

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 working papers

3ie working papers cover a range of content. They may focus on current issues, debates and enduring challenges facing development policymakers, program managers, practitioners and the impact evaluation and systematic review communities. Policy-relevant papers in this series synthesize or draw on relevant findings from mixed-method impact evaluations, systematic reviews funded by 3ie, as well as other rigorous evidence to offer new analyses, findings, insights and recommendations. Papers focusing on methods and technical guides also draw on similar sources to help advance understanding, design and use of rigorous and appropriate evaluations and reviews. 3ie also uses this series to publish lessons learned from 3ie grant-making.

About this working paper

Authors of this paper “Half-time for SDG evidence generation: Insights from the Development Evidence Portal,” provide an overview of state of effectiveness evidence across the 17 Sustainable Development Goals. The authors have analyzed impact evaluations and systematic reviews available in 3ie’s Development Evidence Portal published between 1990 to 2022. This report is produced by 3ie with funding from the German Institute for Development Evaluation (Deval) and the Federal Ministry for Economic Cooperation and Development (BMZ).

The content of this report is the sole responsibility of the authors and does not represent the opinions of DEval, BMZ, 3ie, its donors or its Board of Commissioners. Any errors or omissions are also the sole responsibility of the authors. Please direct any comments or queries to the corresponding author, Mark Engelbert, at mengelbert@3ieimpact.org

Suggested Citation: Engelbert, Mark, Zafeer Ravat, Cem Yavuz, Durgadas Menon, Constantin Grywatz, Kathrin Wolf, Half-Time for SDG Evidence Generation, 2023. 3ie Working paper 59. New Delhi: International Initiative for Impact Evaluation (3ie). Available at: <https://doi.org/10.23846/WP0059>

Designed by: [Green Communication Design Inc](#)

© International Initiative for Impact Evaluation (3ie), 2023

HALF-TIME for SDG EVIDENCE GENERATION

Insights from the Development Evidence Portal

Mark Engelbert*, International Initiative for Impact Evaluation (3ie)
and University of East Anglia (UEA)

Zafeer Ravat, 3ie and UEA

Cem Yavuz, 3ie and UEA

Durgadas Menon, 3ie

Constantin Grywatz, German Institute for Development Evaluation (DEval)

Kathrin Wolf, DEval

December, 2023



With financial support from the:



Abstract

The Sustainable Development Goals (SDGs) provide the organising agenda for global development.

To achieve the SDGs, the international community will need rigorous evidence about which programmes and policies are most effective in improving indicators related to the goals. The Development Evidence Portal (DEP), maintained by the International Initiative for Impact Evaluation (3ie), is a database that gathers such evidence for low- and middle-income countries, in the form of impact evaluations and systematic reviews. In this report, we analyse data from the DEP to answer the following questions: What is the state of the evidence base for the SDGs, and where are the gaps in the evidence? Given resource limitations, the DEP's data on health sector interventions are incomplete, and this report therefore covers only non-health interventions (though we are still able to analyse the state of evidence for the impacts of non-health interventions on health outcomes).

We provide an overview of the evidence base across the SDG agenda, followed by a series of dashboards summarising the state of the evidence for each of the 17 SDGs between 1990 and 2022. We find that there are substantial evidence bases, built up over several decades, for SDGs focusing on poverty (SDG 1), health (SDG 3), and education (SDG 4). SDGs focused on environmental sustainability (SDGs 12–15) have received limited attention in the impact evaluation literature, apart from a recent surge in evaluations of environmental policies in China. Across nearly all SDGs, there is a limited supply of high-quality evidence synthesis. We also find a striking lack of attention to gender and equity considerations in evaluation research for all SDGs.

We recommend that the SDGs be given a more central role in setting research priorities and coordinating research activities in global development. Just as the SDGs set firm targets against which progress is monitored, global producers and users of evidence should set and monitor progress towards targets for creating a robust evidence base for interventions to achieve the SDGs. To create such an evidence base, researchers and research commissioners will need to prioritise and dedicate resources to efficiently generating evidence for neglected SDGs (e.g., SDGs 6, 7, 11, and 14), conducting high-quality synthesis, and gathering evidence on whether global development interventions reduce or exacerbate existing inequalities. Such efforts will ensure that the evidence base has sufficient coverage of the core themes of sustainable and equitable development to support sound decision making, even after the SDG agenda closes in 2030.

Table of contents

Abstract

Key findings: SDG evidence at a glance

Recommendations

List of abbreviations

1. Why the world needs evidence for the Sustainable Development Goals
2. How we used the DEP to create this report
3. The big-picture view of SDG evidence
 - 3.1 Evidence for the “People” and “Peace” SDGs is on a different trajectory from “Planet” and “Prosperity”
 - 3.2 Where is the evidence?
 - 3.3 What interventions and outcomes are studied?
 - 3.4 Who is funding the research?
 - 3.5 Does the evidence base align with aid spending?
 - 3.6 Most synthesis evidence is subject to methodological limitations
 - 3.7 Attention to gender and equity is lacking in the evidence base
 - 3.8 Evidence and state fragility
4. The state of the evidence for the 17 SDGs
 - SDG 1: No poverty
 - SDG 2: Zero hunger
 - SDG 3: Good health and well-being
 - SDG 4: Quality education
 - SDG 5: Gender equality
 - SDG 6: Clean water and sanitation
 - SDG 7: Affordable and clean energy
 - SDG 8: Decent work and economic growth
 - SDG 9: Industry, innovation and infrastructure
 - SDG 10: Reduced inequalities
 - SDG 11: Sustainable cities and communities

SDG 12: Responsible consumption and production

SDG 13: Climate action

SDG 14: Life below water

SDG 15: Life on land

SDG 16: Peace, justice and strong institutions

SDG 17: Partnerships for the goals

References

Appendix A: Methods used to create this report

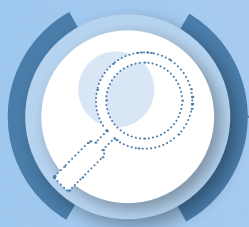
- A.1 The types of evidence contained in DEP
- A.2 How the DEP is updated
- A.3 Why this report excludes evidence from the health sector
- A.4 How we classified studies’ interventions and outcomes for this report
- A.5 How we capture studies’ attention (or lack of attention) to gender and equity
- A.6 How we assess the quality of systematic reviews
- A.7 How we assessed the availability of evidence where it is most needed
- A.8 ODA estimates for the SDGs

Appendix B: Some caveats to this report

- B.1 Studies from non-L&MIC countries are mostly excluded
- B.2 Studies might be categorised to SDGs in multiple ways
- B.3 Many studies have been categorised to multiple SDGs
- B.4 Good country-level data on SDG-related indicators are sometimes unavailable
- B.5 Limitations to comparisons between evidence and other indicators

Appendix C: Common SDG combinations

Key findings: SDG evidence at a glance



Evidence is often lacking where it is most needed.

We compared the amount of evidence available for specific SDGs in particular countries to key country-level indicators related to those SDGs. For example, we compared the availability of evidence related to SDG 1 (No Poverty) with the percentage of a country's population living in extreme poverty. A frequent pattern in these comparisons is that there tends to be more evidence from countries where conditions are worse; however, evidence is often scarce from countries where problems are at their most extreme. This is likely because countries with the worst challenges are environments where rigorous impact evaluations (IEs) are difficult to conduct. Nevertheless, this pattern has the unfortunate implication that the global community frequently lacks evidence that can support decision-making in contexts where evidence-informed interventions are most essential for attaining the SDGs. (See the dashboards for [SDG 1](#), [SDG 2](#), [SDG 4](#), [SDG 5](#), [SDG 10](#), [SDG 13](#), and [SDG 16](#) for these comparisons.)



The SDGs in the “People” group¹ (SDGs 1–5) have substantial evidence bases that have been built over the last three decades.

Four of the top five SDGs, in terms of the number of IEs, are from the “People” group, each with upwards of 1,000 IEs. The evidence bases for “Planet” (SDGs 6 and 12–15) and “Prosperity” (SDGs 7–11) have grown rapidly in the last three to five years, but research has focused narrowly on regulatory regimes in China. There is still very little evaluation research related to SDGs 12 (Responsible Consumption and Production), 14 (Life Below Water), and 15 (Life on Land). (See [Section 3.1](#).)

¹ The “5Ps” framework groups the 17 SDGs into five thematic categories: People, Prosperity, Planet, Peace, and Partnerships [1].



However, in the last few years, there has been a levelling off or decline in the production of evidence on SDGs in the “People” group.

These are SDGs for which many interventions happen at a local level, meaning they are often evaluated using randomised evaluation methods that involve primary data collection. We may be seeing a dip in production due to the COVID-19 pandemic, which curbed much primary data collection. (See [Section 3.1](#).)



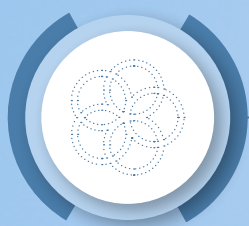
SDGs in the “Planet” (6 and 12–15) and “Prosperity” (7–11) groups have been traditionally less studied, but have seen recent explosions of research, mainly focused on China.

In the last four to five years, there have been rapid increases in the production of evidence related to SDGs 7 (Affordable and Clean Energy), 8 (Decent Work and Economic Growth), and 13 (Climate Action). This explosion in research concerning certain SDGs is almost entirely due to a meteoric rise in studies on Chinese policies, conducted by researchers based in China and funded by Chinese government agencies. The interventions related to these SDGs are typically evaluated using quasi-experimental methods, which can often be employed with secondary datasets. (See [SDG 7](#), [SDG 8](#), and [SDG 13](#).)



SDG 16 (“Peace”) has a moderately sized evidence base, and has also seen a recent dip in evidence production.

Like the SDGs in the “People” group, SDG 16 was on an upward trend of evidence production, but has seen a decline since 2020. Notably, SDG 16’s share of the overall evidence base falls significantly below its share of aid spending. (See [Sections 3.1](#) and [3.6](#).)



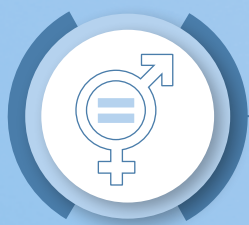
We identified very little evidence for SDG 17 (“Partnerships”).

The Development Evidence Portal contains only 26 IEs classified to SDG 17, and no systematic reviews (SRs). We have adopted a conservative approach to classifying studies to SDG 17, whereby studies must have a specific focus on international partnership (see [Appendix B.2](#) for details). We find that little such evidence exists, likely because many partnerships take the form of one-off agreements that are difficult to evaluate using counterfactual-based methods.² (See [SDG 17](#).)



Few SDGs have large pools of systematic reviews to provide synthesised findings, and there is an even more limited supply of high-quality reviews.

Some of the SDGs with larger IE literatures – such as SDGs 1–4 and SDG 6 – also have multiple reliable SRs. However, even for SDGs with larger synthesis literatures, the pool of available SRs is insufficient to cover the range of interventions represented in the IE literature. Other SDGs – particularly 11–13 – have few reliable SRs at all. Across the full evidence base, most SRs are subject to major methodological limitations. (See [Section 3.5](#).)



There is limited attention to gender and equity considerations in evaluation research.

Only a small proportion of studies take steps like conducting subgroup analysis to determine whether interventions benefit the most vulnerable or including measures of inequality among the study’s outcomes. For every SDG apart from SDG 5 (Gender Equality), over 65 per cent of studies do not account for gender or equity in their research design. For nine of the SDGs, more than four in five IEs do not address gender or equity. (See [Section 3.7](#).)

² Other resources exist that consider the evidence base for SDG 17 from a different angle. The SDG Synthesis Coalition has recently conducted a systematic review of evidence for SDG 17, having set a broader scope for studies related to the SDG [2].



Evidence is concentrated in countries with moderate levels of fragility.

When countries are grouped into quintiles based on their levels of state fragility, 83 per cent of IEs have been conducted in countries in the middle three quintiles. However, there are significantly more studies (941 studies; 12%) conducted in the most fragile countries than in the least fragile countries (405 studies; 5%). (See [Section 3.8.](#))



Across the full dataset, the most studied intervention type is agricultural extension and training (643 IEs; 9% of all IEs), most of which interventions have taken place in Sub-Saharan Africa.

Other frequently evaluated interventions include environmental regulation (609; 8%, mostly Chinese policies); cash transfers (conditional [489; 7%] and unconditional [340; 5%]); and access to credit (370; 5%).³ (See [Section 3.3.](#))



The most commonly measured outcomes are those related to household income and assets (1,228; 17% of IEs), followed by employment and productivity measures (708; 10%), food security (653; 9%), agricultural production (634; 9%), and household consumption (548; 7%). (See [Section 3.3.](#))

³ Because evaluations of health interventions are not included in the data for this report, this is a list of the most commonly evaluated non-health interventions.

Recommendations



Global development efforts need an “evidence for sustainable development” agenda.

A likely reason for the lopsided nature of the SDG evidence is that there has been no unifying agenda to generate evidence for the SDGs (or the Millennium Development Goals before them), in the same way that there is an agenda for achieving the SDGs. Going forward, the SDGs should be more central to prioritisation discussions among evaluation researchers and funders focused on global development. A dedicated “evidence for sustainable development” agenda, adopted by researchers and research funders, could be the necessary coordinating force. Such an agenda should be grounded in the SDGs, but flexible enough to ensure a well-rounded evidence base for sustainable development beyond Agenda 2030.



Greater attention is needed to producing evidence to support the “Planet”-focused SDGs.

The recent explosion of China-focused evaluations on environmental topics has provided a template for robust research programmes in this area: quasi-experimental methods, leveraging administrative and other secondary data sources, can be used to evaluate the effects of regional and national environmental policies. This approach can be applied much more broadly, and combined with traditional randomised evaluation methods, to build an evidence base for environmental interventions across a variety of contexts.



Innovative, low-cost evaluation methods must see widespread adoption to fulfil the evidence needs of the SDG agenda.

The evidence gaps for many SDGs are quite large. Traditional randomised controlled trials (RCTs) are an important source of rigorous evidence, but tend to be expensive and are difficult to conduct in many contexts. Thus, RCTs are unlikely to be able to provide all the evidence needed to support the SDG agenda. The significant dip in “People”-focused IEs in the years since the onset of the COVID-19 pandemic may indicate that the evaluation field has been overly reliant on primary data collection for studying interventions related to certain SDGs. Alternative approaches, such as remote sensing, can be leveraged to expand the evidence base in a cost-effective way, even as primary data collection and attendant evaluation designs become feasible again.



Achieving the SDGs requires dedicated efforts to generate evidence where it is needed the most.

Our analysis finds that countries struggling the most with SDG-related indicators often have the most limited evidence bases. It is especially important that policy and programming decisions in these contexts be supported by rigorous evidence. Conducting evaluations in these often challenging contexts will require concerted efforts on the part of evaluators and research funders. This, again, may require exploring innovative ways to conduct rigorous evaluations in contexts where traditional evaluations are difficult to implement. Also, local researchers will often be best placed to lead evaluations in challenging contexts, but may face barriers such as limited access to funding sources (currently, two-thirds of authors on studies conducted in high-fragility countries are based in high-income countries). Initiatives to develop and tap into local research capacity will help close these types of evidence gaps.



There is a need to invest in high-quality systematic reviews, especially beyond the health field.

There are several SDGs for which the number of SRs is particularly low relative to the number of IEs, and may, therefore, be good candidates for synthesis. These include SDGs 8 (Decent Work and Economic Growth), 9 (Industry, Innovation, and Infrastructure), 11 (Sustainable Cities and Communities), and 13 (Climate Action). Such synthesis could add significantly to our understanding of challenges related to these SDGs. Because recent China-focused studies dominate the IE literature for these SDGs, it would be necessary to exercise caution in extrapolating from Chinese studies to other contexts.



Much greater attention is needed to considerations of gender and equity in evaluation research.

Researchers should follow established guidelines for equity-sensitive research [3]–[5], and commissioners of research should require that evaluations and syntheses consider disparate impacts across groups. Equity considerations should be incorporated into study designs from the earliest stages: Researchers should consider how an intervention might affect vulnerable populations differently, and ensure that sample sizes provide adequate power for meaningful subgroup analysis to identify any disparate effects on those populations (this, in turn, requires that funders commit adequate resources to support equity-sensitive research). Adopting an equity-sensitive analytical framework – for example, a theoretical model of how an intervention might alleviate or exacerbate existing inequalities – can help identify important inequality-related outcomes that should be measured. Finally, mixed-methods research combining quantitative and qualitative approaches can shed light on the mechanisms by which interventions affect different populations differently.



List of abbreviations

3ie	International Initiative for Impact Evaluation
CRS	Creditor Reporting System
DAC	Development Assistance Committee
DEP	Development Evidence Portal
GHGs	Greenhouse gases
HICs	High-income countries
IEs	Impact evaluations
L&MICs	Low- and middle-income countries
LAC	Latin America and the Caribbean
MENA	Middle East and North Africa
NGO	Non-governmental organization
ODA	Official development assistance
OECD	Organization for Economic Cooperation and Development
SDGs	Sustainable Development Goals
SRs	Systematic reviews
SSA	Sub-Saharan Africa
USAID	United States Agency for International Development
WaSH	Water, sanitation, and hygiene

Why the world needs evidence for the Sustainable Development Goals

Since their adoption in 2015, the SDGs have provided the basis for the global community's efforts to eradicate poverty and ensure peace and prosperity for all.

As we are now at the midway point of the 15-year timeline set for achieving most targets under the SDGs, it is an opportune time to evaluate progress towards the SDGs. Progress has been inadequate and many SDG targets remain far out of reach [6]. Observers are calling for renewed efforts to make progress towards the SDGs, and in particular to place the SDGs related to environmental sustainability on equal footing with those emphasising economic growth and well-being [7].

With achieving the SDGs estimated to require investments of USD 2.5 trillion or more per year [8], it is crucial that resources invested in attaining the SDGs be spent wisely. Resources should go towards policies and programmes that actually induce changes in the direction of the SDG targets, rather than going to interventions that seem intuitively promising but fail to yield their intended effects. Thus, to achieve these goals, policymakers

will need reliable evidence about the effectiveness of various strategies for promoting sustainable development. Key sources of such evidence are rigorous impact evaluations (IEs) and systematic reviews (SRs) of these evaluations. Recognising the need for rigorous effectiveness evidence on the SDGs, the United Nations Development Program has called for syntheses of "rigorous evidence on what has worked, why and where to advance and accelerate the SDG achievements" [9].





Decisions based on evidence can magnify the impacts of development interventions by helping decision makers choose what to implement and how to implement it.



For example, in 2011 a microfinance institution in Vietnam conducted a pilot to combine microfinance with business trainings targeted to women. When an IE of the pilot found positive impacts on business performance, the microfinance institution scaled up the practice of integrating gender and business trainings with their regular meetings, thus exposing a larger population of women to the benefits of the intervention [10].

Evidence can also support decisions that prevent scarce resources from being wasted.



In 2010 an NGO launched a pilot to train women in Ghana on building and using cook stoves whose design was expected to reduce fuel use and indoor air pollution. But when a randomised evaluation found no effect on these outcomes, the NGO declined to expand the pilot and began looking for a better stove design [11].

When ineffective interventions are discontinued, valuable resources are freed, creating the opportunity to invest in something with greater impact.



Why impact evaluations and systematic reviews?

Evidence comes in many forms, all of which can be valuable tools for decision-making. Existing mechanisms for monitoring progress toward the SDGs provide valuable information about where efforts are falling short. However, an inherent limitation of such monitoring data is that they do not reveal which alternative approaches would yield better progress. Therefore, we emphasise two sources of evidence that can provide this crucial information to support decision-making in global development.

The first is **impact evaluations**, which are studies evaluating the effects of a particular policy or programme. In particular, we focus on quantitative IEs that use a counterfactual design [12].





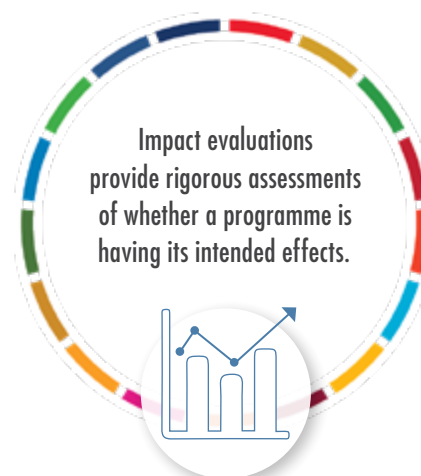
These evaluations use randomisation and other approaches to provide a rigorous answer to the question: *what would have happened if this programme/policy had not been implemented?*⁴ Thus, these evaluations can indicate whether development interventions are having their intended effects. This information is crucial in deciding which interventions are the best use of scarce resources. Just as drugs and other medical interventions must be rigorously scientifically tested before being approved for widespread use, the effectiveness of social and economic interventions should also be assessed using scientific methods [13].

The second key source of evidence is **systematic reviews**, which apply rigorous scientific methods to the process of collecting all the literature on a topic and analysing it systematically. The results of individual IEs of the same intervention often differ, even when conducted in seemingly similar contexts. Thus, it is often valuable to consider evaluations of similar interventions in concert with one another to be able to draw a holistic picture of the intervention's effects. When conducted according to best practices, SRs provide the best basis for drawing general conclusions about how well an intervention has worked in the past, and how well it can be expected to work in the future, given the presence or absence of particular contextual features.

A substantial body of high-quality SRs, covering a broad range of interventions and outcomes, is therefore an essential feature of a mature evidence base for decision making in a thematic area.

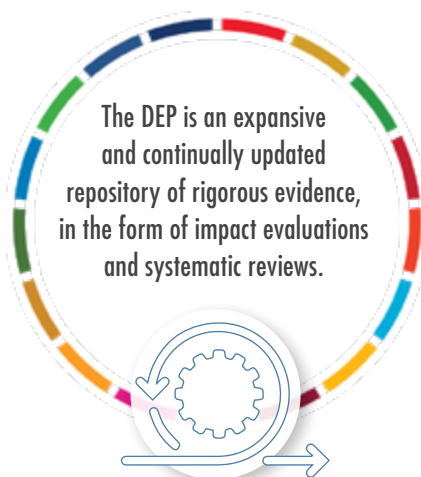
A resource for evidence-informed decision-making

The [Development Evidence Portal \(DEP\)](#) from the International Initiative for Impact Evaluation (3ie) brings together existing evaluations and



syntheses on the effectiveness of social and economic development interventions in low- and middle-income countries (L&MICs). The DEP is an expansive and continually updated repository of rigorous evidence, and forms a key part of the architecture 3ie has developed over the last decade to support evidence-informed decision-making in international development. The DEP is a curated database that aggregates evaluation and synthesis research from a diverse range of sources.⁵

With over 12,000 IEs and 1,000 SRs as of September 2023, the DEP constitutes an important source of the kind of evidence necessary to support achieving the SDGs.



⁴ The most common and most straightforward approach to evaluation is to attempt to discover what would have happened if the intervention were absent altogether. But IEs can also answer questions such as: *What would have happened if the intervention were implemented differently?*

⁵ More information on 3ie's methods for populating and updating the DEP is available in [Appendix A: Methods used to create this report](#).



How we used the DEP to create this report

In this report, we analyse the evidence base for attaining the SDGs, drawing on the data collected for the DEP.

We highlight trends in the growth of the evidence base over time, discuss major gaps, and offer recommendations for how the field can build the evidence base the SDG agenda requires. Our focus in this report is on what evidence is and is not available, rather on what the evidence tells us about how to achieve the SDGs. However, we highlight key insights from recent, reliable SRs in the DEP on the effectiveness of particular interventions relevant to specific SDGs.

The DEP study coding procedures have created a rich dataset with information on each study's thematic area(s), geographic focus, funding sources, interventions, outcomes, and attention to gender and equity. We developed this data collection schema to capture many of the most important

features of the evidence base. This allows us to identify many key trends and patterns. However, it is important to acknowledge that a number of important topics are not reflected in our dataset, and therefore fall outside the scope of this report.

The results presented in this report are divided into two main sections: [Section 3](#) provides an overview of patterns in the evidence across the 17 SDGs. This is followed by a series of 17 dashboards summarising key insights from our data on the evidence for each of the SDGs.

Further details on the methods used in this report, and on some of the report's limitations, can be found in [Appendix A](#) and [B](#).



The big-picture view of SDG evidence

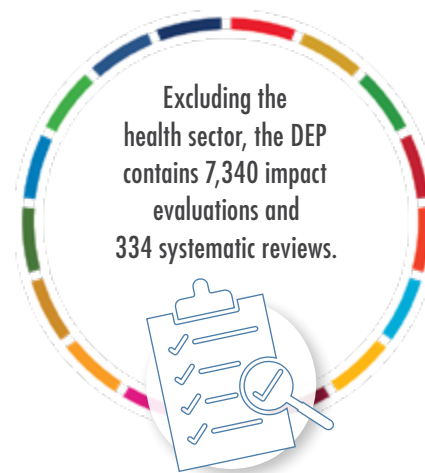


3.1 Evidence for the “People” and “Peace” SDGs is on a different trajectory from “Planet” and “Prosperity”

As of August 2023, the DEP contains, excluding the health sector, 7,340 IEs and 334 SRs. The amount of evidence available varies widely across SDGs (Figure 1). There are over 1,000 IEs each for SDGs 1, 2, 4, and 8. These SDGs also tend to have the largest number of SRs: the number of reviews for these SDGs ranges from 38 to 97, whereas most other SDGs have fewer than 20 SRs. A notable exception is SDG 6 (water & sanitation), which has 50 SRs and 360 IEs, a much higher ratio of synthesis than observed for any other SDG.

After many years of steady growth, the production of IEs related to the “People” and “Peace” SDGs has declined since 2020 (Figure 2). Beginning in 2016, there has also been a rapid acceleration of evidence generation for the “Prosperity” and “Planet” SDGs. The number of IEs for “Partnerships” (SDG 17) is too small to discern any meaningful patterns in the production of evidence over time.

The rapid decline in “People”- and “Peace”-focused evaluations, without a similar decline in “Prosperity” and “Planet”, may be partially explained by the fact that the sorts of interventions related to the “People” and “Peace” SDGs are those that are implemented at a local level, using primary data collected by the research team. The recent dip may result from the limitation of primary data collection activities during the COVID-19 pandemic.⁶ The interventions relevant to the “Planet” and “Prosperity” SDGs, on the other hand, are broader policies that are often evaluated using quasi-experimental designs applied to secondary datasets; the pandemic would not have curtailed these evaluations in the same way.

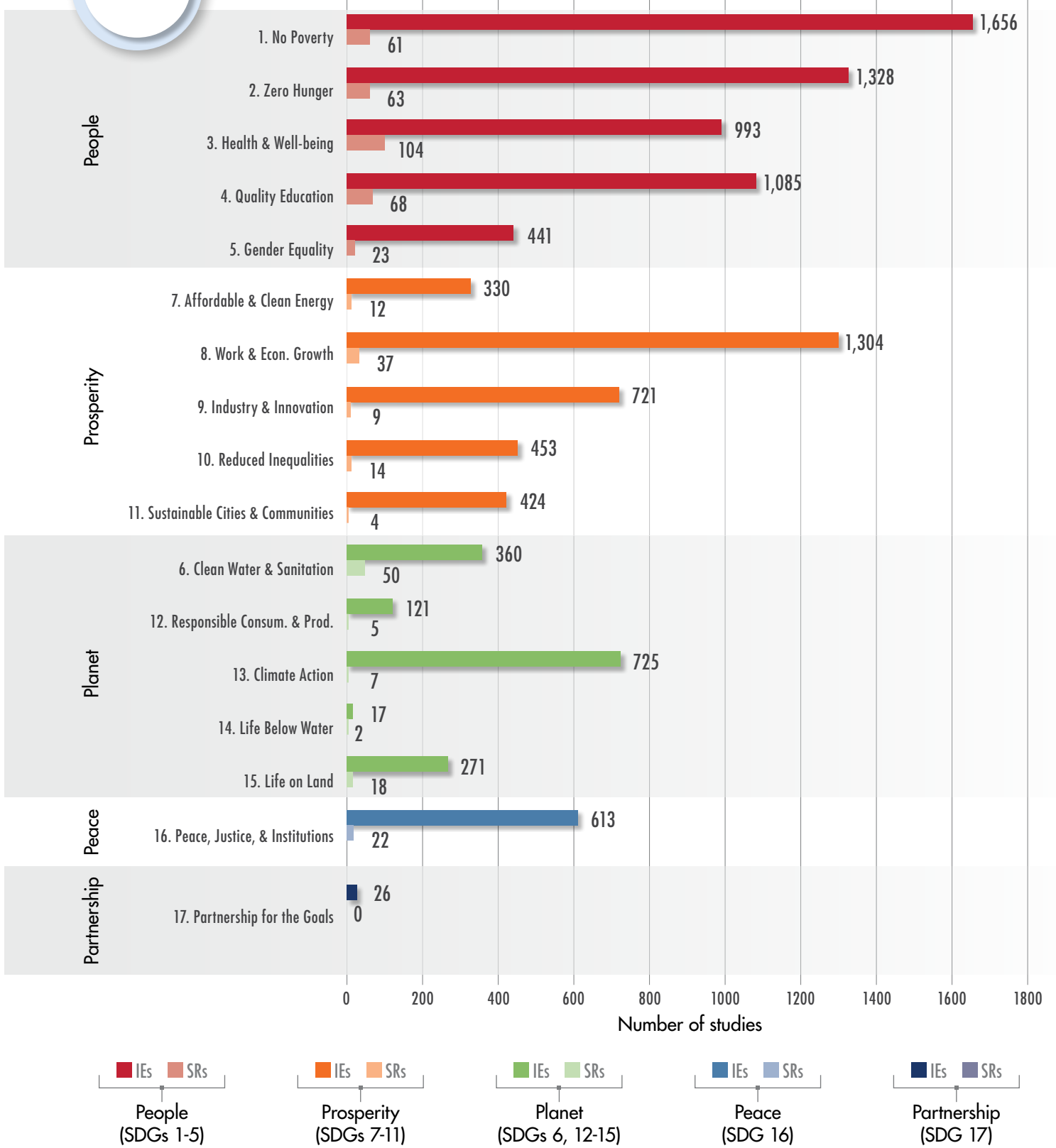


⁶ This is not the only possible explanation for the observed pattern. For example, it could also be that a perception has emerged among researchers and research funders that research on “People”-related topics is “saturated” and there is now less to be gained from further research. In general, our data do not permit us to adjudicate among different hypotheses about the causes of observed trends.



Figure 1

Large discrepancies in evidence bases across the SDGs



Source: DEP data as of 23 August 2023. Note: some studies are classified with more than one SDG, so the sum of studies by SDG is greater than the number of unique studies in DEP.

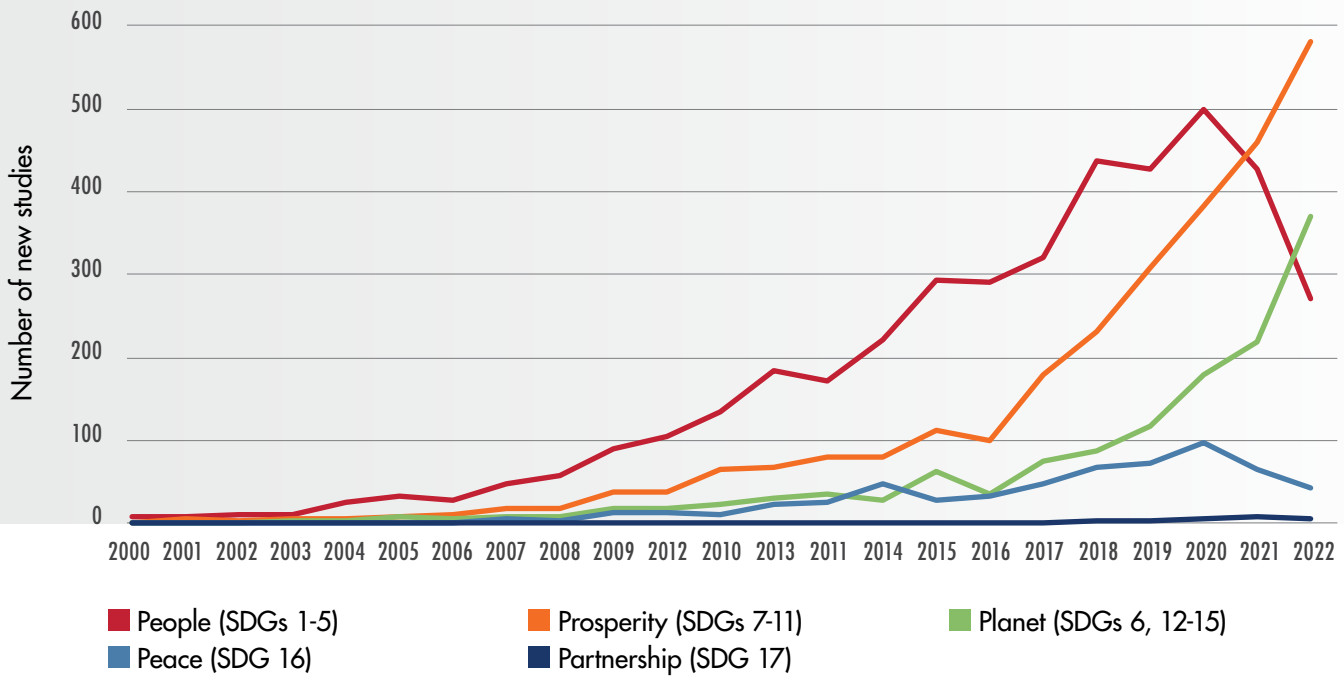


The number of SRs per SDG published each year is generally too small to detect meaningful trends over time – the number of SRs relevant to an SDG in a given year is rarely more than ten, and usually less than five. However, there is no clear increase in the production of SRs – either as a whole or within any

SDG or group of SDGs – in the last four to five years. This is despite the fact that the primary literature has increased for many topics, and that the pandemic should not, in principle, have posed any serious barriers to conducting rigorous evidence synthesis.

