



CENTER
FOR
GLOBAL
DEVELOPMENT

BREAKTHROUGH TO POLICY USE

Reinvigorating Impact Evaluation for Global Development



Final Report of CGD's
Working Group on New
Evidence Tools for Policy
Impact

Chairs:
Amanda Glassman and
Ruth Levine



CENTER FOR GLOBAL DEVELOPMENT

2055 L Street, NW Fifth Floor
Washington, DC 20036

1 Abbey Gardens
Great College Street London
SW1P 3SE

www.cgdev.org

Center for Global Development. 2022.
Creative Commons Attribution-NonCommercial 4.0



CENTER
FOR
GLOBAL
DEVELOPMENT



BREAKTHROUGH TO POLICY USE

Reinvigorating Impact Evaluation for Global Development

Final Report of
CGD's Working Group
on New Evidence Tools
for Policy Impact

Authored by Julia Kaufman, Amanda Glassman, Ruth Levine, and Janeen Madan Keller

Explore more content in the digital report:
www.CGDev.org/evidence-to-impact

Contents

- Working Group on New Evidence Tools for Policy Impact v**
- Acknowledgements vi**
- Note vii**
- Foreword viii**
- 1 Impact Evaluation and the Broader Evidence Agenda Since the 2000s:
Significant Progress, Persistent Challenges 1**
 - 1.1. Motivation 1
 - 1.2. Two Decades of Progress 2
 - 1.3. Examining Outstanding Challenges 9
 - 1.4. Why This CGD Working Group? A Note on Background, Scope, and Mandate 13
- 2 The Way Forward: A Renewed Agenda for Investments in Impact Evaluation
and the Broader Evidence Ecosystem 15**
 - Recommendation 1. Design Evaluations that Start from the Policy Question and Decision Space 17
 - Recommendation 2. Harness Technology for Timely, Lower-Cost Evidence 19
 - Recommendation 3. Advance Locally Grounded Evidence-to-Policy Partnerships 21
 - Recommendation 4. Strengthen Evidence Use through New Incentives and Structures 23
 - Recommendation 5. Invest in Evidence Leaders and Communities 25
- 3 From Ideas to Action: Recommendations for a Selection of Development Funders 26**
- Appendix A. A Look Back at Two Decades of Progress in the Impact Evaluation Landscape 27**
- References and Resources 29**
- Boxes**
 - Box 1. Strengths, weaknesses, and examples of data enabling faster impact evaluation 5
 - Box 2. Cost analysis is overlooked 7
 - Box 3. Impact evaluations are still relatively rare 12
 - Box 4. Understanding the return on investment 15
 - Box 5. Examples of evidence use and its impact on policy and lives at scale 16
 - Box 6. Indicative checklist for funders considering impact evaluation proposals with the aim of
decision responsiveness 18
 - Box 7. Examples of methodological innovations for rapid impact evaluation 20

Figures

Figure 1. Number of published impact evaluations from 1990 to 2020	2
Figure 2. Number of impact evaluation authors in each country and names of select institutions with impact evaluation expertise	3
Figure 3. Maturity of methods in the responsive evidence toolkit	4
Figure 4. Distribution of impact evaluations and systematic reviews by sector	7
Figure 5. Studies with equity-specific analyses in 3ie portal	10
Figure 6. Reported main funders of impact evaluations in 3ie portal	11
Figure 7. Analytical framework on the elements of evidence use in development decision making	24

Table

Table 1. The role of cost in evidence generation	11
--	----

Working Group on New Evidence Tools for Policy Impact

Working Group Co-Chairs

Amanda Glassman, Center for Global Development

Ruth Levine, IDinsight

Working Group Members

Tania Alfonso, US Agency for International Development

Norma Altshuler, Open Philanthropy and formerly William and Flora Hewlett Foundation

