Rule of Law and Justice
An evidence gap map

May 2023
About 3ie

The International Initiative for Impact Evaluation (3ie) develops evidence on how to effectively transform the lives of the poor in low- and middle-income countries. Established in 2008, we offer comprehensive support and a diversity of approaches to achieve development goals by producing, synthesizing and promoting the uptake of impact evaluation evidence. We work closely with governments, foundations, NGOs, development institutions and research organizations to address their decision-making needs. With offices in Washington DC, New Delhi and London and a global network of leading researchers, we offer deep expertise across our extensive menu of evaluation services.

3ie evidence gap map reports

3ie evidence gap maps are thematic collections of information about impact evaluations or systematic reviews that measure the effects of international development policies and programmes. The maps provide a visual display of completed and ongoing systematic reviews and impact evaluations in a sector or sub-sector, structured around a framework of interventions and outcomes.

The evidence gap map reports provide all the supporting documentation for the maps, including the background information for the theme of the map, the methods and results, protocols, and the analysis of results.

About this evidence gap map report

This report presents the findings of a systematic search to identify and map the evidence base of impact evaluations and systematic reviews of interventions that aim to strengthen rule of law and access to justice. The EGM was developed by 3ie and made possible with generous support from the United States Agency for International Development (USAID)’s Center for Democracy, Human Rights and Governance. All of the content of this report is the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or its Board of Commissioners. Any errors and omissions are also the sole responsibility of the authors. Please direct any comments or queries to the corresponding author, Douglas Glandon at dglandon@3ieimpact.org.


Executive editor: Birte Snilstveit
Managing editor: Tanvi Lal
Publications and web design: Akarsh Gupta
Production manager: Mallika Rao

© International Initiative for Impact Evaluation (3ie), 2023
Rule of Law and Justice: an evidence gap map

Ada Sonnenfeld  
International Initiative for Impact Evaluation (3ie)

Tomasz Kozakiewicz  
3ie

Lina Khan  
Independent consultant

Jennifer Doherty  
3ie

Katherine Garcia  
3ie

John Eyers  
3ie

Rami Zalfou  
Independent Consultant

Douglas Glandon  
3ie

Evidence Gap Map Report 19  
April 2023
Executive Summary

The Pathfinders Task Force on Justice (2019) estimates that 5.1 billion people have no access to effective justice. Nearly 60 percent of existing justice problems remain unresolved and over 253 million people live in situations of extreme injustice, including conditions of slavery, statelessness and high levels of insecurity. Understanding the rule of law (RoL) evidence landscape can help support evidence informed decision-making on how to invest resources in policy implementation and research. Rigorous evidence, or lack thereof, can help us to identify opportunities for evidence synthesis that address questions on policy effectiveness.

This evidence gap map (EGM) report presents the findings of a systematic search to identify and map the evidence base of impact evaluations (IEs) and systematic reviews (SRs) of interventions to improve outcomes in the areas of rule of law and justice around the world.

The EGM draws on an extended USAID definition of rule of law as:

A principle of governance in which all persons, institutions and entities, public and private, including the State itself, are accountable to laws that are publicly promulgated, equally enforced and independently adjudicated, and which are consistent with international human rights norms and standards. Effective rule of law ensures that these laws and the justice institutions that support them protect individual rights and are responsive to and inclusive of the needs of all people in society.

(USAID 2010; HiiL 2018; UN SDG16; Stabilisation and Humanitarian Aid Department of the Government of the Netherlands 2015).

Rule of law is often framed as a means of ensuring or pursuing justice. As such, there is substantial overlap between interventions aiming to strengthen rule of law and those aiming to strengthen systems of and access to justice. Access to justice is conceptualized as relating to access to courts and legal assistance, as well as access to legal and judicial outcomes that are just and equitable (UNDP 2004). For this EGM, we conceptualize justice systems as comprising a country’s criminal and civil justice systems as well as informal and traditional mechanisms of justice.

The RoL EGM provides a snapshot of the evidence base for programming that targets justice systems worldwide. The EGM covers a comprehensive set of interventions that have been implemented across multiple geographies, and plots the evidence base for their effects on outcomes of rule of law and justice.

Methods

To identify all potentially relevant studies, we implemented a comprehensive search and systematic screening process. Relevant studies evaluated the effectiveness of interventions aiming to strengthen rule of law and improve access to justice. We extracted descriptive and bibliographic data from all included studies. For systematic reviews, we critically appraised the methods applied and extracted the implications for policy and practice from high-confidence reviews.
Using 3ie’s EGM software, we created an online, interactive matrix that maps all included studies according to the interventions evaluated and the outcomes reported. The platform provides additional filters so that users can further explore the available evidence. For example, users can search for evidence by global regions, country income levels, or population. The EGM can be viewed at: https://gapmaps.3ieimpact.org/evidence-maps/rule-law-egm.

Main findings

We identified over 50,000 studies through our searches, from which we included 637 completed IEs, 17 ongoing IEs, 107 completed SRs and 11 ongoing SRs in the RoL EGM. The search identified studies dating back to 1990, but most were published after 2000, with an increase in publication of studies evaluating interventions in low- and middle-income countries (L&MICs) starting in 2009.

The distribution of the evidence base is very uneven across geographies. Most included studies evaluated programs implemented in high-income countries (HICs) and the United States in particular: 71 percent of all included IEs were undertaken in the United States alone. By contrast, less than a fifth of included impact evaluations took place in L&MICs. These findings highlight important differences in geographic trends in the rigorous evidence base and raise questions regarding the generalizability of findings to other contexts.

Studies are unevenly distributed across intervention groups. Most studies are concentrated on a quarter of the 29 intervention categories in the EGM. These are primarily focused on creating new services or expanding coverage, access and/or quality of existing legal and justice services. Amongst included studies from North America and Europe, programs focused on capacity building and reform of police are the most frequently evaluated, followed by a) diversion of populations out of the criminal justice system, such as into probation; b) rehabilitation of ex-offenders, such as through skills-building or other interventions to help prisoners integrate effectively into society; and c) crime prevention initiatives, such as economic, mental health support or education for those at risk of engaging in crime. Similar to studies from HICs, most of the studies in L&MICs focus on police reform and engagement, though we also identified a cluster of evidence from L&MICs for legal registration interventions. Most studies focus on the general adult population, with additional focus on urban or peri-urban areas, youth, and women. There is a gap in rigorous evidence for rule of law initiatives targeting youth in L&MICs. Outcomes measuring changes in crime, violence and prison population numbers were reported twice as frequently as any other outcome category.

We identified many methodological gaps in the evidence base, where best practices for ethical and rigorous research on program effects were not clearly followed. Less than 20 percent of impact evaluations report having received approval for the research from an independent ethics review board. Ensuring appropriate consideration of ethics is a core requirement of human-subjects research, particularly when the research is dealing with vulnerable populations, as many of the included studies do, and when there is a risk that interventions may have adverse effects on individuals' wellbeing.
Few studies undertook analysis of disaggregated effects, despite important questions regarding relationships between treatment within the criminal justice sector, access to justice and characteristics of ethnicity, socioeconomic status or education, and disability. This is related to questions of ethics and ensuring that human-subject research does no harm. Few included studies drew on both quantitative and qualitative data, and fewer still incorporated theories of change, particularly among studies from the United States and other HICs. These methods can help identify the relevant contextual factors that explain why an intervention was or was not effective at achieving impacts on outcomes of interest, for different sub-population groups. This in turn may enable more effective application of research findings. This is particularly important given the uneven distribution of studies across geographies noted above. Finally, few impact evaluations undertook cost-effectiveness or cost-benefit analyses, despite questions of cost being important for many stakeholders.

In addition to extracting data on the characteristics of the evidence base, we further critically appraised the methods used in included systematic reviews. Overall, our appraisals suggest low confidence in the findings of 52 percent of the 107 completed SRs. This was typically due either to limited search and screening processes or a lack of adequate assessment of risks of bias in included primary studies. Limitations to the search and screening process may lead to situations in which relevant evidence is not identified and included, which may bias the findings. Adequate risk of bias assessments are necessary for enabling appropriate interpretation of findings on effects.

From the 38 high-confidence reviews, we extracted data on key findings and implications for policy and programming. Overall, the SR findings highlight the effectiveness, including cost-effectiveness, of non-punitive approaches to preventing future justice problems, such as building positive support networks for at-risk individuals and mentoring programs for youth. Further, the reviews identified strong evidence that cognitive behavioral therapy (CBT) is an effective treatment for preventing reoffending, but not necessarily for all types of offenders.

**Conclusion and implications**

Overall, the EGM identifies a fractured evidence base for rule of law and justice programming, which is highly concentrated in a few intervention types and populations within the United States. There are multiple gaps in the evidence base for key populations and geographies of interest, such as for women and girls, and youth programming in L&MICs. Nonetheless, the large body of evidence identified in the EGM and its findings can be utilized by policymakers and implementers during program design, to identify relevant rigorous evidence from both IEs and SRs. The findings highlight multiple implications for research within the rule of law and justice sector, including the importance of seeking independent ethics approval for primary research, and of using approaches such as mixed (qualitative and quantitative) research methods, sub-group analyses and theories of change. These considerations can improve future understanding of what works, for whom, and under what conditions to strengthen rule of law and justice.
Contents

Executive Summary ............................................................................................................... i
List of figures and tables ................................................................................................. v
List of Abbreviations and acronyms............................................................................... vi
1. Introduction ................................................................................................................ 1
   1.1 Background and rationale .................................................................................... 1
   1.2 Study objective and questions ............................................................................ 7
   1.3 Structure of the remainder of this report .......................................................... 7
2. Scope ............................................................................................................................ 7
   2.1 Key concepts and definitions ............................................................................ 7
   2.2 EGM theoretical framework ............................................................................ 9
3. Methods ....................................................................................................................... 13
   3.1 Overall methodological approach .................................................................... 13
   3.2 Criteria for including and excluding studies in the EGM ............................... 14
   3.3 Search, screening and data extraction .............................................................. 16
   3.4 Presentation of the map .................................................................................... 17
   3.5 Analysis and reporting ...................................................................................... 19
4. Findings ....................................................................................................................... 19
   4.1 Volume of evidence .......................................................................................... 19
   4.2 Characteristics of the evidence base ............................................................... 20
   4.3 Findings from high confidence SRs ................................................................. 29
   4.4 Discussion: evidence gap analysis ................................................................... 33
5. Conclusions and implications ................................................................................... 40
   5.1 Implications for policy and practice ............................................................... 42
   5.2 Implications for future research ...................................................................... 43
Online Appendixes ......................................................................................................... 44
References ....................................................................................................................... 45
List of figures and tables

Figure 1: Conceptual framework for the EGM................................................................. 8
Figure 2: PRISMA diagram of systematic search and screening ..................................... 20
Figure 3: Count of RoL impact evaluations published over time, by country income group* ............................................................................................................. 21
Figure 4: Impact evaluations published by country .......................................................... 22
Figure 5: Count of included studies by populations targeted, region & study type* .......... 22
Figure 6: Count of included studies by intervention, country income-group (IEs),
   confidence level (SRs), region & study type................................................................. 24
Figure 7: Count of studies by outcomes reported , country income group (for IEs) &
   confidence level (for SRs).......................................................................................... 25
Figure 8: Counts of included studies by equity consideration, region & study type .......... 28
Figure 9: Percent of included systematic reviews assessed at each confidence level ... 29

Table 1: EGM research questions ................................................................................... 7
Table 2: Key concept definitions .................................................................................... 12
Table 3: Summary PICOS criteria for studies to be included in the RoL EGM ............... 14
Table 4: Definition of EGM filters ................................................................................... 18
Table 5: Distribution of impact evaluations by intervention domain and geography ....... 37
## List of Abbreviations and acronyms

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>3ie</td>
<td>International Initiative for Impact Evaluation</td>
</tr>
<tr>
<td>EGM</td>
<td>Evidence gap map</td>
</tr>
<tr>
<td>HIC</td>
<td>High-income country</td>
</tr>
<tr>
<td>IE</td>
<td>Impact evaluation</td>
</tr>
<tr>
<td>L&amp;MICs</td>
<td>Low- and middle-income countries</td>
</tr>
<tr>
<td>QED</td>
<td>Quasi-experimental design</td>
</tr>
<tr>
<td>RCT</td>
<td>Randomized controlled trial</td>
</tr>
<tr>
<td>SR</td>
<td>Systematic review</td>
</tr>
<tr>
<td>USAID</td>
<td>United States Agency for International Development</td>
</tr>
</tbody>
</table>
1. Introduction

1.1 Background and rationale

This EGM draws on an extended USAID definition of rule of law as:

A principle of governance in which all persons, institutions and entities, public and private, including the State itself, are accountable to laws that are publicly promulgated, equally enforced and independently adjudicated, and which are consistent with international human rights norms and standards. Effective rule of law ensures that these laws and the justice institutions that support them protect individual rights and are responsive to and inclusive of the needs of all people in society.

(USAID 2010; HiIL 2018; UN SDG16; Stabilisation and Humanitarian Aid Department of the Government of the Netherlands 2015).

Rule of law is often framed as a means of ensuring or pursuing justice. The UN Sustainable Development Goal (SDG) 16.3 calls for both rule of law and justice for all as a joint objective (UN General Assembly 2015). As such, there is substantial overlap between interventions aiming to strengthen rule of law and those aiming to strengthen systems of and access to justice. Access to justice is conceptualized as relating to access to courts and legal assistance, as well as access to legal and judicial outcomes that are just and equitable (UNDP 2004). For this EGM, we conceptualize the justice system as comprising a country’s criminal and civil justice systems as well as informal and traditional mechanisms of justice. The justice sector is often the primary mechanism for shaping, enforcing and strengthening rule of law within a society. A focus on people-centered justice recognizes that rule of law is a necessary component of supporting justice for all, while prioritizing as a primary objective the goal of solving justice problems for all people within a society (Pathfinders Task Force on Justice 2019). Given the closeness with which the concepts of rule of law and justice are linked, we focus in this EGM on both rule of law and justice programming.