Figure 2

Evidence for “People” SDGs is slowing as “Prosperity” and “Planet” surge





3.2 Where is the evidence?

In the sections below dedicated to particular SDGs, we often present the breakdown of IEs by region. These analyses will be more informative when they are compared to the breakdown of studies by region in our entire dataset. The region with the most IEs in our dataset is East Asia, followed closely by Sub-Saharan Africa (Figure 3).⁷ Latin America and South Asia have moderately sized evidence bases, while there is little evidence from Europe/Central Asia or the Middle East and North Africa (MENA).

The rate of evidence production across regions has varied over time (Figure 4). Prior to 2012, Latin America and the Caribbean (LAC) was the most popular region for IEs, and the production of evidence in LAC has experienced a steady increase from about 15–30 IEs per year in the mid-2000s to 100 or more in 2018–2020. The production of studies focused on Sub-Saharan Africa (SSA) began rising rapidly around 2008: the number of IEs from SSA rose from 17 in 2008 to a peak of 247 in 2018, and it was the region with the most IEs from 2012 to 2019.

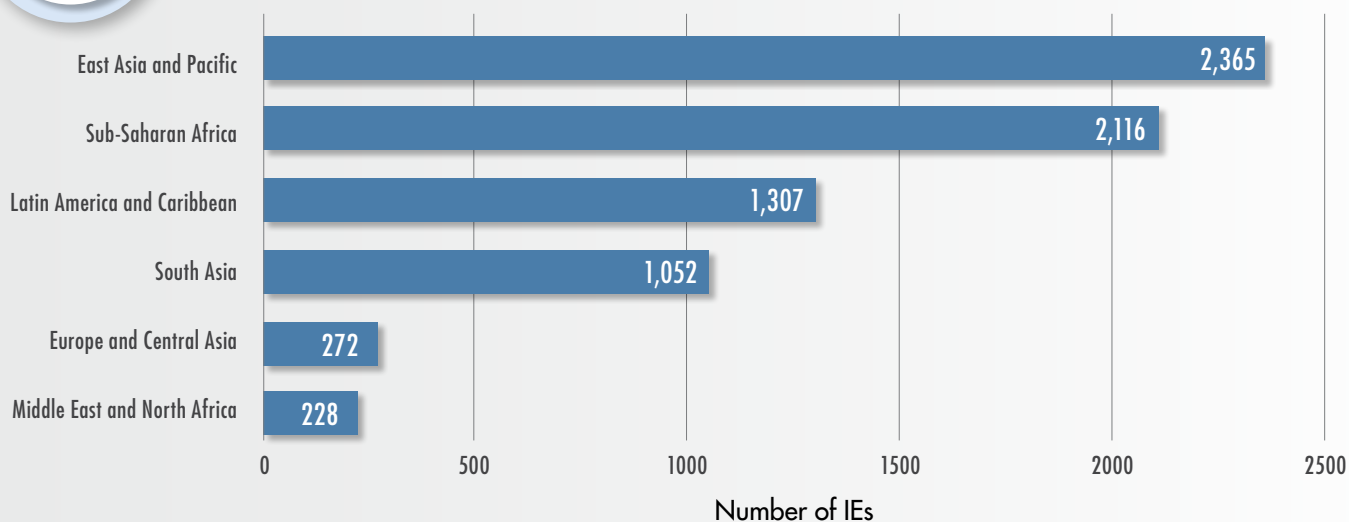
East Asia and the Pacific saw moderate increases in the number of IEs per year between 2008 and 2016, and has seen an extremely rapid increase since then (from 107 IEs in 2017 to 692 in 2022).

Among individual countries, China has by far the greatest number of IEs (1,835), more than three times as many as the second-place country – India, with 562 (Figure 5). While there have been many IEs conducted in Sub-Saharan Africa as a whole, this evidence is heavily concentrated in a handful of countries, including Ethiopia, Kenya, Uganda, Ghana, Nigeria, and Malawi. Many other countries in Sub-Saharan Africa have limited evidence bases. Disparities between countries are also apparent in other regions. In East Asia, only China, Indonesia, the Philippines, and Vietnam have more than 50 IEs. In Latin America, evidence is concentrated in Brazil, Mexico, Colombia, and Peru.



Figure 3

Evidence is concentrated in East Asia and Sub-Saharan Africa



⁷ Recall that the dataset used in this report excludes Health sector studies; if these studies were counted, the evidence base from Sub-Saharan Africa would be larger than that from East Asia by about 1,000 IEs.



Figure 4

After a recent surge, East Asia has supplanted Sub-Saharan Africa as the region with the most IEs

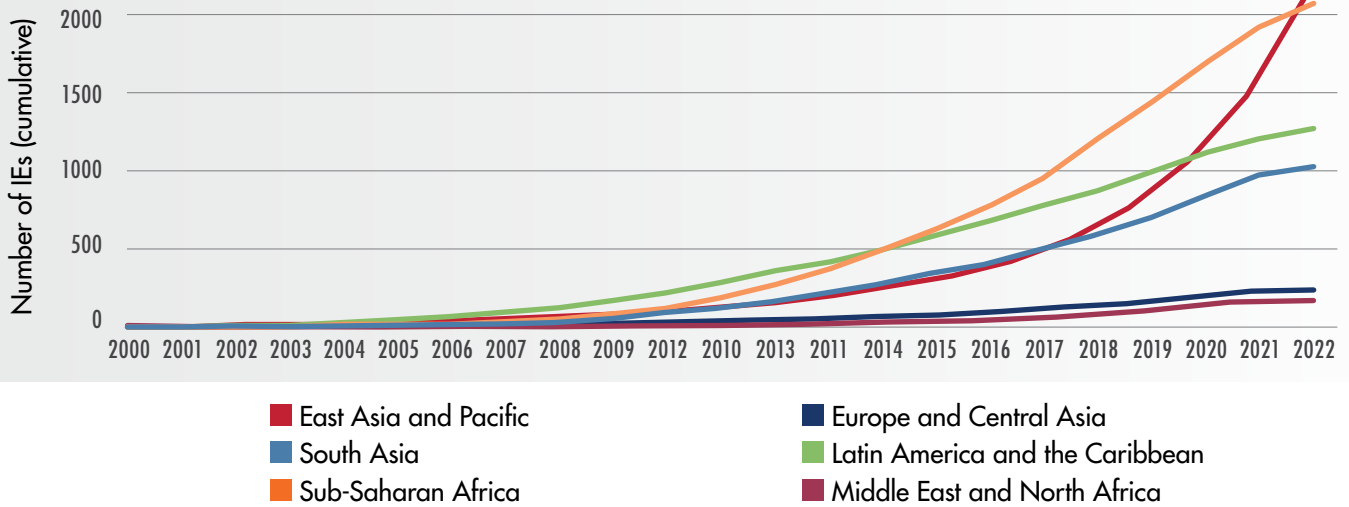
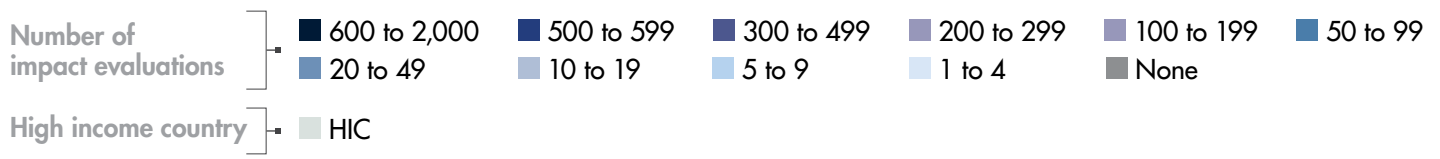
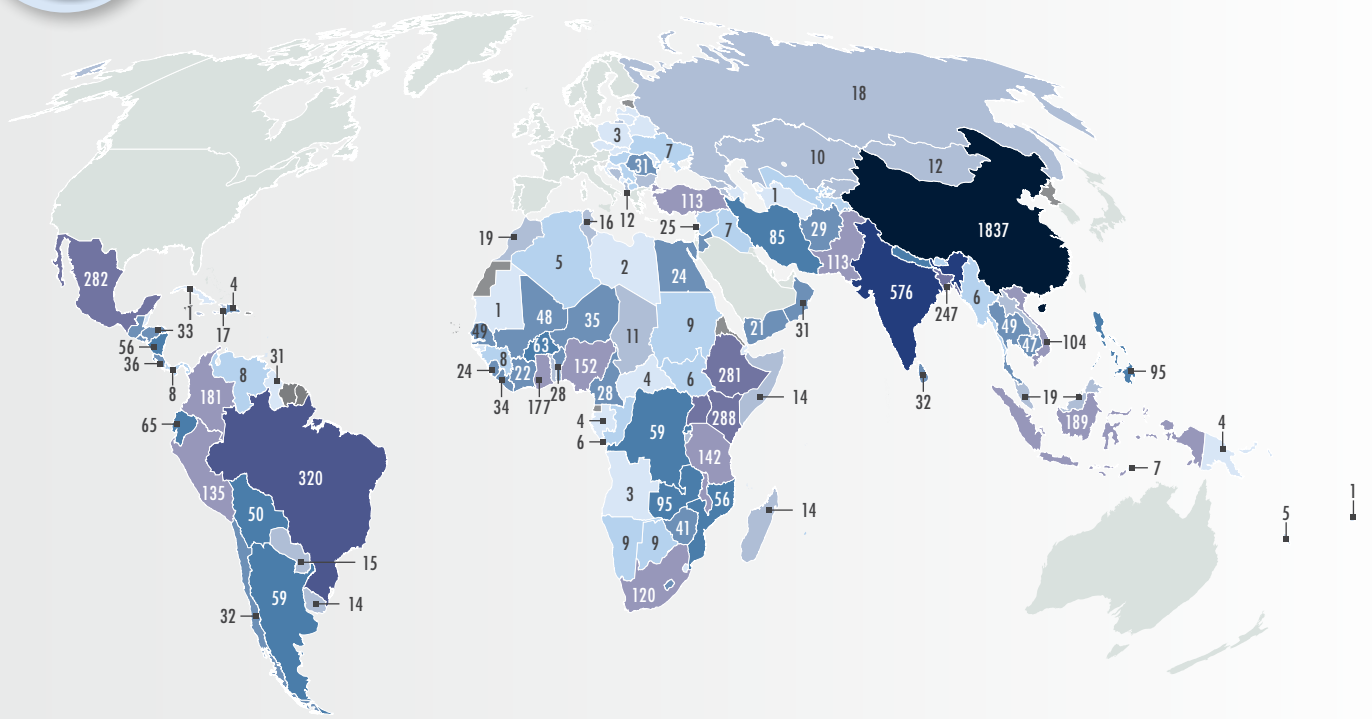


Figure 5

Large discrepancies in evidence across countries





3.3 What interventions and outcomes are studied?

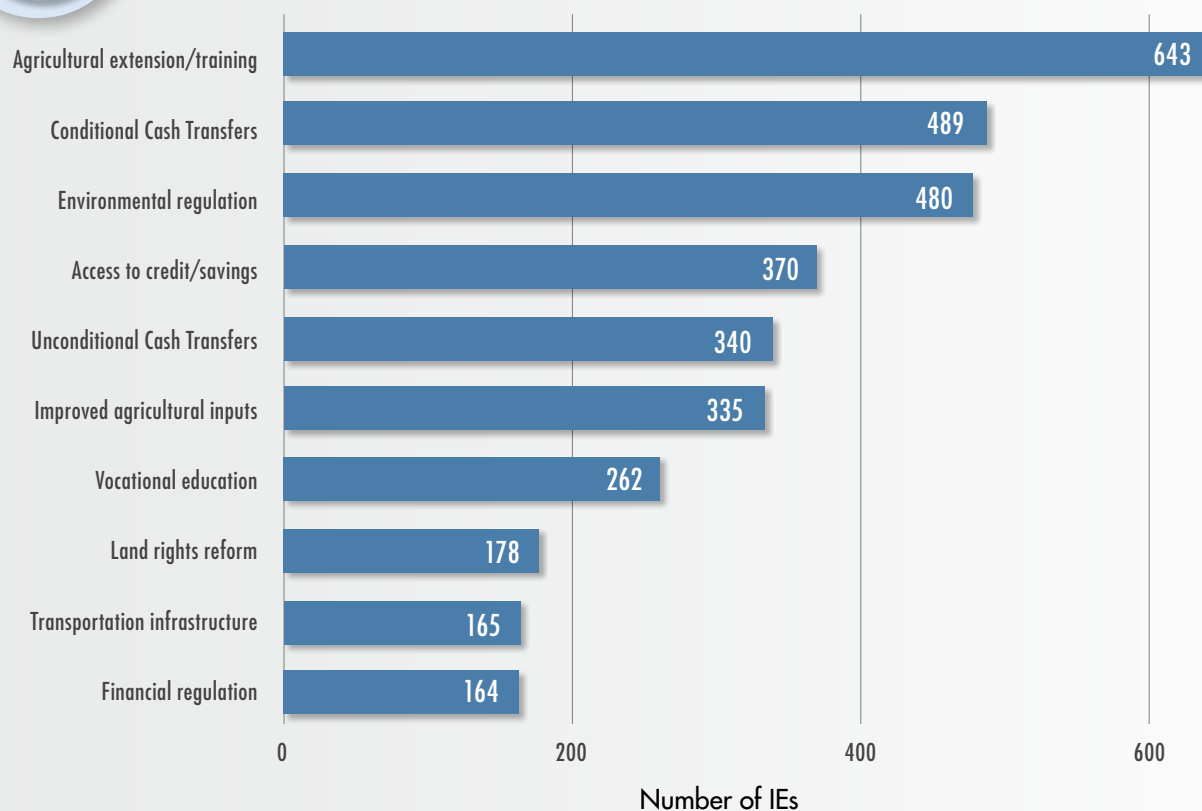
A handful of intervention types account for a significant proportion of the IE literature.⁸ About 40 per cent of IEs (3,016 studies) evaluate one of the 10 interventions shown in Figure 6.⁹ The interventions most commonly studied in IEs cover a number of themes, including agricultural production, regulatory schemes, and education. The most commonly evaluated single intervention types include agricultural extension or training, cash transfers, and environmental regulation.

Agricultural extension and cash transfers are also frequently considered in SRs, but environmental regulations are not.

As with interventions, a large share of IEs measure outcomes from a relatively small set of popular categories: 45 per cent (3,310 studies) measure one of the outcomes shown in Figure 7. The most commonly measured outcomes in IEs include household income/assets, agricultural production, food security, and employment. These outcomes are also among the most commonly included outcomes in SRs.

Figure 6

Studies evaluate a diverse range of interventions, with considerable attention to agricultural production



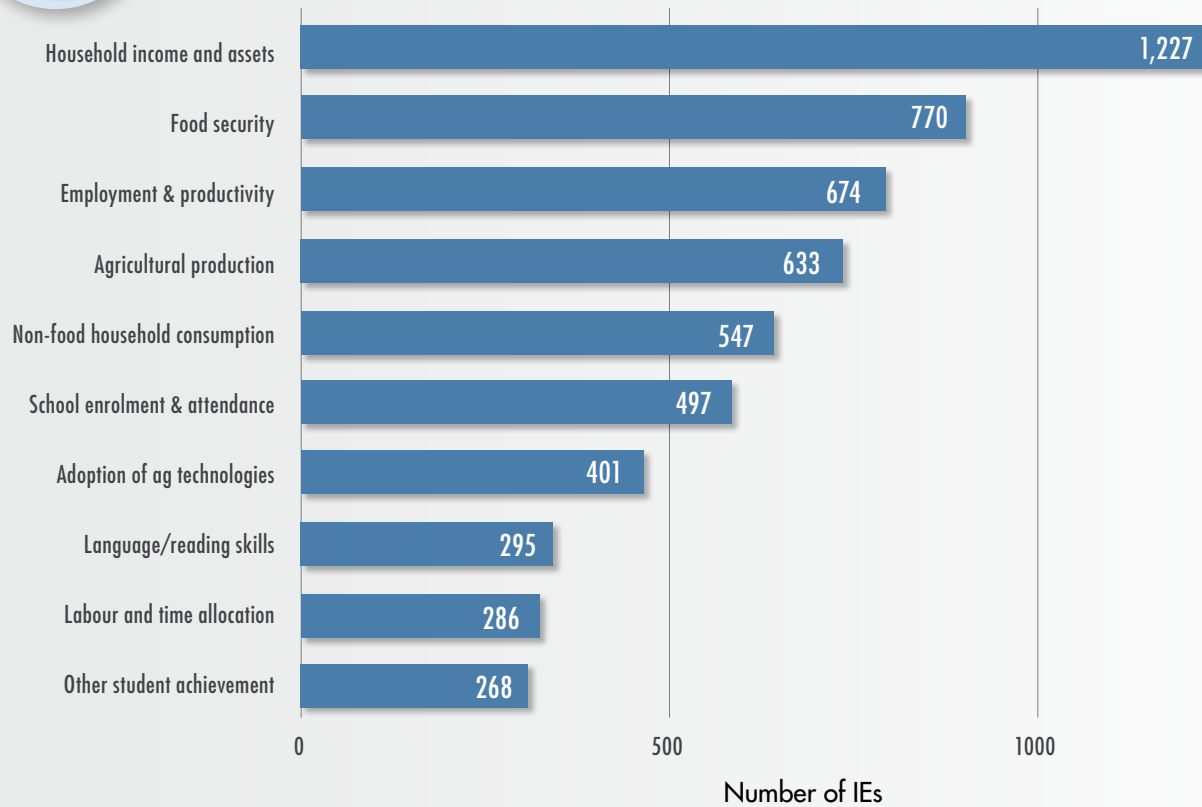
⁸ Note that for this report, we have aggregated specific intervention and outcome categories in the DEP's taxonomy into larger groups to make patterns more apparent. For details, see [Appendix A.4](#).

⁹ Note that we are reporting occurrences of intervention components (see [Appendix A.4](#)). Thus, the number of unique studies captured in [Figure 5](#) (3,016) is smaller than the total number of components shown (3,426), because some of these studies assess more than one of these components.



Figure 7

The most frequently measured outcomes relate to household well-being, agricultural production, and education



3.4 Who is funding the research?

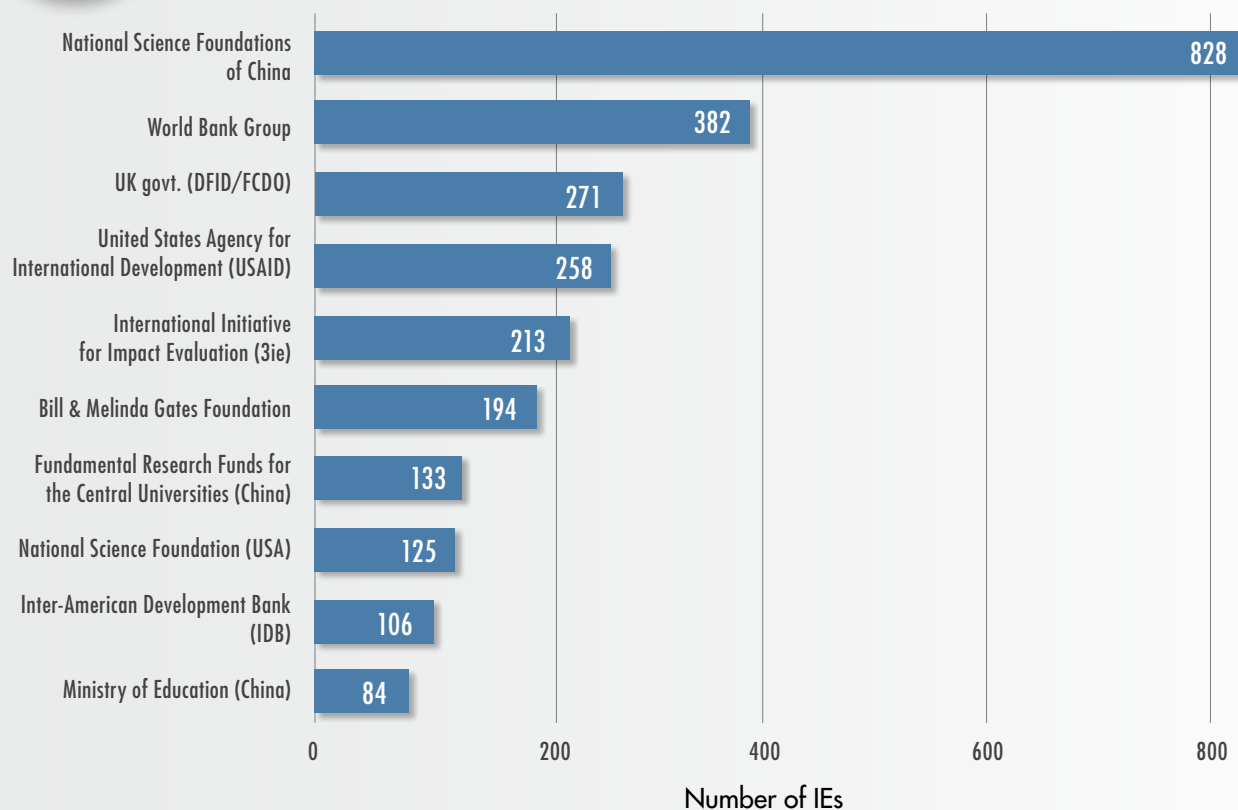
The National Science Foundations of China¹⁰ have funded by far the largest number of IEs in our dataset, with two other Chinese government agencies also among the top 10 funders (Figure 8). Impact evaluations in most other countries are funded by a variety of sources, but among studies focused on China, funding from Chinese government agencies dominates: such agencies have funded approximately 75 per cent of all China-focused studies. Other top funders globally include the World Bank, the UK government, the US Agency for International Development (USAID), and 3ie.



¹⁰ In this category, we have grouped together China's National Natural Science Foundation and National Social Science Foundation to make it analogous with other national science agencies like the US National Science Foundation.

Figure 8

The Chinese government is by far the largest funder of IEs



3.5 Does the evidence base align with aid spending?

Drawing on data from the Organization for Economic Cooperation and Development's (OECD's) [SDG Financing Lab](#),¹¹ we have analysed the concentration of evidence focused on each SDG with the amount of official development assistance (ODA) allocated to projects related to that SDG between 2012 and 2019.¹² This allows us to detect any major discrepancies between where resources are spent and where evidence is available.

Several large discrepancies are evident ([Figure 9](#)). However, some of these discrepancies likely arise from differences in how development projects and evaluation studies frame their objectives. Therefore, they may not reflect genuine misalignments between spending and evidence. In particular, the SDG Financing Lab appears to have taken a more “conservative” approach than DEP when classifying projects to SDG 1 (i.e., they have applied a more restrictive set of conditions for a project to be classified to SDG 1), and a more “liberal” approach than DEP when classifying projects to SDG 10.

¹¹ The SDG Financing Lab applied a machine learning algorithm to project descriptions, which then tagged each project, and the amount spent on it, with one or more SDGs.

¹² This is the full length of time covered by the SDG Financing Lab's data, giving the most general picture available of financing for the different SDGs.



Thus, although the share of ODA for SDG 1 is much smaller than the share of evidence, and the share of ODA for SDG 10 is much larger, we do not feel confident in identifying these discrepancies as genuine misalignments. Further discussion of these apparent misalignments is available in [Appendix B.5](#).

Evaluation research focuses on health far more than aid spending does. The evidence base appears skewed towards health. Even though the DEP dataset used for this report excludes studies with health interventions (approximately 40% of all studies in DEP), the percentage of IEs related to SDG 3 (9.1%) nearly matches the percentage of ODA for SDG 3 (10.5%). The SR evidence is even more skewed towards health. Education, food security, and water, sanitation, and hygiene (WaSH) are also over-represented in the SR evidence, relative to their share of ODA.

The concentration of synthesis on a limited range of topics means that for many SDGs, their share of SR evidence is well below their share of ODA disbursements.

Peace and justice interventions are understudied. We see several SDGs for which the share of ODA far outstrips the share of evidence. This pattern is most evident for SDGs 10, 16, and 17, each of which account for around 11 per cent of ODA, but less than 6 per cent of the IE evidence. This pattern is not surprising for SDG 17, where interventions are likely difficult to evaluate, but it is concerning that, for example, there is a limited evidence base to support decision-making around peace and justice interventions, despite these interventions absorbing such a sizable share of ODA spending.

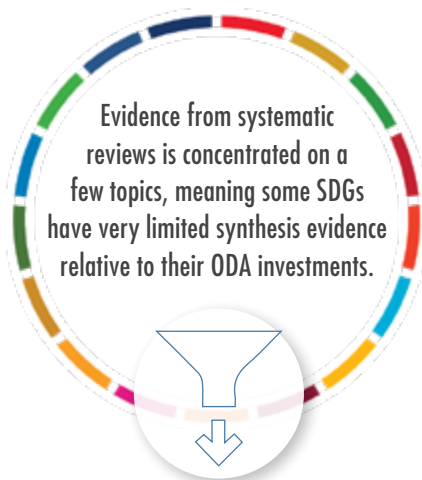
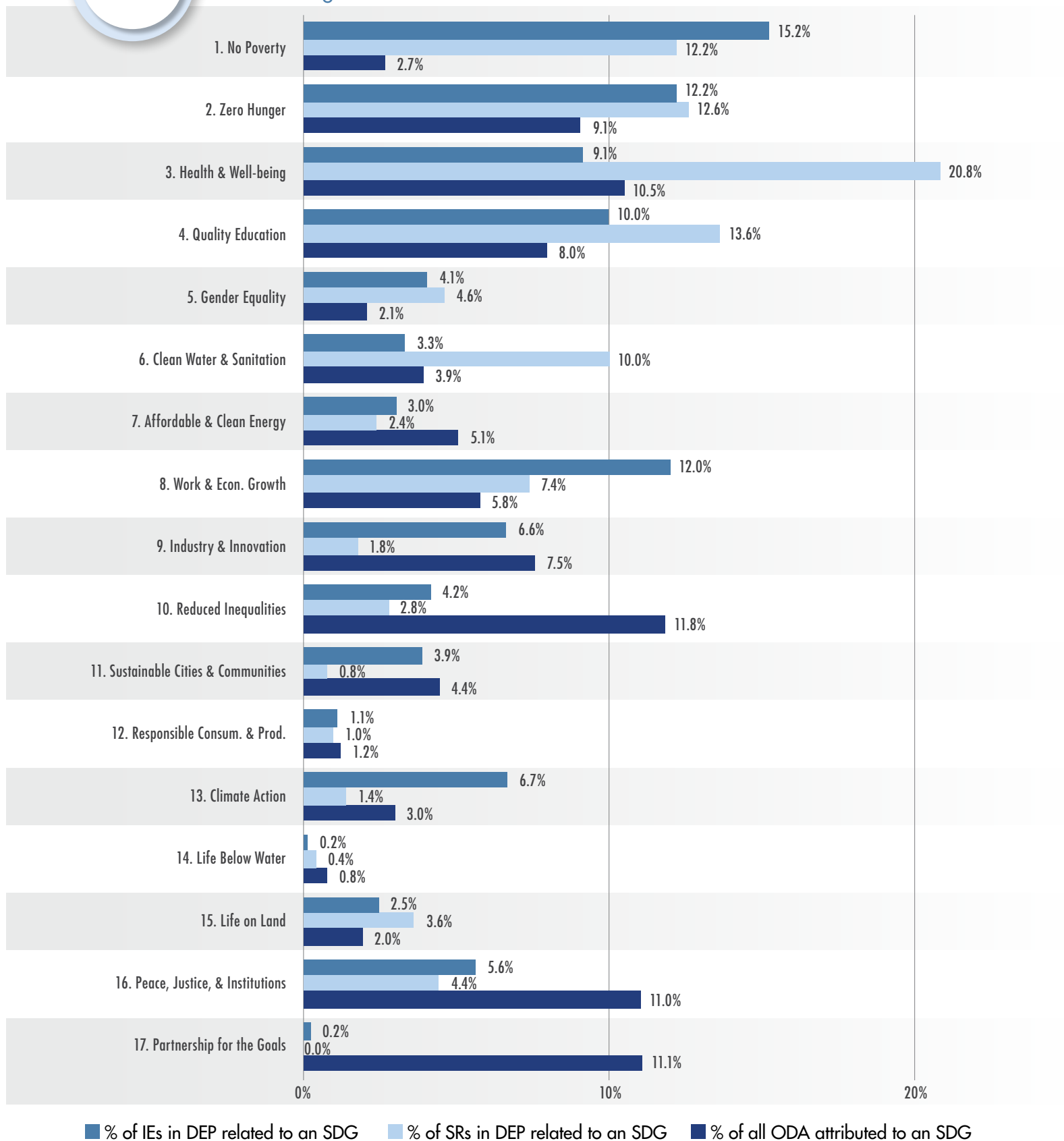




Figure 9

The SDGs that receive the most investment do not always have the largest evidence bases



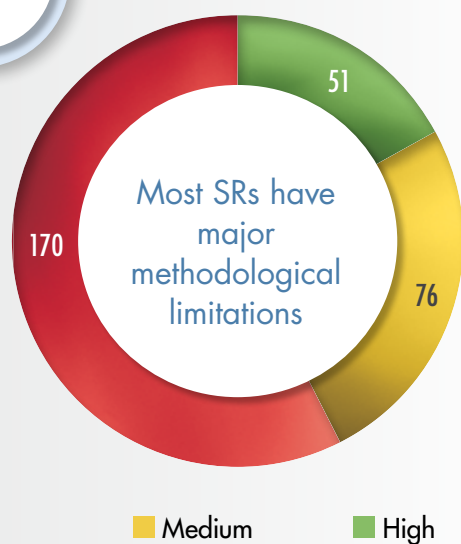


3.6 Most synthesis evidence is subject to methodological limitations

Most SRs in the DEP have been assessed as “low confidence” (Figure 10).¹³ Of the 249 (non-health) SRs in the DEP, 170 (57%) have a confidence rating of “low”. The remainder is roughly equally split between high-confidence (51; 17%) and medium-confidence (76; 26%) reviews. Low-confidence reviews can provide valuable insights, but there is a greater risk that methodological shortcomings undermine the review’s conclusions.

The distributions of SR confidence ratings within each SDG are relatively similar across the SDGs. The proportion of low-confidence reviews ranges between 40 and 60 per cent for most SDGs, and most SDGs have between 15 and 25 per cent high-confidence reviews. The reliability of synthesis evidence for SDG 8 (Work and Economic Growth) is particularly low, with over 60 per cent of SRs being low-confidence, and only 6 per cent high-confidence.

Figure 10



3.7 Attention to gender and equity is lacking in the evidence base

Very few studies in our sample take account of equity considerations in their research designs (Figure 11). Across the entire dataset, over three-quarters of IEs and over two-thirds of SRs do not address gender or equity. This pattern is consistent across the SDGs: for nearly every SDG besides SDG 5 (Gender Equality), the majority of IEs and SRs – typically above 70 per cent – do not address gender or equity.