Jeannie Annan, International Rescue Committee

Essaïd Azzouzi, Millennium Challenge Account-Morocco

Kelly Bidwell, Office of Evaluation Sciences, US General Services Administration

Cynthia Bosumtwi-Sam, Innovations for Poverty Action

Baboucarr Bouy, Effective Intervention

Annie Chumpitaz, formerly Ministry of Education, Peru

Cláudia Costin, Center for Excellence and Innovation in Education Policies, Getulio Vargas Foundation

Iqbal Dhaliwal, Abdul Latif Jameel Poverty Action Lab

Casey Dunning, Millennium Challenge Corporation

Peter Evans, U4 Anti-Corruption Resource Centre

Marie Gaarder, International Initiative for Impact Evaluation (3ie)

Seth Garz, Bill & Melinda Gates Foundation

Ashu Handa, University of North Carolina-Chapel Hill and the Transfer Project

Daniel Handel, 3ie

Gonzalo Hernández Licona, Multidimensional Poverty Peer Network

Michael Hiscox, Harvard University

Prudence Kaoma, Ministry of Finance and National Planning, Zambia

Andrew Karlyn, formerly Living Goods

Janeen Madan Keller, Center for Global Development

Megan Kennedy-Chouane, Development Cooperation Directorate, Organization for Economic Co-operation and Development

Catherine Kyobutungi, African Population and Health Research Center

Arianna Legovini, Development Impact Evaluation, World Bank

Ida Lindkvist, Norwegian Agency for Development Cooperation

Timothy Lubanga, Office of the Prime Minister, Uganda

Laurenz Mahlanza-Langer, Africa Centre for Evidence, University of Johannesburg

Santhosh Mathew, Bill & Melinda Gates Foundation

Mushfiq Mobarak, Yale University

Nompumelelo Mohohlwane, South African Department of Basic Education

Gulzar Natarajan, Government of Andhra Pradesh, India

Paul Niehaus, University of California, San Diego

Amos Njuguna, Network of Impact Evaluation Researchers in Africa

Radha Rajkotia, Innovations for Poverty Action

Ferdinando Regalia, Inter-American Development Bank

Mauricio Santamaría, Asociación Nacional de Instituciones Financieras—Centro de Estudios Económicos

Neil Buddy Shah, Clinton Health Access Initiative

Russell Siegelman, Stanford Business School

Working Group Staff

Julia Kaufman, Center for Global Development

For full member profiles, visit www.CGDev.org/evidence-to-impact.

Acknowledgements

The authors thank all working group members for their engagement and thoughtful comments, critiques, and suggestions throughout the working group process, which lasted from December 2020 to June 2022. The authors are especially grateful to Norma Altshuler, Chris Maloney, and Sarah Rose for their feedback. The report draws from working group discussions and background research conducted by CGD experts and commissioned from external partners. The project benefited immensely from members' collective knowledge of policy

processes and incentives, evaluation studies, and evidence resources and was further aided by consultations with nearly 100 policymakers and evidence experts from development institutions and the development research and evaluation communities. This work was also informed by a private roundtable discussion with philanthropic funders in October 2021. Many thanks are due to Emily Schabacker and Sarah Allen for supporting report production and Stephanie Donohoe, Jeremy Gaines, and Steven Love for designing digital media.

Note

This report was written by Julia Kaufman, Amanda Glassman, Ruth Levine, and Janeen Madan Keller and is a product of the Center for Global Development. Amanda Glassman and Ruth Levine served as co-chairs of the working group. The report's content is based on the deliberations and input from working group members, background research, and consultations. The working group was composed of experts from low- and middle-income country governments, bilateral aid agencies, multilateral organizations, NGOs, and academia. All working group members participated in their individual capacities and have had the opportunity to review and provide input to this report. However, members do not necessarily endorse all components of this report, nor do its contents constitute

a policy commitment by any party. All errors and omissions are those of the authors.