1.1.1 Challenges to the rule of law and justice

The ideal reflected in definitions of the rule of law is far from the reality experienced by many populations globally. The Pathfinders Task Force on Justice (2019) estimates that 5.1 billion people have no access to effective justice and that nearly 60 percent of existing justice problems remain unresolved. These problems affect some populations more than others. For example, people may be unable to access justice due to a lack of legal identity. This issue is widespread; for example, low rates of birth registration mean that 46 percent of children under the age of five in sub-Saharan Africa have no official identity (High-Level Political Forum on Sustainable Development 2019). Structural inequalities may be reflected in, and exacerbated by, lack of access to justice for marginalized groups. For example, women, people with disabilities, and displaced people may all face challenges in accessing justice (OECD 2019). Other groups may also be disadvantaged in their access to justice by being members of minority ethnic or religious groups (Pathfinders Task Force on Justice 2019).

The justice problems people face around the world are varied. The Pathfinders Task Force on Justice (2019) estimates that over 253 million people live in situations of
extreme injustice, including conditions of slavery, statelessness and high levels of insecurity. People also face detention without trial; the High-Level Political Forum on Sustainable Development (2019) estimates that 30 percent of the global prison population is comprised of people who have not been sentenced, primarily individuals in pre-trial detention. People in low- and middle-income countries (L&MICs) also face several common justice and legal problems relating to family, employment, land, crime and neighbor disputes (HiIL 2018). These problems tend to be exacerbated for people living in contexts where life is already difficult, including in countries that have been affected by conflict, disasters or political transition (ibid.). These problems are not, however, restricted to L&MICs. Challenges to accessing justice, particularly in policing and incarceration, also exist in high income countries and often disproportionately affect specific population groups, such as women and historically marginalized communities (OECD 2019).

Justice problems have direct negative effects on people’s lives, including through increased violence, mental health challenges and loss of livelihoods (HiIL 2018). They also cause broader development challenges for society. Lack of effective and consistently applied rules and frameworks for businesses and financial markets can reduce their ability to function effectively, producing negative effects for the economy (OECD 2019). Lack of clear laws that are equally applied can also lead to instability when people resort to conflict and violence to redress injustice and inequality (Pathfinders Task Force on Justice 2019). The related uncertainty, violence and worries about safety also have negative spillover effects on health and wellbeing and the ability of youth to learn in school (Laurito 2019). These negative health effects are in addition to the strains on health care, education, and other public services caused by poor regulation and high-levels of corruption that can flourish without effective rule of law (Pinzon-Rondon et al. 2015). Indeed, without inclusive justice other sustainable development goals are unlikely to be met (High-Level Political Forum on Sustainable Development 2019).

Several key barriers to ensuring effective rule of law and justice exist. Depending on the context, these may relate to the capacity of institutions and actors within the legal and justice systems and services; contextual factors; political interference; or limited ability and access for people to seek and gain effective justice.

Justice institutions and actors in L&MICs may have capacity, resource or access constraints that limit their ability to process cases effectively, particularly in remote areas. There is a need to provide more resources and build both technical and institutional capacity to reduce the strain on formal justice providers to better enable them to support people’s access to justice (HiIL 2018). While most countries have signed international treaties that protect specific elements of human rights,¹ in many contexts

¹ These may include a) treaties and declarations of international human rights, such as the Universal Declaration of Human Rights of 1948, the International Covenant on Economic, Social and Cultural Rights of 1966, the Convention on the Rights of the Child of 1989 and the United Nations Declaration on the Rights of Indigenous Peoples of 2007; b) the Geneva Conventions on international humanitarian law; c) industry and trade agreements, such as those governed by the World Trade Organisation; and specific rule of law and justice agreements such as those that govern cooperation within Interpol.
the legal frameworks and institutions that are charged with enforcing them are not sufficiently adapted, capable or necessarily invested in upholding those commitments (Neumayer 2005).

**Transplanting traditional Western forms of law and justice institutions to other countries may not be appropriate.** There has been a growing recognition that these efforts are ineffective because they do not consider the justice institutions that already exist in a country, whether formal or informal. They also fail to consider the culture and customs of the country that help to determine whether these systems will be viewed as legitimate or useful (Desai and Woolcock 2015; Department for Stabilisation and Humanitarian Aid 2015). Thus, rigorous RoL evidence from HICs may not be applicable or generalizable to L&MIC contexts. There is a greater need to focus on interventions that work with non-state or informal justice institutions and actors, while also considering the interaction and synergies between them and formal and state institutions and actors (USAID 2019).

**Politics intrude on justice systems.** Literature has underscored the challenge presented to rule of law promotion efforts by entrenched interests and inequality within institutions, among elites and within society (Bellin and Lane 2016). New or reformed justice systems face challenges in equality, independence and fairness. They can mean that only some actors are held accountable to the law, or the law is used to support or oppress only specific members of society. The previous history between society and the government may also have an influence on whether power wielded by the state—in terms of laws and their enforcement—is viewed as legitimate by all members of society. As such, laws and justice institutions may not be trusted or considered credible. This can be a particular problem in countries that are transitioning from authoritarian regimes or those recovering from civil wars (O’Connor 2015). Tools such as ‘Thinking and Working Politically’ through applied political economy analysis can help identify context-specific bottlenecks and openings for interventions to strengthen rule of law and justice (Rocha Menocal et al. 2018).

**Many members of society cannot seek justice or access appropriate services.** Traditionally, efforts to strengthen effective rule of law focused on building the capacity of institutions without considering the role and needs of individuals and groups in society who require legal protection and justice solutions (Golub 2003). Without support to become legally empowered, people may not be able to understand and navigate the justice system in order to claim their rights, solve disputes and redress grievances. They may also not be able to hold justice institutions and actors to account to ensure they are effective, responsive and appropriate to their needs (OECD 2019). There are, however, challenges to ensuring access to justice for all people due to the difference in the needs and capacities of different people in society (HiiL 2018). For example, some individuals and groups may lack the necessary legal identity required in a country to access justice services, while others may not have access to the money, time and education required to understand and use the system. Alternatively, some people may be located in remote places where services are not provided, or they may not speak the languages used in formal proceedings. There may also be groups of individuals who do not receive equal protection and treatment in society and are unlikely to trust justice systems and services to treat them fairly, which in turn undermines the legitimacy of those institutions (Department for Stabilisation and Humanitarian Aid 2015). This presents challenges for
intervention design; the myriad justice needs require equally myriad approaches, and thus a process of needs identification and prioritization is required, and no single intervention is able to meet the justice needs of everyone in the target area. These diverse constraints across the globe mean that the legal and justice needs of many people are not currently met.

1.1.2 Why is this map important for policy?
In response to these challenges, there has been increased focus among the international community to focus on strengthening the rule of law and providing justice to all. This has been underscored by the UN SDG 16 commitment to promoting just, peaceful and inclusive societies (UN General Assembly 2015). Programming in this area focuses on a broad range of interventions, including professional development and oversight of actors in the formal justice system, strengthening access to justice services by improving coverage and appropriateness of these services, and building the legal literacy and empowerment of excluded and vulnerable groups to access and claim justice (Goodwin 2017). Increasingly, policy and programming are focusing on a people-centered justice approach to rule of law promotion. This has included research to understand the legal problems and justice needs challenges faced by people in different countries and providing interventions to address these problems and meet their justice needs and to ensure that justice institutions and services are accountable to justice seekers (HiIL 2018; OECD 2019).

The U.S. State Department and USAID support democracy programs around the world, including programs to strengthen the rule of law. They also allocate resources to support countries to develop their criminal justice systems and capabilities with the aim of reducing international crime (U.S. Department of State 2020). USAID has been a long-term supporter of rule of law assistance, including launching Administration of Justice programs in Latin America in the 1980s to strengthen the stature of the judiciary in emerging democracies in the region (Blair and Hansen 1994). Starting in the 1990s, its work expanded to Asia, Africa and Eastern Europe (ibid.). The Agency currently supports rule of law programming as a cornerstone of its Democracy, Rights and Governance Strategy (USAID 2013). The strategy supports work towards multiple objectives related to rule of law, first and foremost, to “foster greater accountability of institutions and leaders to citizens and to the law.” (USAID 2013, USAID 2014). It also recognizes rule of law as part of the objective to “improve development outcomes through the integration of DRG principles and practices across USAID’s development portfolio.”

Today, USAID funds rule of law strengthening activities in dozens of countries around the world, including in areas of the administration of justice, access to justice and legal empowerment, legal profession and education reform, and citizen securing and crime and violence prevention. Funded activities can range in size, duration, and scope from US$ 102 million over 2-3 years to upwards of US$ 23 million over 4-5 years in some countries. The World Bank has also been supporting rule of law programming over a long period. Over the past 25 years it has funded over 800 justice and development projects around the world. The World Bank was also early to recognize that rule of law programming should be responsive to the needs of society. As such, their Justice for the Poor program began over a decade ago with the aim of promoting legitimate and
equitable systems for managing disputes, particularly over land, natural resources and public spending.

Several other international development donors are actively working towards strengthening the rule of law in LMICs by dedicating resources for interventions in this sector. For example, through the Somalia Security and Justice Program, FCDO has allocated £34.7 million to strengthen security and justice institutions by providing training to police and equipping justice agencies with basic infrastructure. Similar programs funded by FCDO have been undertaken in Nigeria, Kenya, Malawi and Tanzania to build regional capability and support development of law enforcement institutions. Additionally, the World Justice Project and the Global Rights Grant provided by the Bill and Melinda Gates Foundation have also collectively contributed US$2.35 million for promotion of rule of law and increased access to justice for poor and marginalized populations in the developing world.

Rule of law promotion has been and continues to be an important component of programming and funding undertaken by many multilateral and bilateral donor agencies and charitable foundations. However, at the same time, donor support for strengthening the rule of law around the world has fallen by 40 percent in the past four years (Manuel et al. 2019). Compared to other key sectors, there are particularly large funding gaps in support for the justice sector: over the past decade, only 1.8% of total donor-funded aid from OECD DAC donors has been allocated to justice system support, compared to 13% for health and 8% for education (ibid.). The overall level of donor support is also limited to a few donors and is heavily skewed to a handful of recipient countries (OECD 2019). Facilitating access to and use of rigorous evidence and data can help demonstrate the value of investments in rule of law assistance and inform decisions about how that funding can be most effectively spent.

The policy response described above indicates the range and scale of interventions implemented around the world to strengthen the rule of law and improve access to justice. There is, however, a need to better understand the RoL evidence landscape. A fundamental step in informing policy decisions is understanding what current evidence exists, where interventions are implemented, and where there are clusters and gaps in that evidence. This type of mapping facilitates the use of evidence from individual studies by making them more accessible to policy makers and practitioners. This in turn may result in increased evidence-informed rule of law-related policy making. The EGM may also inform future research investments, as it identifies primary evidence gaps (where there are no or very few impact evaluations) and synthesis gaps (clusters of impact evaluations, with no recent or high-quality systematic reviews) with respect to specific interventions and outcomes. By understanding the current state of the evidence base, research commissioners can avoid duplication of efforts and prioritize areas where limited to no rigorous evidence has been carried out to date.

Existing mapping initiatives capture some aspects of rule of law programming, but important gaps remain. For example, 3ie produced an EGM of state-society relations that included some studies related to justice institutions but focused more broadly on a range of state institutions and public services (Cameron et al. 2015). This map also only covered evidence from 2000 to 2015 and focused only on L&MICs. It therefore does not capture recent studies or potentially relevant learning on justice challenges in HICs. 3ie
also published an EGM on interventions aiming to build peaceful societies in fragile contexts in April 2020 (Sonnenfeld et al. 2020). Again, the map examined some interventions related to rule of law, including dispute mechanism processes and policing. This map, however, was restricted to studies implemented in fragile contexts in L&MICs.

There are multiple studies that aim to help make sense of the evidence around rule of law programming. These studies typically do not draw on comprehensive search and systematic screening processes, however, and thus focus on a non-comprehensive subset of the evidence. For example, some studies are limited to evidence from a single organization or database, yet the smaller scope enables them to dive more deeply into the findings, and they are thus useful resources. Nolan and colleagues (2019) summarized lessons from selected experimental or quasi-experimental studies on reducing crime, violence, and conflict in L&MICs. They further identify a limited set of systematic reviews of police approaches in largely HIC contexts, but the review was not an exhaustive search of studies about police capacity and other rule of law topics. FCDO summarized evidence from 215 studies of security sector reforms and organizational capacity building for delivering citizen security in L&MICs (Denney and Valters 2015). This non-exhaustive review was carried out in 2015 and drew primarily on studies within a single database from the Governance and Social Development Resource Centre (GSDRC) (Denney and Valters 2015).

There are also studies that aim to map a broader set of evidence beyond studies of effectiveness. In 2019, a rapid evidence mapping process was undertaken that searched for a wide range of evidence types in the security and justice sector for interventions implemented in L&MICs. The research identified numerous observational studies, yet the rapid search returned only fifteen impact evaluations (Jackson et al. 2019). A useful mapping process was also conducted on legal empowerment by Goodwin and Maru (2017). They examined a broad range of evidence, including qualitative interviews, using multiple data gathering techniques that included key informant interviews, scoping existing literature reviews and grey literature searching. Their search for studies, however, concluded in 2013. Their work represents a useful examination of one of the key elements of rule of law programming. Finally, a title was registered for an evidence map with the Campbell Collaboration’s Crime and Justice Coordinating Group in 2018, but the project has not yet been completed (White et al. 2018).

Our EGM complements this work by adopting a comprehensive search strategy that incorporates both grey literature sources and multiple academic databases, and by updating the work to identify new evidence. As this EGM focuses specifically on research about program effects, these complementary sources are useful for research that answers different questions, such as evidence around barriers and facilitators to programming. In undertaking our search for relevant studies, we reviewed existing mapping efforts to identify potentially relevant impact evaluations and systematic reviews. Our RoL EGM provides an updated picture of the evidence base on intervention effects, covering a comprehensive set of rule of law and justice interventions and expanding the scope beyond L&MICs to capture the global evidence base.
1.2 Study objective and questions

The aim of this EGM is to improve the understanding of and access to evidence on the effects of rule of law interventions on justice outcomes among relevant policy, research and implementation stakeholders. This EGM identifies, describes and consolidates the available evidence in a clear and structured way. This in turn may facilitate the use of evidence to inform research, policy and program design decisions.

