

Enhancing accountability for police use of lethal force: Global monitoring and comparative benchmarking

Stephen Skinner 

Law School, University of Exeter, UK. E-mail: S.J.Skinner@exeter.ac.uk

ABSTRACT

Policing and law enforcement bodies and the public authorities that govern them should be accountable for all deaths connected with a use of force. Accountability is a requirement of international human rights law and related provisions on the use of force, and foundational to legitimate, constrained, and consensual policing in liberal democracies. However, divergence in normative compliance and types of state system means that accountability for lethal force is inconsistent around the world. This paper discusses how a global system for monitoring lethal force could enhance accountability by making effective use of comparative benchmarking.

INTRODUCTION

The use of lethal force—that is any use of force that results in or is connected with the loss of human life—in policing and law enforcement has increasingly been a cause of global concern in recent years. Achieving accountability for lethal force is a foundational aim of international human rights law and other international provisions that prohibit arbitrary killing and require state agents to report and investigate deaths resulting from policing and law enforcement activities.¹ Seeking to ensure that states acknowledge and answer for any such deaths is a way of upholding the value of human life, ensuring that states' coercive power is exercised within limits, and ultimately striving to reduce and prevent fatalities. Although it is legally, politically, and morally arguable that states should seek to minimize deaths connected with any of their responsibilities—such as health care, transportation, or environmental protection, for example—it is particularly important that they should do so in relation to policing and law enforcement. This is because these are the activities through which states exercise their domestic monopoly of legitimate force to uphold law and order over those whom they govern and supposedly protect within the national territory. In the context of liberal democracies, it is generally understood that accountability is foundational to legitimate and consensual policing (e.g. [Markham and Punch, 2007](#)), in connection with which the

use of force should always be lawful, necessary, proportionate, and publicly acknowledged ([Bittner, 1975](#); [Casey-Maslen and Connolly, 2017](#); [Punch, 2011](#); [Punch et al., 2016](#); [Skinner, 2019](#)). However, the use of lethal force and degrees of compliance with international law vary around the world (e.g. [Flores et al., 2021](#)). Not all states are liberal democracies and among those that are, not all of them govern or regulate policing and law enforcement in the same ways, adhere consistently to principles of legitimate and consensual policing, or face the same contextual challenges. This means that, from a global perspective, establishing a sense of the extent to which lethal force is used, and accountability for deaths relating to policing and law enforcement is achieved, is fraught with complexity.

This paper discusses how accountability for deaths connected with policing and law enforcement could be enhanced through the establishment of a global system for monitoring lethal force, including its use, regulation, investigation, and the collation and publication of data about it. Specifically, the paper examines how such a monitoring system, run by non-state actors, could usefully incorporate internationally comparative benchmarking. The paper begins by considering the challenges of engaging with policing and law enforcement on a global scale and how existing studies have begun to address the use of lethal force worldwide. It then engages with the foundational issues of what is meant by monitoring and how it can ground accountability. Highlighting accountability's relational, substantive, normative, and processual dimensions, the paper shows how the latter raises the need for benchmarking. The paper subsequently explores the conceptual and practical dimensions of benchmarking, both in the sense of using

¹International Covenant on Civil and Political Rights 1966, Article 6; UN Code of Conduct for Law Enforcement Officials, 1979; UN Principles on the Effective Prevention and Investigation of Extra-Legal, Arbitrary and Summary Executions, 1989; UN Basic Principles on the Use of Force and Firearms by Law Enforcement Officials, 1990; UN Manual on the Effective Prevention and Investigation of Extra-Legal, Arbitrary and Summary Executions, 1991; and regional provisions such as the European Convention on Human Rights 1950, Article 2.

benchmarks and gauging relative performance through comparative analysis. Finally, the paper explores how a benchmarking process could be built into international lethal force monitoring, and argues that gauging relative performance would be the most appropriate approach.

CHALLENGES AND EXISTING STUDIES

To begin with, the meaning of 'policing and law enforcement' is not uncontested. The term 'policing' does not have a simple definition and is generally recognized to encompass a range of activities including the maintenance of public order, the protection of public safety and security, crime prevention, the apprehension of suspected offenders, and other operations involving interaction with the public that fall within the wide spectrum of responsibilities attributed to 'the police' and related (state and non-state) agencies across the globe (Loader, 2000; Waddington, 1999; Wright, 2002). Using 'policing' in conjunction with 'law enforcement' is partly a recognition of the terminology used in international provisions on the use of force and arbitrary killing by state officials,² and partly an acknowledgement of the fact that around the world, force is used in the name of domestic law and order not only by agencies classified as police but also by 'third force' paramilitary as well as military bodies. Although the use of force in policing and law enforcement could potentially include actions by private security providers and situations in which a state agent might exercise custodial power (i.e. exert control over a detainee), the conception of policing and law enforcement used in this paper is focussed on state agencies (or 'policing by government' in the terms used by Loader, 2000), excluding those responsible for managing places of custody, such as prisons, youth detention facilities, migrant holding centres, or secure hospitals.