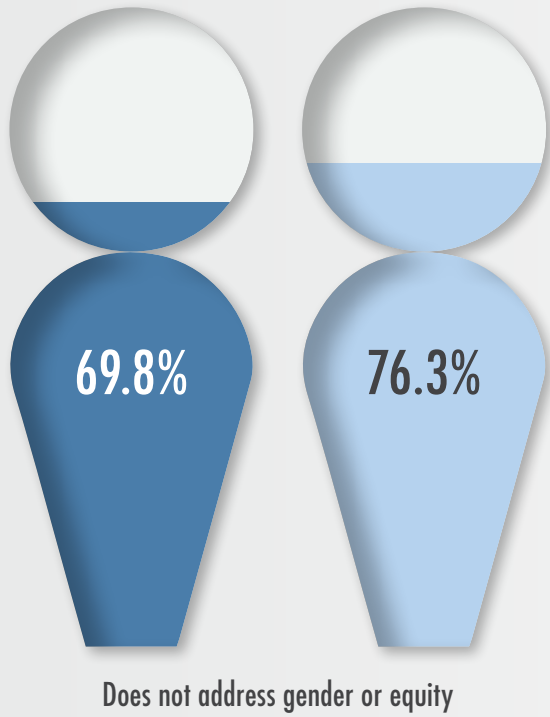
Our equity coding framework captures various ways in which studies might account for gender or equity in their research methods. These include performing sub-group analysis to assess an intervention’s effects on different populations or assessing the intervention’s effects on an inequality-related outcome (such as income inequality or intra-household decision-making). These types of analyses are critical for understanding whether development interventions are promoting the SDGs in an equitable way, or are in fact reinforcing existing inequalities. This lack of attention to gender and equity considerations means that, for the most part, the global community currently lacks the evidence needed to ensure that no one is left behind in the SDG agenda.

¹³ See [Appendix A.6](#) for a description of how we assign confidence ratings to systematic reviews.

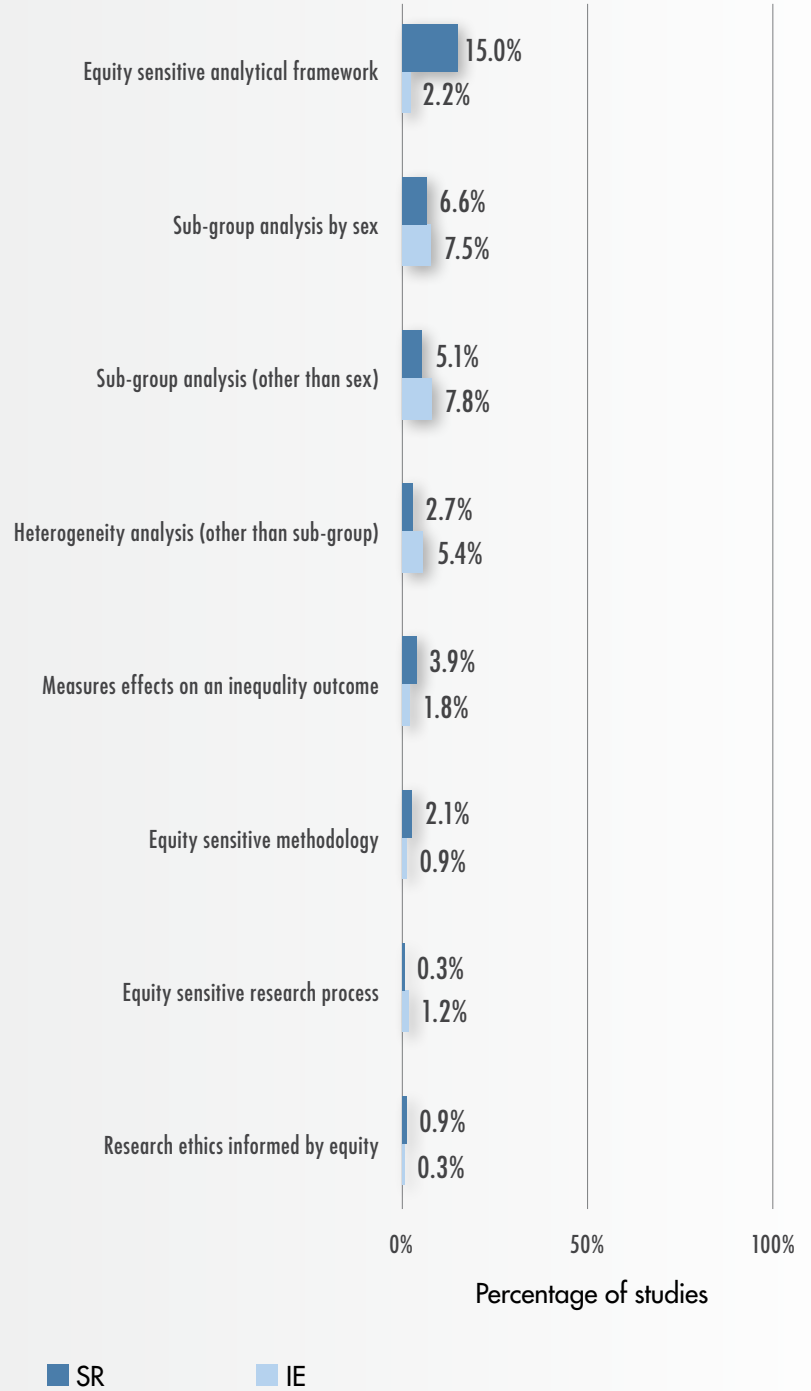


Figure 11

A large majority of IEs and SRs do not consider gender or equity in their research designs



When studies do consider gender or equity, it is usually through sub-group analysis





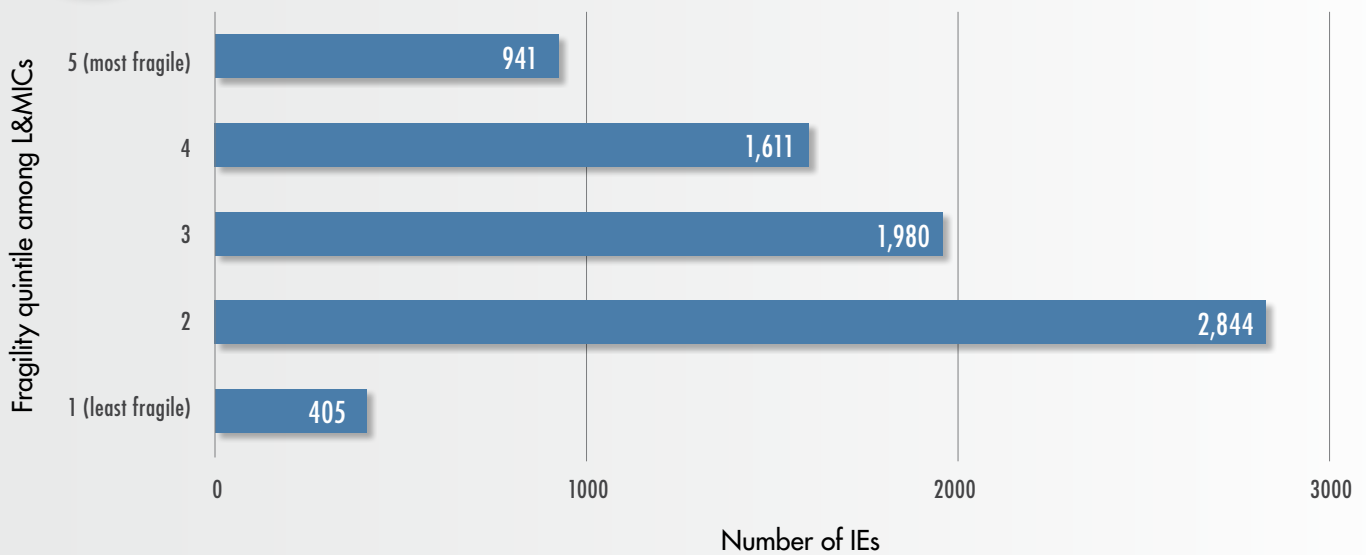
3.8 Evidence and state fragility

When we group countries by fragility level,¹⁴ we find that most evidence is concentrated in the middle three quintiles of the fragility distribution (Figure 12).

Interestingly, there is more evidence from countries with the highest fragility levels than those with the lowest. This may be a promising sign that the research community is prioritising conducting research in challenging contexts, at least to an extent.

Figure 12

Countries with the highest and lowest levels of fragility have the least evidence



¹⁴ Country fragility data are from the Fund for Peace's Fragile States Index [14].



The state of the **evidence** for the **17 SDGs**



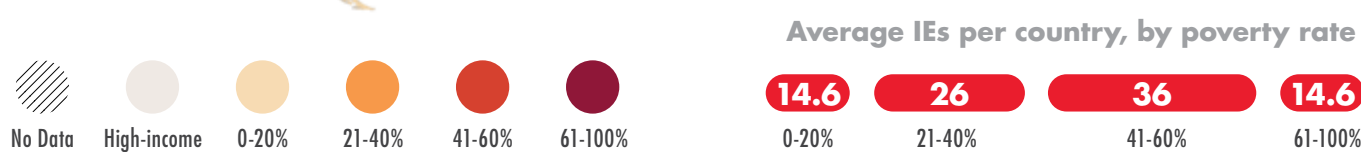
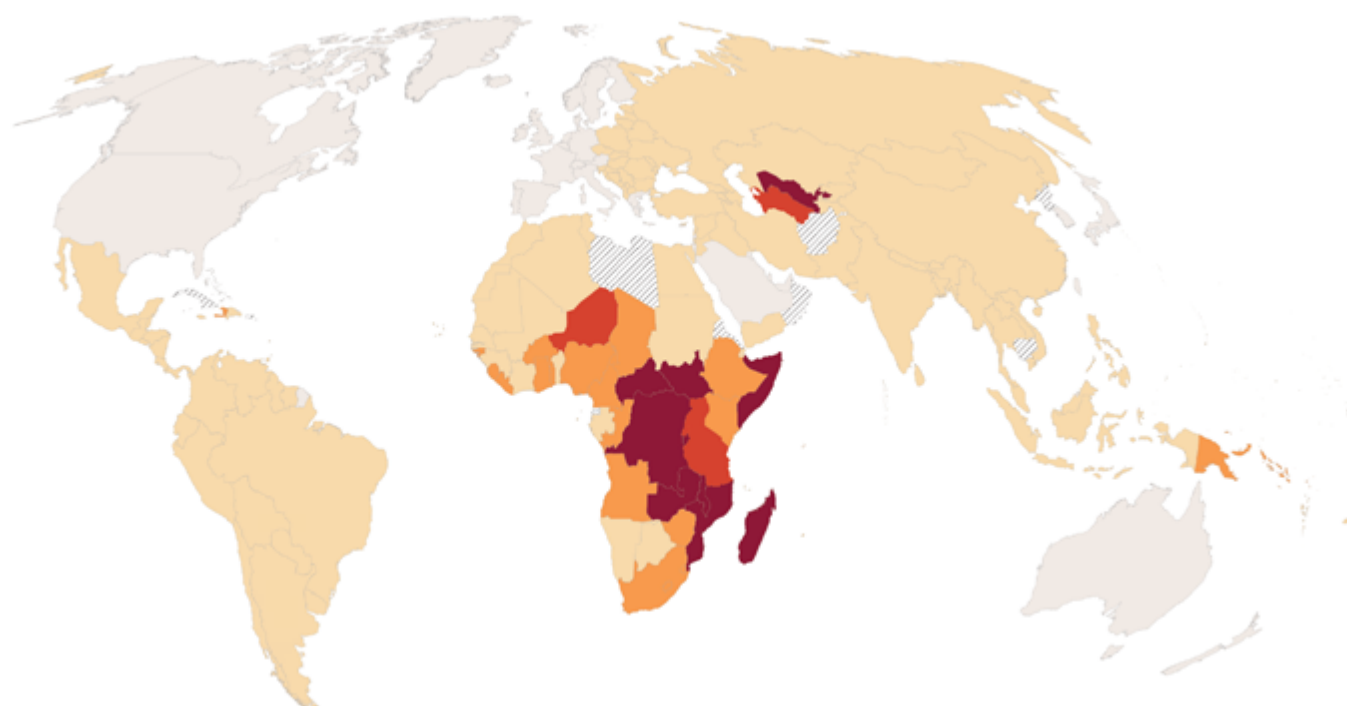


SDG 1 has the largest evidence base of the 17 SDGs, with over 1,600 impact evaluations and 61 systematic reviews. SDG 1-related research tends to focus on countries with moderate, rather than very high, rates of extreme poverty. Evidence is concentrated in Sub-Saharan Africa, with over twice as many studies from that region as in the next most studied region (Latin America). Conditional cash transfers are a major focus in Latin America, while agricultural productivity interventions dominate in Sub-Saharan Africa, and access to credit/savings is well studied in South Asia.

As SDG 1 interventions are well studied in IEs, the needs of the SDG 1 evidence base mirror those of the evidence base as a whole – there is a need for more high-quality synthesis and greater attention to gender and equity.

1,656 IMPACT EVALUATIONS
61 SYSTEMATIC REVIEWS

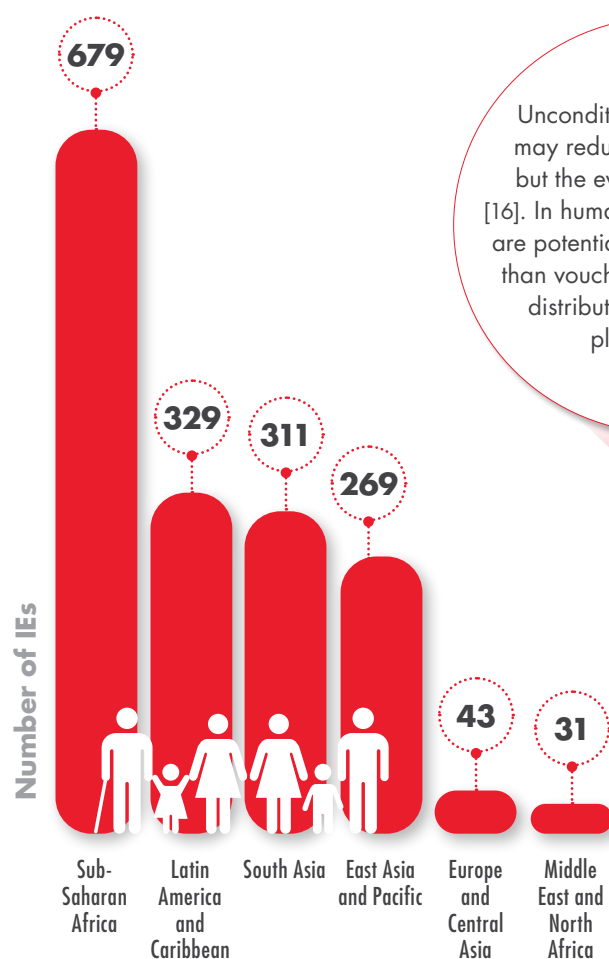
Evidence is scarce from countries with the highest extreme poverty rates



% of country's population below international poverty line (most recent year available)

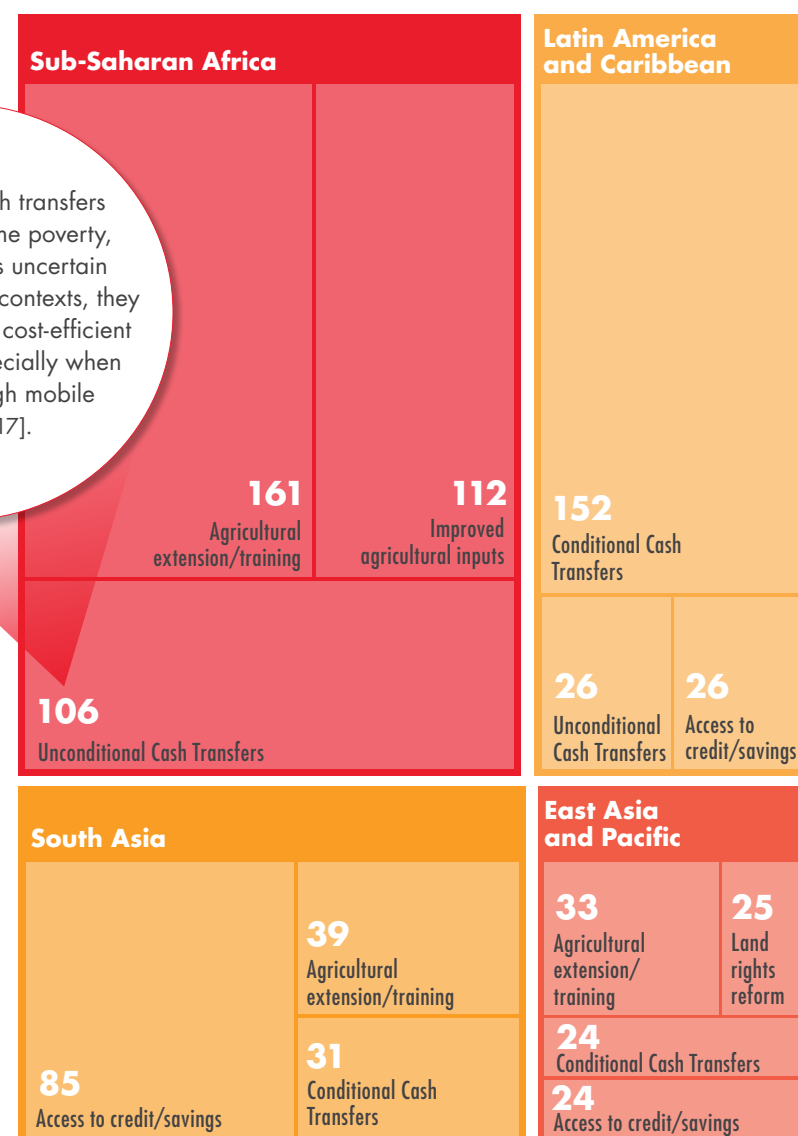
Extreme poverty is defined as living in a household with an income or expenditure per person below the International Poverty Line of USD 2.15 per day in 2017 prices. Source for poverty data: Our World in Data [15].

Evaluation of anti-poverty interventions is concentrated in Sub-Saharan Africa



Unconditional cash transfers may reduce extreme poverty, but the evidence is uncertain [16]. In humanitarian contexts, they are potentially more cost-efficient than vouchers, especially when distributed through mobile platforms [17].

Different anti-poverty interventions are evaluated in different regions



Note: some IEs cover multiple regions.

Number of impact evaluations evaluating each intervention in each region. Not shown: Europe/Central Asia and MENA regions, which have very few IEs for this SDG.



SDG 2 has the second largest evidence base, with 1,328 impact evaluations and 63 systematic reviews. Most evidence focuses on agricultural production, and evidence is particularly skewed towards Sub-Saharan Africa, where about 55 per cent of all IEs have taken place. There is a serious lack of evidence from the most hunger-stricken countries – those with “Alarming” hunger severity ratings have, on average, fewer than 5 IEs. However, there is a stronger evidence base for countries where hunger severity is rated “Serious” – there are about 27 evaluations per country from this group.

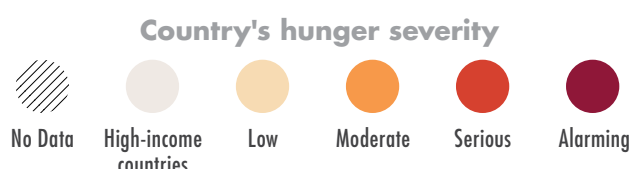
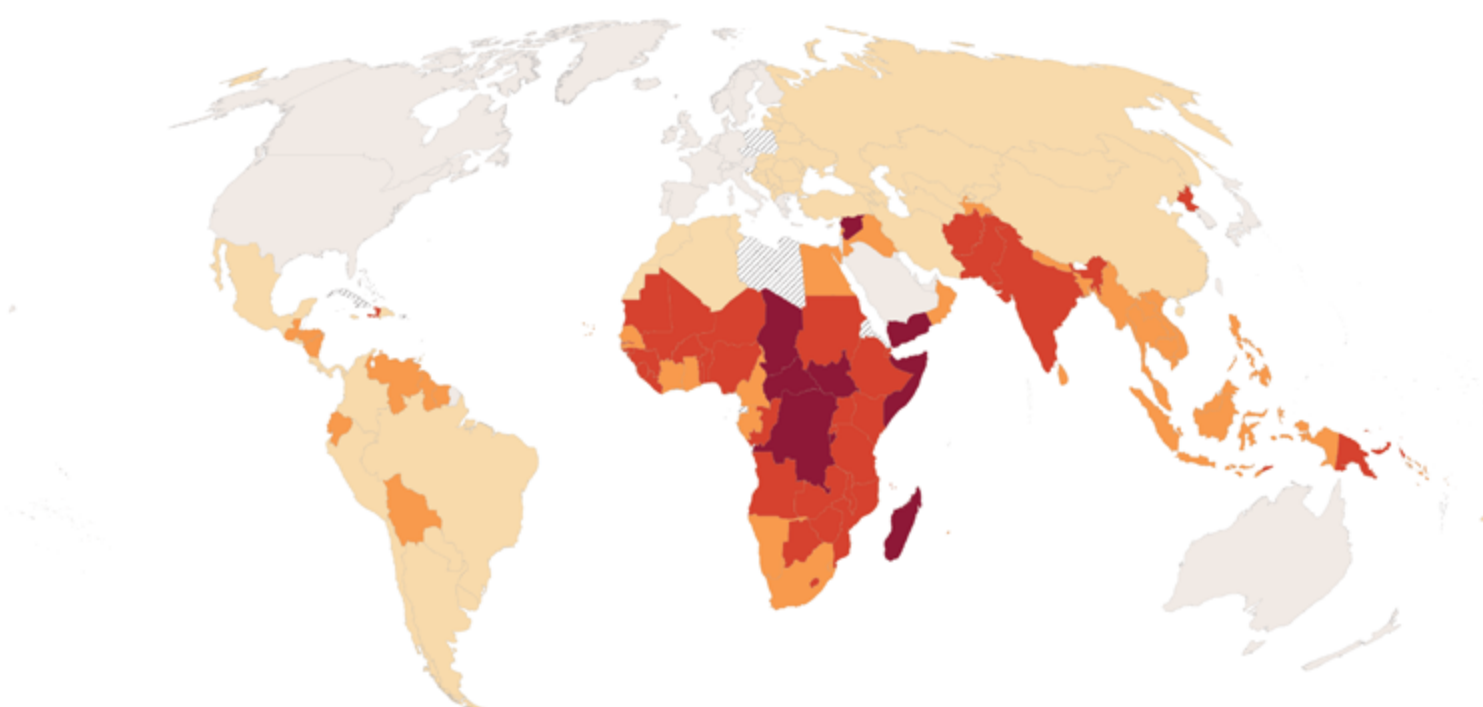
The SDG 2 evidence base will benefit from greater attention to determinants of food insecurity beyond agricultural production, such as transport/storage, food availability, and dietary behavior.

1,328 IMPACT EVALUATIONS

63 SYSTEMATIC REVIEWS

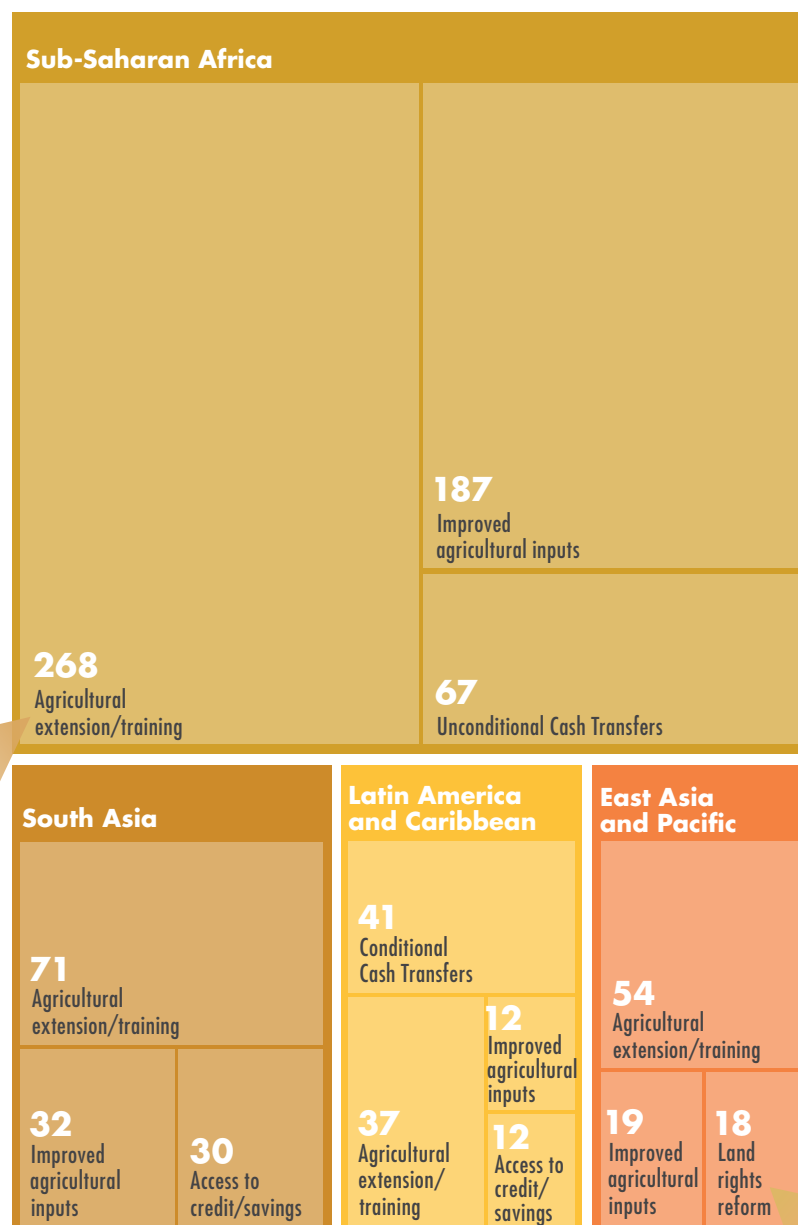
Aquaculture and home food production elevate income and nutrition [20].

Evidence is severely lacking in the most hunger-stricken countries

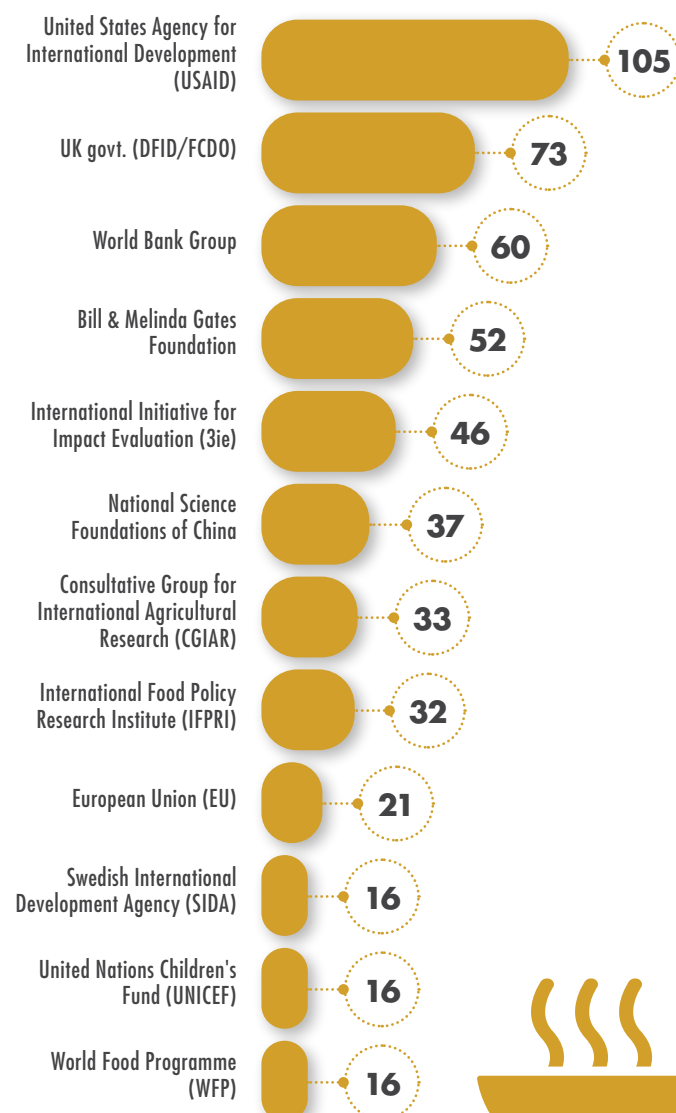


Source for hunger data: Global Hunger Index [18].

Agricultural extension is the most evaluated intervention everywhere except Latin America



USAID leads the world in evidence generation on food security



Tenure formalisation boosts productivity in Asia and Latin America [19].

Number of impact evaluations evaluating each intervention in each region. Not shown: Europe/Central Asia and MENA regions, which have very few IEs for this SDG.



Our analysis of the evidence base for SDG 3 is unique, in that we are looking at the evidence for effects of non-health interventions that nevertheless measure health outcomes. Even without health-focused interventions, the evidence base for SDG 3 is sizable, with nearly 1,000 IEs, and the largest pool of SRs, with just over 100. The evidence base skews towards Sub-Saharan Africa, but the difference between this and other regions is less stark than for SDGs 1 and 2.

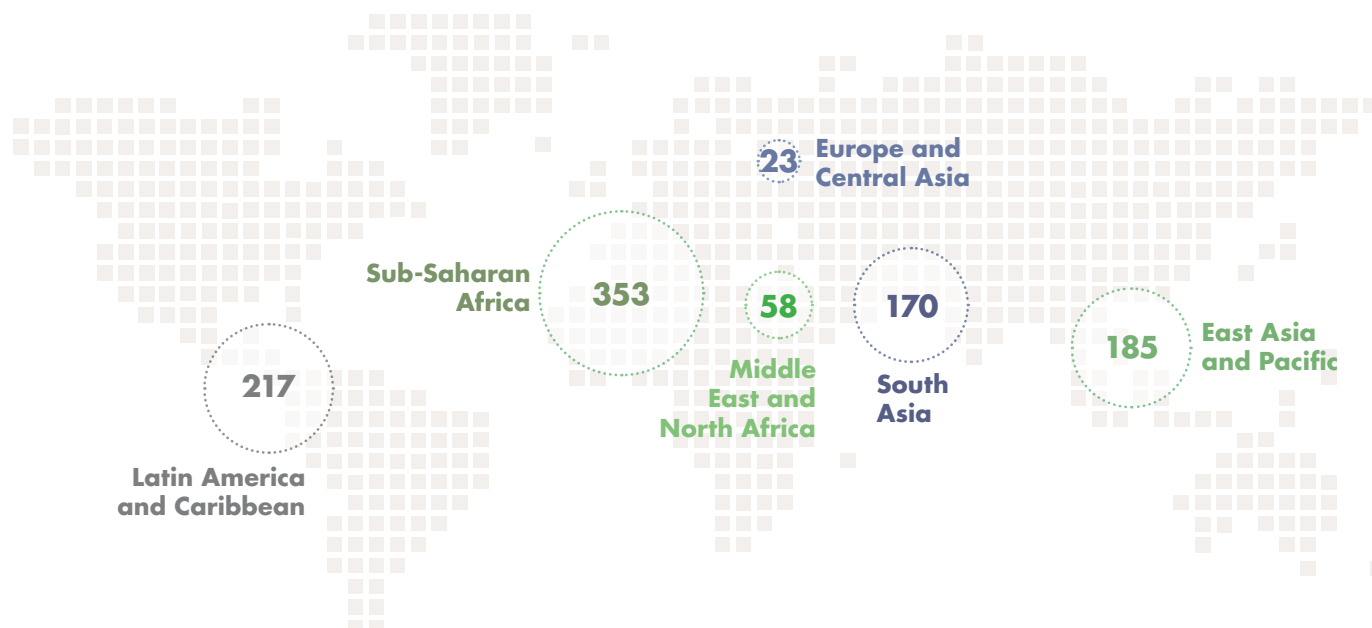
Cash transfers top the list of (non-health) interventions studied for SDG 3, and are widely studied across regions (most heavily in Latin America). Many of the other frequently studied interventions relate to food production/nutrition and sanitation/hygiene. Food security and anthropometric measurements (such as length/height and weight) are the most frequently measured outcomes among SDG 3 studies, indicating that many studies measure food security and health outcomes in tandem.