The specific objectives of this EGM are twofold:
- Identify, describe and summarize the evidence on the effects of rule of law interventions on justice outcomes around the world
- Identify potential primary evidence and synthesis gaps

To meet these objectives, we addressed the research questions shown in Table 1 below.

Table 1: EGM research questions

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ1 What is the extent and what are the characteristics of empirical evidence on</td>
<td>Coverage</td>
</tr>
<tr>
<td>the effectiveness of selected rule of law interventions on justice outcomes around</td>
<td></td>
</tr>
<tr>
<td>the world?</td>
<td></td>
</tr>
<tr>
<td>RQ2 What are the major primary and synthesis evidence gaps in the literature?</td>
<td>Gaps</td>
</tr>
<tr>
<td>RQ3 What intervention/outcome areas could be prioritized for primary research and</td>
<td>Research</td>
</tr>
<tr>
<td>or evidence synthesis?</td>
<td>needs</td>
</tr>
</tbody>
</table>

1.3 Structure of the remainder of this report

The remainder of this report is structured as follows:
- Section 3 defines key terms and describes the conceptual framework adopted for the EGM and the types of studies included;
- Section 4 describes the methods applied in the systematic search, screening, data extraction and analysis of the identified studies;
- Section 5 presents the results; and
- Section 6 outlines implications for policymakers, programmers and researchers, and concludes the report.

2. Scope

2.1 Key concepts and definitions

We developed a conceptual framework for rule of law interventions that encompasses both (i) programming supporting formal state institutions, traditionally the more common focus, and (ii) programming focused on informal/non-state institutions and the role that people in society play in supporting, accessing and providing justice.

Figure 1, below, depicts the conceptual framework. It outlines a stylized theory of change with effective rule of law depicted in the center. In this conceptualization, effective rule of law relies on the functioning of three different domains – systems, services and
societies\textsuperscript{2} – and the existence of supportive interactions and well-functioning feedback loops between them. These interactions and feedback loops (shown in red) serve to underscore the fact that creating effective rule of law is a dynamic, evolving and ongoing process that requires adaptation and responsiveness to the changing needs of society and the specific context in which it operates.

**Figure 1: Conceptual framework for the EGM**

![Conceptual framework for the EGM](image)

Source: Authors

Below, we define the three domains in depth.

**Systems:** Systems in this conceptualization represent both formal and informal institutions, actors, frameworks and processes related to law and justice. They are the basis for providing effective rule of law for a given context and are the infrastructural foundations for producing legal and justice services for society. For example, the systems domain includes laws and policies that govern the functioning of society. It also includes the internal functioning of the justice sector. Importantly, these systems refer not only to the formal/state laws and institutions but also to informal/non-state laws and institutions. Some examples of informal systems include customary law and traditional authorities.

The central illustration in Figure 1 depicts systems on the left of the Venn diagram in blue. The top circle displays the formal/state systems and the bottom circle displays the informal/non-state systems. The overlap between them shows areas where they work together as hybrid systems. The red arrows leading from one systems circle to another depicts effective interactions between formal and informal systems that originate from both sides. The red arrow leading from the systems to the yellow society circle indicates the responsiveness of the systems to the needs of justice seekers. More responsive systems should help to protect and to provide appropriate services to all of society, including those who are typically underserved and previously legally estranged.

**Services:** Services are points of interaction where formal and informal legal and justice institutions and actors come into contact with members of society to deliver legal support, protection or to uphold the law. Some examples of services

---

\textsuperscript{2} The suggestion to focus on a “Systems, Services and Society” framework was provided by Rule of Law advisors from USAID.
include policing, courts, correctional facilities and legal advisory services. These services may be provided by formal or informal institutions and actors. For example, legal advice centers could be staffed by lawyers or paralegals. The services may also be provided by a mixture of state and non-state actors, which, similar to the discussion above with systems, could create integrated or hybrid services.

The central illustration in Figure 1 depicts services in the vertical overlapping segments of the Venn diagram in green. They show where the blue systems circles and the yellow society circle come together. The top segment displays the formal/state services and the bottom segment displays the informal/non-state services. The overlap between them shows areas where these services integrate and operate as hybrid services. The red arrows that depict interactions between the systems and society circles should help to improve these services and ensure they are responsive to the needs of society.

**Society:** Society encompasses all the people, private entities and non-government organizations within a particular context. For example, society contains heterogeneous individuals and groups with different capacities and needs but it also contains businesses, religious institutions, media and civil society organizations. As such, this is a diverse sphere.

The central illustration in Figure 1 depicts a legally empowered society in the yellow circle on the right of the diagram. The red arrows leading to and from the society circle to the formal and informal systems circles indicate interaction, feedback, and accountability occurring between the systems and society. These are important to ensure that the systems and the services that they support are responsive and appropriate to the needs of all members of people in society.

### 2.2 EGM theoretical framework

The above discussion focuses on the central illustration of effective rule of law in Figure 1. We now describe the theory of change that can lead to the creation of effective rule of law and also the outcomes produced by effective rule of law. This is represented by the grey boxes and arrows that move horizontally from the left of Figure 1 to the right.

Interventions that focus on strengthening the three domains described above and the interactions between them are depicted on the far left of the diagram. They are inputs into the system. When these interventions are designed and implemented effectively, they should lead to changes in intermediate outcomes that support the effective functioning and interaction between the three domains.

#### 2.2.1 Interventions

Below, we provide examples of interventions to strengthen the three domains and to ensure effective interaction between them. The list of interventions below is not exhaustive; included interventions are detailed further in Section IV.B.2, and the complete list of included interventions is defined in Appendix A.1.

Interventions focused on systems may work to make laws fairer and more equitable or help them to conform to international human rights standards. Interventions may also
seek to build institutional and human capacity or provide oversight and accountability to these systems on both the formal and informal side. In addition to working specifically on either side of the formal-informal division, interventions may also seek to improve the integration of these two types of systems to create a more effective hybrid system. Interventions may also support linkages between them to ensure successful handling of justice problems that relate to both systems or need to be referred from one to the other. Indeed, while we refer to the ‘formal-informal division,’ we recognize that the boundaries between them are porous and the norms that support them may be intertwined.

Interventions focused on improving **services** may seek to strengthen coverage of and access to existing services or create new quality services that better meet and respond to the needs of a diverse society. Some examples include new or adapted techniques to tackle crime, to protect the vulnerable, to solve legal disputes or to integrate ex-criminals or victims back into society. Interventions in this area may also help to create more choice between the services that people can access to solve their legal problems. For example, while formal state services may be appropriate to solve the justice needs of some people in society for one specific problem, informal/non-state services may be more suitable to serve other members of the population or to address other specific legal problems and justice needs.

Interventions focused on **society** tend to have an overarching goal of ensuring society is legally empowered. Legal empowerment refers to “the use of legal services and related development activities to increase disadvantaged populations’ control over their lives” (Golub 2003, 3). Programming in this area is typically focused on working through civil society but may also engage with justice institutions (ibid). Some examples of interventions focused on society include increasing people’s legal literacy through public education and materials that target specific groups, supporting people to hold justice services to account, strengthening society-led protection and reporting mechanisms and supporting people to use legal techniques to solve their problems and to increase access to quality services.

### 2.2.2 Outcomes

**Intermediate outcomes** are the initial changes that are intended to come about as a result of these inputs. For example, as a result of those interventions there would ideally be changes in terms of knowledge and behaviors of actors as new ideas and processes are adopted. There may also be changes in the coverage and accessibility of services at this intermediate level, in addition to improved integration of formal and informal systems. These intermediate outcomes are an essential step in the causal chain to create the type of effective rule of law depicted in the center of the image where all three domains are working effectively and the interactions between them are functioning well.

If effective rule of law is achieved as a result of shift in these intermediate outcomes, there should be changes in primary outcomes. These outcomes are shown immediately to the right of the central rule of law picture in Figure 1. For the rule of law sector, **primary outcomes** are related to improved prevention and increased satisfactory solving of crime.

---

3 These are examples of outcomes. A more comprehensive list is provided in the methodology section of this protocol.
and justice problems. Some examples of primary outcomes include reduction in crime and violence rates, increased social integration of ex-criminals and improved rates of resolution of justice problems with satisfactory processes and outcomes.

If changes in these primary outcomes are achieved and there are improvements in both the prevention of crime and the solving of justice problems, then there should be positive improvements in final outcomes. These are depicted in the far right of Figure 1 and represent the end of the causal chain. Final outcomes are related to broader socio-economic changes. For example, final outcomes include improvements to economic growth and development, security, health and education. These contributions are theorized to occur because the creation of clear rules and more legal certainty may lead to an increased willingness to invest and trust in doing business; creation of effective and inclusive ways of solving disputes peacefully may help to reduce instability; creation of greater freedom from violence should support people to live healthy and safe lives or enable children to better learn in schools; and creation of improved access to quality services through reduced corruption may also lead to better health and education outcomes.

It should be noted that generally, effective rule of law will only contribute to these final outcomes rather than ensure them. There are likely other elements that are also required to support sustainable changes in socio-economic outcomes. For example, effective rule of law can contribute to security in a country but achieving it may difficult if the context is subject to other factors such as external terrorism, economic shocks, or climate change. Similarly, while effective rule of law may promote greater access to education services, sector specific interventions are likely to be required to improve quality in a sustainable way.

While Figure 1 depicts the three domains and the causal chain that are described above, there are two additional factors that should be considered when interpreting the conceptual framework. The first is related to the principles required for effective rule of law programming and the second is related to assumptions, tensions and non-linearity in the process.

**Principles of effective rule of law programming:** In the above discussion, we refer to interventions being “designed and implemented effectively.” It is important to outline what we mean by that phrase. The interventions that act as inputs to the conceptual framework, whether they be policies, projects or programs, should be designed and implemented with some core principles. The first five principles are drawn from the USAID and UN definition of effective rule of law. The next three principles are adopted from the literature on legal empowerment and people-centered justice (Golub 2003; HiiL 2018; OECD 2019). The final two principles are based on literature on the need to move away from isomorphic mimicry (Carothers 2009; Desai and Woolcock 2015) and suggestions from our expert advisory group that rule of law programs need to be both contextually locally appropriate and locally owned. To create the effective rule of law model in the above diagram, interventions should ideally be designed and implemented to foster the following principles of the rule of law and legal empowerment, as defined in Table 2 below.

---

4 The division between these two types of outcomes is based upon work by The Task Force on Justice.

5 Again, the list of outcomes is not exhaustive. A more comprehensive list of primary outcomes can be found in the Appendix A of detailed methodology of this report.
Table 2: Key concept definitions

<table>
<thead>
<tr>
<th>Principle</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equality</td>
<td>All people and entities should be treated equally before the law</td>
</tr>
<tr>
<td>Consistency</td>
<td>The law should be consistently enforced and applied</td>
</tr>
<tr>
<td>Independence</td>
<td>Actors in the justice system should be independent from improper or private influence</td>
</tr>
<tr>
<td>Prevention</td>
<td>Justice systems should prevent injustice, insecurity and rights abuses.</td>
</tr>
<tr>
<td>Transparency</td>
<td>Laws, processes and relevant information should be public, clear and accessible</td>
</tr>
<tr>
<td>Human rights</td>
<td>Rule of law should uphold and protect international human rights norms and standards</td>
</tr>
<tr>
<td>Responsiveness</td>
<td>Justice systems and services should be responsive to the needs of society and evolving contexts</td>
</tr>
<tr>
<td>Accountability</td>
<td>Justice systems, services and actors should be held responsible for executing their roles to a set standard.</td>
</tr>
<tr>
<td>Inclusivity</td>
<td>All people, especially marginalized individuals and groups, should be legally empowered to access and use appropriate justice systems and services</td>
</tr>
<tr>
<td>Contextual appropriateness</td>
<td>Rule of law innovations should be designed and implemented in ways that take the legal tradition and local contextual factors, such as institutions and culture, into account</td>
</tr>
<tr>
<td>Local ownership</td>
<td>Local actors should be involved in designing and implementing changes in justice systems and services</td>
</tr>
</tbody>
</table>

Non-linearity, tensions and assumptions

While Figure 1 depicts a fairly linear process of movement from inputs to final outcomes, this is for the sake of visual simplicity rather than a reflection of how rule of law interventions are likely to influence change in reality. The arrows that indicate interactions between the different domains serve as a reminder that effective rule of law is a constant process of feedback and redefining needs and solutions.

Rule of law reform is inherently political. As such, it relies on interactions with powerful elite actors with potentially entrenched interests and heterogeneous members of society for which changes may have varied levels of acceptability (Bellin 2016; O’Connor 2015; Department for Stabilisation and Humanitarian Aid 2015). This can lead to shifts between improvements and lapses over time rather than a simple movement in one direction. This is because expectations, resistance and support for changes may alter over time for different actors.

Additionally, as portrayed in the diagram, under the ‘systems’ domain there are competing and interacting forms of institutions, actors and norms that relate to differences in international laws, formal state law and the informal or non-state customary law. There can be tensions between these different ‘systems’ and also between the systems and programming principles that can make the process of reforming law complicated. For example, while the rule of law programming principles suggest that equality is an important component of effective interventions, this may be a difficult principle to enact when working with the unequal and hierarchical power
structures that can be the basis of some non-state traditional justice institutions. These various factors underscore the need for a strong understanding of individual contextual situations of specific countries to be able to determine which interventions are likely to operate effectively and what effective rule of law should look like in each place.

**Rule of law and justice institutions are shaped by the political and legal culture in which they are situated.** In addition to the tensions and non-linearities within the rule of law framework, there are also several factors that can support or hinder rule of law reforms. For example, without a sufficient level of openness to democratic principles it may be difficult for actors to pursue access to fair and accountable justice. Additionally, when justice institutions are not independent from other sectors of the state, rule of law interventions alone may be insufficient to ensure justice actors will behave consistently with respect to upholding the law for different individuals. As such, other supportive contextual factors and other complementary interventions not directly related to rule of law may be important for ensuring that effective rule of law can function. As the underlying drivers of poor rule of law vary greatly across contexts, interventions targeting these bottlenecks and aiming to strengthen the conditions for good rule of law will equally be highly context dependent. For example, in a situation where frontline justice sector actors struggle to support their families, they may rely on bribes even though they know the practice to be illegal and unethical. In such a situation, long-term support for broader economic policies may be required such that through economic growth, the government can increase tax revenue and afford to properly pay civil servants. Indeed, it is important to keep in mind that all rule of law programming is implemented in contexts that fall short of the maximalist definition and principles that we have described above.