In addition, policing and law enforcement activities are organized and regulated differently around the world. Whereas their governance in some countries is predominantly centralized (e.g. France), in others it is predominantly local, with some central government intervention and some variation by regional jurisdiction (e.g. the UK). Countries with federal systems have a combination of national and sub-national governance structures (such as the USA and Brazil). Variations in governance are further complicated by the ways in which states regulate policing and law enforcement and in particular the resort to force. International law and soft law provisions on these matters are intended to establish minimum standards and a framework for accountability despite country-specific differences, but even the 'lowest common denominator' approach of international norms faces challenges, as degrees of state compliance and the scope and specificity of national laws on the use of force vary widely (e.g. Flores *et al.*, 2021).

Policing and law enforcement are also carried out in diverse operational contexts. For example, levels of organized and violent crime, threats to domestic security, and internal stability differ around the world. Geographical and spatial contexts are similarly diverse, raising operational challenges subject to the nature of a particular country's territory and the urban or rural

conditions of its population. Economic prosperity, resourcing, and infrastructure of policing and law enforcement activities, whether agents are routinely armed, and if so with which sorts of weapons, are further contextual variables. Other dimensions of state activity and socio-political culture can also affect the operational realities of policing and law enforcement bodies, such as national regulatory frameworks for gun control and attitudes to gun ownership which can shape the extent to which force is used by policing and law enforcement bodies.

Against this challenging backdrop, accountability (if it exists at all in some systems) for deaths connected with policing and law enforcement is in large part a system-specific matter that is contingent on the legal, institutional, and socio-political frameworks existent within countries. Under international law it is the state (government) that is the signatory (contracting party) to international normative frameworks and so questions of compliance with international law, such as human rights, ultimately focus on the responsibility and accountability of the central state. In more general and practical terms though, accountability for deaths is a matter for policing and law enforcement bodies themselves, their oversight institutions (where they exist), and the relevant level of public authority with governance responsibilities for these activities. Whilst an increase in police accountability has been identified within some individual systems (Stone, 2007), the above conceptual and organizational challenges could be interpreted to indicate the impossibility of achieving any sort of meaningful or feasible comparative framework for enhancing accountability for lethal force that could be effective across different countries. Alternatively, these challenges could be seen to highlight the need for endeavours to devise an accountability framework for loss of life in relation to policing and law enforcement that is separate from and capable of coherently spanning country-specific systems and contexts.

As recent work on gathering and interpreting information relating to lethal force around the world has shown, this is an urgent international concern for global civil society, which needs to try to make meaningful comparisons across countries, whilst recognizing their organizational and contextual differences. Developing a comparative analysis is an important starting point for assessing countries' relative compliance with international law, determining the scope and limits of existing accountability systems within countries, identifying good (and bad) practice, learning lessons to improve policing and law enforcement activities and, ideally, reducing the number of deaths (Osse and Cano, 2017; Rappert *et al.*, 2022).

Endeavours to gather and evaluate information about lethal force in policing and law enforcement across countries and jurisdictions have so far addressed different aspects of state agencies' activities and involved different approaches. For instance, in *Monitor of Use of Lethal Force in Latin America*, the research team focussed on recording and analysing data about the incidence of lethal outcomes of law enforcement activities. They used statistical analysis to measure and compare total uses of force and apparent abuse of force with reference to context (Pérez Correa *et al.*, 2019). The measurement of apparent abuse drew on internationally recognized data analysis of the 'ratio between civilians killed and civilians wounded in interventions by public security agents', known as the 'lethality index' (Pérez Correa *et al.*, 2019). This involved calculating

²See footnote 1.

comparable numerical indicators of incidences of deaths to assess relative levels of fatal outcomes across Latin American systems.

Adopting a different focus, the *Law on Police Use of Force Worldwide* web resource provides a summary, review, and academic analysis of national legal regimes governing the use of force by law enforcement officials around the world.³ By detailing the nature and scope of national legal provisions, this resource seeks to publicize the extent to which states comply with international standards and have domestic legal frameworks in place for regulating the resort to force. This approach does not adopt comparative numerical or qualitative indicators, but instead provides comparable summary outlines of states' domestic laws and adherence to international norms.

The challenge of developing a more qualitative comparative analysis of lethal force is apparent in another project with a different approach. In *Police Lethal Force and Accountability: Monitoring Deaths in Western Europe*, the focus was not on the use or regulation of lethal force, but on what happens after deaths occur in England and Wales, the Netherlands, Belgium, and France (Rappert et al., 2020). This involved identifying and evaluating national procedures, policies, and practices relating to the compilation and publication by law enforcement agencies and (where they exist) their oversight bodies of data about deaths connected with policing and law enforcement activities, as well as investigative and lesson-learning processes. While elements of good practice were identified in three of the jurisdictions, the report noted that in the period studied, France in particular lagged behind in terms of the collation and publication of data about lethal force. The report's evaluations of these activities were derived from comparative assessment and calibration by the research team, due to 'the absence of any international benchmarking standards' (Rappert et al., 2020). Similarly, the *Toward a Lethal Force Monitor* report extended this investigation and approach to South Africa and Kenya, finding various shortcomings in data collection and investigation procedures (Gandhi et al., 2021). Again, the research team determined the systems' relative performance levels 'based on reasonable, evidence-based evaluations of existing practices' (Gandhi et al., 2021). Overall, these studies highlight the challenge of linking the gathering of information about national procedures and practices with a meaningful international indication of relative performance (Rappert et al., 2022). Each of them sought to establish a basis for evaluative comparison but without an established benchmarking approach, each remains a 'stand-alone' endeavour.