This working group was funded by the William & Flora Hewlett Foundation and individual CGD funders. CGD is an independent and nonpartisan research institution. There are no conditions or limitations on CGD's independence in research, findings, conclusions, or resulting publications. Where appropriate, CGD may welcome and consider comments or views from funders, but CGD retains total discretion and final decision-making authority regarding program and project research topics, speakers, participants in activities, and report content.

Suggested Citation:

Kaufman, Julia, Amanda Glassman, Ruth Levine, and Janeen Madan Keller. *Breakthrough to Policy Use: Reinvigorating Impact Evaluation for Global Development*. Washington, DC: Center for Global Development, 2022.

BY THE PRESIDENT OF THE CENTER FOR GLOBAL DEVELOPMENT

Foreword

In 2006, when a CGD working group, led by Ruth Levine, Bill Savedoff, and Nancy Birdsall, published its report *When Will We Ever Learn? Improving Lives Through Impact Evaluation*, very few social programs benefitted from studies that could determine whether or not they actually make a difference. Since then, the world has seen tremendous progress in harnessing better evidence to inform public policy decision making, especially from impact evaluations of programs in low- and middle-income countries. But COVID-19 has put a spotlight on an unfinished agenda, underscoring the need for high-quality, timely, and context-specific evidence—for both effectiveness and political credibility of the response. The pandemic has demonstrated the cost in lives and livelihoods lost when policymakers make decisions based on incomplete or outdated evidence and data.

Given the potential real-world benefits, why have decision makers within governments, aid agencies, multilateral organizations, and NGOs not yet fully harnessed the value of evidence—including from impact evaluations—for better public policies? Looking ahead, how can the development community renew momentum and broaden bases of support for impact evaluation and the wider evidence agenda? These questions were the focus of a CGD Working Group on New Evidence Tools for Policy Impact, which set out to understand why these social benefits continue to go unrealized and to chart out a renewed funding agenda for greater value in government policymaking.

The working group brought together a diverse group of policymakers and experts to review recent progress and identify

outstanding challenges in the use and utility of evidence for global development, with a focus on impact evaluation. The working group's final report highlights how far the field has come in addressing persistent critiques about the scale, generalizability, and policy utility of impact evaluation methods, reflecting a nuanced and more productive global conversation. It also offers recommendations on “what and how to fund to deliver on the promise of impact evaluation and bolster the broader evidence ecosystem” as two intertwined goals. The report urges governments and development partners to do more to integrate evidence and learning into routine operations and programming, while also emphasizing how researchers can elevate implementation, delivery, and cost analyses alongside impact evaluations for greater policy relevance. By spotlighting dozens of resources and examples of good practice and policy impact, the report helps ensure we are benefiting from—and not rehashing—well-developed contributions. In doing so, this report serves as a key resource for funders, practitioners, and students.

Across the recommendations, a core theme is the importance of shifting agenda-setting power and resources to those who best understand local policy contexts and priorities. One related recommendation is to advance equitable and lasting partnerships to support policy-immersed, but still independent, research and evidence groups. To this end, the report proposes a consortium to channel aid and philanthropic funding toward long-term, flexible institutional support for evidence organizations in lower-income countries, enabling them to move beyond short-term consultancies and

time-limited projects towards sustained engagement with policymakers.

The working group's final report is also timely within CGD as we celebrate our 20th anniversary. Since our founding in 2001, CGD's research and analysis have centered on the need for rigorous evidence on development impact; what works to improve lives and health in low- and middle-income countries; and how to enhance the effectiveness of public and aid spending. CGD was built on the conviction that better and more open data, analytics, and evidence applied to policy can make a large difference for development outcomes. This focus has driven lasting changes in development policy and spurred major new initiatives, including the creation of 3ie—a leading provider and broker of impact evaluation evidence—following CGD's Evaluation Gap Working Group.