3. Methods

3.1 Overall methodological approach

EGMs are tools to help policymakers and researchers working in a sector or thematic area make evidence-informed decisions. They make existing evidence more accessible and ease the prioritization of future research by mapping existing studies in a field on a framework of interventions and outcomes. We followed the standards and methods for EGMs developed by 3ie (Snilstveit et al. 2016; Snilstveit et al. 2017).

The map is populated by systematically searching and screening all relevant completed, and ongoing, impact evaluations and systematic reviews. An impact evaluation measures the effects on targeted outcomes that can be attributed to a particular program or intervention; systematic reviews extract and synthesize data from multiple impact evaluations of similar interventions to generate more robust conclusions about their effectiveness than could be provided by a single study. Using 3ie’s EGM software, we created an online, interactive matrix that maps all included studies according to the interventions evaluated and the outcomes reported. This provides a visual display of the volume of evidence for intervention-outcome combination, the type of evidence (impact evaluation, systematic reviews, completed or ongoing), and a confidence rating of the quality for systematic reviews. The platform provides additional filters so that users can further explore the available evidence, for example by global regions, income levels, or population. The EGM can be viewed at https://developmentevidence.3ieimpact.org/egm/rule-of-law-evidence-gap-map.
This report serves as an accompaniment to the interactive map. In this report, we address the key research questions through analysis of the characteristics of the available evidence and key trends (i.e., number of impact evaluation published over the time, geography, focus on interventions and outcomes, targeted audiences).

Evidence gap maps highlight both primary evidence gaps, which should be filled with new impact evaluation studies, and synthesis gaps, wherein a cluster of impact evaluations are ready for new systematic reviews and meta-analyses. EGMs are envisioned as a global public good, and this allows them to be used as a tool which facilitates access to high-quality research.

3.2 Criteria for including and excluding studies in the EGM

Here we summarize the detailed criteria for including and excluding studies according to relevant populations, interventions, comparators, outcomes and study designs (PICOS). In defining the scope of relevant interventions and outcomes, our aim was to be as comprehensive as possible whilst setting a feasible scope that was not too broad to present in a clear and interpretable manner. The complete criteria are reported in Appendix A.1.

Table 3: Summary PICOS criteria for studies to be included in the RoL EGM

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population</td>
<td>We included studies targeting any population type, implemented in any country.</td>
</tr>
<tr>
<td>Interventions</td>
<td>We included interventions that directly intervened to improve aspects of access to and execution of rule of law and justice. We developed a typology of intervention approaches organized into three primary domains: systems, services and society. The intervention categories within each domain are designed to identify the complete range of interventions that specifically aim to strengthen rule of law and access to justice. In the detailed methodology described in Appendix A.2, Table 6 shows the specific intervention groups covered within each of the three core domains described above that are included in the EGM. We included studies that evaluated the effectiveness of at least one intervention included in Table 6. In some cases, studies may evaluate multi-component interventions, wherein at least one but not all of the components correspond to one of the interventions listed in Table 6. We included interventions of this type if at least one of the subcomponents evaluated a rule of law or justice intervention.</td>
</tr>
<tr>
<td>Outcomes</td>
<td>The primary outcomes of interest were justice outcomes that relate to changes in both preventing and solving justice problems. We also included studies that reported effects on intermediate outcomes, including those related to attitudes, behaviors and coverage of justice institutions, services and actors. We included studies that reported at least one primary or intermediate outcome listed and briefly defined in Section III.B.2 above. The complete outcome list is in Table 5 in Appendix A.1 of the</td>
</tr>
<tr>
<td>Criteria</td>
<td>Definition</td>
</tr>
<tr>
<td>---------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
|                | detailed methodology.  

**Study designs**  
We included impact evaluations and systematic reviews that measured the effects of a relevant intervention on outcomes of interest. For impact evaluations, we included counterfactual studies that used an experimental, quasi-experimental or qualitative design and/or analysis method to measure the net change in outcomes that were attributed to an intervention or policy. We included randomized and non-randomized studies that were able to take into account confounding and selection bias. For systematic reviews, we included effectiveness reviews that synthesized the effects of an intervention on outcomes of interest. We excluded reviews that only described programmatic approaches or synthesized findings on barriers and facilitators to implementation.

**Language**  
Studies published in any language were included, although the search terms used were in English only.

**Publication date**  
Studies were included if their publication date was 1990 or after.

**Status of studies**  
We included ongoing and completed impact evaluations and systematic reviews. For on-going studies, we included prospective study records, protocols and trial registries. Providing an indication of the prevalence and characteristics of on-going evaluation evidence is expected to enrich the analysis of current evidence gaps and support decision making in relation to evidence generation.

---

6 This table includes final outcomes that relate to ultimate changes that may come about due to improvements in rule of law and justice, including security, economic development and health outcomes, are included on the EGM matrix. This enabled us to extract data on all outcomes reported by included studies, in order to show the full range of the evidence base for rule of law interventions. However, these were not considered sufficient for inclusion; for example, studies that only reported effects on health outcomes were excluded.

7 Qualitative research designs were not initially included in the scope of the EGM. We conducted a second search for qualitative impact evaluations of RoL interventions that establish a plausible counterfactual after the first version of the map was published in March 2021. The second search followed the same steps outlined in Section IV.C. Additional details on the second search, including the full search strategy, are provided in Appendix A.
3.3 Search, screening and data extraction

3.3.1 Search strategy
This project implemented a sensitive search strategy\(^8\) primarily constructed by a combination of intervention and study design terms. The strategy was developed by an information specialist and an example of the strategy developed for the Social Sciences Citation Index (SSCI) is provided in Appendix A.3.e. The strategy was translated\(^9\) according to the requirements and functionalities of different databases. The search for evidence was conducted using a range of different sources of academic and grey literature, including bibliographic databases (a combination of general social science and environmental-focused databases), repositories of impact evaluations and systematic reviews, specialist organizational databases and websites of bilateral and multilateral agencies. We further conducted forward citation searches of all included studies to identify further potentially relevant impact evaluations and systematic reviews. Where possible, the review team contacted key experts and organizations through our review advisory group (presented in Appendix A.9) and published a blog post soliciting inputs of relevant studies to identify additional studies that meet the inclusion criteria. A full list of sources searched and the detailed process followed can be found in Appendix A.3.

3.3.2 Screening
The selection of studies for data extraction as part of the review was managed using EPPI-Reviewer 4 software (Thomas et al. 2020). Studies were imported into EPPI-Reviewer and, following the removal of duplicates, the titles and abstracts were screened in duplicate by two team members. We utilized EPPI-Reviewer’s machine learning capabilities to streamline the process and efficiently remove clearly irrelevant studies. The full texts of studies that appeared to relate to an impact evaluation or systematic review of a relevant intervention were then screened by two reviewers against the inclusion criteria outlined above. A full list and details of each step can be found in Appendix A.4.

3.3.3 Data Extraction and critical appraisal
We systematically extracted data from all included studies using the data extraction tools available in Appendix B. We converted the tools into XLSForms for use in KoBo Toolbox, which is a useful software for facilitating and quality-assuring data extraction. The data covered the following broad areas:

- **Basic study and publication information**: This coding focused on capturing the general characteristics of the study, including authors, publication date and status, study location, intervention type, outcomes reported, definition of outcome measures, population of interest, study and program funders, time periods for delivery and analysis.

- **Topical cross-cutting issues**: We extracted data on a number of cross-cutting issues, including gender, equity and cost-effectiveness.

---

\(^8\) Sensitive search strategy: sensitive it here synonym of comprehensiveness in relation to the types of studies that can be captured in a search strategy. An increase of sensitivity of a search will reduce its precision and will retrieve more non-relevant articles (Higgins et al 2011).

\(^9\) The search strategy run in different databases make us of strings of key words, often truncated and wildcards variations of the same terms, linked between them with Boolean operators (AND, OR, NOT or proximity operator (N3, N5 etc.). These operators are different for each database so they need to be 'translated'.
Critical appraisal: We critically appraised the methods undertaken in all completed systematic reviews. The critical appraisal assesses the review with regards to how the search, screening, data extraction and synthesis were conducted, and covers all of the most common areas where biases in the study design and analysis are introduced (3ie n.d.). Based on the appraisal, each review was rated as high, medium or low confidence, indicating the level of confidence we have in the findings of the review based on the methods the authors used. A review classified as high-confidence means that the methods undertaken in the review were in line with best practices. These capture the core function of a systematic review of intervention effects as a methodology: that the search process was sufficient to identify all potentially relevant studies, bias was avoided in the selection of studies, and appropriate methods were applied to assess risks of bias in included impact evaluations and synthesize the findings on effects. We extracted the findings of the reviews rated as high confidence. Reviews rated as low or medium confidence may identify useful evidence, and/or have useful descriptive information or framing of an issue, but as we do not have high confidence in the methods undertaken for one of the steps noted above, we do not extract and present the findings.

Due to the size of the evidence base, a streamlined appraisal process was undertaken. First, all systematic reviews published through the Cochrane or Campbell Collaboration were marked as high-confidence, given the peer-reviewed process on which they are published and the alignment between 3ie’s critical appraisal tool and best practices adhered to within these publications. However, as 3ie’s tool is more stringent than the MECCIR checklist requirements for Campbell Collaboration reviews (Methods Group of the Campbell Collaboration 2019), there is a risk that this decision may overstate the confidence of some SRs.

For all SRs that were not Campbell or Cochrane reviews, a rapid appraisal was undertaken to screen each additional systematic review for basic methods quality. Reviews that could quickly be identified as failing a key criterion of the appraisal tool were marked as low confidence, while all others were put forward for the full critical appraisal. Both the rapid and full appraisal were carried out using a standardized checklist (3ie, n.d.). One reviewer conducted the initial critical appraisal, and an SR methods expert conducted a final review of all appraisals. For further details on the critical appraisal process, see Appendix B.3. Due to the size of the IE evidence base, we did not critically appraise included IEs.

3.4 Presentation of the map

We present the results graphically on an interactive online platform. The main framework is a matrix of interventions and outcomes, with grey and colored circles representing IEs and SRs. The SRs follow a traffic-light system to indicate confidence in their findings: green for high, orange for medium, red for low and blue for protocols. The size of the bubble indicates the relative size of the evidence base for that intersection of intervention and outcome. The bubbles within each box of the matrix represent studies reporting effects for that intervention/outcome configuration. Clicking on any bubble will display a list of the studies with hyperlinks to the full text.10

10 Where possible, we have linked to the full text directly; however, for studies behind paywalls, the hyperlink goes to the study landing page that typically, at a minimum, provide the abstract and references.
The interactive aspect of the EGM allows users to filter the results based on key variables, thereby facilitating efficient, user-friendly identification of relevant evidence. The filters and their definitions are provided in Table 4.

Table 4: Definition of EGM filters

<table>
<thead>
<tr>
<th>Filter</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region</td>
<td>This filter identifies studies according to the geographic region in which the interventions were implemented, using the regions as defined by the World Bank.</td>
</tr>
<tr>
<td>Country</td>
<td>This filter allows users to identify the evidence base from a specific country.</td>
</tr>
<tr>
<td>Income Level</td>
<td>This filter allows users to identify the evidence base from a particular country income group, as classified by the World Bank, and to identify evidence from LICs, LMICS or MICs. The income level is based on the status of the country in the first year of intervention, or if not available then the publication year.</td>
</tr>
<tr>
<td>Electoral Democracy</td>
<td>This filter allows users to identify evidence base from a particular country electoral democracy categorization. It uses categories from the V-Dem Electoral Democracy Index ordinal (D) (e_v2x_polyarchy) based on the status of the country in the first year of intervention, or if not available then the publication year.</td>
</tr>
<tr>
<td>Fragility, Conflict and Violence (FCV) Status</td>
<td>This filter allows users to identify the evidence base from countries that are affected by fragility and conflict, as defined by the World Bank’s list of fragile and conflict-affected situations from 2006-2021. It is based on the status of the country in the first year of intervention, or if not available then the publication year.</td>
</tr>
<tr>
<td>Population</td>
<td>This filter enables users to identify studies that contain specific results for a range of key population groups: LGBTQI+5 sexual and gender minorities; ethnic, racial, caste-based, and religious groups; survivors of large scale violence/displacement (includes refugees and internally displaced populations); survivors of gender-based violence; survivors of trafficking; people living with disabilities and chronic health conditions; people with substance use issues; incarcerated people and those re-entering society; sex workers; and dissidents. In case this information were not explicitly specified, the ‘unspecified’ option was chosen; when it was said population of any ethnic group, caste, religious group were included, the option ‘ethnic, racial, caste-based, religious groups-whole population’ was chosen, the same for LGBTQI+-whole population.</td>
</tr>
<tr>
<td>Age</td>
<td>Children, adolescents, youth, adults, older adults, whole population (in case there were no restrictions on the age of the participants)</td>
</tr>
<tr>
<td>Sex</td>
<td>Female, male, whole population</td>
</tr>
<tr>
<td>Setting</td>
<td>Urban, peri-urban, whole population</td>
</tr>
<tr>
<td>Study Design</td>
<td>This filter enables users to identify studies that employed a particular study design, using the list of study designs in Appendix A.1.e.</td>
</tr>
<tr>
<td>Cost Evidence</td>
<td>This filter enables users to identify studies that incorporated cost evidence into their analysis.</td>
</tr>
<tr>
<td>Filter</td>
<td>Definition</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Human Rights</td>
<td>This filter enables users to identify studies that target specific human rights. All 21 human rights outlined in the scope section are included.</td>
</tr>
<tr>
<td>Theme</td>
<td>This filter enables users to identify studies that are included in EGMs from other DRG Center Program Areas: Rule of Law, Civil Society, Independent Media, Governance, Political competition and consensus building.</td>
</tr>
</tbody>
</table>

3.5 Analysis and reporting

To answer Research Question 1 regarding the extent and characteristics of the evidence base, data were extracted on the dates, intervention(s) studied, outcomes reported and population coverage, including regions, countries and specific population groups. For high confidence SRs, we further extracted summaries of the key findings for policy implications. To answer Research Question 2 regarding gaps in the evidence, we assessed the distribution of studies across the included interventions and outcomes. We differentiated between primary evidence gaps, where no IEs exist, and synthesis gaps, where no up-to-date or high confidence SRs exist despite a cluster of IE evidence. Finally, to answer Research Question 3, we shared the draft findings with stakeholders at USAID and the Advisory Group and solicited input regarding policymaker and practitioner priorities for future research.