MONITORING AND ACCOUNTABILITY

Monitoring deaths connected with policing and law enforcement is primarily a contribution to ensuring the accountability of the agencies involved in these activities and the public authorities responsible for them (Rappert et al., 2020). To monitor is to maintain a record of activities with a view to drawing attention to any results that appear to indicate a cause for concern, especially in terms of non-compliance with an identifiable standard or an incidence of outcomes that can be interpreted as problematic. Policing and law enforcement agencies, as well

as their oversight bodies, are involved in monitoring their own activities to the extent that they keep records about them. For present purposes though, the idea of international monitoring with a view to enhancing accountability is intended to indicate a process undertaken by members of civil society to raise awareness of deaths connected with policing and law enforcement and to foster ways of reducing them.⁴ In this sense the emphasis is on accountability as a function of monitoring, in that record-keeping and identifying issues arising are complementary parts of 'holding to account'. The nature of accountability needs closer attention though to highlight its principal components and objectives and to establish how it can be enhanced by monitoring involving benchmarking.

Accountability in general terms involves an obligation on an individual, organization, or political entity to maintain a record of activities and to divulge, explain, and justify it to another or others with an entitlement to know, or with a legitimate interest in knowing, about those activities (Bovens et al., 2014). Where such an obligation is not deemed to exist by the individual, organization, or entity in question, or where the obligation is not completely or consistently complied with, accountability can involve other actors in recording and evaluating the activities in question so as to draw public attention, raise awareness and provide a basis for discussion, that is, to hold the individual, organization, or entity to account where it does not, or does not adequately, do so itself. In the context of liberal democracies under the rule of law, accountability is both a positive attribute and an essential part of what it means to be a democratic state rooted in and constrained by constitutional principles and the obligation to answer to the electorate, as well as an instrument or mechanism for ensuring state power is in practice constrained as legitimate authority (Bovens, 2010; Mulgan, 2003). In the context of systems that have weak liberal democratic attributes, or are not liberal democracies, accountability may be generated externally as a means of trying to establish some influence over state practices that are not otherwise grounded in or restricted by internal politico-legal norms or mechanisms, such as electoral concerns.

Although forms and aims of accountability may vary according to systems' politico-legal character and categorization, it is generally considered to involve three main dimensions. These are (1) a relational dimension of one body accounting, or being held to account, to another (e.g. a state to its citizens or the wider international community) (Mulgan, 2000); (2) a substantive dimension, in terms of the activities being considered; and (3) a normative dimension, in terms of principles, rules, or standards (e.g. principles of liberal democracy or national or international law) on the basis of which accountability is deemed to be required and its results evaluated (Harlow, 2014; Philp, 2009; Rock, 2020). The latter point about evaluating results can be seen to engage a fourth element, namely (4) a processual dimension, or how the substance of the accountability relationship is evaluated with regard to the normative dimension. This is the requirement of establishing an evaluative decision-making process that allows for the accurate assessment of accountability data in relation to the chosen principle, rule,

³See <https://www.policinglaw.info/>.

⁴A similar sense of monitoring has received informative academic attention in the context of prisons, for example, Padfield, 2018 and Rogan, 2021.

or standard and, where more than one state or organization is involved, ensuring comparisons are as meaningful as possible in terms of data equivalence and reliability. This processual dimension linking the substantive and normative dimensions of accountability raises the question of comparative evaluation and its possible foundations.

Before turning to those comparative standards, it is important to note that while accountability can be considered an end in itself, in ensuring that state institutions are answerable for their conduct and its outcomes to those with a legitimate interest in them, it can also be a means to an end. This is because the results of an accountability process, in identifying, explaining, justifying, and evaluating elements of institutional activity, are of limited practical utility if they are not fed back into the system, to confirm and consolidate good practice or as a basis for learning and, where needed, change. Ideally, states and institutions that are held accountable need to be required, persuaded, or encouraged to reflect on the process, to identify what is going well and what needs to be improved (e.g. Skinner, 2019). Consequently, comparative evaluation can be important not only in securing accountability but also supporting lesson learning. This is arguably especially important in relation to law enforcement agencies, whose legitimacy and public consent should ideally require not only accountability but also responsiveness to calls for change.

BENCHMARKING

The recording and evaluation of activities by states, other public bodies, and private actors, such as companies, is a well-established feature of national and international monitoring and accountability processes in numerous fields. Such processes have varied aims, which range from core public accountability objectives, such as compliance with political or legal standards and policy targets (e.g. employment rights,⁵ health care provision,⁶ and environmental protection⁷), to developmental, economic, or business aims (including organizational improvement and efficiency,⁸ corporate social responsibility and informed consumer choice⁹). Where monitoring processes involve the comparative evaluation of data about the activities in question, this is generally called

‘benchmarking’. Benchmarking can be conducted in various ways (Broom and Quirk, 2015), but the two principal forms involve (1) assessing performance data in relation to evidence-based points of reference (set standards or targets that are generally referred to as ‘benchmarks’); or (2) comparing performance data without fixed benchmarks in order to gauge relative performance.