The report offers a compelling case for greater and better investments in evidence-informed policymaking and in impact evaluation as an essential element of that effort. I look forward to seeing the partnerships, investments, and real-world impact that result from this CGD working group. I hope readers share my sense of reinvigorated commitment for the evidence agenda and translate that support into improved decision making for improved lives.



Masood Ahmed
President
Center for Global Development

Impact Evaluation and the Broader Evidence Agenda Since the 2000s: Significant Progress, Persistent Challenges

1.1. MOTIVATION

In the quest to improve social and economic well-being with limited resources, timely and high-quality evidence on performance and outcomes is indispensable. Policymakers, program managers, and funders need different types of evidence from multiple sources to inform the design, implementation, and evaluation of a policy, program, or intervention. One approach known as impact evaluation establishes the attributable net impact, making it uniquely well suited to answer the key question of what a policy or program is achieving.

The goal of impact evaluation is to rigorously identify the causal relationship between a program, policy, or intervention and its desired outcomes. Other kinds of evidence, including qualitative studies, monitoring data, and cross-sectional surveys, provide information on performance, coverage, and implementation considerations and, in turn, shed light on the causal pathways through which a policy or program affects outcomes. This information helps generate hypotheses and inform decisions about adjustments, improvements, and future implementation strategies. As a complement, impact evaluation detects whether a specific policy or program leads to an observable change in outcomes and if it works better than an alternative approach or *counterfactual*. Impact evaluation methods can also generate crucial elements of cost-effectiveness analyses.

Despite all its potential uses for public policy, impact evaluation has also elicited concerns. Impact evaluation is often seen as too costly and time-consuming, bringing excessive attention to questions about specific interventions rather

than underlying structural reforms to drive development outcomes; questions also have been raised about external validity and generalizability of findings (Bédécarrats et al. 2020; Cohen and Easterly 2010; Deaton 2020; Ogden 2017; Teele 2014; Vivaldi 2020).

Over the past decade, a good deal of ink has been spilled on the topic of whether impact evaluation using experimental methods is a savior or a scourge in global development. After the 2019 Nobel Prize in Economics was awarded to Michael Kremer, Esther Duflo, and Abhijit Banerjee for their pioneering work in development economics, a raft of critiques and responses were written—many of which are included in the World Development special issue in 2020 on experimental approaches in development and poverty alleviation (Rodgers et al. 2020). Some of these discussions are among scholars and reflect academic debates about different methodologies. But others are centered on practical policy and resource implications: Do impact evaluations, which consume scarce capacities, time, and money, yield sufficient benefits that respond to the genuine needs of policymakers and result in reforms to policy and practice? If not, are there ways to increase the relevance and timeliness of impact evaluations, while decreasing the costs?

Against this background, the onset of the COVID-19 crisis in 2020 has underscored the need for high-quality, timely, and context-specific evidence—for both the effectiveness and the political credibility of the response. Beyond COVID-19, widespread calls to reckon with—and disrupt—the ways in which development research and institutions perpetuate inequities

between and within countries and their development partners (Abimbólá et al. 2021; Erundu et al. 2021) have spurred reflection about the locus of power within research and evaluation communities.

These dynamics provide an opportunity to take stock of impact evaluation as a policy tool in and of itself, and to examine recent evolution within the field. This is also an opportunity to renew and broaden the bases of support for the evidence agenda. Doing so allows us to address concerns and identify the remaining challenges that limit the uptake of evidence by policymakers for real-world impact. By charting a renewed agenda for more useful, responsive, and relevant impact evaluation that elevates the perspectives of government policymakers and other evidence users around the world, we can reinvigorate policy commitment to, and funding for, rigorous evaluation. With more and better funding, we can harness the full value of impact evaluation to improve lives through improved policy decision making.

