it needs to manage suspects and offenders effectively. | 7.4.1 The force understands the benefits and outcomes from managing offenders effectively as well as the impact/costs associated with offenders, not just to policing but to other agencies; it uses this understanding to inform its allocation of resources in consultation with partner agencies. |
| 8. How good is the force at disrupting serious and organised crime? | 8.1 The force makes good use of all available intelligence to identify, understand and prioritise | 8.1.1 Strategic analysis directs and prioritises SOC intelligence collection.  
8.1.2 The force assesses threat, harm, risk and vulnerability |
<table>
<thead>
<tr>
<th>Core question</th>
<th>Indicators</th>
<th>Associated metrics</th>
</tr>
</thead>
</table>
| **serious and organised crime (SOC) and inform effective decision making.** | 8.2 The force has the right systems, processes, people and skills to tackle SOC and keep the public safe. | 8.2.1 Effective strategic management and planning meet SOC demand.  
8.2.2 The force effectively manages and co-ordinates its response to serious and organised crime threats.  
8.2.3 The force aims to continually improve and develop how it responds to SOC threats. |
| **8.3 Disruptive activity reduces the threat from SOC (Pursue).** | 8.3.1 The force maximises disruption of serious and organised crime by considering a range of overt and covert tactics; this includes using the powers of partner agencies to frustrate organised criminals. |
| **8.4 The force prevents people from engaging or re-engaging in organised crime (Prevent).** | 8.4.1 The force has a consistent and structured approach to identifying those people at risk of being drawn into SOC or whose offending is likely to intensify and become more serious.  
8.4.2 The force works with public and private sector partners to effectively deliver prevent initiatives and diversionary schemes.  
8.4.3 SOC offender management prevents those people in the criminal justice system from continuing to offend. |
| **8.5 Communities, organisations and individuals are resistant and resilient to the impact from serious and organised crime (Protect and Prepare).** | 8.5.1 Police and partner agencies reduce risk and vulnerability in local communities. |
| **9. How good is the force at meeting the requirements of the Strategic Policing Requirement (SPR)? (UNGRADED QUESTION)** | 9.1 The force understands its expected contribution to the strategic policing requirement (SPR) threats and plans accordingly.  
9.1.1 The force has researched and analysed the SPR threats; they sit with local priorities in a logical manner.  
9.1.2 The force works effectively with neighbouring forces, blue light services and other organisations to address the threats in the SPR (connectivity).  
9.2 The force assures itself it continues to have the capacity and capability to respond to the SPR threats.  
9.2.1 The force routinely assesses its capacity and capabilities to address the SPR threats and adjusts accordingly. (capacity and capability).  
9.2.2 The force’s workforce development programmes include measures to build capacity and capability to meet the SPR threats. (capacity and capability).  
9.3 The force plans effectively to meet changing future demands posed by the six SPR threats. (Using the lens of civil emergencies that need a joined-up response across police force boundaries to test this).  
9.2.1 The force routinely assesses its capacity and capabilities to address the SPR threats and adjusts accordingly. (capacity and capability).  
9.2.2. The force’s workforce development programmes include measures to build capacity and capability to meet the SPR threats. (capacity and capability). |
| 10. How good is the force at protecting communities against armed threats? | 10.1 The force has a good understanding of its current and future operational requirements  
10.1.1 Strategic research and analysis identifies the future requirements for armed policing. |
10.1.2 Workforce plans show how the force is building capacity and capabilities in accordance with projected demand.

10.2.1 Effective systems are in place that demonstrate leadership and management oversight of the deployment, command, equipping and training of armed officers (and the use of less lethal weapons).

10.3.1 The selection and acquisition of weapons, specialist munitions, ammunition and less lethal weapons is authorised by the designated chief officer (or in the case of less lethal weapons ‘other’ chief officer where appropriate).

10.3.2 The force’s use of firearms, ammunition and specialist munitions relates to the threats identified in the APSTRA.

10.4.1 Collaborative arrangements between forces are mature and benefits, such as additional capacity/capability and reduced costs, can be quantified.

10.5.1 Where specific threats have been identified contingency plans have been developed to respond to them.

10.6.1 The force routinely identifies learning points and areas for improvement in armed operations at local, regional and national level; these are routinely translated into organisational development.

11.1.1 The force has an ethical culture and environment where the workforce understands what is expected of it, the force is reflective, and the public are treated fairly.

11.1.2 The force tries to make sure that the people in its workforce have a sense of belonging and feel included within the organisation; this aims to create a fair, positive and ethical working environment for all.

11.2.1 The force has a comprehensive and informed understanding of the wellbeing of the workforce.

11.2.2 The force works with a comprehensive and achievable plan to improve the wellbeing of its workforce.

11.3.1 The force improves the wellbeing of the workforce by providing a good range of preventative and supportive measures, with a strong focus on prevention and early intervention.