Academic discussion of benchmarking is wide ranging, but this paper draws mainly on international relations studies of normative compliance by states and other public and private actors, with some reference to other areas. The use of benchmarking to assess state compliance with normative standards has been said to involve ‘comparative metrics of performance’ to ‘neutralise and universalise’ evaluation aims and processes, which generally focus qualitatively on one or more aspects of the conduct, design, or outcomes of the activities in question (Broom and Quirk, 2015). Benchmarking processes have become increasingly popular with states, non-state actors, and international organizations over recent years and are particularly prevalent in the area of human rights, to bridge the gap between states’ claims about commitment and their demonstrable compliance (Harrison and Sekalala, 2015).

In its most specific sense, benchmarking involves the establishment of benchmarks as points of reference in relation to which data can be assessed with a view to improving related practices (Riis, 1995). Academic analysis suggests that developing effective benchmarks requires them to be considered ‘salient’ (relevant and important) by those concerned, to engage the ‘will’ of those using them to make them effective, and to be established on a basis of sufficient expertise to be perceived as credible and reliable (Seabrooke and Wigan, 2015). In addition, benchmarks need to be clearly defined (Smith *et al.*, 2021) to avoid blurred boundaries (García-Aracil and Palomares-Montero, 2010), apparent overlap, or a perception of data duplication or occultation within overly broad categories. Even where benchmarking involves data comparison to gauge performance without fixed benchmarks, to be meaningful and to inspire confidence it should arguably also focus on activities and data that meet similar requirements. Consequently, it is suggested here that the key requirements for benchmarking processes, in general, are their relevance, reliability, clarity, and ‘traction’, that is, their ability to be considered significant and worthy of respect both by those using them and those whose conduct is evaluated through them.

Benchmarking processes are in large part shaped by a comparative method. Benchmarking that considers information about a particular system or activity in relation to a benchmark, or fixed reference point, involves comparative evaluation to decide whether the system or activity in question is above, below, or in line with the benchmark. Benchmarking that considers information about one system or activity by reference to information about another involves comparison to gauge relative performance. Comparison also occurs when the outcome of such forms of benchmarking is interpreted in relation to other data, for example, from additional systems or a previous period (to assess continuity or change). As a process of comparison, benchmarking shares the interrogative aims of

⁵For example, the UK Institute of Employment Rights reports on the compliance of UK employment law with international norms: <https://www.ier.org.uk/product/benchmarking-freedom-of-association-the-uks-non-compliance-with-international-standards-hard-copy/>.

⁶For example, the UK’s National Health Service ‘benchmarking network’ carries out audits of health care provision, using a database of performance indicators assessed in relation to established quality standards: <https://www.nhsbenchmarking.nhs.uk/national-audits>. The UK’s National Institute for Health and Care Excellence (NICE) similarly provides benchmarks to help define and measure quality in health, public health, and social care: <https://www.nice.org.uk/standards-and-indicators>.

⁷For example, the World Resources Institute (WRI) Environmental Democracy Index evaluates compliance with environmental standards: <https://www.wri.org/insights/best-and-worst-countries-environmental-democracy> and <https://www.environmental-democracyindex.org/>. Rather than focus on norm compliance, the Organization for Economic Co-operation and Development (OECD) offers an ‘environmental policy stringency (EPS) indicator’ which ‘aggregates information on selected environmental policies to create a composite measure of relative policy stringency across countries and over time’: <https://www.oecd.org/economy/greeneco/How-stringent-are-environmental-policies.pdf>.

⁸For example, the UNESCO International Bureau of Education and the US Education Commission of the States use international comparative benchmarking to improve organizational performance: <http://www.ibe.unesco.org/en/glossary-curriculum-terminology/i/international-benchmarking-and-curriculum> and <http://www.ncee.org/wp-content/uploads/2010/11/IBtoolkit3-26-09.pdf>.

⁹For example, the World Benchmarking Alliance provides comparative data about the private sector’s role in meeting the UN sustainable development goals: <https://www.worldbenchmarkingalliance.org>.

comparative analysis more generally. By expanding analysis from a single example on its own terms, it becomes possible to open new interpretive perspectives and to identify and evaluate similarities, differences, patterns, and relative merits or limitations (e.g. Fletcher, 1998). To be meaningful and reliable though, comparative analysis must consider circumstantial and institutional context (e.g. Legrand, 1996; Ridgeway and MacDonald, 2014).

As noted above, benchmarking can support a range of aims, which for present purposes notably include accountability and learning (and if need be, change) through evidence-based comparison (e.g. OECD, 2018). In accountability terms, benchmarking objectives include the dissemination of information about institutional or governmental performance so as to raise public and media awareness, generate discussion and scrutiny, and require explanation, justification, or improvement by the body being monitored. With regard to supporting learning and achieving change where necessary, benchmarking can constitute an evidence-based and persuasive way of informing reflection, challenging current understanding, or exerting pressure so as to influence organizational behaviour (Graafland *et al.*, 2004; Riis, 1995). This can be on the basis that evidence-based learning is more robust and a way of sharing good practice (Ammons and Roenigk, 2015), or that comparative benchmarking provides a competitive motivation for improvement.