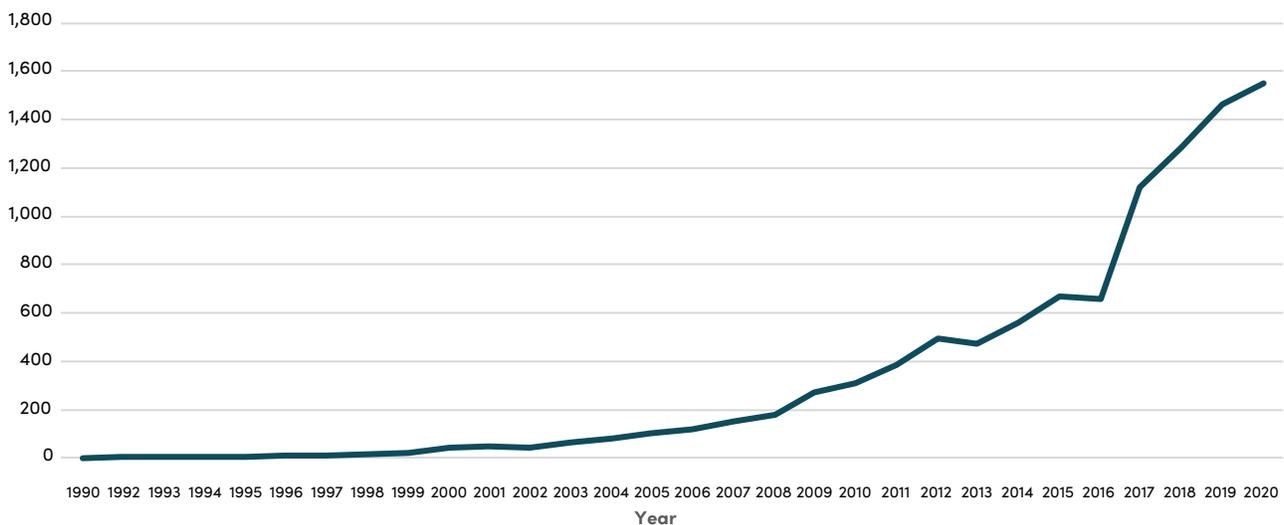
1.2. TWO DECADES OF PROGRESS

Since the 2006 release of the Center for Global Development report *When Will We Ever Learn: Improving Lives Through Impact Evaluation*, and building on evaluations of cash transfer programs in the 2000s, there have been nearly two decades of progress in generating and using evidence for public policy decisions and development programs. See Appendix A for a detailed timeline that looks back at two decades of progress and action in the impact evaluation and evidence-based policymaking landscape.

Spotlight on Key Areas of Progress

1. The amount of available funding and number of published impact evaluation studies has significantly increased. As of April 2022, the International Initiative for Impact Evaluation (3ie) evidence portal includes over 10,000 impact evaluation records. This growth and waves of related progress in evidence use tools and practices has been characterized as the “evidence revolution” (White 2019).