It would be useful to have evidence for the impact of a broader range of non-health interventions on health outcomes. For example, the DEP contains only eight IEs on the effects of transportation infrastructure interventions on health outcomes, despite the fact that road accidents are major sources of disability and mortality in L&MICs [18].

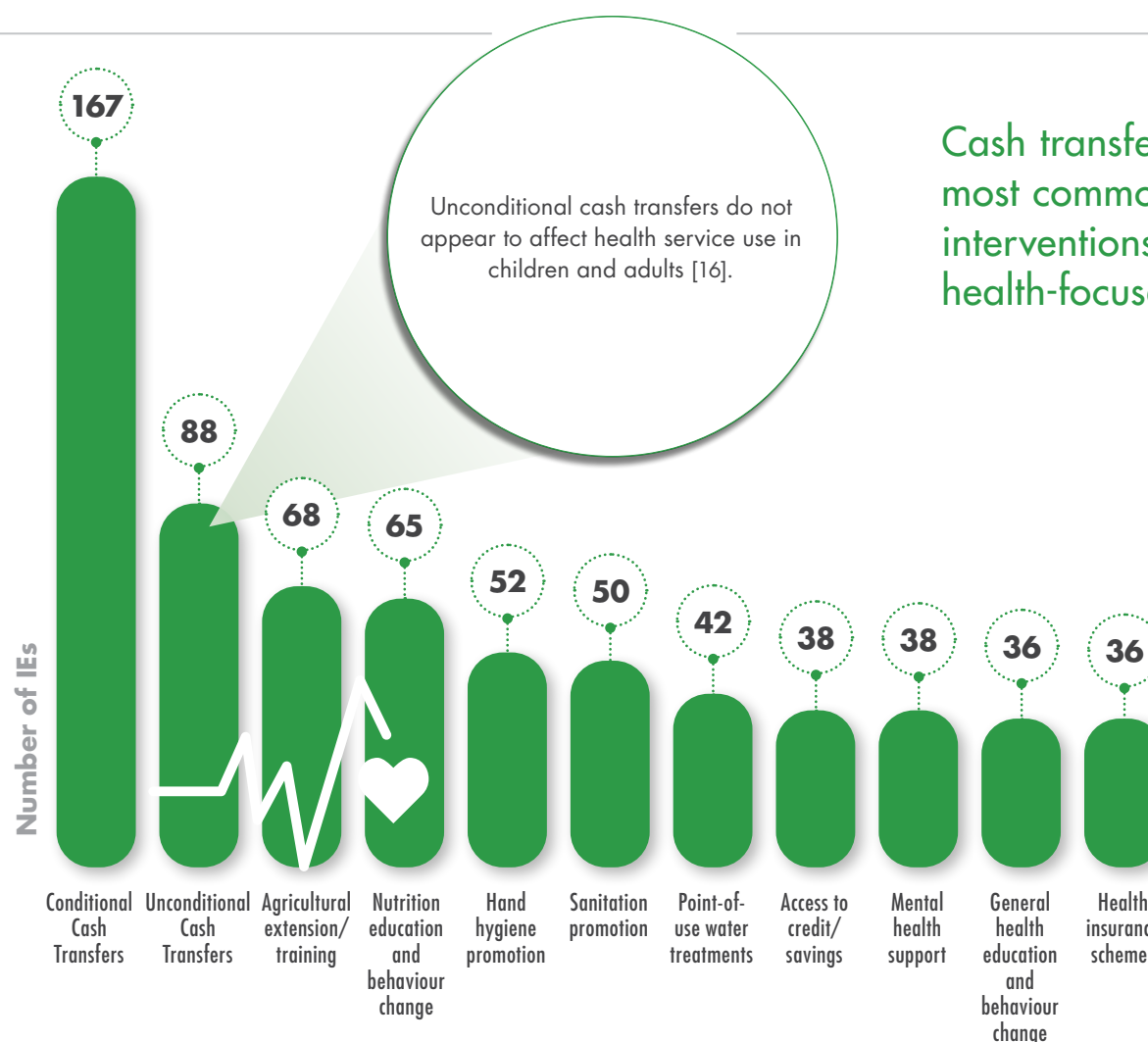
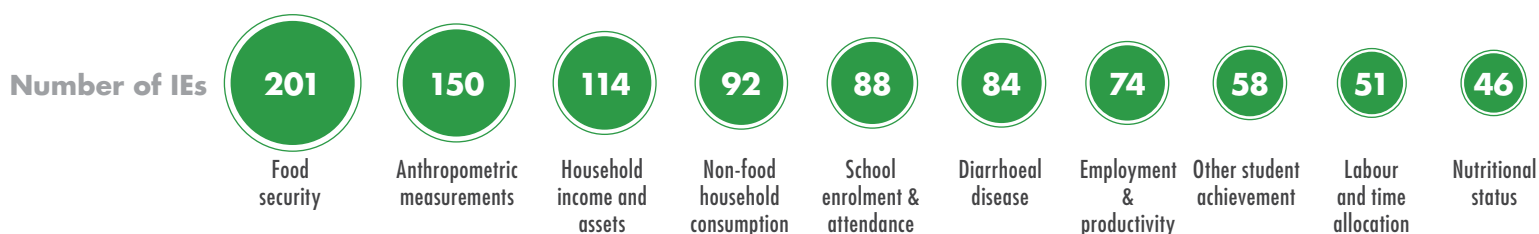
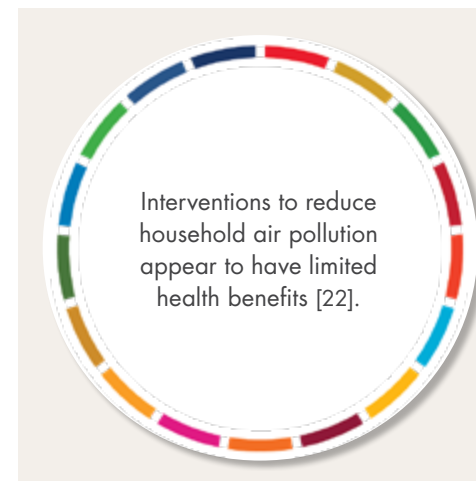
993 IMPACT EVALUATIONS

104 SYSTEMATIC REVIEWS

SDG 3 evidence on non-health interventions is most prevalent in Sub-Saharan Africa



Food security and anthropometrics are commonly measured health-related outcomes of non-health interventions



Unconditional cash transfers do not appear to affect health service use in children and adults [16].

Cash transfers are the most common non-health interventions evaluated in health-focused studies



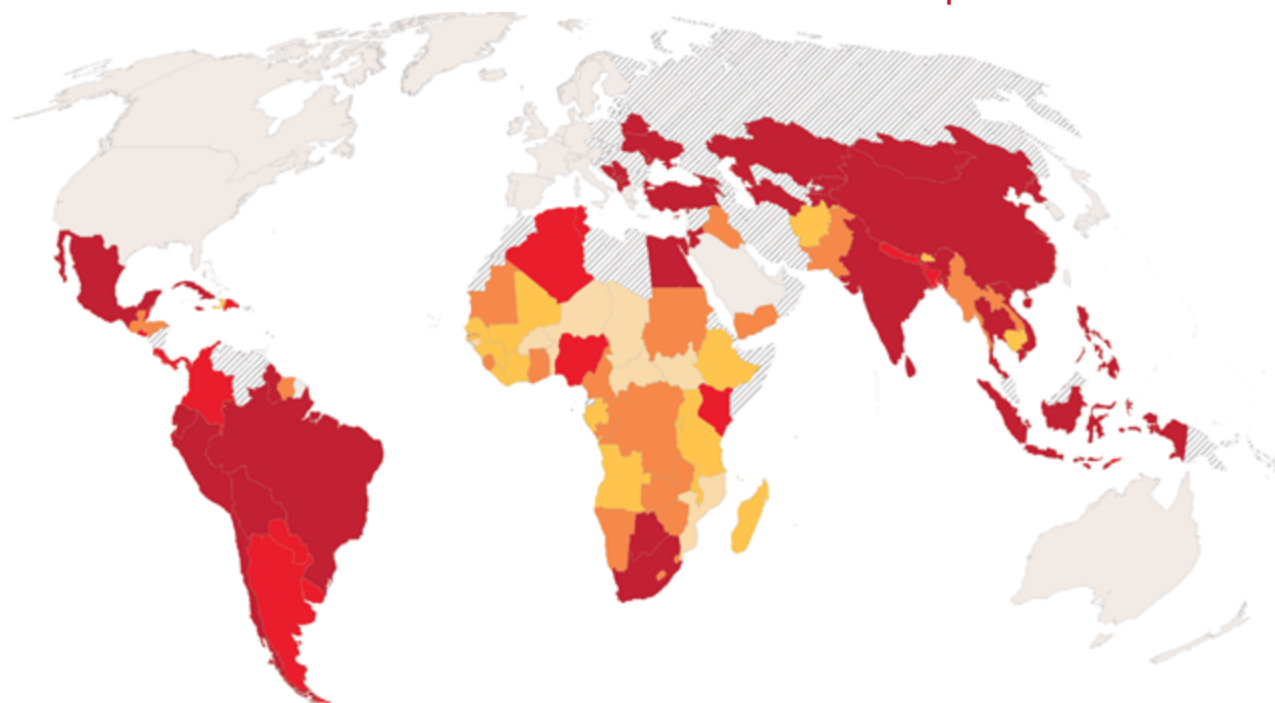
SDG 4 has a large evidence base, with over 1,000 impact evaluations and nearly 70 systematic reviews. It is the only SDG for which Latin America has the most evidence. This is due, in part, to the large number of evaluations of conditional cash transfers in Latin America and their effects on education-related outcomes. Other commonly evaluated interventions focus on reducing financial and other barriers, strengthening the teacher corps, and improving pedagogy.

The World Bank is among the top funders of many SDGs, but has been particularly dedicated to supporting education research – the World Bank has funded nearly three times as many education-related IEs as the next biggest funder (the UK government).

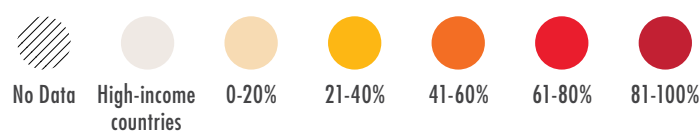
Unfortunately, the evidence base per country gets thinner as completion rates of lower secondary schools decrease. The evidence base for education-related interventions needs to be expanded for countries struggling with low completion rates.

1,085 IMPACT EVALUATIONS
68 SYSTEMATIC REVIEWS

Evidence is most scarce for countries with the lowest school completion rates



Country's lower secondary school completion rate



Average IEs per country, by lower secondary completion rate

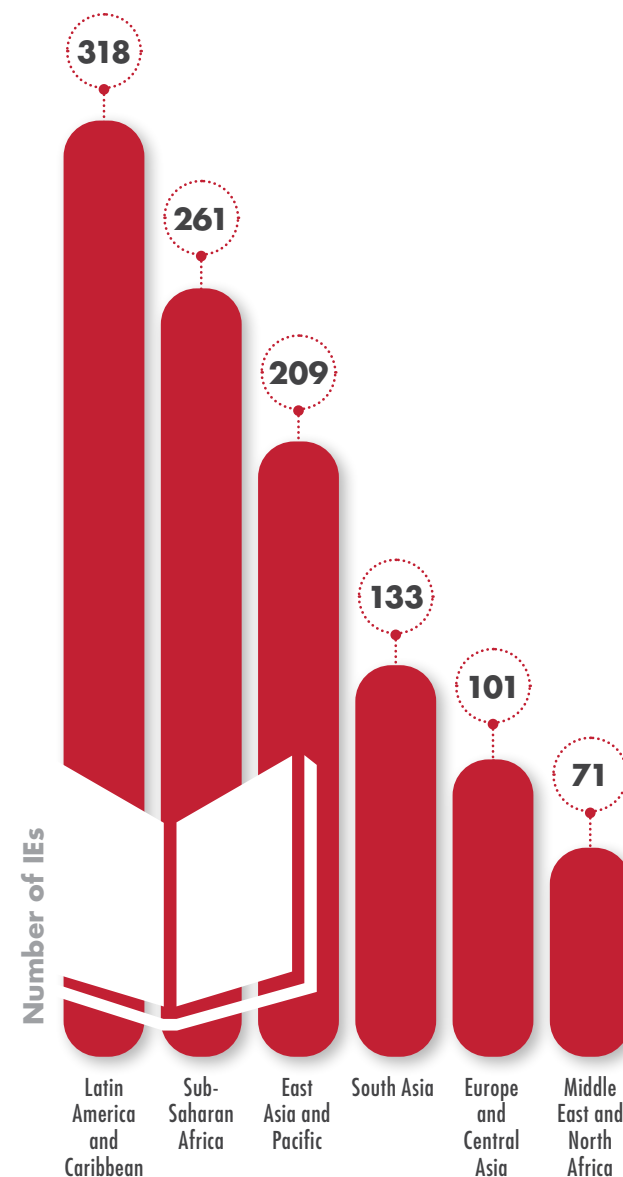


Source for completion data: UNICEF [23].

Conditional cash transfers and reducing financial barriers are the most evaluated interventions in most regions

SDG 4 is one of only two SDGs where Latin America has the most evidence

Latin America and Caribbean		
Cash transfers boost school attendance but don't affect learning [25].		31 Teacher hiring & development
88 Conditional Cash Transfers		28 Reducing financial barriers to education
Sub-Saharan Africa		East Asia and Pacific
40 Reducing financial barriers to education	30 Lowering education access barriers for vulnerable groups	22 Education governance
		18 Conditional Cash Transfers
30 Teacher hiring & development		16 Reducing financial barriers to education
South Asia		
18 Lowering education access barriers for vulnerable groups	14 Reducing financial barriers to education	13 Conditional Cash Transfers



Number of impact evaluations evaluating each intervention in each region. Not shown: Europe/Central Asia and MENA regions, which have few IEs for this SDG.

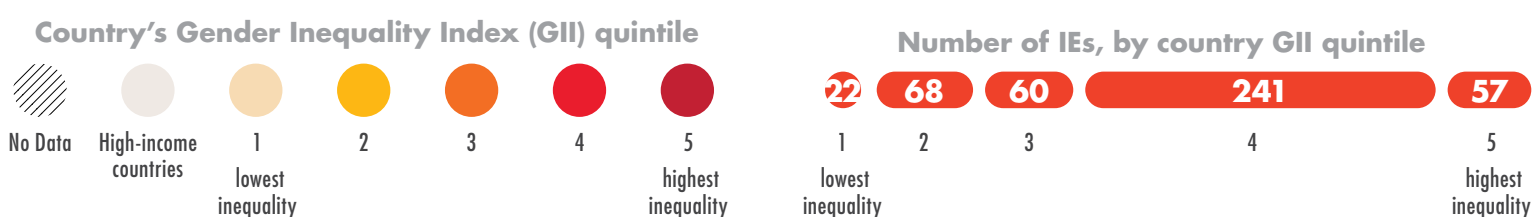
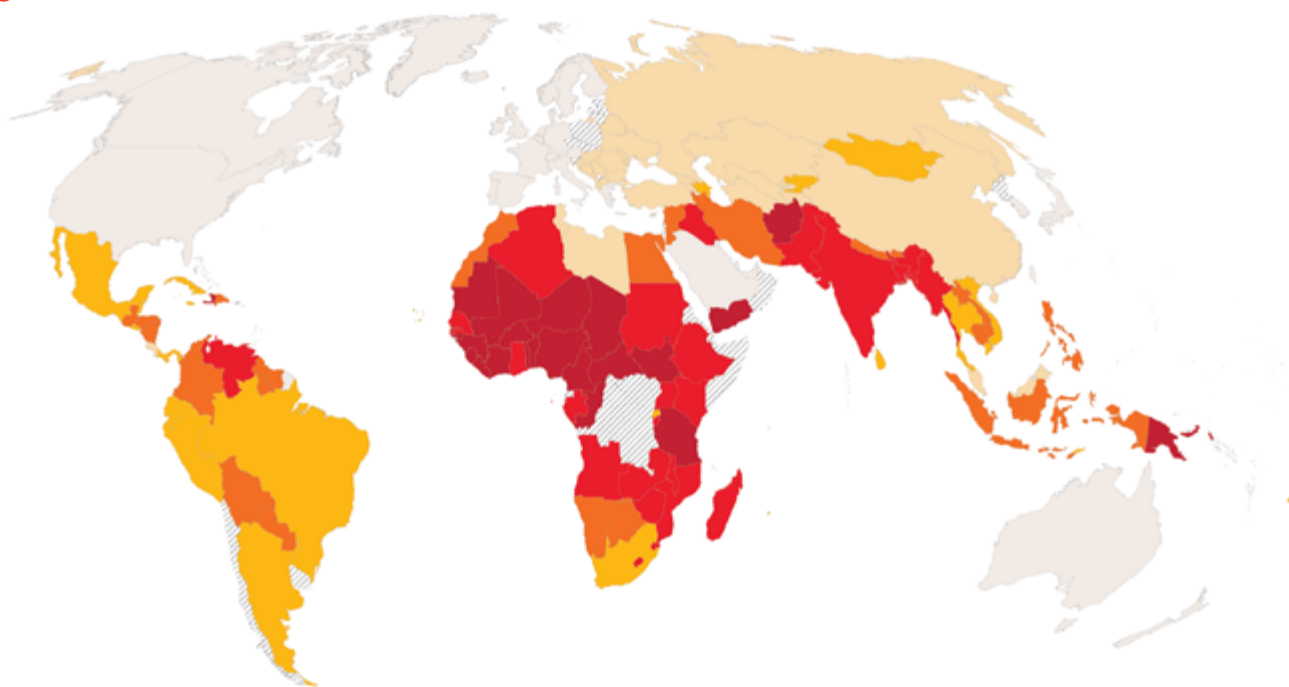


SDG 5 has a moderately sized evidence base, with 441 impact evaluations and 22 systematic reviews. As we might expect, there is greater attention to gender and equity considerations among SDG 5 studies. Still, even here, more than a third of both IEs and SRs do not address gender or equity. When L&MICs are grouped into quintiles by Gender Inequality Index (GII), we find limited evidence from the countries with the highest GII scores, although there is a sizable evidence base from countries in the fourth quintile (indicating high GII, but not the highest). The evidence base would benefit from more research in the most gender-unequal countries.

Some interventions in SDG 5 studies focus specifically on gender issues, while others target education and general economic well-being. The most commonly measured outcomes span themes of economic well-being, education, food security, empowerment, and safety from violence.

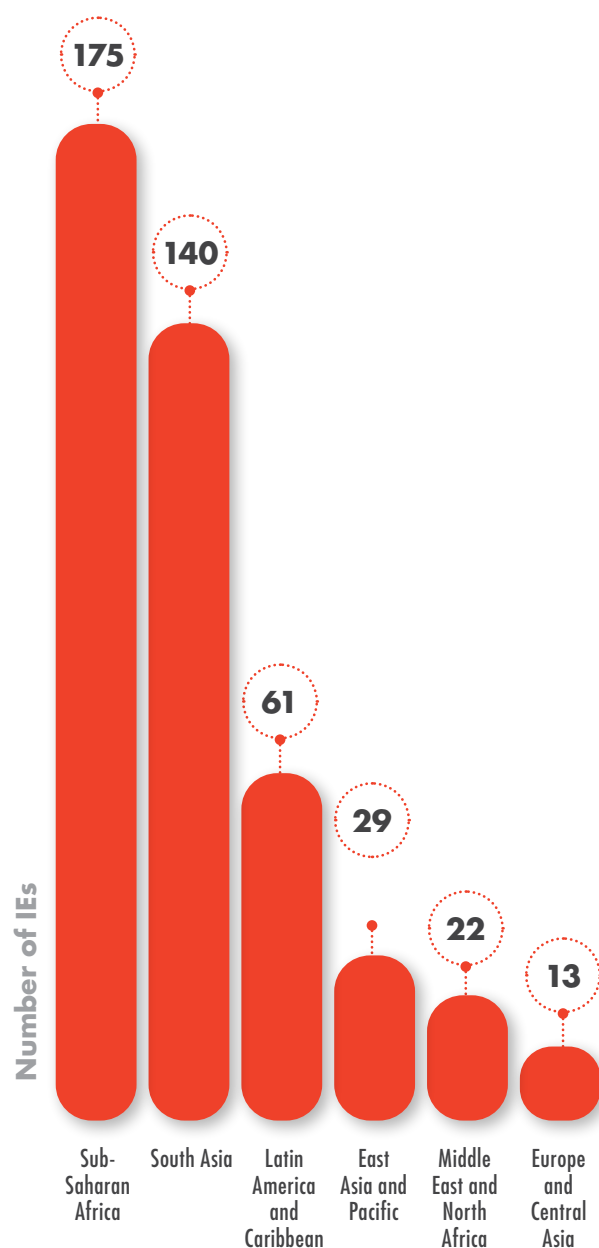
441 IMPACT EVALUATIONS
23 SYSTEMATIC REVIEWS

Countries with high (but not the highest) gender inequality have the most evidence for SDG 5

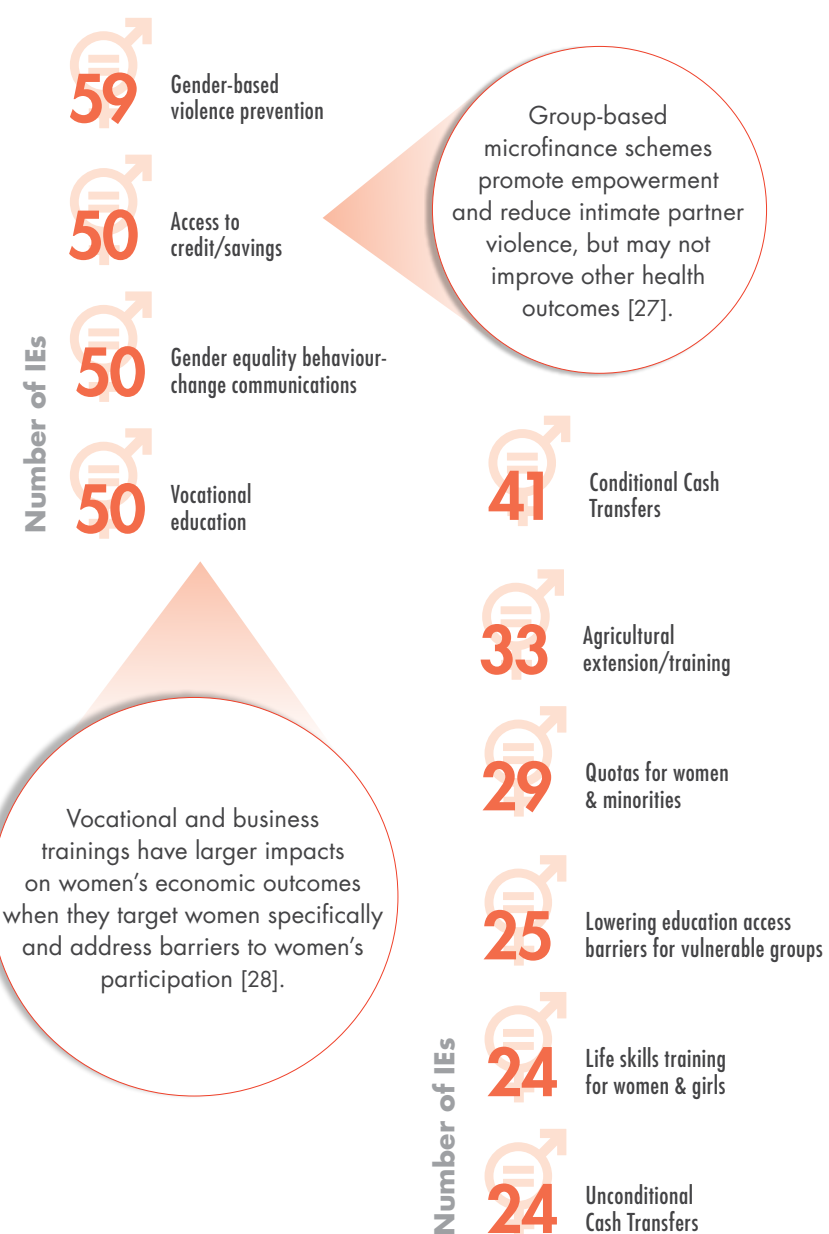


Source for inequality data: UNDP Gender Inequality Index [26]. Sample includes all countries that have held L&MIC status at any point in the last 10 years, and for which GII data are available.

Evidence for SDG 5 is concentrated in Sub-Saharan Africa and South Asia



SDG 5 studies evaluate a mix of gender-specific interventions and more general educational and economic programmes





SDG 6 has a relatively small evidence base of impact evaluations (360), although it has a higher number of systematic reviews (48) relative to the number of impact evaluations than other SDGs. Sub-Saharan Africa, South Asia, and East Asia all have substantial shares of the SDG 6 evidence.

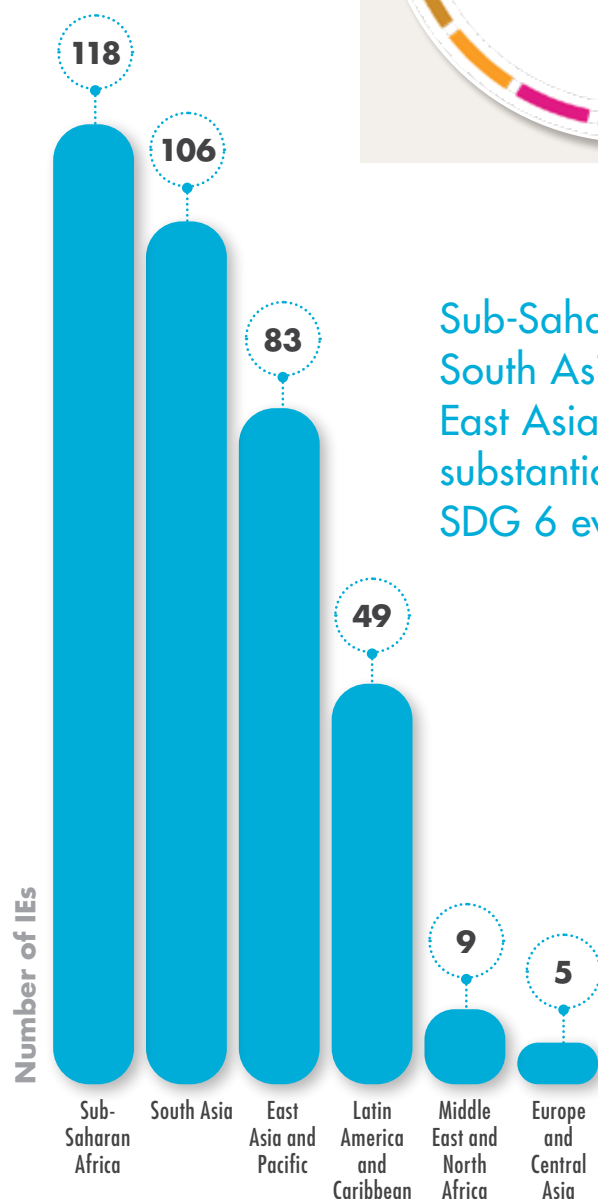
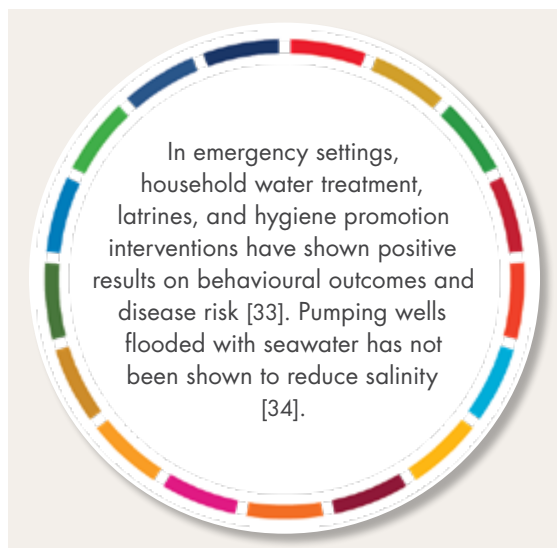
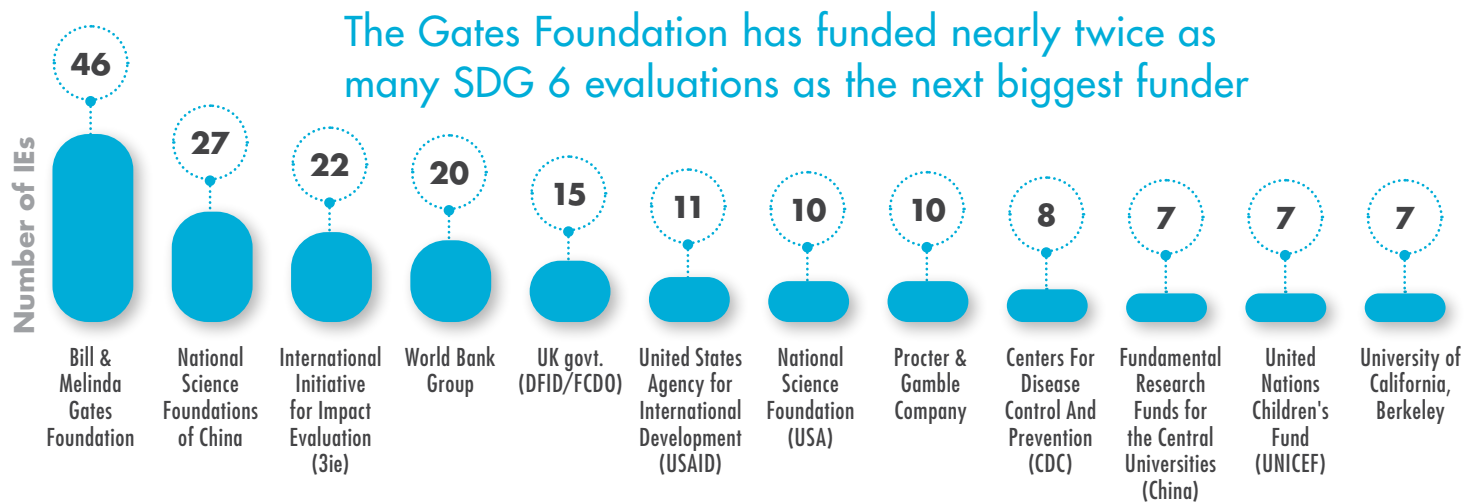
Evaluation research tends to focus on the promotion of general sanitation and hand hygiene, along with water treatments. Diarrhoeal disease is the most measured outcome.

Additional IEs are needed to bolster the relatively limited SDG 6 evidence base. In the meantime, decision makers can take advantage of SDG 6's sizable pool of quality synthesis.

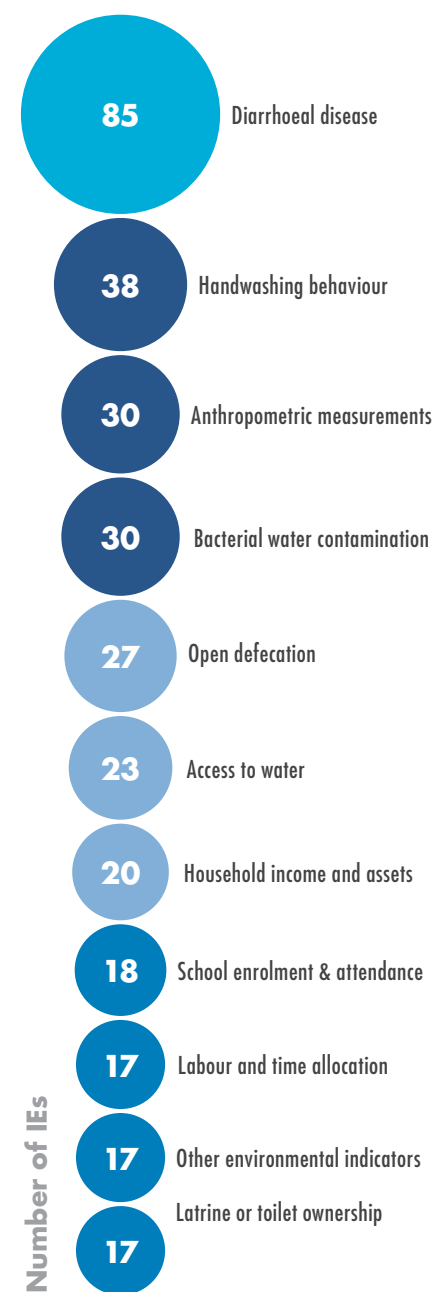
360 IMPACT EVALUATIONS
50 SYSTEMATIC REVIEWS

Promotional interventions can boost handwashing, latrine use, and safe waste practices, but different approaches work best in different contexts [32].