Some of these approaches have been used in relation to policing and law enforcement. For example, the Law Enforcement Benchmarking and Performance Analytics Portal of the International Association of Chiefs of Police enables police forces to compare their own data about variables in their activities with data from similar forces and so evaluate their relative performance.¹⁰ The International Police Science Association World Internal Security and Police Index similarly involves data-based comparative ranking of countries in terms of perceived internal security effectiveness and public confidence, to support improvement by encouraging countries to 'compete positively'.¹¹ Adopting a more qualitative approach, Transparency International UK undertook a 'benchmarking police integrity' study of actual and potential corruption in UK police forces to develop 'twelve pillars of integrity' to improve police performance, based on a comparative analysis of good practice in the UK and globally.¹² In contrast, a UK drone monitoring body undertook a survey of police forces' use of drones in order 'to benchmark' current usage by presenting comparative data, explicitly to foster transparency and public accountability.¹³ Regarding lethal force specifically, as noted in the second section above, the study of Latin America involved comparative benchmarking in the form of ratios of the use and abuse of force (Pérez Correa *et al.*, 2019), and the studies of countries in Western Europe and Africa proposed comparable qualitative evaluations to indicate relative performance levels (Gandhi *et al.*, 2021; Rappert *et al.*, 2020). These examples of benchmarking in the policing and law enforcement field have involved different approaches but have all been oriented towards institutional improvement or accountability.

Reliance on benchmarking processes is not without difficulties, however. A common concern is that the reliance on data to produce evaluative models of state or institutional practices can result in selective analysis that masks its incompleteness, conceals detail, or privileges measurement over engagement with context (Broom and Quirk, 2015; Lebaron and Lister, 2015). Another concern is that reliance on benchmarks can limit stakeholder engagement and development by facilitating convergence or 'levelling' around a readily achievable target, which can end up undermining effective reflection and replacing learning with more mechanistic 'box-ticking' (Knutsson *et al.*, 2012; Preece, 2013). Even without benchmarks, benchmarking in the sense of gauging relative performance can vary in its effectiveness. For instance, benchmarking can range from simply publishing comparative data, to performance ranking, to a more critically engaged process of highlighting unfavourable results ('naming and shaming'¹⁴). These can all have different impacts according to the extent to which the organization or state held to account is persuaded, pressured, or antagonized by the approach adopted. Consequently, benchmarking arguably has an important role to play in monitoring and accountability processes, specifically in relation to lethal force, but must be carefully deployed.

ENHANCING ACCOUNTABILITY FOR LETHAL FORCE

This section explores how accountability for lethal force in policing and law enforcement activities could be enhanced through international monitoring and benchmarking. Although the lawfulness or unlawfulness of deaths must remain a matter for national and international courts, other aspects of lethal force could be covered by international monitoring, including, for example, overall and relative numbers of deaths, national authorities' production, and publication of data about fatal incidents, investigations, and procedures for institutional reflection and lesson-learning. As previously noted, the aim of enhancing accountability for lethal force through international monitoring is a concern of global civil society. It is envisaged here that such monitoring could appropriately be carried out by academic and/or independent researchers working in collaboration with non-governmental organizations and international bodies where appropriate. In this way, monitoring would be independent from national policing and law enforcement institutions and would not need to be reliant on existing international organization frameworks, which are already stretched in terms of resources.

In terms of the processual dimension of accountability, how the substance of the accountability relationship is to be evaluated lacks a definitive normative basis due to the absence of specific international parameters for acceptable numbers of deaths or, for instance, how data about deaths should be collected and published. Even so, the fundamental significance of the right to life in international human rights law and the requirements of international provisions on the use of force

¹⁰<https://www.theiacp.org/benchmarking>.

¹¹<http://www.ipssa-police.org/ProjectInfoDetails/world-internal-security-and-police-index>.

¹²<https://www.transparency.org.uk/integrity-standards-uk-police-service>.

¹³<https://dronewars.net/2020/11/02/benchmarking-police-use-of-drones-in-the-uk/>.

¹⁴Note Pérez Correa *et al.* (2019) on the idea of 'hurting' the image of a law enforcement agency or state.

in law enforcement¹⁵ can arguably be understood as starting points that support endeavours to enhance accountability for lethal force, but this entails going beyond the scope of current international norms. Importantly, international provisions outline minimum standards on which states can, if they are willing, build and improve, with human rights norms, in particular, representing a metaphorical floor, not a ceiling (Rappert *et al.*, 2020). The normative dimension of accountability in this area is thus partly rooted in international law but remains elastic and open to interpretation.

Consequently, the role of benchmarking is particularly important. Internationally comparative studies that have so far focussed on lethal force in policing and law enforcement (as outlined in the second section) constitute an essential evidence base on which future monitoring and an appropriate benchmarking process can be developed (Rappert *et al.*, 2022). This development requires consideration of general and specific factors. In general terms, any global benchmarking process for monitoring lethal force must meet the criteria identified above of relevance, reliability, clarity, and traction.

Relevance

A benchmarking process must address pertinent elements of policing and law enforcement agencies' activities. Any benchmarks that are established must be relevant or, without set benchmarks, any gauging of relative performance must similarly focus on relevant aspects of those activities. For example, benchmarking relating to the availability of data about deaths connected with policing and law enforcement would need to address key issues such as how data is collated, how the collation procedure is organized (e.g. issues of method), and its results (e.g. publication, accessibility, and accuracy).