FIGURE 1 Number of published impact evaluations from 1990 to 2020

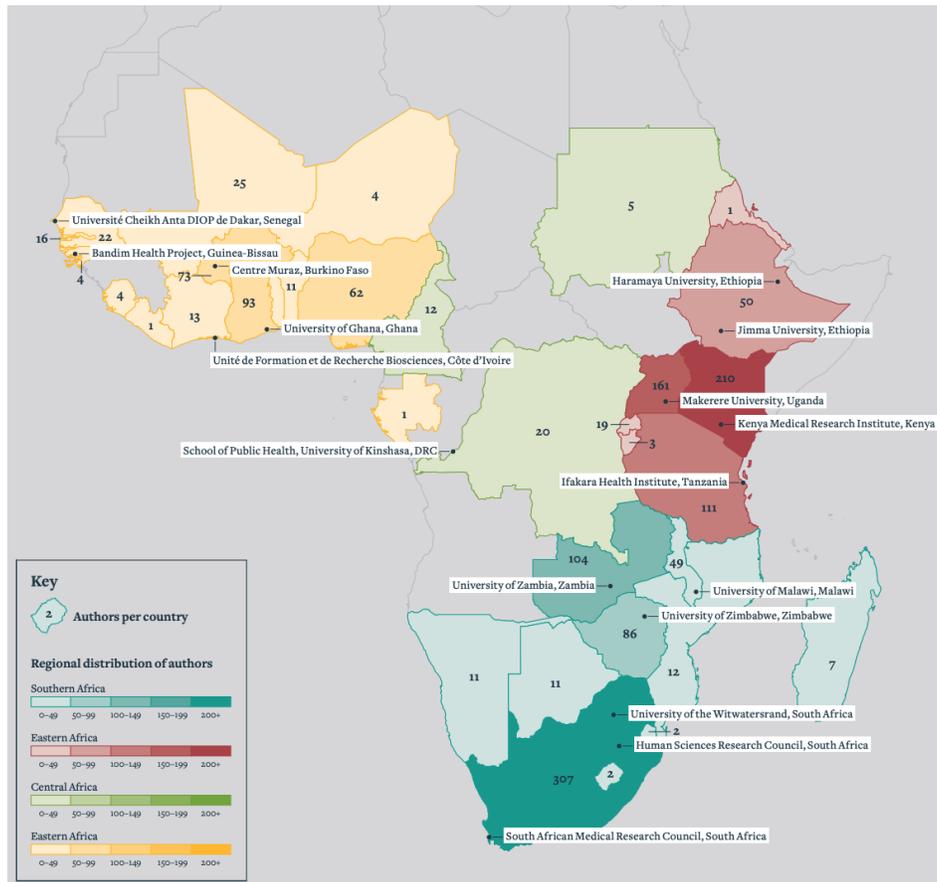


Source: 3ie Development Evidence Portal (data as of March 2022).

2. The global community of researchers and organizations conducting impact evaluations continues to grow, including those based in low- and middle-income countries. Evidence-to-policy partnerships that link researchers familiar with the local context to policy opportunities within that specific context are increasingly seen as a key mechanism for strengthening the *demand for* and *use of* policy-relevant evidence (Campbell et al. 2009; DuMont 2019; Oliver et al.

2014). Evidence-to-policy partnerships offer different benefits based on their specific aims and structure, including evaluation units within or adjacent to governments, temporary secondment or pairing schemes, nongovernmental knowledge brokers and translators, and networks (Taddese 2021). Many partnerships also involve building institutional research capacity (Ezeh and Lu 2019).

FIGURE 2 Number of impact evaluation authors in each country and names of select institutions with impact evaluation expertise