4. Findings

4.1 Volume of evidence

As the PRISMA diagram (Figure 2) shows, the systematic search process returned 52,324 records, with 41,273 records remaining for screening after duplicates were removed. We used a combination of machine learning and manual screening at the title and abstract stage to identify 1,630 studies that looked possibly relevant for inclusion, for which we then screened the full texts. Of these, we included 637 completed and 17 ongoing IEs and 107 completed and 11 ongoing SRs. Searches of academic databases for quantitative research designs were completed in July 2020, and reference checks and grey literature searches were concluded in October 2020. We conducted a second search for qualitative research designs in March 2021.
4.2 Characteristics of the evidence base

4.2.1 Publication over time

Few rule of law studies that met our criteria were published in the 1990s, but the number of rule of law publications increased steadily from about 2003 onwards. Most of the included rule of law studies evaluated interventions in the services domain. Included studies related to the systems and society domains were more frequently published from 2009 onwards. As Figure 3 shows, publications of rule of law impact evaluations in
L&MICs increased steadily from about 2009 to 2020. We included data only up through 2019, as it was the last year for which complete data on publications is available.¹¹

Almost all the systematic reviews focus on interventions in the services domain (n = 91, with only a small number focused on society and systems interventions (n = 2, n = 7, respectively).

**Figure 3: Count of RoL impact evaluations published over time, by country income group**

4.2.2 Population

Most rule of law impact evaluations included in our map (n = 554, 85 percent) took place in high-income contexts (Figure 4). Less than a fifth (n = 101) evaluated interventions in L&MICs, with 39 impact evaluations in upper-middle income countries, 36 impact evaluations lower-middle income countries, and 28 impact evaluations in low-income countries.

Included studies evaluate interventions implemented in North America (n = 479 IEs, n = 64 SRs), followed by Europe and Central Asia (n = 52 IEs, n = 39 SRs).¹² Among predominantly-L&MIC regions, Sub-Saharan Africa has the highest count of included IEs (n = 42), though only six SRs include studies from the region. This is followed by Latin America and the Caribbean (n = 37 IEs, n = 11 SRs); East Asia and the Pacific (n = 27 IEs, n = 24 SRs); and South Asia (n = 18 IEs, n = 7 SRs).

At the country level, the United States has the highest number of included IEs (n = 462), followed by the United Kingdom (n = 28), Australia (n = 13) and Canada (n = 13). Among L&MICs, Colombia has the most IEs (n = 11), followed by Liberia (n = 7), Brazil (n = 6), India (n = 6), Mexico (n = 6) and Uganda (n = 6).

¹¹ As most of the searches for the EGM were carried out starting in July 2020, only studies published in the first half of the year may have been identified. This would skew the trend to suggest a decrease in publication that was due to incomplete data rather than an actual decrease, and as such, the year was dropped.

¹² Because some studies, and systematic reviews in particular, evaluate interventions across multiple countries and regions, the total number of individual studies noted is higher than those included in the EGM.
We coded each included study according to the key characteristics of the target population. Most IEs and SRs focus on the general adult population (Figure 5) (n = 523), followed by adolescents (ages 12-18) (n = 136), youth (younger than age 35) (n = 164) and women (n = 58). About one-third of IEs target urban or peri-urban populations (n = 276). Compared to HICs, a higher proportion of studies in L&MICs focus on rural populations and community leaders.

These population characteristics are often overlapping, and studies were coded for all relevant characteristics.
4.2.3 Interventions

Studies are unevenly distributed across interventions, with most concentrated on a quarter of the 29 intervention categories in the EGM and within the services domain (Figure 6). Because of the high proportion of studies from the United States, for IEs we disaggregate results for evidence from the United States, other HICs and L&MICs. For SRs, we disaggregate results by SR confidence level.¹⁴ Studies are unevenly distributed across interventions, with most concentrated on a quarter of the 29 intervention categories in the EGM and within the services domain. No studies were found in the following five intervention categories:

- Selection and certification process reform
- Professional association capacity building
- Participatory constitutional development
- Protection
- Support to civil society and the media

Within the services domain, the most frequently evaluated intervention category across IEs and SRs is capacity building and system reform of police (n = 165). This is followed by diversion (n = 126); rehabilitation and reintegration programs for ex-offenders (n = 121); and crime prevention (n = 103). The proportions are comparable across high-income countries and L&MICs, and between low and high confidence SRs. The exception is diversion programs, for which no evidence from L&MICs was identified. While compared to the evidence base overall, relatively few studies focus on legal registration interventions and the introduction of new mechanisms to increase access to justice, most of these studies take place in L&MICs (n = 20, n = 8, respectively).

¹⁴ As over 90 percent of SRs (n = 109) cover interventions in the services domain, and less than 15 percent (n = 17) report any data from L&MICs, the vast majority of SR evidence relates to effects on services interventions from HICs. Within SRs reporting data from HICs, 23 only report data from the United States, while 84 report data from the United States and at least one other HIC. Within SRs including data from L&MICs, only 6 report findings exclusively for L&MICs; others include data from at least one HIC. This makes it difficult to disaggregate SR evidence between U.S., other HIC and L&MIC data. As such, we felt a more useful interpretation of differences in the SR evidence base would by by SR confidence level.
The systems domain has the next highest frequency of studies, though the evidence base is substantially smaller than the services domain. The most frequently studied intervention in this domain, court devolution and specialization initiatives, was evaluated in only 39 IEs. Most interventions evaluated in this domain overall were implemented in the United States, though as a proportion of its share in the overall evidence base, there are not many (n = 68, 15 percent of U.S. IEs). Proportionally, there is a greater concentration of the L&MIC evidence in this domain, though the overall number of studies is smaller (n = 21, 21 percent of L&MIC IEs).

The evidence base is smallest for interventions within the society domain (n = 49, 7 percent of all IEs). However, unlike other domains, the majority of IEs in this domain, over 60 percent, evaluate interventions implemented in L&MICs (n = 34). The most frequently evaluated intervention among IEs is legal registration (n = 23). The U.S. evidence base follows different patterns compared to L&MIC evidence within this domain as well: no legal registration interventions are evaluated in the United States, while the most frequently evaluated society intervention within the United States, appearance notification systems, is not evaluated in any L&MIC context. Only two SRs synthesize findings for society interventions, including one for legal registration and one for society-led crime prevention and reporting initiatives.
4.2.4 Outcomes
Almost 550 studies reported outcomes in the prevention of justice problems category (n = 549), over double the next most-reported outcome category, behaviors, for which 177 studies reported outcomes (Figure 7). This reflects the large proportion of included studies that focus on crime and the criminal justice system. Many studies reported outcomes on beliefs, attitudes and norms (n = 114); health and wellbeing (n = 92); and solutions for justice problems (n = 81). The least reported outcomes are intermediate outcomes related to access and functioning of the justice system: accountability of justice actors (n = 6); transparency (n = 3); diversity and representation of minority groups (n = 3); and integration (n = 3).

Figure 7: Count of studies by outcomes reported, country income group (for IEs) & confidence level (for SRs)\textsuperscript{15}

\textsuperscript{15} Studies can appear more than once when they report on multiple outcomes, so the counts in the figure are higher than the total number of studies included. However, each study is counted only once per outcome type (e.g. a study reporting two measures of crime would only be counted once under ‘prevention of justice problems’). The evidence for ongoing studies is likely incomplete, as not all ongoing studies included published pre-analysis plans with exhaustive lists of outcomes to be analyzed. The report only includes the outcomes for which details are available.
4.2.5 Study design

Impact evaluations

Of the 654 impact evaluations included in the EGM, less than half were experimental, randomized controlled trials (RCTs) (n = 304). The proportion of RCTs was higher among studies in L&MICs (n = 61, 60 percent) than it was among HICs (n = 243, 43 percent). Among quasi-experimental methods, 184 studies used matching designs, including propensity score matching, the majority of which were from HIC contexts (n = 177). The next most common methods used in both L&MICs and HIC contexts were difference-in-difference approaches (n = 42, n = 23 respectively). In HICs, interrupted-time-series designs were also common (n = 40). Two studies were based on synthetic control methods, both from HICs.

Overall, just under 20 percent of impact evaluations adopted a mixed-methods approach incorporating both qualitative and quantitative components (n = 145). Proportionally, the use of mixed-methods was slightly higher among L&MIC studies (38 percent) compared to HIC studies (19 percent).

Six studies used qualitative research methods, including qualitative comparative analysis (n = 2), realist evaluation (n = 2) and process tracing (n = 2). Two studies evaluated interventions in L&MICs. Many qualitative studies reviewed for inclusion in this report did not address causal questions or did not use methods that could plausibly establish causal impact.

We extracted data on whether studies included an articulated, clearly defined theory of change. The results of this analysis show that overall, the use of a theory of change to inform the study design and analysis is very infrequent, particularly for studies from HIC contexts, where 13 percent (n = 65) incorporated a theory of change. By contrast, nearly 40 percent of impact evaluations from L&MICs (n = 38) included a theory of change.

Finally, we analyzed the extent to which included studies reported findings of cost effectiveness or cost benefits. Less than 10 percent of all impact evaluations (n = 44) reported some form of cost analysis, most of which comprised of cost-benefit analyses that estimated the return on program investments in the form of savings from future crimes prevented or costs avoided. A similar proportion of studies from L&MICs undertook such analyses (n = 5, 9 percent of all L&MIC IEs).

Systematic reviews

All included systematic reviews undertook at least one form of synthesis of quantitative effects (except where no studies were identified for the effectiveness portion of the review). The most common, among reviews of all confidence levels, was meta-analysis (n = 76). This was followed by narrative description of effect sizes, which sometimes led

---

16 We coded studies as mixed-methods where the findings drew on both qualitative and quantitative data, in order to answer questions not only related to ‘what works’ but also ‘why’ and/or ‘how’ an intervention works. Studies that only drew on qualitative data to inform intervention design or validate research tools were not coded as mixed-methods.
to vote counting, particularly among low-confidence reviews (n = 15). Finally, we further identified 11 reviews that undertook narrative synthesis of quantitative findings. One third of completed systematic reviews included in the EGM carried out a qualitative synthesis (n = 36). The most common type of qualitative synthesis was narrative or thematic synthesis of qualitative or descriptive information (n = 25). Four reviews undertook comparative analysis, and two reviews undertook realist synthesis.

4.2.6 Equity dimensions and focus
We coded impact evaluation studies according to whether and the means through which they reported on ethics approval and incorporated equity-sensitive methods. This includes a variety of means through which a study may consider questions of equity, equality, gender and ethics, such as presenting subgroup analyses for specific population groups to identify disaggregated effects for different populations, or by reporting receipt of ethics approval for the research. Overall, 19 percent of impact evaluations (n = 122) reported receiving approval from an independent review board or similar.

The proportion of studies reporting ethics approval is lowest among interventions implemented in the United States (16 percent, n = 72). Among studies carried out in other HIC contexts, the proportion is slightly higher (22 percent, n = 20). The proportion was highest among studies from L&MICs, as 30 percent (n = 30) of all IEs that targeted L&MICs reported receiving ethics approval. However, generally speaking, ethical approval reporting was low.

The proportion of studies incorporating equity considerations was also highest among interventions implemented in L&MICs, wherein close to half of included impact evaluations addressed equity in some form (47 percent, n = 47). Among studies in HIC contexts, less than 20 percent addressed equity (n = 101), and there was little difference between studies in the United States or other HICs.

The most common way equity was considered in impact evaluations and systematic reviews was through evaluating the effects of interventions specifically targeting a vulnerable population (IEs: n = 65; SRs, n = 15) (Figure 8). This was particularly the case for studies evaluating interventions in the United States. For studies in L&MIC contexts, the most frequent way in which equity was addressed was through sub-group analysis that disaggregated effects by sex (n = 15). This was followed by studies about interventions that targeted a vulnerable population or adopted equity-sensitive analytical frameworks (n = 12 each).

17 Vote counting is where conclusions regarding intervention effectiveness are drawn based on the number of studies that identify a positive effect of an intervention, vs the number that identify a null or negative effect. This approach does not take into account important differences in methodological quality within the primary evidence or important considerations around intervention and study scope, sample, method, and the precision of effect, all of which may have important implications for the interpretation of the findings. These implications may require studies to be weighted differently when interpreting the strength of the evidence and the confidence with which an effect may be identified on average.
The two least common approaches to equity applied in studies included in the RoL EGM include ensuring research ethics were informed by equity (n = 2) and by measuring the effects of interventions on an inequality outcome (n = 2). Because of the low reporting of ethics approval, it is difficult to ascertain whether the studies’ approaches to ethics were informed by gender and equity considerations. Only two studies from L&MIC contexts measured effects on inequality outcomes.

The most commonly considered dimension of equity in included studies was sex, particularly among impact evaluations in L&MIC contexts, wherein a third of the included studies took participants’ sex into account in analyzing findings (n = 35). Among studies in the United States, after sex (n = 45), the most frequently considered dimensions were ethnicity (n = 37) and age (n = 36). While these numbers are larger than those for L&MICs, as proportions of the overall IEs included in the EGM from the United States, these numbers are very low and represent less than 10 percent of IEs. Very few studies took into account other equity considerations such as disability (IEs: n = 11; SRs: n = 8), education (IEs: n = 9; SRs: n = 0), and socioeconomic status (IEs: n = 14; SRs: n = 1).

### 4.2.7 Results of SR critical appraisal

For each completed systematic review, we undertook a critical appraisal to assess the rigor of the review’s methodology. We rated reviews either as low, medium or high confidence, indicating our confidence in the review findings based on the methods used to arrive at those findings.