Reliability

A benchmarking process must inspire confidence and be perceived as credible. This would partly be ensured by the expert knowledge and understanding of the researchers undertaking the process, ideally working in cross-disciplinary teams and in consultation with appropriate organizations as noted above. This would also require benchmarking to be based on identifiable methods that can support reproducible results and to satisfy criteria of transparency and neutrality, necessitating disclosure of affiliations and funding.

Clarity

This is partly a matter of focus and methods, which are related to relevance and reliability. However, clarity also requires a benchmarking process to be as consistently interpretable and applicable as possible, which would require provision for variable evaluation where appropriate, given some of the contextual complexities of policing and law enforcement (as noted in the second section).

Traction

The extent to which a benchmarking process has traction in supporting accountability and (if appropriate) change depends

on the above three factors and the willingness of those held to account to accept and respond to this process. This will partly be determined by the degree to which policing and law enforcement agencies and the governance bodies in charge of them respect the monitoring and accountability process in principle and are prepared to find an appropriate balance between it and their own security and control objectives in practice. Responsiveness to benchmarking is also a matter of capacity and practicability within the constraints of operational and resourcing contexts.¹⁶

In more specific terms, monitoring lethal force could conceivably involve a benchmarking process that uses set benchmarks. Indications of what sorts of benchmarks could be used are already apparent to some extent in existing studies but would need refinement in light of the above criteria. For example, the *Police Lethal Force and Accountability* report focussed on the collection and publication of data, its quality, the nature and independence of investigations into deaths, and the analysis of data gathered with a view to learning lessons (Rappert *et al.*, 2020; see also Rappert *et al.*, 2022). Monitoring these matters in the different systems studied involved (1) noting whether or not a particular activity was carried out and (2) evaluating how and how well it was done.

The first sort of assessment could be benchmarked. For instance, identifying whether data about deaths resulting from policing and law enforcement is collected in a jurisdiction, whether it is made publicly available, whether this is a legal requirement, and whether personal and demographic details are included, could involve benchmarks in the form of minimum expectations to be met, perhaps based on examples of good practice in established systems. The second sort of assessment, evaluating how (well) an activity was carried out, would require more qualitative benchmarks. For them to address matters such as the independence of investigations, international human rights law, and related guidelines could provide a concrete basis for establishing reasonable expectations about institutional conduct, as well as relevant contextual factors to be taken into consideration.¹⁷ For example, the extent to which a national police oversight body relies on current or former police personnel to gather and process data about deaths resulting from law enforcement activities can be evidence of a lack of independence, but an accurate comparative benchmark would need to make allowances for specific local factors such as operational roles and control structures, training, or verification processes. Other benchmarks relating to data quality and accessibility (or 'useability': see Rappert *et al.*, 2022) have so far taken the form of relative 'good-medium-poor' descriptors based on the evaluators' reasonable expectations and cross-system calibration among the small group of countries monitored (Gandhi *et al.*, 2021; Rappert *et al.*, 2020). Turning these sorts of minimum expectations and more evaluative calibrations into benchmarks would require more detailed explanation than the existing studies have

¹⁶Note, for example, the 'Police Accountability – Towards International Standards' project (<https://anr.fr/Project-ANR-20-ORAR-0002>) on the need to contextualize police performance for purposes of standard-setting.

¹⁷For example, the UN Manual on the Effective Prevention and Investigation of Extra-Legal, Arbitrary, and Summary Executions, 1991 or the case law of the European Court of Human Rights on independent investigations under Article 2 of the European Convention on Human Rights.

¹⁵See footnote 1 and related text.

provided in terms of their bases, interpretation, and application to achieve an international standard in line with the above criteria of relevance, reliability, and clarity.

However, establishing set benchmarks that could be globally applicable in the same way would be challenging given the considerable contextual differences in policing and law enforcement arrangements around the world. For instance, benchmarks grounded in Western European experiences of law enforcement would lose relevance and traction if applied in an unmodified form to less developed systems. Conversely, minimal benchmarks established to suit less developed law enforcement systems would be insufficiently precise for global application or would risk facilitating a process of convergence around minimal criteria (Knutsson *et al.*, 2012; Preece, 2013). While an ideal aim might be benchmarking based on universally applicable benchmarks, tempered with appropriately contextualized comparative methods, it is suggested here that international monitoring will need to focus (at least initially) on the second form of benchmarking identified above, namely gauging relative performance.

This sort of benchmarking, by recording, and evaluating data about lethal force and institutional responses to deaths connected with policing and law enforcement in individual systems, presenting the findings contextually and comparatively, would provide a more fluid, flexible, and discursive basis for international monitoring. One way of doing this and addressing concerns about the comparability of findings would be to undertake benchmarking by type of system or by region. For instance, states with law enforcement agencies that operate within oversight frameworks involving specialized independent bodies tend to have similar approaches to gathering and reporting data about incidents of lethal force, which provides a reasonable basis for comparison in that regard, but even so, not all are equally developed or reliable so contextual variation would need to be addressed (Gandhi *et al.*, 2021; Rappert *et al.*, 2020, 2022). Alternatively, benchmarking could be developed incrementally by region, to the extent that social, political, legal, and economic similarities are sufficiently identifiable. Comparable regions and sub-regions could be identified around the world (e.g. Western Europe, Central and Eastern Europe, and Latin America), although variations within geographical divisions would need consideration, such as transition towards or away from democracy and degrees of military involvement in policing and law enforcement, to ensure benchmarking takes adequate account of factors affecting local practices.