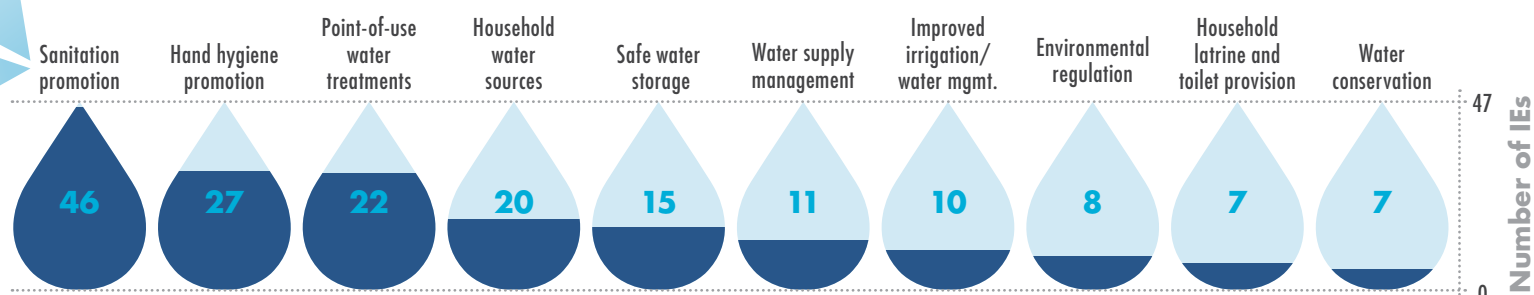
Evidence on the effectiveness of interventions that promote safe disposal of child faeces is limited and inconclusive [31].



Outcomes related to diarrhoeal disease are measured in about a quarter of all SDG 6 studies



The most evaluated SDG 6 interventions include the promotion of sanitation, hand hygiene, and point-of-use water treatments





SDG 7 has a relatively small evidence base of 330 impact evaluations and 12 systematic reviews. SDG 7 is among the SDGs for which a large proportion of the evidence has emerged very recently and is focused on China. There are now more studies focused on China than on all other L&MICs combined, and the vast majority of these studies have been published in the last four years.

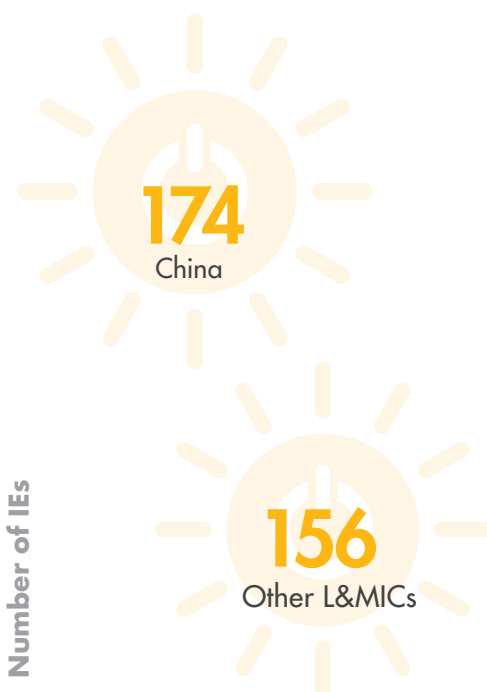
China-focused studies tend to measure the effects of environmental regulation, market incentives, and energy infrastructure. The predominant outcome measure is household energy use, though air pollution and greenhouse gas emissions are also measured.

There is a need to balance the evidence base by studying energy-related interventions in a wider range of contexts. This SDG also has a low ratio of SRs to IEs, suggesting a need for additional synthesis of energy-related research. While SRs exist for interventions like clean household energy, market incentives, and energy infrastructure, there are, notably, no SRs on emissions trading schemes.

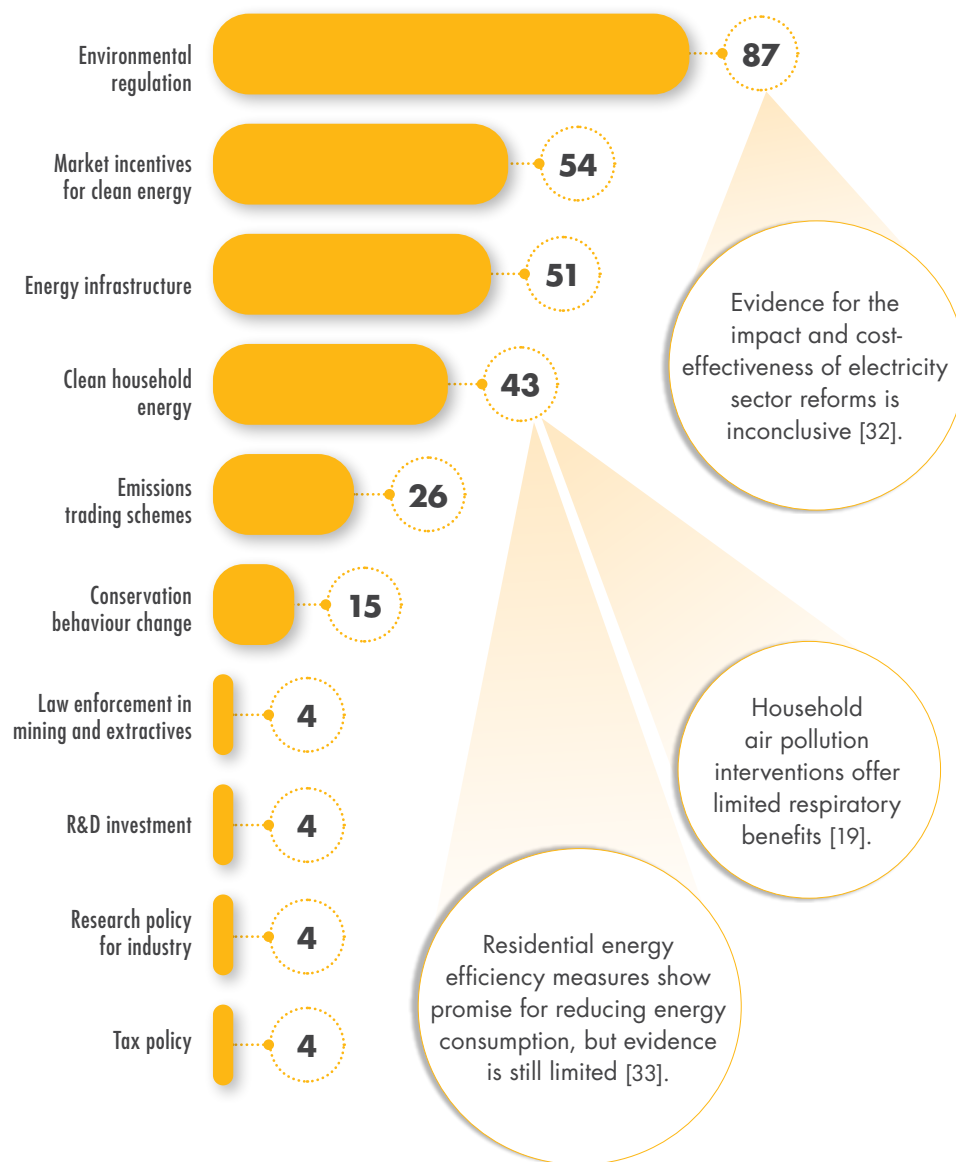
330 IMPACT EVALUATIONS

12 SYSTEMATIC REVIEWS

There are more SDG 7 studies focused on China than all other L&MICs combined

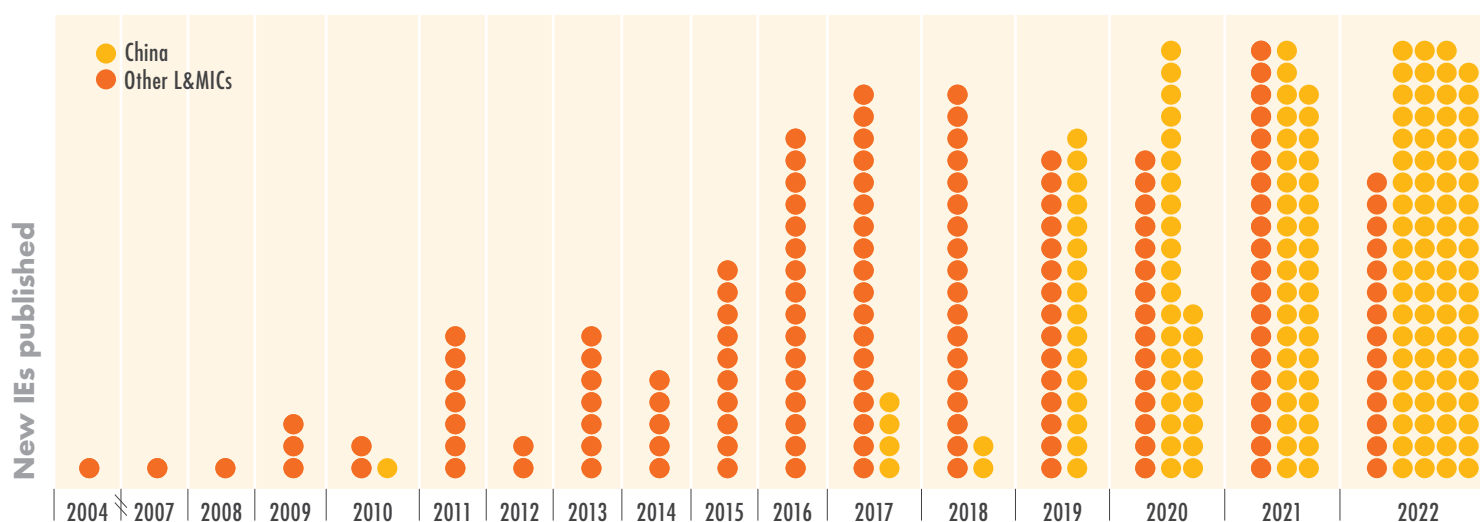


About a third of the SDG 7 studies measure the effects of environmental regulations



Roughly a third of SDG 7 studies measure the outcome of household energy use

China's production of SDG 7 evidence has exploded in the last several years





SDG 8 has one of the largest evidence bases with over 1,300 impact evaluations. Studies focused on Chinese policies have come to dominate the literature in recent years. However, unlike other SDGs where this is the case (e.g., 7, 9, and 11-13), a sizable existing literature from other regions had accumulated before the recent explosion of China-focused research.

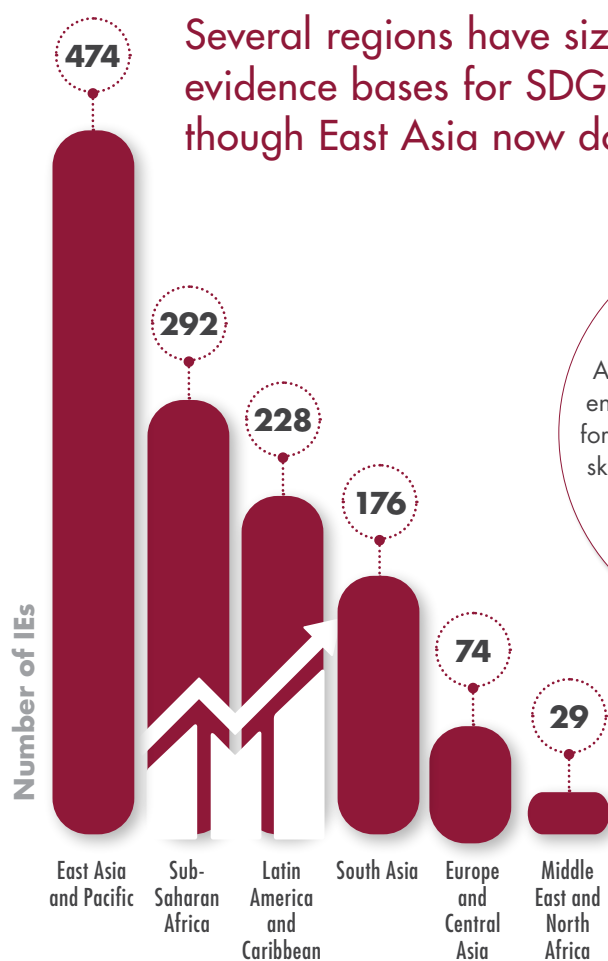
The number of SRs (37) is quite low relative to the number of IEs, indicating a need for additional synthesis. For instance, there are no SDG 8 SRs synthesising the IE literatures on financial regulation or inflation control.

1,304 IMPACT EVALUATIONS
37 SYSTEMATIC REVIEWS

China's production of SDG 8-related evidence has quickly caught and surpassed the rest of the world

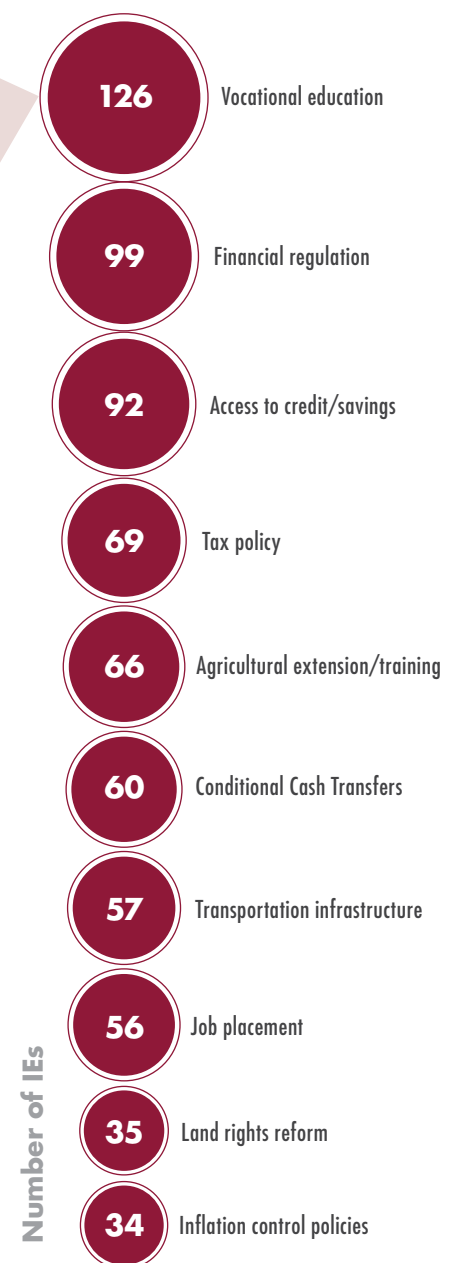


Several regions have sizable evidence bases for SDG 8, though East Asia now dominates



Active labour market programmes enhance job and income prospects for youth, with entrepreneurship and skill-building approaches being the most effective in L&MICs [37].

A broad range of interventions are evaluated for SDG 8, including localised programmes and broad policies



China-focused studies evaluate regulatory regimes, while a broader range of interventions are evaluated elsewhere

East Asia and Pacific			
62 Financial regulation		34 Tax policy	
		33 Transportation infrastructure	
South Asia		Sub-Saharan Africa	
21 Access to credit/savings		46 Vocational education	
		38 Agricultural extension/training	
		36 Access to credit/savings	
Latin America and Caribbean			
18 Conditional Cash Transfers			
11 Vocational education		31 Vocational education	
		22 Conditional Cash Transfers	
		18 Tax policy	

Number of impact evaluations evaluating each intervention in each region. Not shown: Europe/Central Asia and MENA regions, which have very few IEs for this SDG.



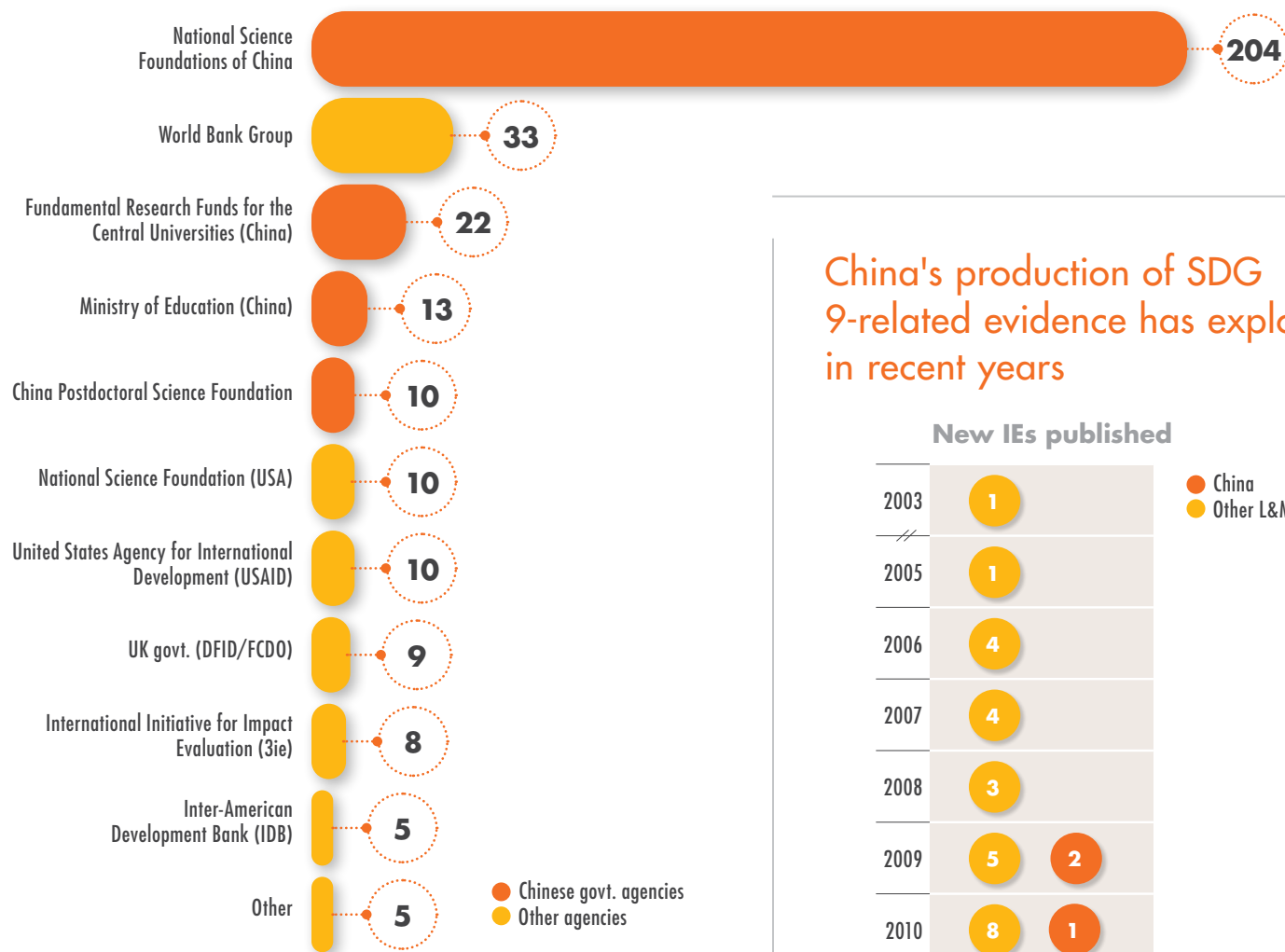


While SDG 9 has one of the larger evidence bases in terms of impact evaluations with over 700, it has one of the smallest evidence bases in terms of systematic reviews, with only 9. The effectiveness literature on industry and infrastructure would benefit from additional evidence synthesis.

The evidence for SDG 9 is particularly skewed towards recent China-focused studies: approximately 60 per cent of SDG 9 studies focus on China, primarily regulatory regimes and transportation infrastructure. Not surprisingly, Chinese government agencies lead by far in funding these studies.

721 IMPACT EVALUATIONS
9 SYSTEMATIC REVIEWS

Chinese government agencies have funded much of the SDG 9 effectiveness literature



China's production of SDG 9-related evidence has exploded in recent years

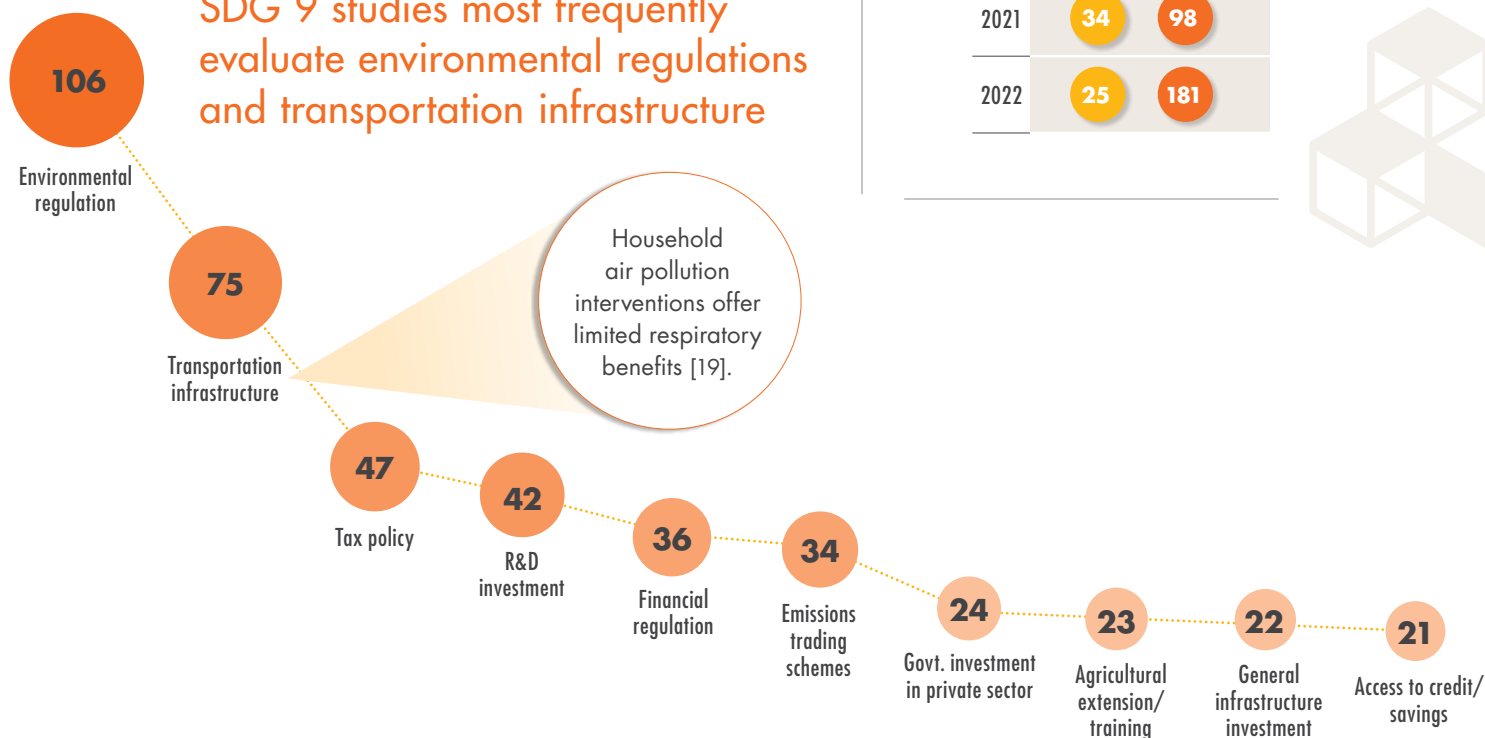
New IEs published

Year	Other L&MICs	China
2003	1	0
2005	1	0
2006	4	0
2007	4	0
2008	3	0
2009	5	2
2010	8	1
2011	10	0
2012	12	1
2013	18	4
2014	9	1
2015	9	3
2016	13	2
2017	23	9
2018	36	14
2019	40	41
2020	39	53
2021	34	98
2022	25	181

Legend: ● China, ● Other L&MICs



SDG 9 studies most frequently evaluate environmental regulations and transportation infrastructure





The size of the evidence base from impact evaluations for SDG 10 is at the median of our dataset, with just over 450 impact evaluations. The distribution of evidence across regions is much more balanced than for other SDGs.

When we compare the availability of SDG 10 evidence with country-level economic inequality (as measured by the Gini Coefficient), we find that evidence is most abundant among countries with moderate levels of inequality. However, we do not see the same acute lack of evidence from countries with the most severe challenges as we do for other SDGs. This means there is at least some relevant evidence to draw upon for decision-making on programming that targets inequality, even in contexts where income inequality is severe.

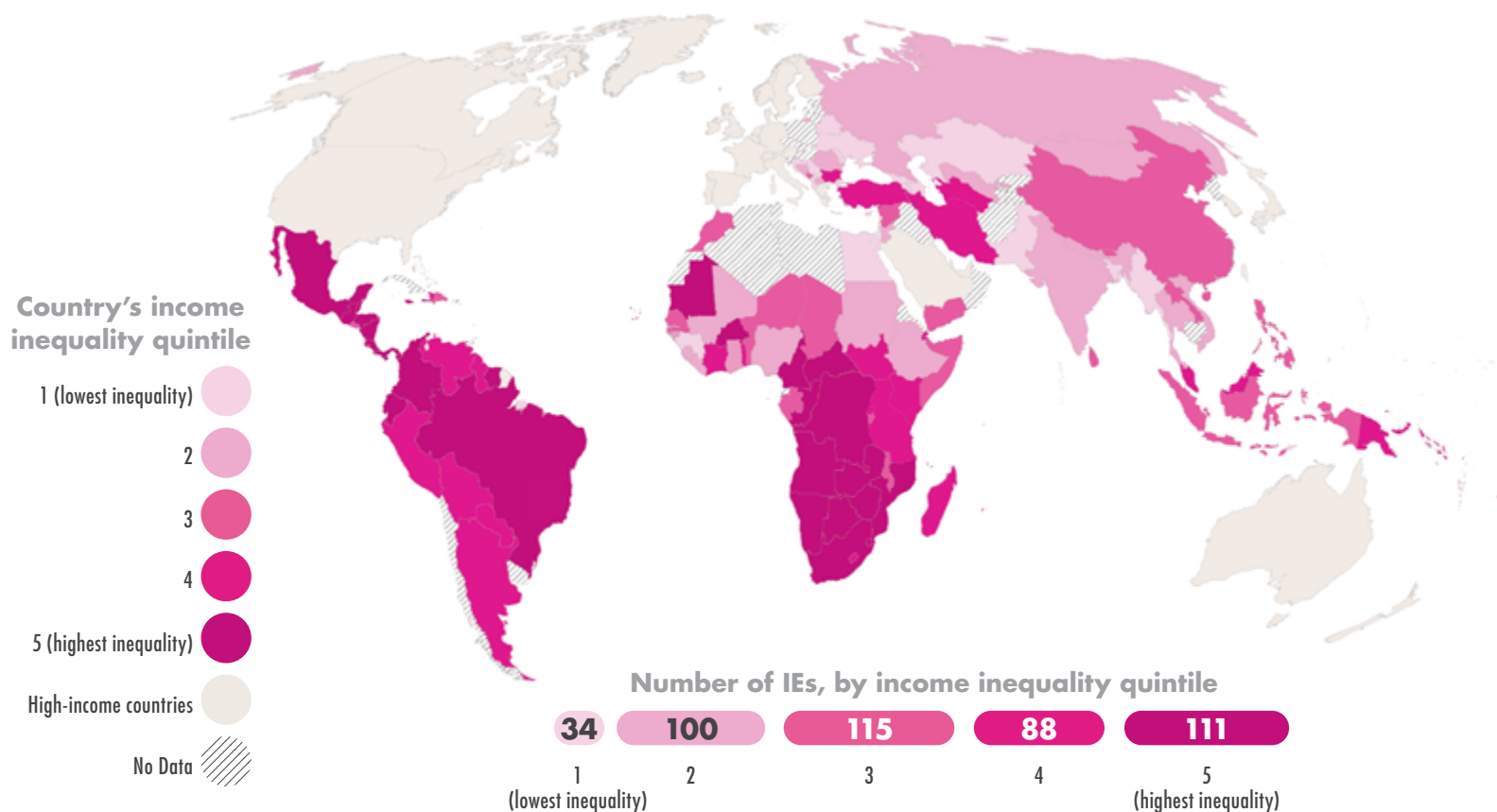
Conditional and unconditional cash transfers are both among the most frequently evaluated interventions, along with access to credit. The most frequently measured outcomes are economic indicators; several educational outcomes also appear among the top 10, but they are measured much less frequently.

With only 14 SRs, SDG 10 has one of the lowest ratios of SRs to IEs of any SDG. Additional synthesis of inequality-related evidence would help ensure the evidence base provides actionable guidance.

453 IMPACT EVALUATIONS

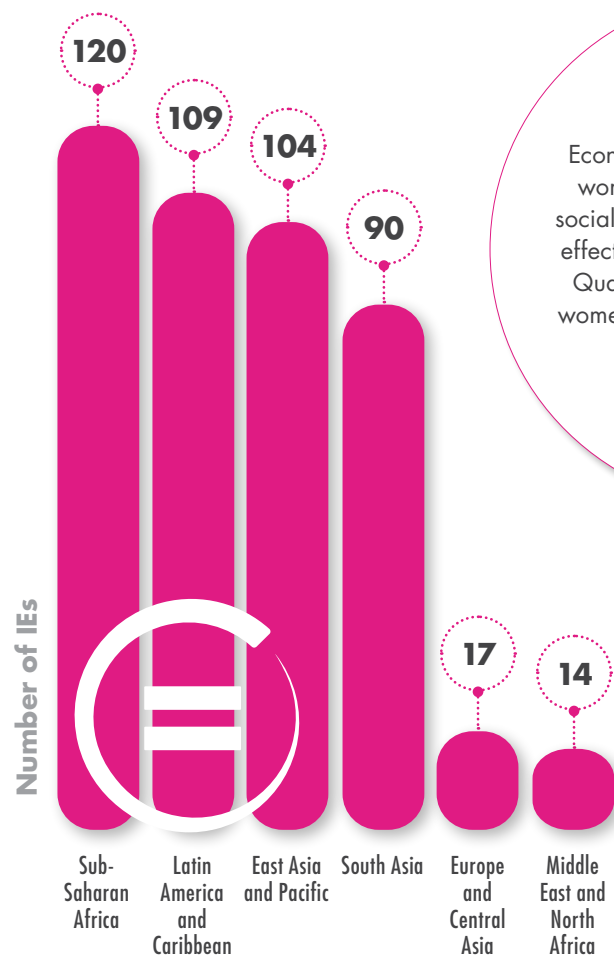
14 SYSTEMATIC REVIEWS

SDG 10 evidence is concentrated in countries with moderate levels of economic inequality



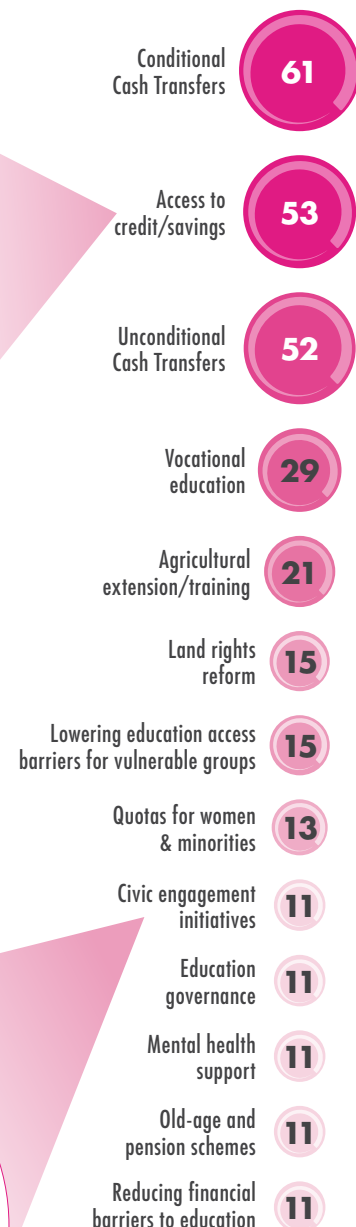
Income inequality is measured by the Gini Coefficient. Source for income inequality data: Our World in Data [38].