11.3.2 The force improves the wellbeing of its workforce through supportive leadership at all levels.

11.3.3 The force makes sure members of its workforce are well equipped to maintain their own wellbeing, and gives
<table>
<thead>
<tr>
<th>Core question</th>
<th>Indicators</th>
<th>Associated metrics</th>
</tr>
</thead>
</table>
|               |them the necessary time and space to do so.  
11.3.4 The force understands how its wellbeing provision is making a difference. | 11.3.4 The force understands how its wellbeing provision is making a difference. |
| 11.4 The force is building its workforce for the future. | 11.4.1 The force understands its recruitment needs and has an effective plan to meet them.  
11.4.2 The force is taking effective action so that its workforce better reflects its communities.  
11.4.3 The force understands what factors influence retention and is making good progress with its strategy to encourage people to stay. | 11.4.1 The force understands its recruitment needs and has an effective plan to meet them.  
11.4.2 The force is taking effective action so that its workforce better reflects its communities.  
11.4.3 The force understands what factors influence retention and is making good progress with its strategy to encourage people to stay. |
| 11.5 The force is developing its workforce to be fit for the future. | 11.5.1 The force understands what learning and development is required to match predicted future needs and has a plan to achieve them.  
11.5.2 The force is making progress in its plan to achieve the requirements of the Policing Education Qualifications Framework (PEQF) initial entry routes. | 11.5.1 The force understands what learning and development is required to match predicted future needs and has a plan to achieve them.  
11.5.2 The force is making progress in its plan to achieve the requirements of the Policing Education Qualifications Framework (PEQF) initial entry routes. |
| 11.6 Proactive and disruptive action taken by the force and effective vetting management reduce the threat and risk posed by police corruption. | 11.6.1 The force manages the vetting of its workforce effectively.  
11.6.2 The force has IT monitoring software that operates across all its IT systems.  
11.6.3 The force’s counter corruption units understand and act successfully on the threat and risk posed by police corruption. | 11.6.1 The force manages the vetting of its workforce effectively.  
11.6.2 The force has IT monitoring software that operates across all its IT systems.  
11.6.3 The force’s counter corruption units understand and act successfully on the threat and risk posed by police corruption. |
| 12. How good is the force at planning and managing its organisation efficiently, making sure it achieves value for money, now and in the future? | 12.1 The force has an effective strategic planning and performance framework, making sure it tackles what is important locally and nationally.  
12.2 The force manages current demand well.  
12.3 The force makes sure it has the capability and capacity it needs to meet and manage current demands in the most efficient manner.  
12.4 The force understands future demand and is planning to make sure it has the right resources in place to meet future needs.  
12.5 The force makes the best use of the finance it has available, and its plans are both ambitious and sustainable. | 12.1.1 The force has a comprehensive understanding of the needs and expectations of its communities; effective governance and performance management systems are in place that make sure it translates its strategic vision into well-run services.  
12.2.1 The force has a comprehensive understanding of all sources of demand.  
12.2.2 The force has an operating model that supports it to respond to priorities and meet demands.  
12.3.1 The force has a good understanding of the capability and capacity of its workforce and other assets, across all business areas. Where it has identified gaps, there are plans in place to address them.  
12.3.2 The force demonstrates it is making the best use of the allocation of its resources to manage demand and is providing the public with services which represent good value for money.  
12.4.1 The force has analysed future demand and is proactive in identifying emerging trends and patterns.  
12.4.2 The force’s plans make sure that the operating model will meet future demands and provide value for money.  
12.4.3 The force can demonstrate a well-evidenced rationale for the priority areas selected for investment of additional officers, recruited as part of the ‘uplift’. It is clear what improved outcomes will be achieved.  
2.5.1 Resources follow priorities; the medium-term financial plan is comprehensive, realistic and aligns with the organisational change programme and other force plans. The force has a sound understanding of service costs and outcomes. | 12.1.1 The force has a comprehensive understanding of the needs and expectations of its communities; effective governance and performance management systems are in place that make sure it translates its strategic vision into well-run services.  
12.2.1 The force has a comprehensive understanding of all sources of demand.  
12.2.2 The force has an operating model that supports it to respond to priorities and meet demands.  
12.3.1 The force has a good understanding of the capability and capacity of its workforce and other assets, across all business areas. Where it has identified gaps, there are plans in place to address them.  
12.3.2 The force demonstrates it is making the best use of the allocation of its resources to manage demand and is providing the public with services which represent good value for money.  
12.4.1 The force has analysed future demand and is proactive in identifying emerging trends and patterns.  
12.4.2 The force’s plans make sure that the operating model will meet future demands and provide value for money.  
12.4.3 The force can demonstrate a well-evidenced rationale for the priority areas selected for investment of additional officers, recruited as part of the ‘uplift’. It is clear what improved outcomes will be achieved.  
2.5.1 Resources follow priorities; the medium-term financial plan is comprehensive, realistic and aligns with the organisational change programme and other force plans. The force has a sound understanding of service costs and outcomes. |
| Core question                                                                 | Indicators                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                   |
|------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|---|
| 12.5.2 The force’s financial plans, including its investment programme, are affordable, sustainable and will support it to continue to meet future demands.                                                                                       | 12.6 The force actively seeks opportunities to improve services through collaboration, and makes the most of the benefits of working collaboratively in line with its statutory obligations.   | 12.6.1 The force is engaged in ambitious and effective collaborations with partner organisations, demonstrably leading to better value for money, greater resilience or having achieved cost savings. It continues to actively look for more opportunities to share services.                  | 12.6.2 The force can clearly demonstrate why it is or isn’t collaborating. It uses effective programme management techniques throughout the process, from identifying the business case to tracking the benefits realised.                                                                                                                 | 12.7 The force can demonstrate it is continuing to achieve efficiency savings and improve productivity.                                                                                           | 12.7.1 The force makes the most of the productivity of its resources and assets.                                                                                                           | 12.7.2 The force continues to improve productivity through digital, data and technology solutions, including mobile working.                                                                 |