Of the 107 completed effectiveness reviews included in the EGM, over half (n = 56) were assessed as being of low confidence (Figure 9). Of these, 41 were identified as low confidence during the rapid appraisal. The most common reasons for which a review’s methods were assessed as resulting in low confidence in the findings were a) that bias was not avoided in the selection of articles (n = 19) or b) that there was no or an unclear method applied for assessing risk of bias in included primary studies (n = 10). The lower proportion of medium confidence reviews may be due in part to the decision taken to automatically code all Cochrane or Campbell Collaboration reviews as high confidence, based on the peer-review process and rigorous quality assurance required for publication of systematic reviews with those institutions.
Similar to reviews appraised as low-confidence through the rapid screening tool, SRs for which a complete appraisal was undertaken were most commonly marked as low confidence due to issues related to search and screening or risk of bias. At the search and screening stage, reviews were marked as low confidence where methodological limitations meant there was a risk that the review had not identified the full evidence base. This was often due either to the exclusion of studies not published in peer-reviewed journals, or to risks of errors in the screening and selection process due to single-reviewer screening. Common issues in SR methods for assessing study quality included either an absence of risk of bias analysis or the omission of key sources of bias in the risk assessment. Further, low-confidence reviews often did not integrate the risk of bias findings with the analysis of effects, such that readers were not easily able to interpret the effects appropriately.

Figure 9: Percent of included systematic reviews assessed at each confidence level

4.3 Findings from high confidence SRs

This section presents key findings and policy implications from the 38 high-confidence reviews represented in this EGM. Out of 38 high-confidence reviews, almost all (n = 34) synthesize the effectiveness of interventions in the services domain. The remaining two domains are represented by four high confidence SRs: two in the society domain and two in the systems domain.

The first section reports the cross-cutting implications for rigorous evidence and research for strengthening the rule of law, and subsequent sections report domain-specific and group-specific key findings from the reviews relevant to policymaking, research and programming. Appendix C includes links to summaries of each of the individual SRs. If summaries could not be found, then links to the full report text are provided.

4.3.1 Cross-cutting implications

Across high-confidence SRs of different intervention types, many of the implications were similar. These include:

- **A need for more evidence from L&MICs.** Only one high-confidence review focused entirely on evidence from L&MICs, while only five included at least one L&MIC study.
● A need for more IEs of established interventions in different settings to examine which approaches are effective in which contexts. Given the high concentration of evidence from the United States, many SRs identified a need for more research focusing on different geographic areas to understand the extent to which findings may or may not be generalizable across contexts. In some cases, the primary evidence base contributing to SR findings was even limited almost entirely to a white male population from a single state within the United States, highlighting a gap in the rigorous evidence base for many population groups of interest.

● There are gaps in measures of impact other than crime rates and gaps in evaluations with longer follow-up periods. A people-centered justice approach recognizes that people’s justice needs go beyond freedom from crime, yet limited reporting of non-crime outcomes limits our understanding of the impact of rule of law and justice interventions. Similarly, limited evidence assessing longer-term impacts inhibits our understanding of the sustainability of rule of law and justice intervention impacts.

● Few impact evaluations disentangle the effects of different elements of multicomponent interventions or use sufficiently rigorous research designs. High-confidence SRs often found that few IEs drew on samples large enough to be representative and facilitate generalization of the findings. Further, few reported economic analyses of costs, a key consideration for many stakeholders.

● A need for more triangulation of data in impact evaluations. Triangulation, wherein data from multiple data sources (e.g. both police records and victim reports/survey data) are used in evaluating program effects, can help strengthen confidence in the findings, particularly where data limitations may make it difficult to address risks of bias through statistical methods alone.

4.3.2 Systems

Human capacity development for informal/non-state justice actors (Based on IE evidence from the United States)

One review assessed the effectiveness of capacity building for non-state justice actors and foster carers (Macdonald and Turner, 2008). The treatment foster care (TFC) intervention was assessed, where child services select foster carers, and give them specialized training, professional support and access to crisis services. Evidence from the SR suggests that compared to usual care, TFC appears to generate worthwhile and statistically significant reductions in antisocial and criminal behavior, as well as less time spent in locked settings such as psychiatric hospitals or youth justice settings.

4.3.3 Services

Capacity building and system reform of police (Based on IE evidence from the United States, United Kingdom, Canada, Australia, Columbia, Lebanon, Libya, Argentina, Denmark, India, Sweden, Trinidad and Tobago)

Eight systematic reviews assess the effectiveness of interventions that build the capacity or reform systems of police (Lum et al. 2006; Koper and Mayo-Wilson 2012; Mazerolle et al. 2013; Higginson and Mazerolle 2014; Braga et al. 2018; Braga et al. 2019a; Braga et al. 2019b; Hinkle et al. 2020; Mazerolle et al. 2020). Overall, the reviews suggest that hot spot policing (Braga et al. 2019a), problem-oriented policing (Hinkle et al. 2020) and focused deterrence strategies (Braga et al. 2018) are effective in reducing crime. All
three are likely to produce a diffusion of benefits into nearby areas. Focused deterrence was particularly successful for gang/group programs. Problem-oriented policing (POP) was effective in reducing crime and disorder at a variety of different units of analysis and for diverse types of interventions though it had limited impacts on fear of crime, police legitimacy, and collective efficacy. It was also found to be cost-effective. Police-led interventions specifically aimed at increasing legitimacy in the eyes of the general public have a significant impact on public satisfaction with and confidence in the police, with a minor effect on reoffending (Mazerolle et al. 2013). At place level, such interventions reduce overall and major (Part 1) crime (Higginson and Mazerolle 2014). Policing disorder is associated with reductions in crime, but only when community and problem-solving tactics are used (Braga 2019b). Street-level policing approaches are effective in reducing drug crime, particularly those involving cooperation with local third parties such as local businesses, government authorities and housing associations (Mazerolle et al. 2020). Evidence from United States and Columbia suggests that gun detection patrols in high-crime areas reduce gun crime at high-risk times (Koper and Mayo-Wilson 2012).

**Deterrence mechanisms (Based on IE evidence from the United States, United Kingdom, Norway, Sweden, Canada)**

Four SRs assess the effectiveness of crime deterrence interventions (Welsh and Farrington 2008a; Welsh and Farrington 2008b; Welsh and Farrington 2009; Simpson et al. 2014). The reviews found that of all area surveillance techniques, improved street lighting appears to be the most effective (Welsh and Farrington 2008b; Welsh and Farrington 2009). CCTV surveillance is effective at reducing vehicle crimes in car parks though it is not possible to disentangle its effects from other components like extra security guards, better lighting or fencing (Welsh and Farrington 2008a). Neither improved street lighting nor CCTV prevents violent crime (Welsh and Farrington 2008a; Welsh and Farrington 2008b). Laws and regulations have only small deterrence effects on corporate crime (Simpson et al. 2014). Legal interventions such as laws and punitive sanctions (e.g. arrest, fines, or a likelihood of prosecution) have a small effect on company non-compliance. Regulatory policies (e.g. company inspections) and non-punitive sanctions by agencies (e.g. cease and desist orders) have a modest but consistent effect on crimes committed by individual employees (ibid.).

**Diversion (Based on IE evidence from the United States, United Kingdom, Australia, Canada, Denmark, Israel, Netherlands, Sweden, Switzerland, Guam, New Zealand)**

Five SRs assess the effectiveness of diversion interventions (Perry et al. 2009; Mitchell et al. 2012a; Petrosino et al. 2010; Villetaz et al. 2015; Wilson et al. 2018). Overall, custodial sentences fare no better or worse than non-custodial sentences in reducing re-offending (Villetaz et al. 2015). Compared to alternative sentencing, processing of juvenile delinquents by the juvenile justice system is associated with more offending behavior, not less (Petrosino et al. 2010). For low-risk youth who come into contact with the justice system, alternatives to court processing (police-led diversion) are more effective in reducing a youth’s future contact with the justice system (Wilson et al. 2018). Perry and colleagues (2009) found that pretrial release with drug testing and intensive supervision had limited effects as compared to routine parole and probation. However, overall drug courts were assessed to have a large effect in reducing recidivism for adults and drunk driving offenders, with a smaller but significant effect observed for juveniles (Mitchell et al. 2012a).
Crime prevention (Based on IE evidence from the United States, United Kingdom, Israel, North Macedonia, Nepal, Philippines, Bermuda)

Seven SRs assess the effectiveness of interventions which target individuals vulnerable to engaging in crime (Welsh and Farrington 2006; Fisher et al. 2008; Egli et al. 2009; Van Der Laan et al. 2011; Petrosino et al. 2013; Tolan et al. 2013; Wilson et al. 2016). Mentoring interventions for at-risk youth are effective in reducing delinquency (Tolan et al. 2013), though juvenile curfews (Wilson et al. 2016) and ‘scared straight’ programs (Petrosino et al. 2013) are not. Among drug substitution programs heroin maintenance significantly reduces criminal involvement as compared to Methadone or another standard treatment (Egli et al. 2009). Buprenorphine and Naltrexone are found to be promising, although evidence is limited (ibid.).

Rehabilitation and reintegration programs for ex-offenders (Based on IE evidence from the United States, Australia, United Kingdom, Germany, Netherlands, Taiwan - China)

Eight SRs assess the effectiveness of rehabilitation and reintegration programs for ex-offenders (Garrido 2007; Lipsey et al. 2007; Feder et al. 2008; Mitchell et al. 2012b; Shmucker and Loesel 2017; Wilson et al. 2005; Visher et al. 2006; Strang et al. 2012). Cognitive-behavioral therapy (CBT) is effective at reducing re-offending regardless where the treatment takes place (prison, community or institution for juvenile offenders). It is particularly effective for participating offenders with high recidivism risk (Lipsey et al. 2007). While for individuals convicted of domestic violence, court-mandated psychoeducational treatments and CBT showed a small positive effect when official measures of repeat violence were examined, no effect was found on victim reports of repeat violence (Feder et al. 2008). Psychosocial treatment for sexual offenders in the community and in forensic hospitals reduces re-offending: particularly CBT and individualized treatment (Shmucker and Loesel 2017). For serious juvenile offenders placed in correctional facilities, apart from CBT which is effective in preventing crime, other programs’ effectiveness is small or doubtful (Garrido 2007).

Employment-focused interventions such as job training and/or job placement for ex-offenders were not found to reduce recidivism (Visher et al. 2006). Incarceration-based drug treatment programs for substance-abusing offenders are modestly effective in reducing criminal behavior and drug use (Mitchell et al. 2012b). While no effects were found for boot camp interventions (Wilson et al. 2005; Mitchell et al. 2012b), therapeutic communities consistently yet modestly were found to reduce recidivism (Perry et al. 2009; Mitchell et al. 2012b) and drug relapse (Mitchell et al. 2012b). Counselling reduces reoffending but not drug relapse, while narcotic maintenance programs reduce drug relapse but not reoffending (ibid.). Face-to-face restorative justice conferences (RJCs) have a small but highly cost-effective impact on decreasing the frequency of subsequent crimes (Strang et al. 2012).

Social services for victims of crime and violence (Based on IE evidence from the United States)

One SR assessed the effectiveness of services for victims of crime and violence (Davis et al. 2008). Second responder programs attempt to prevent future incidents of domestic
violence by providing follow-up interviews (second response) to victims just after incidents of violence occur. More victims report incidents after the follow-up interviews, though no preventative effect was found on the number of new incidents of domestic violence.

### 4.3.4 Society

**Legal registration** *(Based on IE evidence from Cambodia, China, Côte d'Ivoire, Ethiopia, Ghana, India, Kenya, Madagascar, Malawi, Mozambique, Nicaragua, Peru, Rwanda, Somalia, Vietnam, Zambia)*

One SR assessed the effectiveness of legal registration interventions (Lawry et al. 2014). The only high-quality synthesis which focused entirely on low- and middle-income countries, it assessed the effectiveness of strengthening land property rights in rural areas. Land property rights interventions at the individual level increase investment, productivity and farmer incomes in Asia and Latin America but less so on the African continent. While the positive effects are derived from general improved perceived security and resulting long-term investment, qualitative studies find that adverse effects might occur such as reducing the access to land by women or farmers with low socio-economic status.

**Society-led crime prevention** *(Based on IE evidence from the United States, Australia, Canada, United Kingdom)*

One SR assessed the effectiveness of community-led crime prevention activities (Bennett et al. 2008). The review found that neighborhood watch schemes are effective in reducing crime. There is no difference in effects regarding whether the intervention was implemented alone or in combination with additional components such as property marking or security services. Little is known about the factors that moderate whether or not the schemes are effective.

### 4.4 Discussion: evidence gap analysis

There is an increasing trend in the publication of impact evaluations over time, with a noticeable increase starting from around 2009, particularly in L&MICs. Nonetheless, the evidence base is fragmented, with the majority of studies concentrated on a few types of interventions.

Below, we discuss the evidence gaps in more detail. We organize the discussion into primary evidence gaps, where few or no impact evaluation evidence exists for particular interventions, outcomes and populations, and ‘synthesis’ gaps, where we identify a cluster of primary study evidence but a lack of up-to-date, high-quality SRs (Snilstveit et al. 2017). Finally, we also highlight several ‘methodological’ gaps – namely, the low reporting of the use of ethics review boards to ensure studies do no harm, and the low use of methods such as theory of change that can inform and strengthen the interpretation of evaluation findings.

#### 4.4.1 Interpreting patterns in the evidence base

A primary objective of EGMs is to enable fast and easy access to relevant rigorous evidence for donors, implementers and researchers, to facilitate evidence-informed decision-making. An additional contribution of EGMs is to enable more efficient use of research investments, by identifying areas where substantial evidence already exists and where it may be needed.
Evidence gaps
On the interactive online matrix, blank squares indicate intervention-outcome configurations for which no impact evaluation or systematic review evidence was identified. The EGM does not explain what drives these gaps, however, and the relevant explanations will vary across intervention-outcome configurations within the map. Blanks squares can exist for three reasons:

1. **A meaningful gap in the evidence base, which should inform future research agendas.** This is particularly important where a gap exists for an intervention that has been widely implemented with the aim of achieving a particular outcome, despite a lack of rigorous impact evidence to support the causal claim.