Evidence on reducing inequality is evenly distributed across four major regions



Economic self-help groups enhance women's economic, political, and social empowerment, and have larger effects when combined with training. Qualitative evidence indicates that women members perceive themselves as empowered [42].

Cash transfers are the most commonly evaluated interventions in studies focused on inequality



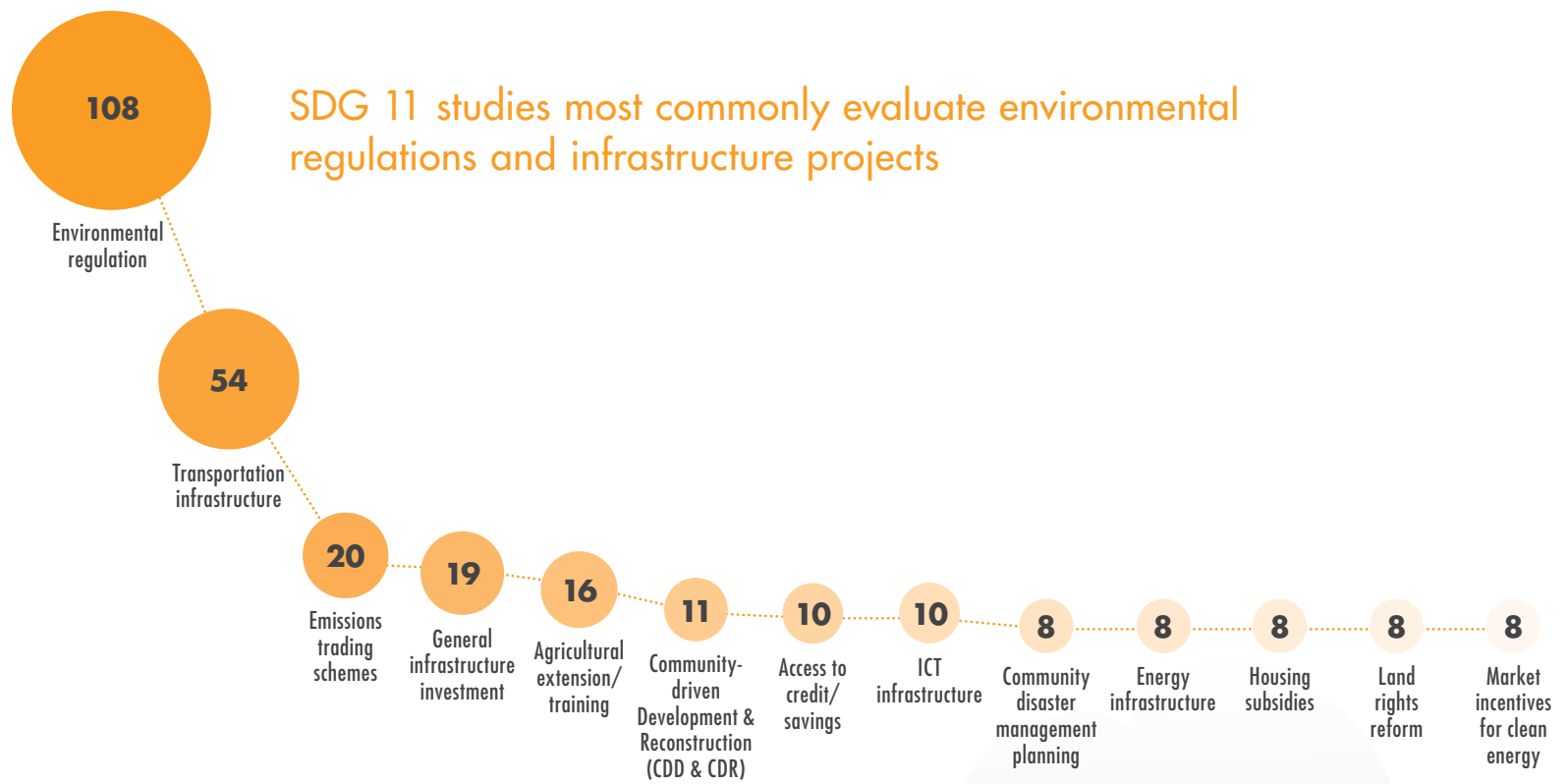
Programmes that foster citizen-provider engagement can enhance service quality. But pushing for political accountability without addressing provider challenges is generally ineffective [41].



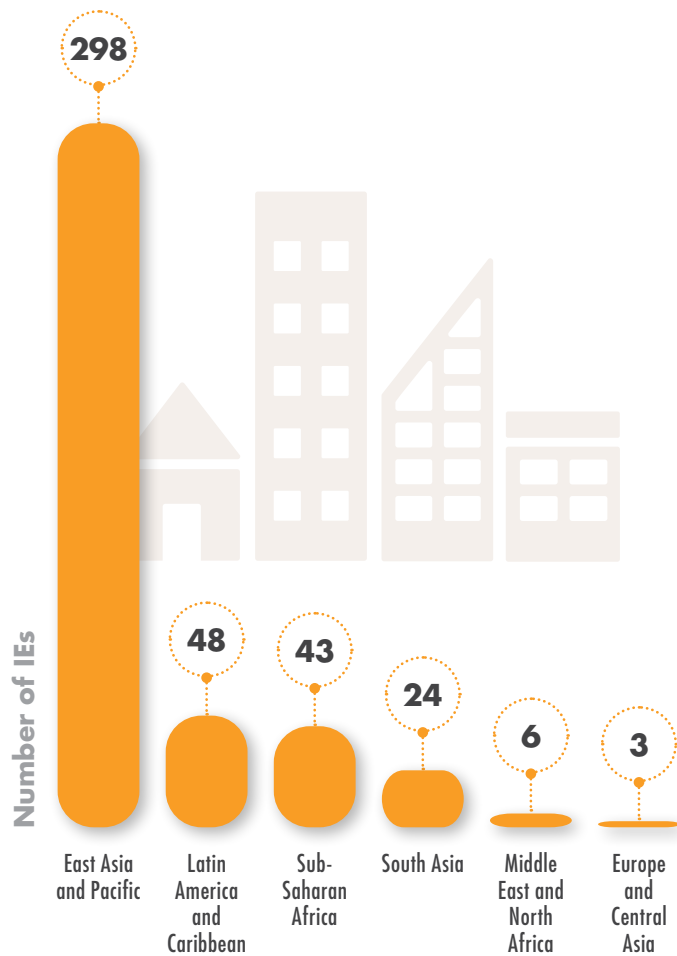
SDG 11 has a moderately-sized evidence base of 424 impact evaluations, and a very limited amount of synthesis, with only 4 systematic reviews. SDG 11 is dominated by recent China-focused studies, the literature being quite scarce for all other regions. Environmental regulation and infrastructure projects are the most commonly evaluated interventions, while air pollution and greenhouse gas emissions are the most commonly measured outcomes.

Additional evidence from a broader range of contexts is needed for a well-rounded evidence base. SDG 11 is also notable for currently having no medium- or high-confidence SRs, which highlights the need for quality synthesis in this area.

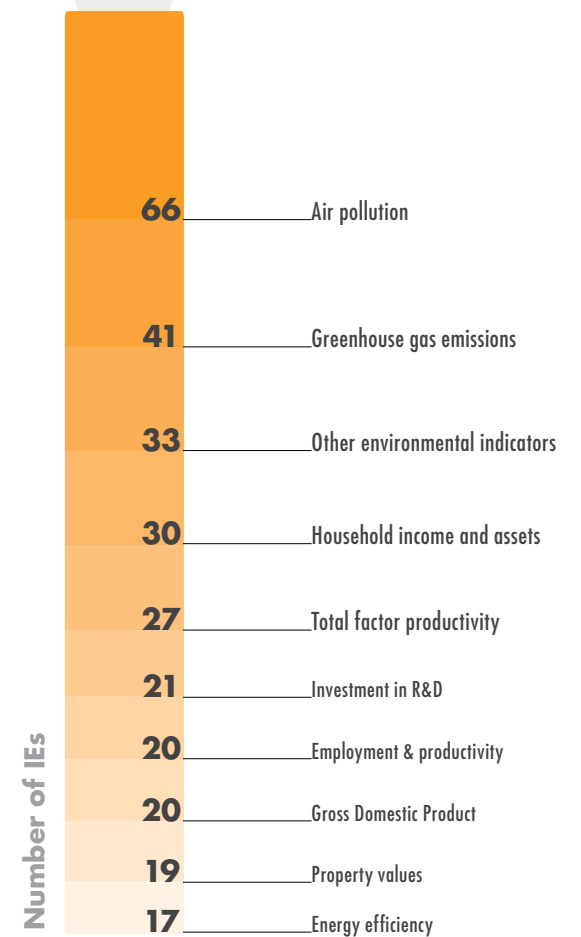
424 IMPACT EVALUATIONS
4 SYSTEMATIC REVIEWS



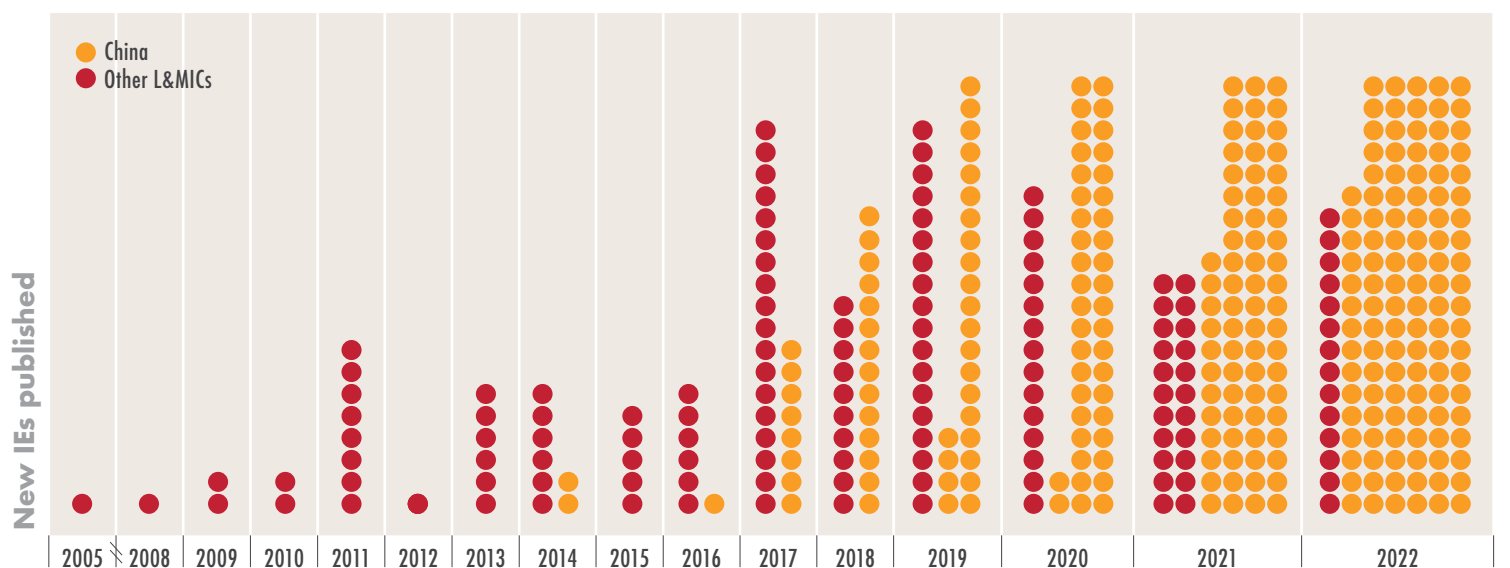
There is little SDG 11-related evidence outside East Asia (China in particular)



SDG 11 studies most commonly measure air pollution and greenhouse gas emissions



China-focused studies have come to dominate the SDG 11-related literature in recent years



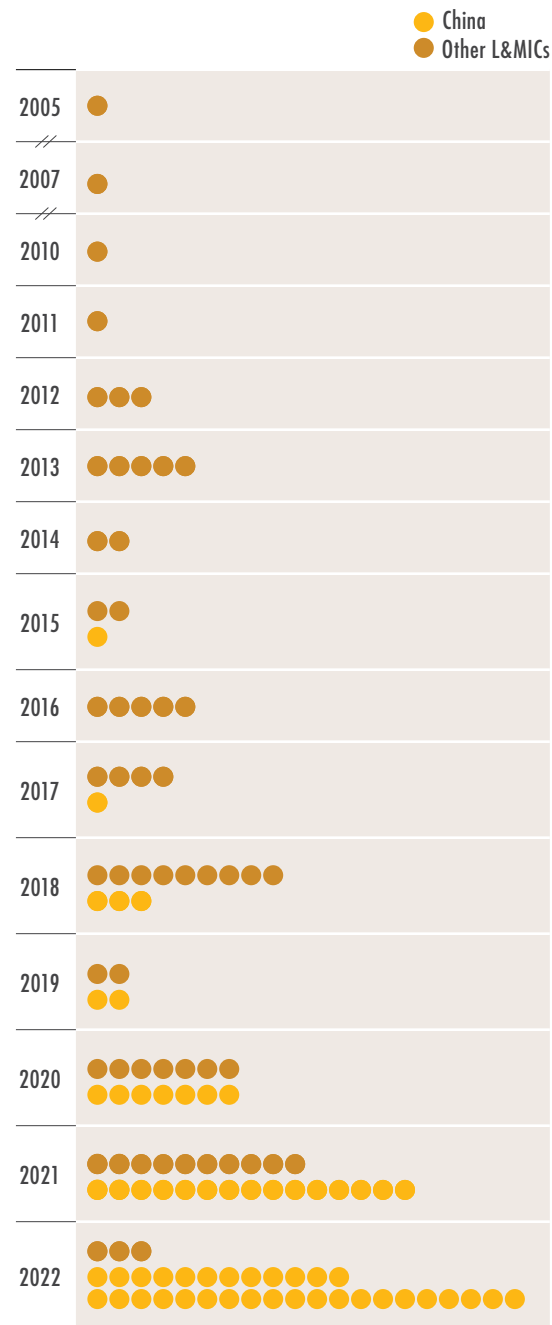


SDG 12 has one of the smallest evidence bases in our dataset, with only about 120 impact evaluations and 5 systematic reviews. This limits our ability to draw firm conclusions about effectiveness research in this area.

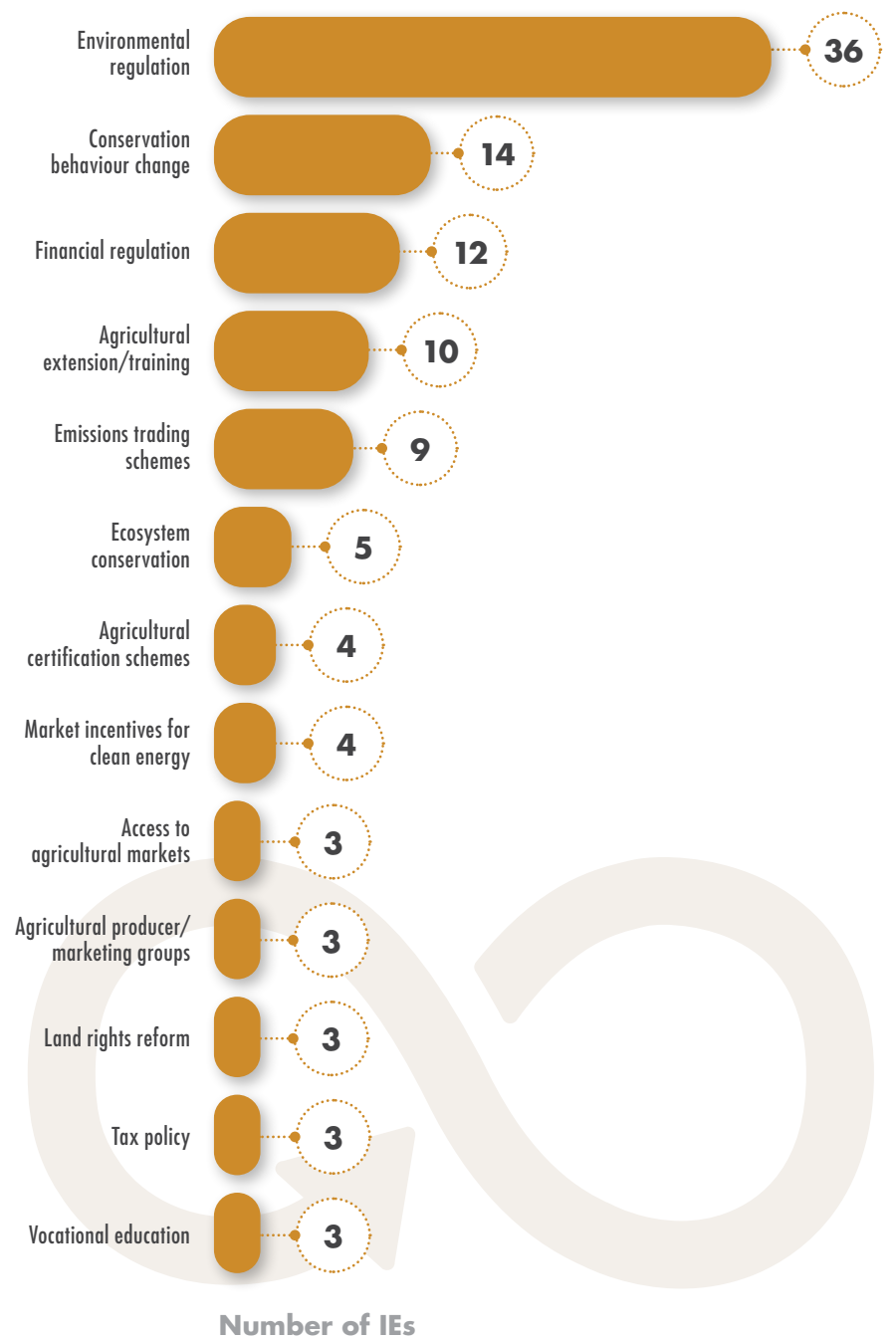
Like many other SDGs, SDG 12 has seen a recent spike in China-focused research, though the numbers are still small in absolute terms. About half of all SDG 12 studies focus on China. India is the only other country with more than five evaluations. As with other literature dominated by recent China-focused studies, the most commonly evaluated intervention is environmental regulation.

SDG 12 needs much more attention from the research community to build a rigorous and actionable evidence base.

China's production of SDG 12-related evidence has spiked in recent years, but the evidence base remains small



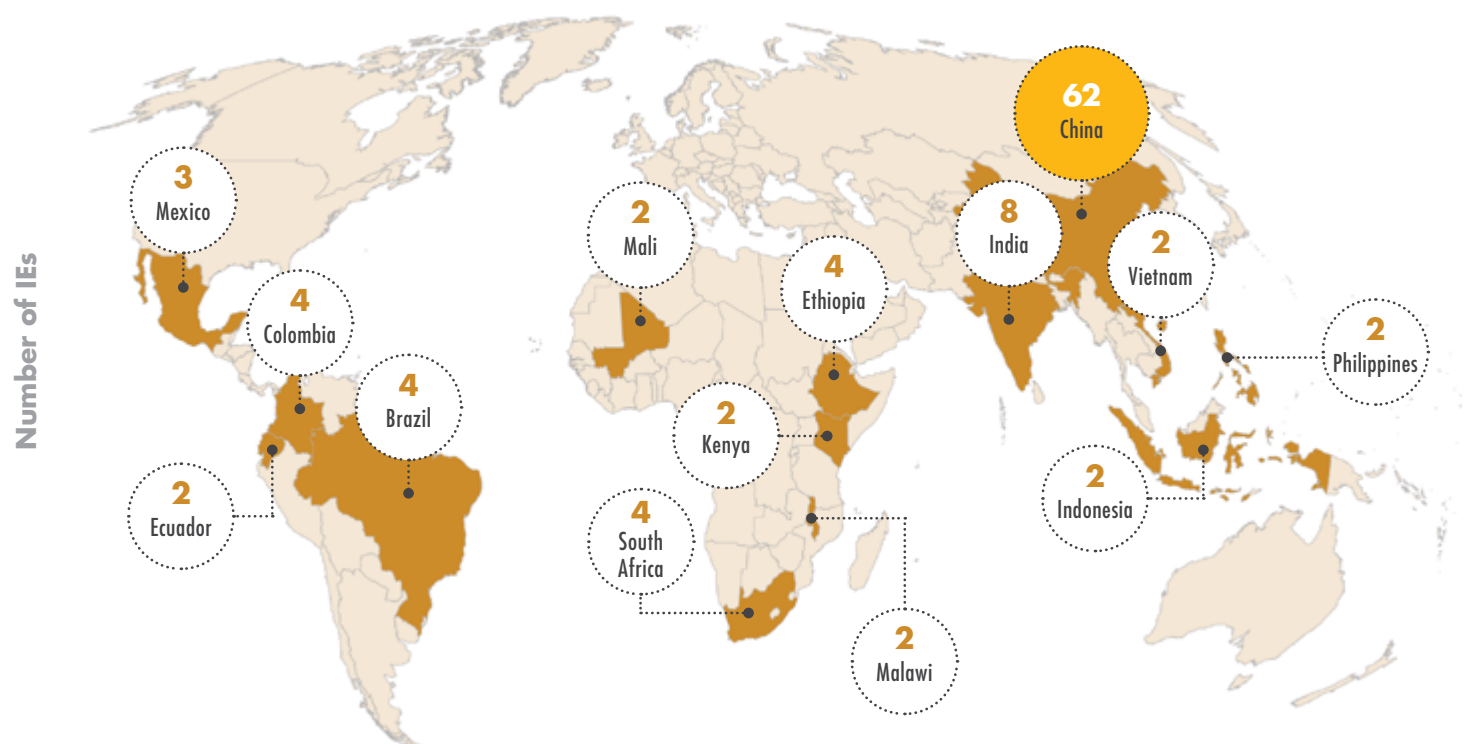
The SDG 12 literature focuses on environmental regulations in China



121 IMPACT EVALUATIONS

5 SYSTEMATIC REVIEWS

China is the only country with a substantial evidence base for SDG 12



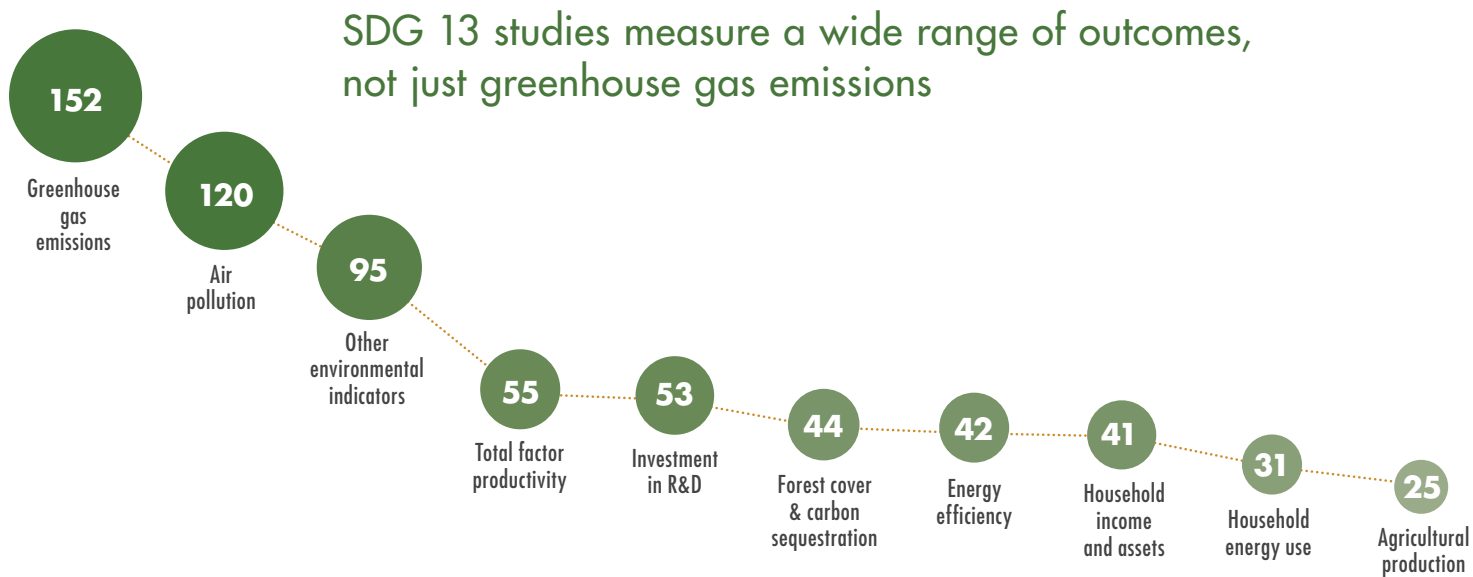


SDG 13 has a fairly large number of impact evaluations in absolute terms (725). Approximately 80 percent of the SDG 13 impact evaluations have focused on China. Of the six L&MICs that are among the world's top 10 greenhouse gas emitters, China is the only one with much of an evidence base. The vast majority of these China-focused studies are quasi-experimental evaluations of Chinese environmental regulations, mostly conducted in the last five years. These studies are nearly all authored by researchers at Chinese institutions, with funding from the Chinese government.

Interestingly, fewer than half of SDG 13 studies measure greenhouse gas emissions as an outcome. Many studies of environmental regulations measure impacts on air pollution and other environmental indicators instead. The presence of total factor productivity and household income among the top 10 outcomes indicates that at least some attention is being paid to the economic, as well as the ecological, effects of environmental policies in China.

Evidence is needed from a broader range of contexts and on a broader range of interventions. SDG 13 also has almost no systematic reviews (7), suggesting this literature is a good candidate for attention from synthesis researchers.

725 IMPACT EVALUATIONS
7 SYSTEMATIC REVIEWS

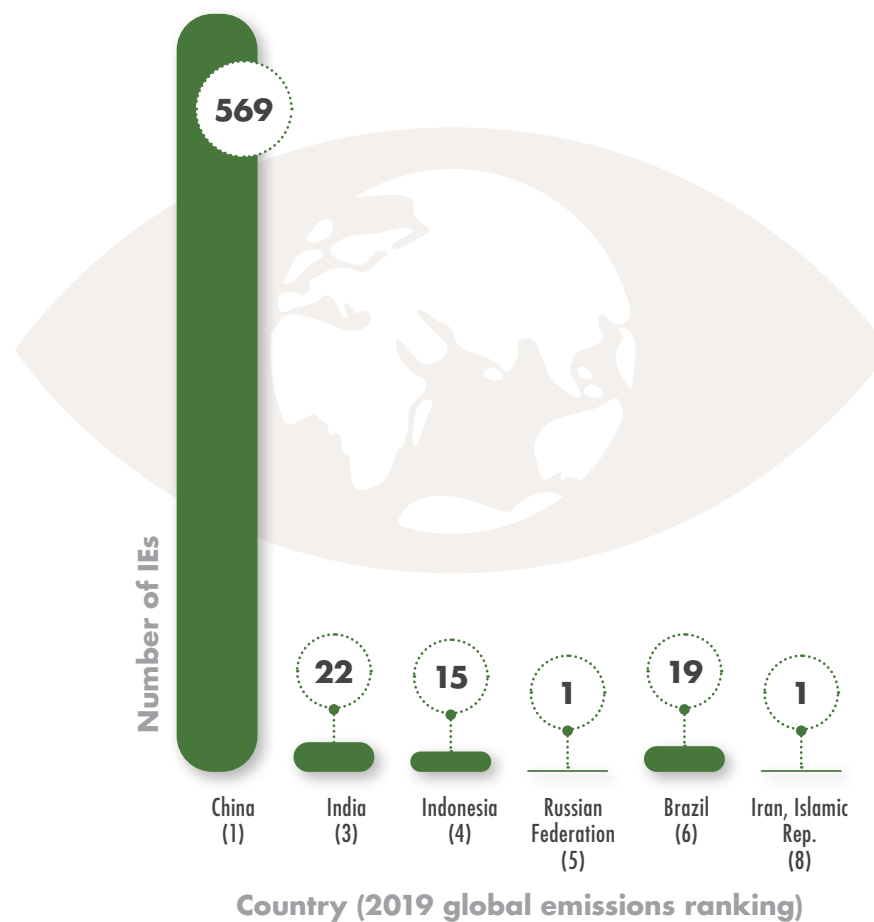


Nearly all SDG 13-related evidence has emerged from China in the last 5 years

Year	Other L&MICs	China
2008	1	0
2009	2	0
2010	1	1
2011	3	0
2012	0	1
2013	2	0
2014	2	3
2015	15	1
2016	8	2
2017	17	9
2018	17	16
2019	21	41
2020	21	72
2021	24	119
2022	20	286

New IEs published

Among L&MICs in the top 10 global greenhouse gas emitters, only China has a sizable evidence base



Emissions data source: World Resources Institute [45].

Why are there so few SDG 13 studies outside China?

It is unclear why there are so few evaluations in other countries – we might expect this to be a popular research topic worldwide given the prominence of climate change in global discourse. The prevalence of China-focused studies seems to reflect two conditions that may be somewhat unique: (1) China has implemented a large number of domestic policies targeting climate change, providing a large body of potential interventions to be evaluated; and (2) as one of the world's largest economies, China has the resources to mobilise a large domestic research effort. It is likely that many L&MICs have not implemented climate-related policies as aggressively, leaving less to evaluate. We also expect there is a large literature evaluating the effects of climate-related interventions enacted in high-income countries (e.g., [43], [44]), but unless these measure outcomes in at least one L&MIC, they would fall outside the DEP's scope.



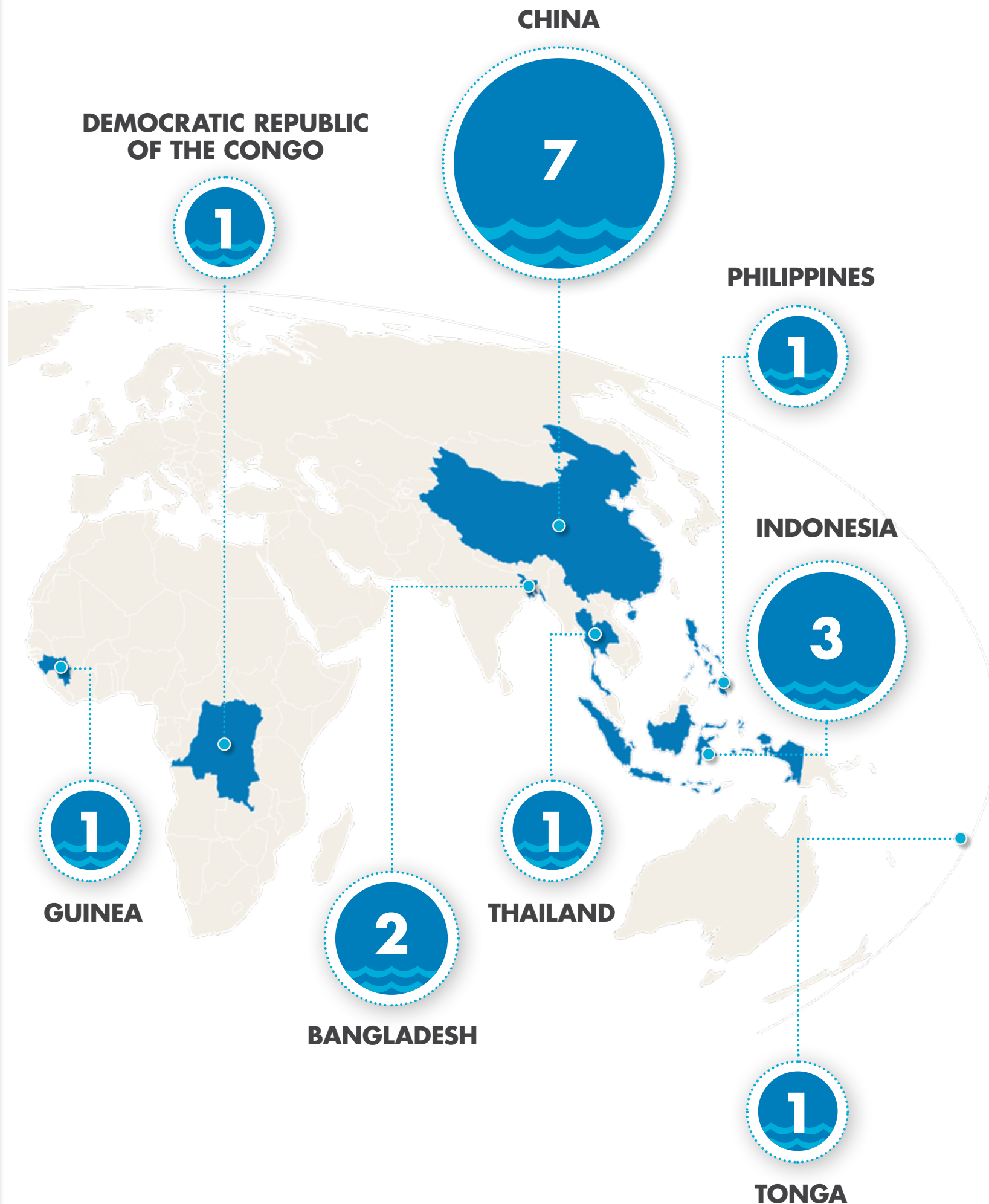
With fewer than 20 impact evaluations and only 2 (low-confidence) systematic reviews, there is little we can say about the evidence base for SDG 14, except that it is inadequate.