Developing a process of relative benchmarking also requires consideration of which approach to take for the systems assessed. The 'high stakes' nature of enhancing accountability for deaths resulting from policing and law enforcement, and ultimately trying to reduce them, could be seen to demand a tough approach, like 'naming and shaming'. However, that could risk alienating essential stakeholders, undermining a perception that benchmarking is reliable and neutral, or weakening the constructive lesson-learning potential of the accountability process. Yet a more cautious approach, such as producing a table of comparative ranking, could risk being ineffectual if state agencies may simply acknowledge or ignore it. A key challenge will therefore be to identify an appropriate approach, even if this means, at least in the short term, different tactics for different

systems or regions. This will also require awareness of differences in capacity and what it is reasonable to expect law enforcement agencies to achieve, because even where they might be willing to engage with benchmarking findings, the practicability and cost of improving local practices, such as data recording and publication, could be significant obstacles to responsiveness.

Finally, it is important to recall that as this paper has argued, monitoring and benchmarking are part of the 'processual' dimension of accountability. As such, they are valuable as ends in themselves due to the process they entail, as much as for any specific measures and outcomes they might involve. Monitoring and benchmarking need to be understood as active processes over time and, as forms of comparative analysis, they have interrogative and discursive forms and aims. This means that even if they begin approximately and require flexibility and adjustments for contextual factors, the process of developing them is useful as a way of raising awareness, attracting attention, and provoking discussion among all stakeholders involved in the accountability relationship.

CONCLUSION

This paper has addressed the need to enhance accountability for deaths connected with policing and law enforcement activities around the world by developing international monitoring. It has outlined examples of recent and ongoing endeavours to monitor lethal force, examined how monitoring constitutes a form of accountability, and underlined how accountability's processual dimension could usefully incorporate benchmarking. Drawing on a cross-disciplinary analysis of academic literature and examples from a range of sectors, including policing, the paper has identified key concerns relating to the establishment and use of different approaches to benchmarking. On those bases, the paper has made general and specific proposals about how benchmarking could be used to support global lethal force monitoring. Despite the complexity of international comparison and evaluation in this field, the paper is grounded in the aspiration that fostering comparative discussion can improve policing and law enforcement practices and ultimately reduce their lethal outcomes.

FUNDING

This article was partly supported by Open Society Foundations project grant 119209 Lethal Force Monitor: Next Steps.

ACKNOWLEDGEMENTS

I would like to thank Brian Rappert and Abi Dymond, as well as this journal's anonymous referees, for their comments and feedback.

REFERENCES

- Ammons, D. N. and Roenigk, D. J. (2015). 'Benchmarking and Interorganizational Learning in Local Government'. *Journal of Public Administration Research and Theory* 25(1): 309–335.
- Bittner, E. (1975). *The Functions of the Police in Modern Society – A Review of Background Factors, Current Practices, and Possible Role Models*. New York: Jason Aronson.