2. **There is limited underlying theory suggesting a causal relationship.** As EGMs aim to capture the full range of interventions implemented and outcomes measured within a sector or sub-sector, it is possible that not every intervention may be expected to influence every outcome. However, examining the strength of the theory for each intervention-outcome configuration on the map is unfortunately beyond the scope of an EGM.

3. **There are methodological limitations that limit the utility of impact evaluations, or ethical considerations that prevent the use of such methods.** For example, in the case of interventions such as witness protection programs, counterfactual impact evaluations may not be ethical. In such cases, alternative methods may be utilized, such as qualitative or process evaluations, which consider the topic of impact but do not fulfill the inclusion criteria for the EGM. Furthermore, the limited evidence base for ‘systems’ and ‘society’ interventions does not mean these areas are not worthy of implementation; rather, that interventions in these domains may be more challenging to evaluate, and should be accompanied by a robust research agenda.

Evidence concentrations
There are two potential explanations for why concentrations of evidence may exist for a given intervention-outcome configuration:

1. **Understanding the intervention’s effectiveness is of particular importance to the rule of law and justice community.** This may be because the intervention is particularly commonly implemented or substantial funding is invested in the intervention. The intervention may address a priority rule of law and justice challenge, such as addressing the relationship between mental health and criminal behaviors. Political priorities may also influence which interventions are evaluated for effectiveness against which outcomes.

2. **There is a commonly recognized theoretical link between an intervention and the outcome.** It is important to note, however, that the EGM only identifies studies that have reported effects on a particular outcome. The existence of many studies of a particular intervention and outcome does not necessarily mean that the intervention is effective at producing impacts on that outcome. Where high-confidence systematic reviews have assessed a relationship between an intervention and outcome, the findings are summarized above in Section V.C. Beyond this, however, questions of which interventions are effective at producing effects on which outcomes are beyond the scope of this report.

---

18Per USAID policy (ADS 201), all USAID-funded projects require a theory of change.
Using the evidence patterns in the EGM
When interpreting the concentrations and apparent gaps in the evidence base identified through the EGM, it is important to consider these different possibilities. Nonetheless, there are many ways in which users may draw on the patterns identified in the EGM to support future work:

1. **Inform research agenda-setting processes.** The EGM findings can help identify priority areas for future research investment, particularly where combined with expertise from diverse stakeholders in order to effectively interpret the different gaps.
   a. Investments in new impact evaluations may be particularly beneficial where they target interventions for which limited evidence exists, or where there is limited evidence for the effects of the intervention on a population or context of interest.
   b. Where large concentrations of primary evidence already exist, particularly for certain populations such as the United States, investments in additional impact evaluations may not provide as much value as investments in evaluations of interventions and outcomes for which little or no effectiveness evidence exists.
   c. Where there are concentrations of impact evaluation evidence, and existing systematic reviews are out of date or do not cover populations of interest, new systematic reviews may help ensure policymaking and programming is informed by the best available evidence. Examples of such opportunities are described below in Section VI.D.3.
   d. Investments in synthesis evidence may not be necessary where multiple high-confidence reviews exist, such as for the effects of diversion programs on outcomes of prevention of justice problems. Systematic reviews represent substantial investments of time and funding, and thus the EGM may enable more efficient allocation of research investments.

2. **Support policy and program design.** Where stakeholders are interested in targeting a particular outcome, they can utilize the EGM to identify which interventions may be likely to impact the outcome of interest. The hyperlinks within the online EGM enable easy access to rigorous evidence that can be consulted when designing new policies and programs, to identify which intervention approaches may be more or less effective at impacting the outcome of interest. Where multiple interventions have been undertaken to influence a particular outcome, stakeholders can use the filters in the EGM to identify which interventions may have been undertaken for geographies or population groups of interest. Similarly, the EGM can enable stakeholders to identify which interventions have targeted a particular population group of interest, such as women and girls, or people with disabilities.

3. **Identify the range of outcomes that have been theorized to be impacted by a particular intervention.** This can support stakeholders in understanding all outcomes that may be influenced through their intervention. This is particularly important when considering potential adverse effects, which may be captured in outcomes related to wellbeing.

4. **Identifying examples of impact evaluations undertaken in a particular context or utilizing a particular method.** This can be useful for identifying potential challenges and strategies applied to address challenges that may
strengthen the quality of future research. For example, stakeholders interested in undertaking impact evaluations in fragile contexts may use the country filter to identify evidence from particularly fragile contexts, to understand the methods, challenges and approaches utilized when undertaking impact evaluations in such complex environments.

4.4.2 Primary study evidence gaps
The evidence base is unevenly distributed among the three intervention domains and among geographic contexts. Table 5 below shows the distribution of impact evaluations across different domains and geographies.

**There are gaps in the evidence base for many systems and society interventions.** The majority of impact evaluations (80 percent, n = 524) evaluated interventions in the services domain. The gap was largest for interventions in the society domain (7 percent of IEs, n = 49). Compared to services interventions, however, the evidence base was also limited for systems interventions (12 percent, n = 81). These findings are dominated by trends in the U.S. evidence base. The 90 percent of IEs from the United States (n = 415) evaluated interventions in the services domain. Only seven studies were identified of impact evaluations targeting society interventions in the United States (2 percent) and 40 targeting systems interventions (9 percent). Within evidence from other HICs, there is also a high concentration of evidence within services interventions, though a higher proportion of studies target interventions in the systems domain compared to the U.S. evidence base (18 percent, n = 16). Among L&MICs, the evidence base is somewhat more evenly distributed, as just over half of impact evaluations targeted the services domain (42 percent, n = 42), with the second half of studies roughly split between systems and society interventions. This suggests that the gap in society interventions among L&MIC contexts is less than among studies evaluating programs in the United States.

We identified no impact evaluations for six intervention groups: for selection and certification process reform, professional association capacity building, and participatory constitutional development among systems interventions; for protection, among services interventions; and support to civil society and the media, and legal empowerment among society interventions. As noted, these may not all necessarily relate to priority evidence gaps. Particularly for protection interventions, there may be ethical considerations that limit the implementation of counterfactual impact evaluation research designs.

**There is a substantial evidence gap for many geographic contexts.** Overall, 70 percent of the evidence base is concentrated in a single country, the United States. This gap is particularly great for evidence from L&MICs, which represent 62 percent of all countries globally but only 15 percent of the impact evaluation evidence base for rule of law and justice interventions.
Table 5: Distribution of impact evaluations by intervention domain and geography

<table>
<thead>
<tr>
<th>Domain</th>
<th>United States</th>
<th>Other HIC</th>
<th>L&amp;MIC</th>
<th>Total by domain</th>
</tr>
</thead>
<tbody>
<tr>
<td>Systems</td>
<td>40</td>
<td>16</td>
<td>25</td>
<td>81</td>
</tr>
<tr>
<td>Services</td>
<td>415</td>
<td>67</td>
<td>42</td>
<td>524</td>
</tr>
<tr>
<td>Society</td>
<td>7</td>
<td>8</td>
<td>34</td>
<td>49</td>
</tr>
<tr>
<td>Total by geography</td>
<td>462</td>
<td>91</td>
<td>110</td>
<td>654</td>
</tr>
</tbody>
</table>

There are practical and ethical constraints that may partially explain limited evidence for certain interventions and outcomes. For example, interventions that target a country-wide population may be challenging to evaluate against a valid control group. In other situations, such as where priority outcomes are longer-term, diffuse or difficult to measure, stakeholders may struggle to identify measurable changes within often-short programmatic cycles. As noted earlier, for interventions such as witness protection, ethical considerations may limit the extent to which experimental or quasi-experimental research designs can be applied. However, the existence of studies across the systems and society domains and in many L&MICs including fragile contexts such as Afghanistan suggest that for some types of interventions, these studies are feasible.

There are also gaps in the evidence base for specific population groups. For example, we found few studies that evaluated the effects on youth, particularly among studies from L&MICs, which reported effects on adolescents in only 6 percent of studies (n = 6). Adolescents and youth ages 15-24 make up a large and growing share of populations in Sub-Saharan Africa, and a relatively large share of populations in South Asia, Latin America and the Caribbean, and Western Asia (United Nations 2018). In addition, the population of adolescents and youth ages 15-24 is expected to rise by over 100 million across 47 low-income countries by 2050 (United Nations Department of Social and Economic Affairs 2019). As such, the low number of rule of law studies focused on adolescents and youth in L&MICs represents a particular gap in the evidence base for a population that is growing.

We also found few studies reporting effects for women and girls (n = 58, 4 percent). This included both a low overall number of studies targeting women and girls, as well as very few studies reporting effects disaggregated by sex. This is a key gap as women may face different justice problems than men, and therefore may need different types of support. As a focus of people-centered justice is shifting the goal of rule of law and justice interventions towards prioritizing efforts to solve the justice problems of people, a core question must be whose justice problems are being addressed (Pathfinders Task Force on Justice 2019). The limited evidence available for women and girls makes it difficult to understand whether interventions are meeting their justice needs. Similarly, very few studies targeted or measured effects for other vulnerable populations such as people with disabilities (n = 19) or forcibly displaced populations (n = 4).

4.4.3 Synthesis gaps

Overall, the pattern of interventions and outcomes reported by systematic reviews broadly follows that of impact evaluations, and there are no obvious major synthesis gaps.
in the evidence base. To enable systematic reviews to draw clear conclusions and identify both generalizable and context-specific findings, a cluster of observations for comparable interventions and outcomes is generally needed. For example, whilst there appears to be a cluster of impact evaluations measuring the effects of the introduction of new mechanisms to increase access to justice (n = 14), the disparate outcomes measured means the largest number of studies available for a single intervention and outcome intersection is six (on solutions to justice problems); as a result, we do not consider this a synthesis gap. As noted earlier, there may be myriad causal pathways hypothesized for the interventions included in the EGM; this heterogeneity could partially explain why there appear to be so many gaps in the evidence base. When combined with the diversity and inconsistency in outcome measurement, these factors make it more challenging for synthesis work to draw clear conclusions that will support policymakers and programs. The relatively fragmented evidence base further limits the number of synthesis gaps, but we identify two topics with potential for evidence synthesis below.

There is a concentration of evidence for legal registration interventions, for which only one high-confidence review exists, which is out of date. Many of the 18 IEs measuring effects of these interventions were published after the review was completed in 2014. While the majority of these interventions report effects on economic outcomes, the specific rule of law outcomes reported across the studies is less consistent, which may pose a challenge for effective synthesis.

Another potential synthesis gap could be identified for the effects of interventions promoting participation and human rights standards. However, the largest concentration is for effects of the intervention on outcomes in the category of ‘prevention of justice problems’, for which less than half of the 24 IEs report effects.

4.4.4 Methodological gaps
Few impact evaluations and systematic reviews of rule of law interventions used mixed-methods and other theory-based approaches such as theories of change to address complexity, particularly in evidence from the United States and other HICs. Such methods can enable evaluations and systematic reviews to ‘open the black box’ to understand why the evaluation may or may not identify effects on outcomes, and identify contextual factors that may influence whether, for which populations and under which conditions the intervention is effective. Understanding why an intervention may or may not have worked in a given context, and how the different outcomes measured relate to each other, can enable more effective interpretation of the evaluation findings. This facilitates the effectiveness of subsequent decisions regarding whether to scale, adopt or adapt the intervention in new contexts. Only 19 percent of studies from HICs and from the United States incorporated mixed-methods analyses (U.S. n = 90; other HIC n = 17). Within the evidence-base from L&MICs, the proportion was significantly higher, as 39 percent adopted mixed-methods research designs (n = 39). Overall, however, this is still relatively low.

The use of theories of change was even lower. Among impact evaluations from the United States, only 11 percent (n = 50) incorporated a theory of change that articulated the steps through which the intervention was expected to affect the outcomes of interest, the relationship between outcomes within the causal chain, relevant contextual factors that may interact with the intervention and assumptions underlying each step. Among
studies from other HICs, 16 percent included theories of change (n = 15). However, over a third of studies from L&MICs incorporated theories of change (n = 38). While another recent evidence mapping study of L&MIC evidence found that the use of theories of change has been increasing in recent years (Sonnenfeld et al. 2020), the use of theories of change within the United States and other HICs over the most recent decade has remained low (13 percent, n = 65). By contrast, 10 of 13 ongoing studies identified that target L&MIC contexts incorporate theories of change, which suggests that this finding may particularly relate to the increased use of theories of change among the L&MIC evidence base. This finding may be explained in part by foreign assistance donor requirements, which commonly include theory of change in project design.

The use of theories of change in impact evaluations helps to provide a framework for data collection and analysis (White 2009), which may facilitate the identification of more useful and relevant findings. When evaluations are not grounded in a clear theory, it is often harder to effectively learn from them and identify where common assumptions may be invalid. Theories of change facilitate interpretation of how different outcomes relate to each other, which can help make sense of mixed results. For example, by looking for evidence of effects across different outcomes and structuring analyses along the causal chain, researchers can identify where the program theory may break down and, in doing so, help explain a lack of effect on final outcomes.

**Few impact evaluations evaluated questions of cost-effectiveness or cost-benefits, despite this being an important question for many stakeholders.** Less than 10 percent of studies overall (n = 56) clearly reported findings on cost-effectiveness or cost-benefits of rule of law interventions. Proportionally, rates were similar across all geographic contexts.

**Very few impact evaluations report having received approval from an independent ethics review board** (19 percent, n = 122). This is particularly the case among studies targeting U.S. populations (16 percent, n = 72), but while highest among studies targeting L&MICs (30 percent, n = 30), it remains under a third. This is despite widespread recognition of the importance of approaches such as ‘do no harm’ within the development community. When considering this finding, however, it is important to note several limitations: first, not all of these studies were based on primary data collected by the authors, which may explain why ethics approval was not sought; second, not all studies may report ethics approval even when it was received; and third, in some situations authors may not have access to an independent review board, but may nonetheless have substantial ethics procedures that they follow. However, we do not expect that these factors explain this low finding overall, particularly given the extent of the evidence base from the United States, where the research infrastructure is well developed. Ensuring appropriate consideration of ethics is a core requirement of human-subjects research, particularly when the research is dealing with vulnerable populations, as many of the included studies do, and when the effects of the interventions have the capacity to do harm to individuals’ wellbeing. Greater reporting of ethics considerations, including whether or not review by an independent board was needed or sought, would help understand what drives this trend.