However, with overfishing and pollution of marine ecosystems reaching crisis levels worldwide, potentially devastating consequences loom not only for the Earth's interconnected marine ecosystems but also for the coastal zones of many L&MICs. We need dedicated efforts to develop a rigorous evidence base for the impacts of a range of interventions on the coastal zones and marine ecosystems of L&MICs.

The lack of SDG 14 studies in our dataset may partially reflect the DEP's focus on L&MICs. Studies that measure the impacts of interventions on the world's interconnected marine ecosystems may not explicitly frame themselves as L&MIC-focused research, and may therefore fall outside of DEP's scope. Initial scoping of this possibility suggests it is more likely that the SDG 14 literature is small than that the DEP is missing a substantial body of evidence. But this requires further testing.

Among the handful of SDG 14 studies, the most commonly evaluated intervention is environmental regulation, with four studies. The the most frequently measured outcomes are household income/assets (four studies) and agricultural production (three studies). The top outcomes suggest that much of the existing SDG 14-related research focuses on the economic impacts of interventions (e.g., those targeting coastal fisheries) on households, rather than on the health and sustainability of marine ecosystems.

There are only a handful of SDG 14 studies, and they are nearly all from the East Asia region



17 IMPACT EVALUATIONS
2 SYSTEMATIC REVIEWS

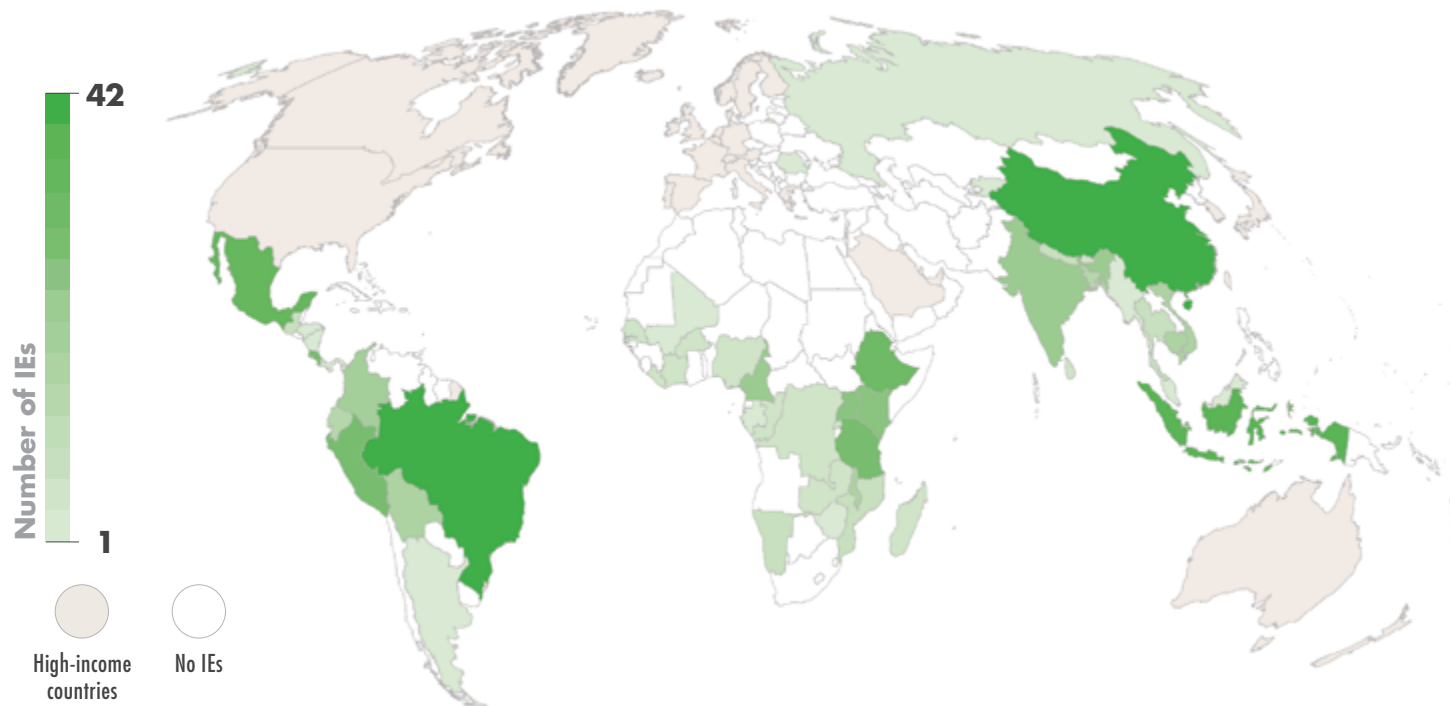


SDG 15 has a relatively small evidence base, with 271 impact evaluations and 18 systematic reviews. These studies are roughly equally distributed across Latin America, Sub-Saharan Africa, and East Asia. Forest conservation policy is the most frequently evaluated intervention, followed by participatory forest management. The most frequently measured outcome is forest cover, with other outcomes having to do with agricultural/agroforestry production and its economic effects.

Given this SDG's very limited evidence base, more studies are needed. Importantly, future research should measure environmental indicators other than forest cover, such as biodiversity.

271 IMPACT EVALUATIONS
18 SYSTEMATIC REVIEWS

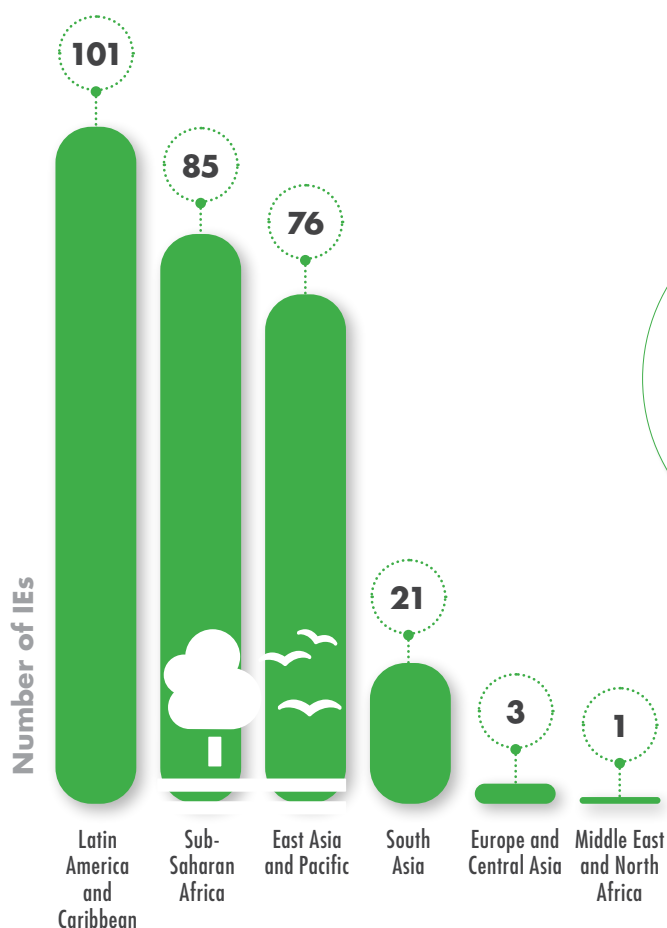
Brazil and China are leading sites of SDG 15 studies



Apart from forest cover, SDG 15 studies also measure agricultural and economic outcomes



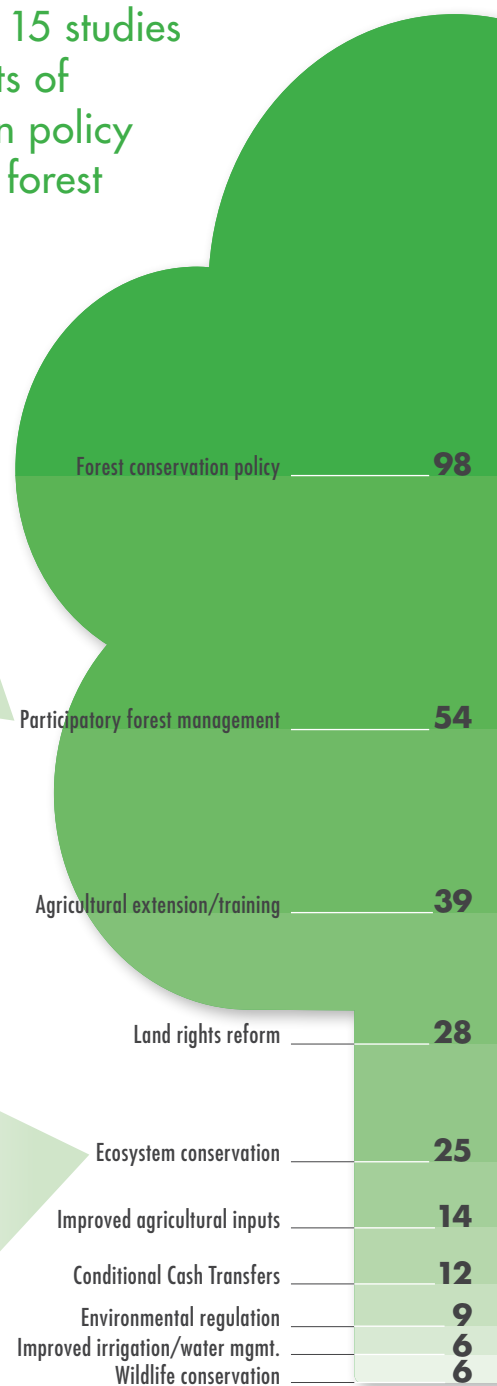
Evidence for SDG 15 is distributed across Latin America, Sub-Saharan Africa, and East Asia



Over half of SDG 15 studies evaluate the effects of forest conservation policy and participatory forest management

Terrestrial protected areas have varied impacts. While decentralised forest management shows modest benefits against deforestation, it might have negative effects on poverty [48], [49]

Payment for ecosystem services (PES) programmes may curb deforestation, enhance forest cover, and increase incomes [46]. PES may only increase incomes among wealthier landowners, and seem to be less effective in poor areas [47].





SDG 16 has a moderately sized evidence base, with 613 impact evaluations and 22 systematic reviews. Sub-Saharan Africa and Latin America are the most studied regions, but South Asia and East Asia have substantial evidence bases as well.

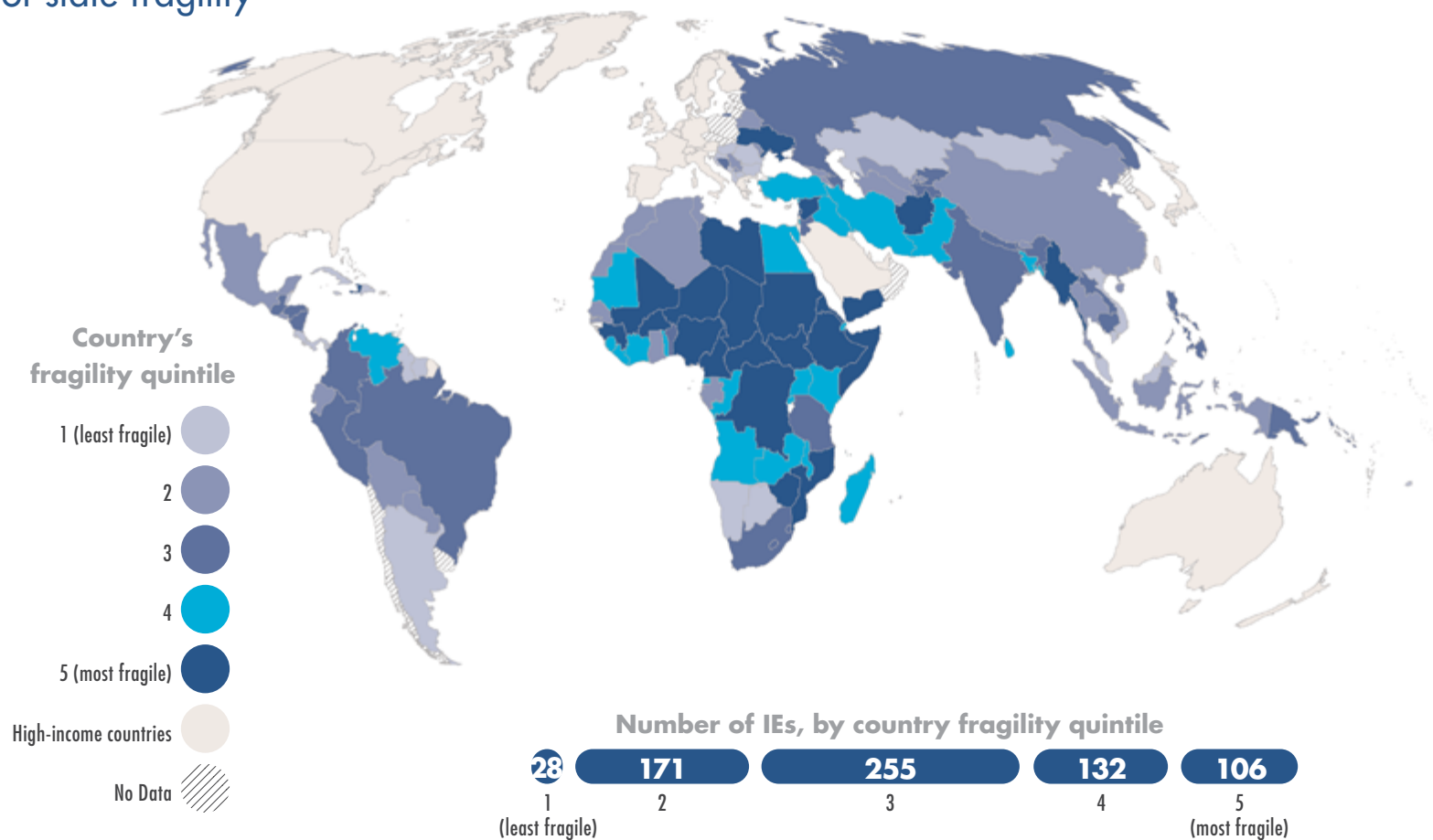
When we compare the availability of SDG 16 evidence with countries' fragility rankings, we find most evidence focused on countries with moderate levels of state fragility. However, there are over 100 IEs from countries with the highest fragility index scores.

SDG 16 studies most commonly evaluate civic engagement initiatives, with some attention to electoral systems and quotas (e.g., in local governments) for women and minorities. Evaluations tend to measure outcomes related to civic knowledge and engagement, as well as crime and violence.

Given the large share of aid spending dedicated to SDG 16 interventions, there is a need for greater attention to quality synthesis.

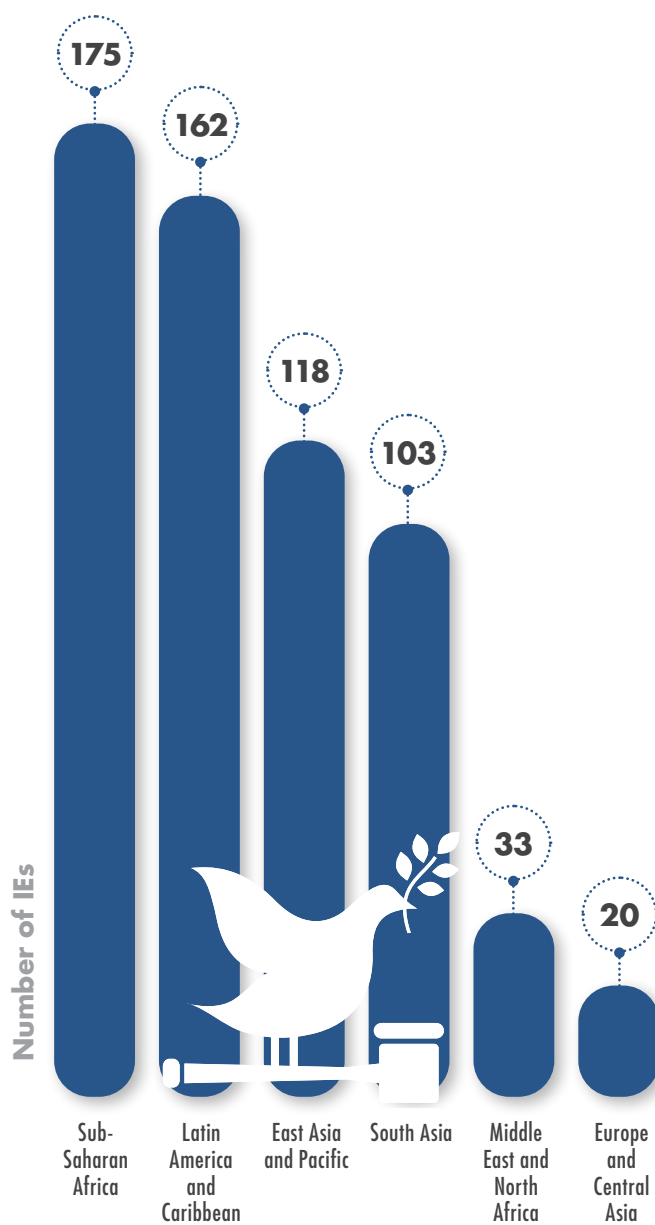
613 IMPACT EVALUATIONS
22 SYSTEMATIC REVIEWS

Evidence for SDG 16 is mostly concentrated in countries with moderate levels of state fragility

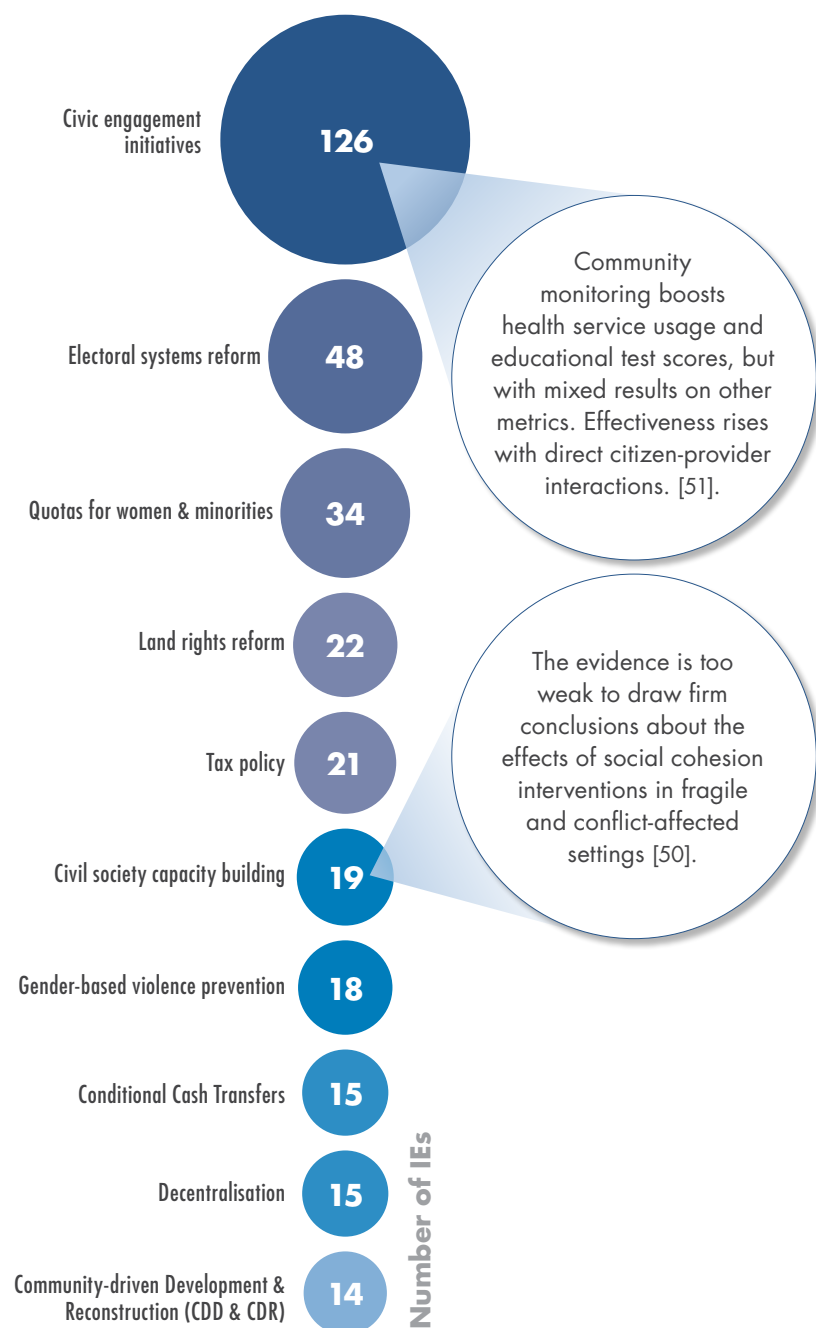


Source for fragility data: Fragile States Index [14].

SDG 16 evidence is fairly evenly distributed across four major regions



A large plurality of SDG 16 studies examine the effects of civic engagement activities





SDG 17 is unique among the SDGs in that it is less a goal in itself than a strategy for pursuing the other 16 goals. There are few impact evaluations related to SDG 17 (only 26), and no systematic reviews. Interventions in this domain are likely to be one-off agreements that are often difficult to evaluate using traditional IE methods.

Among the SDG 17 studies that do exist, the only intervention category to be evaluated more than once is tax policy, which has been the subject of six evaluations. This suggests that the existing research is related to the first indicator for SDG 17: "Strengthen domestic resource mobilization, including through international support to developing countries, to improve domestic capacity for tax and other revenue collection" [50].

26 IMPACT EVALUATIONS

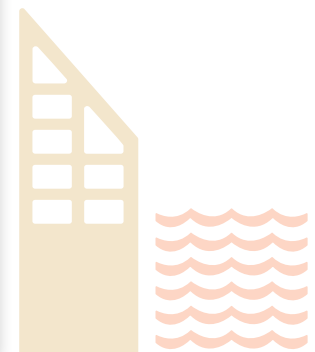
0 SYSTEMATIC REVIEWS

SDG 1
No poverty

1,656 IMPACT EVALUATIONS
61 SYSTEMATIC REVIEWS

SDG 2
Zero hunger

1,328 IMPACT EVALUATIONS
63 SYSTEMATIC REVIEWS



SDG 3
Good health and well-being

993 IMPACT EVALUATIONS
104 SYSTEMATIC REVIEWS

SDG 4
Quality education

1,085 IMPACT EVALUATIONS
68 SYSTEMATIC REVIEWS

SDG 5
Gender equality

441 IMPACT EVALUATIONS
23 SYSTEMATIC REVIEWS

SDG 6
Clean water and sanitation

360 IMPACT EVALUATIONS
50 SYSTEMATIC REVIEWS

SDG 7
Affordable clean energy

330 IMPACT EVALUATIONS
12 SYSTEMATIC REVIEWS

SDG 8
Decent work and economic growth

1,304 IMPACT EVALUATIONS
37 SYSTEMATIC REVIEWS

SDG 9
Industry, innovation and infrastructure

721 IMPACT EVALUATIONS
9 SYSTEMATIC REVIEWS



SDG 10
Reduced inequalities

453 IMPACT EVALUATIONS
14 SYSTEMATIC REVIEWS

SDG 11
Sustainable cities and communities

424 IMPACT EVALUATIONS
4 SYSTEMATIC REVIEWS

SDG 12
Responsible consumption and production

121 IMPACT EVALUATIONS
5 SYSTEMATIC REVIEWS

SDG 13
Climate action

725 IMPACT EVALUATIONS
7 SYSTEMATIC REVIEWS

SDG 14
Life below water

17 IMPACT EVALUATIONS
2 SYSTEMATIC REVIEWS

SDG 15
Life on land

271 IMPACT EVALUATIONS
18 SYSTEMATIC REVIEWS

SDG 16
Peace, justice and strong institutions

613 IMPACT EVALUATIONS
22 SYSTEMATIC REVIEWS

SDG 17
Partnerships for the goals

26 IMPACT EVALUATIONS
0 SYSTEMATIC REVIEWS



References

- [1] ESCWA, "The 5Ps of the Sustainable Development Goals," United Nations Economic and Social Commission for West Asia, Beirut. Accessed: Oct. 29, 2023. [Online]. Available: https://www.unescwa.org/sites/default/files/inline-files/the_5ps_of_the_sustainable_development_goals.pdf
- [2] T. de Hoop *et al.*, "What works to accelerate progress on the Partnership Pillar of the Sustainable Development Goals: A synthesis of evaluative evidence," The Global SDG Synthesis Coalition, New York, 2023. [Online]. Available: <https://www.sdgsynthesiscoalition.org/pillar/partnership-pillar>
- [3] O. Hankivsky, K. W. Springer, and G. Hunting, "Beyond sex and gender difference in funding and reporting of health research," *Research Integrity and Peer Review*, vol. 3, no. 1, Dec. 2018, doi: 10.1186/s41073-018-0050-6.
- [4] V. A. Welch *et al.*, "Extending the PRISMA statement to equity-focused systematic reviews (PRISMA-E 2012): explanation and elaboration," *Journal of Development Effectiveness*, pp. 1–38, Jan. 2016, doi: 10.1080/19439342.2015.1113196.
- [5] V. A. Welch *et al.*, "CONSORT-Equity 2017 extension and elaboration for better reporting of health equity in randomised trials," *BMJ*, vol. 359, p. j5085, Nov. 2017, doi: 10.1136/bmj.j5085.
- [6] J. D. Sachs, G. Lafortune, G. Fuller, and E. Drumm, "Implementing the SDG Stimulus. Sustainable Development Report 2023," Dublin University Press, 2023. doi: 10.25546/102924.
- [7] O. Forestier and R. E. Kim, "Cherry-picking the Sustainable Development Goals: Goal prioritization by national governments and implications for global governance," *Sustainable Development*, vol. 28, no. 5, pp. 1269–1278, 2020, doi: 10.1002/sd.2082.
- [8] United Nations, World investment report 2014: *Investing in the SDGs - an action plan*. in United Nations Conference on Trade and Development (UNCTAD) World Investment Report (WIR). United Nations Conference on Trade and Development, 2014. doi: 10.18356/3e74cde5-en.
- [9] UNDP, "Joint UN-Wide Syntheses of Evidence of SDGs' Achievements in Peace, People, Prosperity, Planet, Partnership." United Nations Development Programme, Oct. 2021. Accessed: Aug. 18, 2023. [Online]. Available: https://procurement-notices.undp.org/view_notice.cfm?notice_id=85080



- [10] 3ie, "Adding training on financial management and gender equality to microfinance in Vietnam." Accessed: Oct. 19, 2023. [Online]. Available: <https://www.3ieimpact.org/evidence-hub/Evidence-impact-summaries/adding-training-financial-management-and-gender-equality>
- [11] 3ie, "Informing a cook stoves programme in Ghana." Accessed: Oct. 19, 2023. [Online]. Available: <https://www.3ieimpact.org/evidence-hub/Evidence-impact-summaries/informing-cook-stoves-programme-ghana>
- [12] P. J. Gertler, S. Martinez, P. Premand, L. B. Rawlings, and C. M. J. Vermeersch, *Impact Evaluation in Practice, Second Edition*. The World Bank, 2016. doi: 10.1596/978-1-4648-0779-4.
- [13] W. D. Savedoff, R. Levine, and N. Birdsall, "When will we ever learn? Improving lives through impact evaluation," Center for Global Development, Washington, DC, 2006.
- [14] Fund for Peace, "Fragile States Index." 2023. Accessed: Aug. 21, 2023. [Online]. Available: <https://fragilestatesindex.org/excel/>
- [15] Our World in Data, "Data on Poverty." Our World in Data, Jul. 16, 2023. Accessed: Aug. 22, 2023. [Online]. Available: <https://github.com/owid/poverty-data>
- [16] F. Pega, R. Pabayo, C. Benny, E.-Y. Lee, S. K. Lhachimi, and S. Y. Liu, "Unconditional cash transfers for reducing poverty and vulnerabilities: effect on use of health services and health outcomes in low- and middle-income countries," *Cochrane Database of Systematic Reviews*, no. 3, Art. no. 3, 2022, doi: 10.1002/14651858.CD011135.pub3.
- [17] S. Doocy and H. Tappis, "Cash-based approaches in humanitarian emergencies: a systematic review," *Campbell Systematic Reviews*, vol. 13, no. 1, Art. no. 1, Jan. 2017, doi: 10.4073/csr.2017.17.
- [18] Global Hunger Index, "Global Hunger Index Scores by 2022 GHI Rank." Global Hunger Index, 2022. Accessed: Aug. 23, 2023. [Online]. Available: <https://www.globalhungerindex.org/ranking.html>
- [19] S. Lawry, C. Samii, R. Hall, A. Leopold, D. Hornby, and F. Mtero, "The Impact of Land Property Rights Interventions on Investment and Agricultural Productivity in Developing Countries: a Systematic Review," *Campbell Systematic Reviews*, vol. 10, no. 1, Art. no. 1, 2014, doi: 10.4073/csr.2014.1.
- [20] C. Gonzalez Parrao *et al.*, "Aquaculture for improving productivity, income, nutrition and women's empowerment in low- and middle-income countries: A systematic review and meta-analysis," *Campbell Systematic Reviews*, vol. 17, no. 4, Art. no. 4, 2021, doi: 10.1002/cl2.1195.
- [21] M. Haghani, A. Behnood, V. Dixit, and O. Oviedo-Trespalacios, "Road safety research in the context of low- and middle-income countries: Macro-scale literature analyses, trends, knowledge gaps and challenges," *Safety Science*, vol. 146, p. 105513, Feb. 2022, doi: 10.1016/j.ssci.2021.105513.



- [22] S. Saleh *et al.*, "Air pollution interventions and respiratory health: a systematic review," *The International Journal of Tuberculosis and Lung Disease*, vol. 24, no. 2, Art. no. 2, Feb. 2020, doi: 10.5588/ijtld.19.0417.
- [23] UNICEF, "Secondary Education and Enrollment Statistics." UNICEF Data, Jun. 2022. Accessed: Aug. 19, 2023. [Online]. Available: <https://data.unicef.org/topic/education/secondary-education/>
- [24] R. Carr-Hill, C. Rolleston, R. Schendel, and H. Waddington, "The effectiveness of school-based decision making in improving educational outcomes: a systematic review," *Journal of Development Effectiveness*, vol. 10, no. 1, Art. no. 1, 2018, doi: 10.1080/19439342.2018.1440250.
- [25] B. Snilstveit *et al.*, "Interventions for improving learning outcomes and access to education in low- and middle-income countries: A systematic review," *3ie*, 2015.
- [26] UNDP, "Gender Inequality Index," *Human Development Reports*. Human Development Reports, United Nations Development Programme, 2022. Accessed: Aug. 19, 2023. [Online]. Available: <https://hdr.undp.org/data-center/documentation-and-downloads>
- [27] L. Orton, A. Pennington, S. Nayak, A. Sowden, M. White, and M. Whitehead, "Group-based microfinance for collective empowerment: a systematic review of health impacts," *Bull World Health Organ*, vol. 94, no. 9, Art. no. 9, Sep. 2016, doi: 10.2471/BLT.15.168252.
- [28] M. Chinen, T. de Hoop, L. Alcázar, M. Balarin, and J. Sennett, "Vocational and business training to improve women's labour market outcomes in low- and middle-income countries: a systematic review," *Campbell Systematic Reviews*, vol. 13, no. 1, Art. no. 1, 2017, doi: 10.4073/csr.2017.16.
- [29] E. Lwamba *et al.*, "Strengthening women's empowerment and gender equality in fragile contexts towards peaceful and inclusive societies: A systematic review and meta-analysis," *Campbell Systematic Reviews*, vol. 18, no. 1, Art. no. 1, 2022, doi: 10.1002/cl2.1214.
- [30] D. Daley, M. McCauley, and N. van den Broek, "Interventions for women who report domestic violence during and after pregnancy in low- and middle-income countries: a systematic literature review," *BMC Pregnancy Childbirth*, vol. 20, no. 1, Art. no. 1, Mar. 2020, doi: 10.1186/s12884-020-2819-0.
- [31] F. Majorin, B. Torondel, G. K. S. Chan, and T. Clasen, "Interventions to improve disposal of child faeces for preventing diarrhoea and soil-transmitted helminth infection," *Cochrane Database of Systematic Reviews*, no. 9, Art. no. 9, 2019, doi: 10.1002/14651858.CD011055.pub2.