- Bovens, M. (2010). 'Two Concepts of Accountability: Accountability as a Virtue and as a Mechanism.' *West European Politics* 33(5): 946–967.
- Bovens, M., Schillemans, T., and Goodin, R. E. (2014). 'Public Accountability.' In Bovens, M., Goodin, R. E. and Schillemans, T. (eds), *The Oxford Handbook of Public Accountability*. Oxford: Oxford University Press, pp. 1–20.
- Broom, A. and Quirk, J. (2015). 'Governing the World at a Distance: The Practice of Global Benchmarking.' *Review of International Studies* 41(5): 819–841.
- Casey-Maslen, S. and Connolly, S. (2017). *Police Use of Force under International Law*. Cambridge: Cambridge University Press.
- Fletcher, G. (1998). 'Comparative Law as a Subversive Discipline.' *American Journal of Comparative Law* 46: 683–700.
- Flores, C., Citro, B., Guruli, N. *et al.* (2021). 'Global impunity: How Police Laws and Policies in the World's Wealthiest Countries Fail International Human Rights Standards.' *Georgia Journal of International and Comparative Law* 49(2): 243–294.
- Gandhi, D., Heyns, C., Maslen, S. *et al.* (2021). *Toward a Lethal Force Monitor*. Licensed under CC-BY-SA.
- García-Aracil, A. and Palomares-Montero, D. (2010). 'Examining Benchmark Indicator Systems for the Evaluation of Higher Education Institutions.' *Higher Education* 60(2): 217–234.
- Graafland, J. J., Eijffinger, S. C. W., and Smid, H. (2004). 'Benchmarking of Corporate Social Responsibility: Methodological Problems and Robustness.' *Journal of Business Ethics* 53(1/2): 137–152.
- Harlow, C. (2014). 'Accountability and Constitutional Law.' In Bovens, M., Goodin, R. E. and Schillemans, T. (eds), *The Oxford Handbook of Public Accountability*. Oxford: Oxford University Press, pp. 195–210.
- Harrison, J. and Sekalala, S. (2015). 'Addressing the Compliance Gap? UN Initiatives to Benchmark the Human Rights Performance of States and Corporations.' *Review of International Studies* 41(5): 925–945.
- Knutsson, H., Ramberg, U., and Tagesson, T. (2012). 'Benchmarking Impact Through Municipal Benchmarking Networks: Improvement or Levelling of Performance?' *Public Performance & Management Review* 36(1): 102–123.
- Lebaron, G. and Lister, J. (2015). 'Benchmarking Global Supply Chains: The Power of the 'Ethical Audit' Regime.' *Review of International Studies* 41(5): 905–924.
- Legrand, P. (1996). 'European Legal Systems Are Not Converging.' *International and Comparative Law Quarterly* 45: 52–81.
- Loader, I. (2000). 'Plural Policing and Democratic Governance.' *Social & Legal Studies* 9(3): 323–345.
- Markham, G. and Punch, M. (2007). 'Embracing Accountability: The Way Forward—Part One.' *Policing: A Journal of Policy and Practice* 1(3): 300–308.
- Mulgan, R. (2000). 'Accountability': An Ever-Expanding Concept?' *Public Administration* 78(3): 555–573.
- Mulgan, R. (2003). *Holding Power to Account: Accountability in Modern Democracies*. Basingstoke: Palgrave Macmillan.
- OECD Public Governance Reviews. (2018). *Benchmarking Civil Service Reform in Kazakhstan*. Paris: OECD Publishing. doi:[10.1787/9789264288096-en](https://doi.org/10.1787/9789264288096-en)
- Osse, A. and Cano, I. (2017). 'Police Deadly Use of Firearms: An International Comparison.' *The International Journal of Human Rights* 21(5): 629–649.
- Padfield, N. (2018). 'Monitoring Prisons in England and Wales: Who Ensures the Fair Treatment of Prisoners?' *Crime, Law and Social Change* 70: 57–76.
- Pérez Correa, C., Silva Forné, C., and Cano, I. (2019). *Monitor of Use of Lethal Force in Latin America: A Comparative Study of Brazil, Colombia, El Salvador, Mexico and Venezuela*. Mexico City: Centro de Investigación y Docencia Económicas.
- Philp, M. (2009). 'Delimiting Democratic Accountability.' *Political Studies* 57: 28–53.
- Preece, D. V. (2013). 'Convergence through Benchmarking and Policy Learning: The Impact of the Open Method of Coordination on Social Policy within the European Union.' In Graefe, P., Simmons, J. M. and White, L. A. (eds), *Overpromising and Underperforming: Understanding and Evaluating New Intergovernmental Accountability Regimes*. Toronto: University of Toronto Press, pp. 169–189.
- Punch, M. (2011). *Shoot to Kill: Police Accountability, Firearms and Fatal Force*. Bristol: The Policy Press.
- Punch, M., Crawshaw, R., and Markham, G. (2016). 'Democratic Principles and Police Fatal Force: Avoidance of Debate, Unresolved Accountability and Human Rights.' *International Journal of Law, Crime and Justice* 46: 2–12.
- Rappert, B., Adang, O., Daillère, A., *et al.* (2020). *Police Lethal Force and Accountability: Monitoring Deaths in Western Europe*. Licensed under CC-BY-SA.
- Rappert, B., Adang, O., De Paepe, J., *et al.* (2022). 'A Global Lethal Force Monitor: Comparative Opportunities and Challenges.' *Homicide Studies* 27(1): 187–203.
- Ridgeway, G. and MacDonald, J. M. (2014). 'A Method for Internal Benchmarking of Criminal Justice System Performance.' *Crime & Delinquency* 60(1): 145–162.
- Riis, J. O. (1995). 'How to Implement Benchmarking.' In Rolstadås, A. (ed), *Benchmarking – Theory and Practice*. Dordrecht: Springer, pp. 377–381.
- Rock, E. (2020). *Measuring Accountability in Public Governance Regimes*. Cambridge: Cambridge University Press.
- Rogan, M. (2021). 'Prison Inspection and Monitoring: The Need to Reform European Law and Policy.' *European Journal on Criminal Policy and Research* 27: 285–305.
- Seabrooke, L. and Wigan, D. (2015). 'How Activists Use Benchmarks: Reformist and Revolutionary Benchmarks for Global Economic Justice.' *Review of International Studies* 41(5): 887–904.
- Skinner, S. (2019). *Lethal Force, the Right to Life and the ECHR: Narratives of Death and Democracy*. Oxford: Hart Publishing.
- Smith, M. R., Tillyer, R., Lloyd, C., and Petrocelli, M. (2021). 'Benchmarking Disparities in Police Stops: A Comparative Application of 2nd and 3rd Generation Techniques.' *Justice Quarterly* 38(3): 513–536.
- Stone, C. (2007). 'Tracing Police Accountability in Theory and Practice: From Philadelphia to Abuja and Sao Paulo.' *Theoretical Criminology* 11(2): 245–259.
- Waddington, P. A. J. (1999). *Policing Citizens*. London: UCL Press.
- Wright, A. (2002). *Policing: An Introduction to Concepts and Practice*. Cullompton: Willan Publishing.