Finally, 15 percent of studies undertook analysis of disaggregated effects by sex (n = 91) or other dimensions (n = 78), despite important questions regarding relationships
between treatment within the criminal justice sector, access to justice and characteristics of ethnicity, socioeconomic status or education, and disability. This is related to questions of ethics and ensuring that human-subject research does no harm.

With regard to study design, our focus on only experimental or quasi-experimental studies excluded other thematically relevant studies, which could partially explain some geographic and primary study evidence gaps. Studies included in Jackson et al. (2019)’s mapping of security and justice interventions in L&MICs, which included non-experimental study designs, can provide insight into questions such as relevant barriers and facilitators that may help or hinder implementation, and information on the application of rule of law interventions in contexts not covered by the impact evaluation evidence base. For example, their EGM identified almost 90 studies taking place in West Africa, whereas we identified only 16 IEs from the region. However, similarly to this EGM, they found few relevant studies in Central Asia, East Asia, Eastern Europe and North Africa (Jackson et al. 2019). Also similar to our review, they found evidence gaps in legal empowerment interventions and for studies that focused on forcibly displaced populations (Jackson et al., 2019). This suggests an absolute evidence gap for these interventions and populations. In terms of outcome evidence, they identified gaps in outcome information about crime and gender-based violence rates in L&MICs. This suggests a distinction from the majority-U.S. evidence included in this EGM; we identified many studies, particularly from the United States, that reported outcomes relating to prevention of justice problems such as crime rates. This suggests that in L&MIC contexts, there may be additional challenges tracking or attributing change in crime rates to justice interventions due to limited data availability.

5. Conclusions and implications

We identified 637 completed IEs, 17 ongoing IEs, 107 completed SRs and 11 ongoing SRs. The evidence base has increased sharply in recent years, particularly from L&MIC contexts. However, the majority of the evidence remains concentrated in HICs, particularly in the United States, which accounts for over 70 percent of included IEs (n = 462). The evidence is further highly concentrated within a small number of intervention categories, primarily within the services domain, across contexts. The proportion of the evidence base that evaluate interventions in the systems domain is somewhat similar across HICs outside of the United States and L&MICs, at 18 percent and 25 percent of the IE evidence base, respectively.

Overall, the most frequently evaluated intervention category across impact evaluations and systematic reviews is capacity building and system reform of police (n = 165). This is followed by diversion (n = 126); rehabilitation and reintegration programs for ex-offenders (n = 121); and crime prevention (n = 103). While these four categories capture the most frequently evaluated interventions in the United States and other HICs, the trend for studies targeting interventions implemented in L&MICs is slightly different. Capacity-building and system reform of police, and legal registration interventions, are the most frequently evaluated interventions in L&MICs (n = 20; n = 20), while only two legal registration intervention studies were identified in the United States and other HICs (n = 2). Crime prevention was the next most common category in L&MICs (n = 11). No IEs were identified from L&MIC contexts evaluating diversion programs, and only one IE was identified for a rehabilitation and reintegration intervention.
While overall, the evidence base is largest for interventions within the services domain across all geographies, the concentration was highest in the United States. This was driven by the fact that over half of the IE evidence from the United States is restricted to three intervention groups: capacity building and system reform of police (n = 111); diversion (n = 89); and rehabilitation and reintegration programs for ex-offenders (n = 82). This may be driven by contextual factors that are particular to the U.S. context. Compared to some L&MIC contexts, the United States also has well established justice systems across the country, which likely leads to a different set of bottlenecks to rule of law or people-centered justice that require different interventions. For example, the United States incarcerates more people, and more people per capita, than any other country (Walmsley 2018). This context may influence the focus on diversion programs within the U.S. rule of law evidence base.

Across all HIC contexts, there was limited evidence from interventions within the society domain (n = 15). Within L&MICs, however, 34 IEs were identified (34 percent), the majority of which evaluated legal registration interventions. Differences in the size of the evidence base for different geographies may also relate to different levels of funding for different population groups. However, it may also, and perhaps more likely, be due to differences in which interventions are more/less relevant for different contexts. For example, in high-income contexts, populations already have high rates of legal registration, which thus makes such interventions unnecessary.

These findings highlight important differences in geographic trends in the rigorous evidence base and raise questions regarding the generalizability of the evidence base. Any intervention aiming to strengthen rule of law and access to justice will interact with local social, cultural, political and economic factors during implementation, which will influence the way in which the intervention is received and the mechanisms it is or is not able to trigger. These factors will influence which interventions will be appropriate for the target population and outcomes, as drivers of rule of law and justice problems vary across contexts. This creates a challenge in understanding which evidence is generalizable for which contexts and populations, and the ways in which interventions must be adapted to be transferable to a given new context. This challenge is compounded by the limited use of mixed-methods and theory of change within the rule of law evidence base, particularly among studies from the United States and other HICs. These methods can help identify the relevant contextual factors that explain why a particular instance of an intervention being implemented in a given context was or was not effective at achieving impacts on outcomes of interest, which in turn enables more effective application of research findings.

Findings from high-confidence systematic reviews highlight the importance of measuring outcomes for different population groups; the importance of cost-effectiveness analysis; and the potential value of incorporating qualitative data alongside quantitative findings. As noted in Section V.C above, high-confidence systematic reviews of diversion programs have found that, on average, they are no more or less effective at reducing re-offending compared to custodial sentencing. For youth populations in particular, diversion programs are more effective at reducing a youth’s future contact with the justice system, compared to court processing. Mitchell et al. (2012a) found that drug courts are more effective for adult populations than youth. While these findings suggest that non-incarceration-based approaches may be more effective for youth populations,
cost-effectiveness data is required to understand whether such approaches may also be preferable for adult populations as well. As diversion programs performed similarly to custodial sentencing in terms of reducing re-offending, understanding the relative costs may have important implications for policy.

In some cases, it may not be possible to design an IE with sufficient power to identify heterogenous effects for sub-groups of the population. In such situations, qualitative analyses may identify important differences in the ways in which different groups experience an intervention. For example, Lawry et al., 2014 identified qualitative evidence suggesting that while land registration interventions were beneficial for target populations on average, they may worsen outcomes for particularly vulnerable populations such as women and households from the lowest socioeconomic groups.

5.1 Implications for policy and practice

The large evidence base has led to a number of synthesis efforts, including 38 high-confidence SRs synthesizing what we know about interventions, particularly for those in the services domain. Although the available evidence often limits the extent to which these reviews can identify clear policy implications, we suggest consulting these syntheses in the design and development of new policies and programs. We summarize the implications from these SRs here, and in more detail in Section V.C, and encourage readers to consult the full reviews for more details.

5.1.1 Implications from SRs

- Efforts to build positive support networks for at-risk individuals may be particularly effective at preventing future justice problems. High-confidence SRs identified multiple instances in which efforts to strengthen support networks, opportunities and non-punitive approaches were more effective at reducing future justice problems compared to custodial or court-based approaches.
- Non-punitive approaches may be particularly effective for preventing or reducing reoffending rates for youth.
- Though these findings draw primarily on evidence from the United States, they highlight the importance of youth-specific programming. This may be particularly relevant for supporting rule of law and access to justice in many L&MICs where youth form the majority of the population. Identifying effective interventions for reducing or preventing criminal behaviors among young people may therefore be valuable in such contexts.
- Cognitive-behavioral therapy is an effective treatment for preventing reoffending, but not for all types of offenders. However, there were inconclusive findings regarding the effectiveness of CBT and other psychoeducational treatments for preventing domestic violence among past abusers. This suggests that such treatments may be more effective for particular types of crimes than others, and more research is needed to understand for which populations such approaches are effective for which outcomes, under which circumstances.

5.1.2 Implications from SRs and IEs

The large body of evidence identified in the EGM and its findings can be utilized by policymakers and implementers during program design, to identify relevant rigorous evidence from both IEs and SRs. The following points may support effective use of the
EGM in support of evidence-informed policy and program design:

- Where no evidence exists for an intervention, or none exists from the relevant geography, consider whether it would be possible to include an IE within the new program, taking into account the implications for research noted below;
- Where relevant SRs exist, consult the studies to inform program design. Even low-confidence reviews may present useful information, such as descriptions of the evidence base or theories of change, although findings related to program effects should be interpreted with caution;
- Where there are existing IEs but no recent or high-quality SRs:
  - Consult the relevant primary studies but use caution when interpreting the findings. Conclusions regarding intervention effectiveness should not be drawn from single studies or by counting the number of ‘successful’ interventions. Further, all results may not be directly transferable to different contexts. Policymakers and practitioners should consult IEs as well as sector and regional specialists when judging the transferability of results; and
  - If the cluster of evidence is large enough, consider commissioning an SR. Since SRs take time to develop, the EGM should be consulted as early as possible in the planning stages when designing new programs or strategies.

5.2 Implications for future research

When commissioning and designing new studies, we suggest that funders and researchers consider the following:

- Prioritizing intervention categories and outcomes that incorporate well-established and/or previously tested theories of change AND for which there are no or few existing IEs (or the existing IEs are not relevant to the target population, e.g., L&MICs, youth, rural, etc.).
- Adopting a common set of outcomes across studies to enhance the value and potential for cross-study lessons and evidence synthesis;
- Incorporating measures of cost-effectiveness and sustainability;
- Employing study designs informed by a mixed-methods, theory-based approach to IE that considers a range of questions relevant to policy and practice, including implementation, contextual factors and costs;
- Applying the most rigorous IE method that is appropriate and feasible, considering both intervention design and context. For intervention types that do not lend themselves to quantitative IE, this may mean applying rigorous ‘small n’ methods for small sample size IE designs (White and Phillips 2012);
- Ensuring research designs and methods are sensitive to inequalities across different population groups. Taking into account diverse experiences, power dynamics and gendered inequality in study design and conduct will ensure new studies are sensitive to the needs of vulnerable groups and the effects interventions may have on them. Methods to strengthen equity-sensitive research include adopting methods allowing for the identification of differential effects amongst different population groups; and
- Ensuring ethical research conduct and protection of research participants, including undertaking and reporting of the review and approval of study protocols and procedures by relevant review boards.
Online Appendixes

Online appendix A: Detailed methodology

Online appendix B: Data extraction tools

Online appendix C: Links to summary findings and recommendations from systematic reviews

Online appendix D: Rule of Law Protocol
References

Included impact evaluations (completed)


Andersen, J. P., Gustafson, H., 2016. A training method to improve police ssf of force decision making: A Randomized Controlled Trial. SAGA Open 6 (2).


Ayoub, L.H., Pooler, T., 2015. Coming Home to Harlem: A Randomized Controlled Trial of the Harlem Parole Reentry Court. Center for Court Innovation.


Baldwin, K., Muyengwa, S., 2014. Impact Evaluation of Supporting Traditional Leaders and Local Structures to Mitigate Community-level Conflict in Zimbabwe. USAID.


Berk, J., 2008. Does work release work?


55


Duwe, G., Goldman, R.A., 2009. The Impact of Prison-Based Treatment on Sex Offender RecidivismThe Impact of Prison-Based Treatment on Sex Offender Recidivism:


Elk, R. van, Steeg, M. van der, Webbink, D., 2013. The effects of a special program for multi-problem school dropouts on educational enrolment, employment and criminal behavior; Evidence from a field experiment. CPB Netherlands Bureau for Economic Policy Analysis.


Killian, D., Agee, J., 2014. The contribution of legal awareness raising toward access to justice and stability in Afghanistan. USAID.


Picard-Fritsche, S., Rempel, M., Reich, W., Farley, E., Kerodal, A., 2016. Implementing Evidence Based Assessment and Treatment Matching. Center for Court Innovation.


Piza, C., Souza, A.P., 2016. Short and long-term effects of a child-labor ban. Sao Paulo School of Economics.


83


https://doi.org/10.1016/j.jcrimjus.2017.06.002


https://doi.org/10.1007/s10940-016-9308-7

https://doi.org/10.1007/s10896-007-9116-9


https://doi.org/10.1037/a0013881


https://doi.org/10.1080/07418825.2019.1568520


Slough, T., Fariss, C., 2017. Impact Evaluation of USAID Haiti PROJUSTICE Program Pretrial Detention Component. USAID.


Yokum, D., Ravishankar, A., Coppock, A., 2017. Evaluating the Effects of Police Body-Worn Cameras: A Randomized Controlled Trial. The Lab @ DC.


Included impact evaluations (ongoing)


Morse, B.S., 2020. Pre-Analysis Plan: Can Community Policing Build Trust and Improve Police Effectiveness?


Included systematic reviews (completed)

Abt, T., Winship, C., 2016. What works in reducing community violence: A meta-review and field study for the Northern Triangle, USAID.


**Included systematic reviews (ongoing)**


Bradshaw, D., McLaughlin, K., Muldoon, O., 2017. Parenting programs for reducing negative outcomes for incarcerated fathers and their families, Campbell Systematic Reviews.


Vigurs, C., Quy, K., Rivas, C., Dickson, K., Schucan-Bird, K., 2018. Motivation enhancing approaches to reduce recidivism and increase program attendance for men court ordered to attend domestic violence perpetrator programs: a systematic review, Campbell Systematic Reviews.

Other references


Center for Evidence-Based Crime Policy. *Evidence-Based Policing Matrix*. George Mason University. Available at: https://cebcp.org/evidence-based-policing/the-matrix/


Other publications in the 3ie Evidence Gap Map Report Series

The following papers are available from http://www.3ieimpact.org//evidence-hub/evidence-gap-maps


An estimated 1.5 billion people have no access to effective justice leaving nearly 60 per cent of justice problems unresolved. Effective rule of law ensures that laws and the justice institutions, actors, and processes that support them are responsive to and inclusive of the needs of all people in society. While there has been an increased investment in programs to promote rule of law, there is a need to better understand the evidence base. The authors of this report analyze interventions that strengthen and build linkages between systems, services, and society, which can improve the prevention of and solutions to justice problems.