- [32] E. de Buck *et al.*, “Promoting handwashing and sanitation behaviour change in low-and middle-income countries: a mixed-method systematic review,” in *Promoting handwashing and sanitation behaviour change in low-and middle-income countries: a mixed-method systematic review*, 2017, pp. 519–519. Accessed: Aug. 23, 2023. [Online]. Available: https://www.3ieimpact.org/sites/default/files/2019-01/SR%2036-Behaviour-change-sanitation_2.pdf
- [33] T. Yates, J. Allen, M. Leandre Joseph, and D. Lantagne, *WASH Interventions in Disease Outbreak Response*. Oxfam, 2017. doi: 10.21201/2017.8753.
- [34] T. Yates, J. Allen, M. Leandre Joseph, and D. Lantagne, “Short-term WASH interventions in emergency response: a systematic review,” *International Initiative for Impact Evaluation (3ie)*, Feb. 2017. doi: 10.23846/SR0033.
- [35] M. Berretta, J. Furgeson, Y. (Nicole) Wu, C. Zamawe, I. Hamilton, and J. Evers, “Residential energy efficiency interventions: A meta-analysis of effectiveness studies,” *Campbell Systematic Reviews*, vol. 17, no. 4, Art. no. 4, 2021, doi: 10.1002/cl2.1206.
- [36] G. Bensch, M. Sievert, J. Langbein, and N. Kneppel, “Effects and mechanisms of market-based reforms on access to electricity in developing countries: a systematic review,” *International Initiative for Impact Evaluation (3ie)*, 2016. Accessed: Dec. 15, 2017. [Online]. Available: <http://www.3ieimpact.org/en/publications/systematic-review-publications/3ie-systematic-review-31/>
- [37] J. Kluge *et al.*, “Interventions to improve the labour market outcomes of youth: A systematic review of training, entrepreneurship promotion, employment services and subsidized employment interventions,” *Campbell Systematic Reviews*, vol. 13, no. 1, Art. no. 1, 2017, doi: 10.4073/csr.2017.12.
- [38] X. Cirera and R. Lakshman, “The impact of export processing zones on employment, wages and labour conditions in developing countries: systematic review,” *Journal of Development Effectiveness*, vol. 9, no. 3, Art. no. 3, 2017, doi: 10.1080/19439342.2017.1309448.
- [39] C. Piza *et al.*, “Business support for small and medium enterprises in low- and middle-income countries,” *International Initiative for Impact Evaluation (3ie)*, Jul. 2016. doi: 10.23846/SR51180cida.
- [40] J. Hine, M. Abedin, R. Stevens, T. Airey, and T. Anderson, “Does the extension of the rural road network have a positive impact on poverty reduction and resilience for the rural areas served? If so how, and if not why not? A systematic review. Protocol.” *Social Science Research Unit, UCL Institute of Education, University of London*, 2016.
- [41] H. Waddington, A. Sonnenfeld, J. Finetti, M. Gaarder, and J. Stevenson, “Does incorporating participation and accountability improve development outcomes? Meta-analysis and framework synthesis,” *International Initiative for Impact Evaluation (3ie)*, New Delhi, *Systematic Review 43*, May 2019. doi: 10.23846/SR00043.



- [42] C. Brody *et al.*, "Economic Self-Help group Programs for Improving Women's Empowerment: A Systematic Review," *Campbell Systematic Reviews*, vol. 11, no. 1, Art. no. 1, 2015, doi: 10.4073/csr.2015.19.
- [43] J. B. Rode, A. L. Dent, C. N. Benedict, D. B. Brosnahan, R. L. Martinez, and P. H. Ditto, "Influencing climate change attitudes in the United States: A systematic review and meta-analysis," *Journal of Environmental Psychology*, vol. 76, p. 101623, Aug. 2021, doi: 10.1016/j.jenvp.2021.101623.
- [44] J. D. Ford, L. Berrang-Ford, and J. Paterson, "A systematic review of observed climate change adaptation in developed nations," *Climatic Change*, vol. 106, no. 2, pp. 327–336, May 2011, doi: 10.1007/s10584-011-0045-5.
- [45] J. Friedrich, M. Ge, A. Pickens, and L. Vigna, "Global Historical Emissions." World Resources Institute, Mar. 02, 2023. Accessed: Aug. 21, 2023. [Online]. Available: <https://www.wri.org/insights/interactive-chart-shows-changes-worlds-top-10-emitters>
- [46] B. Snilsveit *et al.*, "Incentives for climate mitigation in the land use sector – the effects of payment for environmental services (PES) on environmental and socio-economic outcomes in low- and middle-income countries: a mixed-method systematic review," International Initiative for Impact Evaluation (3ie), Jul. 2019. doi: 10.23846/SR00044.
- [47] C. Samii *et al.*, "Effects of Payment for Environmental Services (PES) on Deforestation and Poverty in Low and Middle Income Countries: A Systematic Review," *Campbell Systematic Reviews*, vol. 10, no. 1, Art. no. 1, 2014, doi: 10.4073/csr.2014.11.
- [48] A. S. Pullin *et al.*, "Human well-being impacts of terrestrial protected areas," *Environ Evid*, vol. 2, no. 1, p. 19, Oct. 2013, doi: 10.1186/2047-2382-2-19.
- [49] C. Samii, M. Lisiecki, P. Kulkarni, L. Paler, and L. Chavis, "Decentralised forest management for reducing deforestation and poverty in low- and middle-income countries," 2015, doi: 10.23846/SR41173a.
- [50] A. Sonnenfeld *et al.*, "Strengthening intergroup social cohesion in fragile situations," International Initiative for Impact Evaluation (3ie), Feb. 2021. doi: 10.23846/SR00046.
- [51] E. Molina, L. Carella, A. Pacheco, G. Cruces, and L. Gasparini, "Community monitoring interventions to curb corruption and increase access and quality of service delivery in low- and middle-income countries: a systematic review," *Campbell Systematic Reviews*, vol. 12, no. 1, Art. no. 1, 2016, doi: 10.4073/csr.2016.8.
- [52] United Nations, "Goal 17 targets and indicators," 2015. Accessed: Aug. 21, 2023. [Online]. Available: <https://sdgs.un.org/goals/goal17>



- [53] J. Benn, G. Gualberti, A. Nichols, and V. Thielemans, "Proposal to include an SDG focus field in the CRS database," Organization for Economic Co-Operation and Development, DCD/DAC/STAT(2018)41/REV1, 2018. Accessed: Aug. 17, 2023. [Online]. Available: [https://one.oecd.org/document/DCD/DAC/STAT\(2018\)41/REV1/en/pdf](https://one.oecd.org/document/DCD/DAC/STAT(2018)41/REV1/en/pdf)
- [54] OECD, "DAC and CRS code lists - OECD." Organisation for Economic Co-operation and Development, Jul. 2023. Accessed: Aug. 17, 2023. [Online]. Available: <https://www.oecd.org/dac/financing-sustainable-development/development-finance-standards/dacandcrscodelists.htm>
- [55] J. Benn and V. Gaveau, "Comprehensive review of purpose codes and policy markers in light of the SDGs and post-2015 agenda," Organization for Economic Co-Operation and Development, Paris, DCD/DAC/STAT(2015)9, 2015. Accessed: Aug. 17, 2023. [Online]. Available: [https://one.oecd.org/document/DCD/DAC/STAT\(2015\)9/en/pdf](https://one.oecd.org/document/DCD/DAC/STAT(2015)9/en/pdf)
- [56] A. Pincet, S. Okabe, and M. Pawelczyk, "Linking Aid to the Sustainable Development Goals – a machine learning approach," OECD, Paris, OECD Development Co-operation Working Papers 52, Feb. 2019. doi: 10.1787/4bdaeb8c-en.
- [57] J. Hudson, "Promises kept, promises broken? The relationship between aid commitments and disbursements," *Review of Development Finance*, vol. 3, no. 3, Art. no. 3, Jul. 2013, doi: 10.1016/j.rdf.2013.08.001.

Appendix A: Methods used to create this report

A.1 The types of evidence contained in DEP

In order for a study to be eligible for the DEP as an impact evaluation, it must meet the following criteria:

1. Adopt a counterfactual-based method (experimental or quasi-experimental) for providing a quantitative estimate of an intervention's effect.
2. Evaluate the effects of a social or economic programme or policy on people or environments in one or more L&MICs (this includes studies measuring the effects of policies – such as immigration policy – adopted in high-income countries on residents of L&MICs).

To keep the scope of DEP commensurate with the resources available for maintaining it, the DEP excludes studies focused on mental health (except in populations affected by violence or displacement) and non-communicable diseases. The DEP also excludes clinical trials and agronomic studies that aim to determine the effects of a treatment or procedure under ideal or tightly controlled conditions. Finally, the DEP excludes studies that measure the impact of uncoordinated social or economic activities in the absence of a deliberate intervention (for example, studies that examine differences in outcomes between farmers who do and do not adopt a particular technology, in the absence of any dedicated effort to disseminate or promote the technology).

In order for a study to qualify for the DEP as a systematic review, it must report its inclusion criteria as well as systematic procedures for:

1. Searching for studies
2. Identifying eligible studies from among the search results
3. Analysing the included studies¹⁵

¹⁵ Analysis for systematic reviews may be quantitative or qualitative, as long as the analysis method is clearly reported and systematic.



A.2 How the DEP is updated

We maintain the content of the DEP by using systematic search and screening procedures to identify IEs and SRs focusing on the effects of social and economic interventions in L&MICs. We use a “continuous evidence surveillance” model, whereby we perform monthly searches for new papers across a variety of scholarly databases and institutional portals. We then screen search results using a combination of machine learning classifiers and trained human screeners.

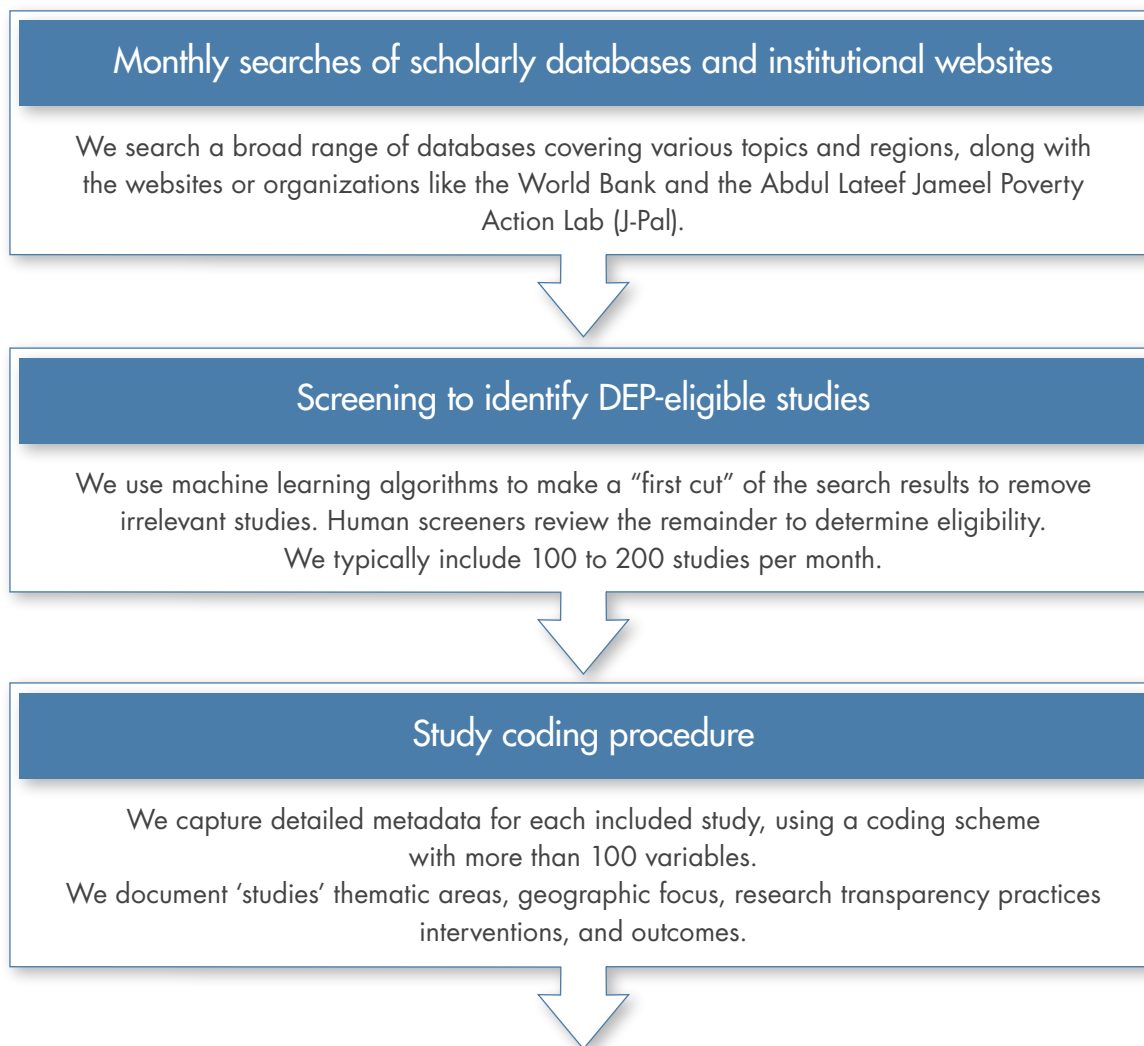
The search terms used to identify for systematic reviews combine two concepts: L&MIC terms and SR methods terms – i.e., the search strategy retrieves items whose metadata contain *both* terms related to L&MICs *and* terms related to SR methods. The search terms used to identify IEs replace SR methods terms with IE methods terms, and add a third concept: interventions, policies, and programmes. That is, the strategy to identify IEs retrieves items that mention all of L&MICs, IE methods, and an intervention.¹⁶

We search the following databases on a monthly or quarterly basis:

- | | |
|---|--|
| - Academic search complete (EBSCO) | - International Political Science Abstracts (Ovid) |
| - Africa-Wide Information (EBSCO) | - MEDLINE (Ovid) |
| - Agricola (Ovid) | - Oxfam Policy & Practice (EBSCO) |
| - CAB Abstracts (Ovid) | - PsycInfo (Ovid) |
| - Communications & Mass Media (EBSCO) | - RePEc (EBSCO) |
| - EconLit (Ovid) | - Science Direct (EBSCO) |
| - Embase (Ovid) | - Scopus (Elsevier) |
| - ERIC (Ovid) | - Web of Science (Clarivate) |
| - Gender Studies (EBSCO) | - WHO Global Index Medicus (WHO website) |
| - Global Health (formerly CAB Global Health) (Ovid) | - World Bank eLibrary (EBSCO) |
| - GreenFILE (EBSCO) | |

¹⁶ The additional requirement for study metadata to mention interventions is used to increase the precision of the search strategy and keep the number of search hits manageable.

Figure A.1: Process for populating the Development Evidence Portal



Once eligible studies are identified and added to the portal, 3ie conducts an extensive study coding process to categorise studies and capture additional information that is made available to users of the portal. Among other things, studies are coded according to applicable sectors (using the World Bank’s taxonomy of sectors), SDGs, research funding sources, and the interventions evaluated and outcomes measured in the study. In addition, for SRs, we conduct detailed critical appraisals and based on the review’s methods, assign each review a confidence rating of high, medium, or low.

Each study is classified into a single sector based on the nature of the intervention evaluated, and classified according to one or more SDGs based on the overall thematic relevance of the study to the SDGs, taking into account the outcomes measured and the stated purpose of the intervention (See [Appendix B.2](#) for additional details about SDG classification decisions).



A.3 Why this report excludes evidence from the health sector

The DEP has been brought up to date, for all sectors except for health, to the end of 2022. Health is the largest sector by a significant margin, accounting for approximately 40 per cent of studies in the DEP. Resource limitations meant that we could not fully code all new IEs and SRs added to the literature in the last few years. Because other resources exist dedicated to evidence in the health sector (e.g., [Epistemonikos](#)), we have de-prioritised the health sector to ensure the DEP could be kept up to date and complete for all other sectors.

Because the DEP's health sector data are incomplete, all health sector studies in the DEP are excluded from the analyses presented in this report. Thus, our findings for SDG 3: Good health and well-being report on the characteristics of evidence for the impacts of non-health interventions on health-related outcomes.

A.4 How we classified studies' interventions and outcomes for this report

We use 3ie's [taxonomy](#) of interventions and outcomes to categorise the interventions evaluated and outcomes measured in each study. For the purposes of this report, we have made some modifications to the coding of interventions and outcomes. First, because the intervention and outcome categories in the taxonomy are very fine-grained, we have grouped similar interventions and outcomes into broader categories to make patterns more evident. For example, interventions such as livestock management training, crop management training, and farmer field schools were among the categories grouped under agricultural extension/training. Similarly, outcome categories such as deforestation, vegetation levels, and forest coverage were grouped under forest cover and carbon sequestration. Through this consolidation exercise, we grouped 362 common intervention types into 57 categories, and 444 common outcomes into 31 categories.

In addition, for this report, our analysis looks at intervention *components*, regardless of whether they are studied in isolation or as part of a multi-component intervention. In general, it is important to consider evaluations of multi-component interventions separately from evaluations of the same components when implemented in isolation. The theory of change for an intervention that combines, for example, a cash transfer with entrepreneurial training will be different from that for a cash transfer intervention alone. For this reason, the DEP coding scheme captures all intervention components and indicates whether they were evaluated separately (e.g., using multiple treatment arms) or as a package. However, this approach leads to a large number of unique combinations of components that may be evaluated in relatively few studies. This makes it difficult to discern patterns in what is being studied, hence our decision to analyse components rather than full intervention packages.



A.5 How we capture studies' attention (or lack of attention) to gender and equity

The DEP coding scheme includes a protocol for capturing whether and how studies account for gender or equity in their research design. The protocol focuses on the research process rather than the intervention. Thus, even if a study evaluates an intervention that targets women or another vulnerable group, it can be coded as “Does not address gender or equity”, if the researchers did not adopt an approach to the research design or analysis that sheds light on the equitability of the intervention’s effects.

A.6 How we assess the quality of systematic reviews

3ie uses a comprehensive [critical appraisal tool](#) to assign confidence ratings to SRs. The tool assesses the methods used to conduct the review against established best practices in synthesis methodology. A review’s rating of “high”, “medium”, or “low” reflects how confident readers can be in the review’s conclusions, based on how it was carried out. Research staff trained in applying the tool conduct appraisals, and a senior staff member with expertise in evidence synthesis reviews them. The appraisal tool assesses questions like the following:

- Was the review’s search for evidence comprehensive?
- Did the review authors take steps to avoid bias and human error in selecting studies to include in the review?
- Did the review use a rigorous and reliable method to extract data from primary studies?
- Was the review’s analysis method (quantitative or qualitative) rigorous and systematic?
- Did the review use a recognised tool to assess the quality of primary studies, and take differences in quality into account in the analysis?



A.7 How we assessed the availability of evidence where it is most needed

In this report, we have drawn on external data to examine whether evidence for particular SDGs is available “where it is needed most” – that is, in countries that face the greatest challenges in attaining those SDGs. For example, we compared the availability of evidence related to SDG 1 (No Poverty) with the percentage of a country’s population living in extreme poverty (i.e., below the international poverty line of USD 2.15 per day). For several SDGs, however, there were no broadly applicable country-level indicators with readily available data or the volume of evidence in the DEP was too small to make meaningful comparisons.

The data on country-level indicators we used for comparisons with availability of evidence were of different types. Therefore, we adopted several approaches to sorting countries into groups to make the comparisons. When the preparers of the dataset had already grouped the indicators into categories (as with the Global Hunger Index’s sorting of countries into hunger severity categories of Low, Moderate, etc.), we adopted those groupings.

When the indicators were expressed as percentages (such as the rate of extreme poverty in a country), and thus readily interpretable, we sorted them into five groups (i.e., 0–20%, 21–40%, etc.).¹⁷ For both pre-defined groupings and percentage groupings, the number of countries per group was not necessarily equal, so we took the average number of IEs per country.

However, the scales for some indicators, such as the Gender Inequality Index or Gini Coefficient, are not readily interpretable, so in these cases we divided countries into quintiles based on their scores. As the quintiles all had the same number of countries, we report the total number of IEs across all countries in each quintile.

A.8 ODA estimates for the SDGs

Allocating projects classified as ODA to SDGs presents a problem. In recent years, following a proposal by the Working Party on Development Finance Statistics, spending has been classified by SDG [53], but there are no official classifications of ODA by SDGs for spending reported to OECD before this change was introduced. ODA has traditionally been reported using the Development Assistance Committee (DAC) and Creditor Reporting System (CRS) purpose codes [54]. In 2015, an attempt was made to map these codes to SDGs as a way of tracking the resources dedicated to each

¹⁷ Note that for the comparison of SDG 1 evidence with extreme poverty rates, we combined the 61–80 per cent and the 81–100 per cent groups, because the only two countries with reported extreme poverty rates above 80 per cent were Madagascar, for which the most recent data were from 2012, and Uzbekistan, for which the most recent data were from 2003.



SDG. The group charged with this task concluded that there were many cases where DAC-CRS codes could not be reliably mapped to SDGs, and some SDGs were not adequately represented in the taxonomy of DAC-CRS codes [55, App. 3].

In 2018, a team of OECD researchers led by Arnaud Pincet applied a machine learning algorithm to assign projects, and their reported budgets, in the CRS system to one or more SDGs [56]. The group found the algorithm to be highly accurate in replicating human judgments in assigning projects to SDGs.

Pincet and colleagues highlight some challenges related to classifying development assistance projects to SDGs. Certain themes cut across several SDGs. Allowing classifications to multiple SDGs can partially address this, as Pincet and colleagues allowed for ODA and as 3ie allows for studies in DEP. However, certain SDGs pose particular challenges. For example, as Pincet and colleagues point out, most, if not all, ODA is intended to some extent to reduce poverty, and is therefore relevant to SDG 1. Likewise, a great deal of aid spending aims, at least indirectly, to reduce inequalities and could be classified as relevant to SDG 10. Pincet and colleagues adopted a strategy of assigning projects to SDG 1 only when they target “explicitly the poorest populations” or provide “basic services” [56, p. 34]. As shown in the ODA data, only a small percentage of projects met this criterion.

In contrast, if a DEP study evaluates an intervention that is framed as a poverty-reduction programme, and if the outcomes measured are related to income or poverty measures, it could be classified under SDG 1 even if the population targeted is poor but not “the poorest”. As we see from the DEP data on outcomes, a great many studies indeed use these outcome measures, and so are classified as SDG 1 for the DEP, but Pincet and colleagues’ algorithm would not classify them the same way. Therefore, we believe the discrepancy between ODA and evidence for SDGs 1 and 10 are more likely attributable to differences in how DAC projects and research studies tend to frame their objectives, than to a genuine misalignment between spending and evidence. We have noted that the DEP coding schema takes a conservative approach to classifying studies to SDG 17 (Partnerships). However, Pincet and colleagues provide little information about how they classified project spending to SDG 17. Consequently, we are unable to say how much the discrepancy between evidence and spending for SDG 17 is also the result of different coding practices.

Note that when comparing evidence to ODA, we used ODA disbursements rather than commitments because committed funds are not always disbursed [57].



Appendix B:

Some caveats to this report

B.1 Studies from non-L&MIC countries are mostly excluded

The DEP search strategy targets L&MIC-focused studies. This works well for SDGs like 1 and 2, because they focus on interventions targeting localised challenges like poverty and food security. But this approach may miss some relevant evidence, particularly research centring on the “Planet”-focused SDGs (6, 12–15), which concern the Earth’s interconnected ecological systems. Some evidence relevant to mitigating the effects of climate change and ecosystem degradation on L&MICs may exist in studies that do not frame themselves as L&MIC-focused studies.

B.2 Studies might be categorised to SDGs in multiple ways

The 17 SDGs are, by design, interconnected, interdependent, and overlapping in themes. As a consequence, many interventions and outcomes are relevant to multiple SDGs, which presents a challenge for consistently categorising research studies by SDG. We have endeavoured to maintain consistency in our classifications by categorising studies according to the main goals of their interventions and the relevance of their outcomes to specific SDG targets. Also, because the DEP’s primary function is to help users find relevant evidence, our classification decisions prioritise discoverability: we aim to code studies to the SDGs where we anticipate users of DEP’s SDG filter will expect to find them.¹⁸

However, there is inevitably some judgment involved on the part of coders in choosing the one or two most applicable SDGs for a study with broad relevance to multiple SDGs. To reduce the variability of such coding decisions, a random sample of 25 per cent of all DEP studies are selected for review by a more senior coder; any differences of opinion are discussed and if necessary, coding is modified to reflect the coders’ consensus. Even so, our decisions about assigning studies to SDGs may differ from the approaches taken by other SDG-focused research (e.g., on the amount of aid spending for the SDGs—see [Section 3.4](#) and [Appendix A.8](#)).

¹⁸ For example, while we typically classify studies to SDGs based on the outcomes measured, we will not always consider every single outcome measured in a study when making SDG classifications. If a study evaluates a teacher training intervention and measures 20 education-related outcomes, but also happens to show one health-related outcome in a results table without any broader discussion of health impacts, we would typically not classify that study to SDG 3, because this study is unlikely to be relevant to a user primarily interested in health evidence.



B.3 Many studies have been categorised to multiple SDGs

The DEP coding scheme allows coders to classify a study with multiple SDGs, because the evidence presented in a particular study will often be relevant in a fairly direct way for multiple SDGs. For instance, many studies evaluate interventions that aim to improve agricultural production and increase household income for smallholder farmers, and measure household-level outcomes of income and food security. Such a study is logically categorised to SDGs 1 (No Poverty) and 2 (Zero Hunger).

Across the full dataset, 2,999 IEs (41% of the sample) have been categorised to multiple SDGs. Of these, 2,473 were categorised to two SDGs, 526 were categorised to three SDGs, and 3 were categorised to four SDGs. The combination of SDGs 1 and 2 is by far the most common combination (446 studies). [Appendix C](#) shows all combinations with at least 25 studies.

Given the prominence of China-focused studies in our analysis, and the concentration of China-focused evidence in a handful of related SDGs, we examined the prevalence of multi-SDG studies on China in particular. Of the 1,837 China-focused studies in our sample, 658 (36%) were categorised to multiple SDGs. Thus, there is considerable “double counting” of China-focused studies in the tallies of studies for particular SDGs, but this is actually less common for China-focused studies than for the dataset as a whole. The most common combinations for China-focused studies are SDG 11 (Sustainable Cities & Communities) with SDG 13 (Climate Action) (81 studies), SDG 9 (Industry & Innovation) with SDG 13 (68 studies) and SDG 8 (Work & Economic Growth) with SDG 9 (57 studies). [Appendix C](#) shows all combinations with at least 10 studies.

B.4 Good country-level data on SDG-related indicators are sometimes unavailable

When we have compared the availability of evidence to SDG-related indicators, we have relied on data collated by leading international organisations such as the World Bank and the United Nations. However, these datasets typically rely on national surveys, some of which are quite out of date for certain countries. When using these datasets, we have used the most recent data available for each country, but in some cases, these may be a decade or more old. Consequently, these comparisons need to be interpreted with some caution.



B.5 Limitations to comparisons between evidence and other indicators

When comparing the availability of evidence with other indicators, such as aid spending, poverty rates, school completion rates, and so on, our data cannot tell us how to interpret these relationships, or what the relationships *should* be. For instance, if we find (as we sometimes do) that more evidence is available in countries with less severe challenges related to the SDGs, we cannot necessarily conclude that this is due to a misalignment between evidence needs and research priorities. It could be that these countries are doing better on these SDG indicators precisely because they have had larger evidence bases to draw upon.¹⁹

In fact, we believe that in most cases, the preferred pattern would be to have more evidence in sectors where more money is spent. This is for the simple reason that when aid is spent on initiatives that aren't backed by evidence, the risk of wasting resources on ineffective projects is greater. Likewise, when countries with the greatest development challenges have the least evidence, we believe the most straightforward explanation is that not enough resources have gone into generating evidence where it is most needed. The alternative explanation, that better-off countries are doing better because of their larger evidence bases, implies that these countries have institutionalised evidence use over a long period, which we believe to be unlikely given the recency of the turn towards use of rigorous evidence in development policy making.

Given the foregoing considerations, in this report we typically frame deviations between evidence volume and other indicators as evidence of misalignment. However, it is important to recognise that alternative explanations are available.

¹⁹ Assessing this hypothesis would require reliable longitudinal data on the indicators, which are frequently not available.

Appendix C: Common SDG combinations

Table C1: Frequent combinations of SDGs for impact evaluations – full dataset

SDG Combination			No. of IEs
1. No Poverty	2. Zero Hunger		446
3. Health & Well-being	6. Clean Water & Sanitation		139
8. Work & Econ. Growth	9. Industry & Innovation		131
11. Sustainable Cities & Communities	13. Climate Action		90
1. No Poverty	8. Work & Econ. Growth		88
1. No Poverty	10. Reduced Inequalities		86
2. Zero Hunger	3. Health & Well-being		80
9. Industry & Innovation	13. Climate Action		73
1. No Poverty	3. Health & Well-being		72
1. No Poverty	4. Quality Education		61
4. Quality Education	8. Work & Econ. Growth		59
3. Health & Well-being	4. Quality Education		58
1. No Poverty	2. Zero Hunger	3. Health & Well-being	57
3. Health & Well-being	5. Gender Equality		49
1. No Poverty	5. Gender Equality		48
13. Climate Action	7. Affordable & Clean Energy		48
2. Zero Hunger	8. Work & Econ. Growth		46
13. Climate Action	15. Life on Land		45
10. Reduced Inequalities	8. Work & Econ. Growth		43
16. Peace, Justice, & Institutions	5. Gender Equality		43
4. Quality Education	5. Gender Equality		38
10. Reduced Inequalities	4. Quality Education		37
5. Gender Equality	8. Work & Econ. Growth		36
13. Climate Action	8. Work & Econ. Growth		33

◀ **Table C1:** Frequent combinations of SDGs for impact evaluations – full dataset (cont'd)

SDG Combination			No. of IEs
12. Responsible Consum. & Prod.	13. Climate Action		31
16. Peace, Justice, & Institutions	8. Work & Econ. Growth		27
1. No Poverty	2. Zero Hunger	8. Work & Econ. Growth	26
1. No Poverty	3. Health & Well-being	4. Quality Education	26
16. Peace, Justice, & Institutions	4. Quality Education		26
1. No Poverty	15. Life on Land		25
11. Sustainable Cities & Communities	9. Industry & Innovation		25

Table C2: Frequent combinations of SDGs for impact evaluations – China studies

SDG Combination			No. of IEs
11. Sustainable Cities & Communities	13. Climate Action		81
9. Industry & Innovation	13. Climate Action		68
8. Work & Econ. Growth	9. Industry & Innovation		57
7. Affordable & Clean Energy	13. Climate Action		39
8. Work & Econ. Growth	13. Climate Action		32
12. Responsible Consum. & Prod.	13. Climate Action		26
1. No Poverty	2. Zero Hunger		24
9. Industry & Innovation	11. Sustainable Cities & Communities		20
7. Affordable & Clean Energy	9. Industry & Innovation		17
7. Affordable & Clean Energy	9. Industry & Innovation	13. Climate Action	16
1. No Poverty	10. Reduced Inequalities		15
7. Affordable & Clean Energy	11. Sustainable Cities & Communities	13. Climate Action	15
9. Industry & Innovation	11. Sustainable Cities & Communities	13. Climate Action	12
3. Health & Well-being	13. Climate Action		12
8. Work & Econ. Growth	11. Sustainable Cities & Communities		10



International
Initiative for
Impact Evaluation