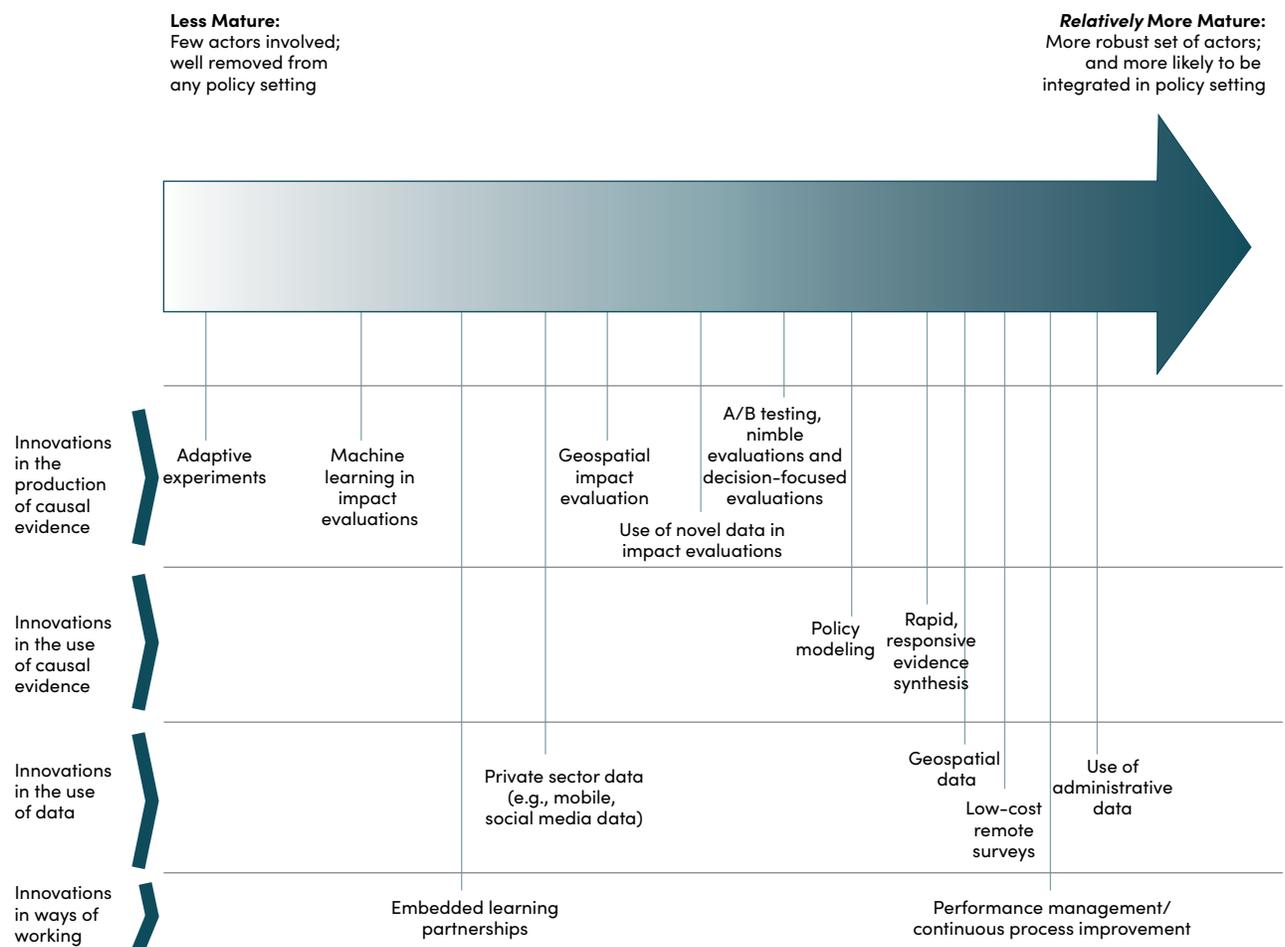


Source: Altshuler and Staats 2019.

3. Notable advances in data and evaluation methodologies and practices enable faster, lower-cost, and/or larger scale evaluations. A range of methodological approaches such as A/B testing with multiple treatment arms, adaptive evaluation, and “surrogate” proxies are enhancing the usability and relevance of experimental evidence for policy decisions. In tandem, technological advances in Wi-Fi, cell phones, GPS, and satellite imagery make it more feasible to gather and share data, while new types of software make data easier to combine, analyze, and use. The RECOVER Research Hub collates findings from 10 organizations on the impacts of COVID-19 in low- and middle-income countries, reflecting progress in speed, rigor, and accessibility.

Relatedly, evaluations are increasingly conducted at large enough scale to credibly inform policy. Evidence at scale helps address policy-relevant questions about attributable impact in the context of real-world implementation challenges (Muralidharan and Niehaus 2017). Scale also allows for measurement of outcomes such as wages, prices, or outputs—first-order considerations for policymakers. Data digitization and scale go hand in hand, enabling “closed-loop” experimentation in which programs can be delivered and evaluated iteratively (Cole, Parienté, et al. 2020).

FIGURE 3 Maturity of methods in the responsive evidence toolkit



Source: Eddy 2021.

BOX 1. STRENGTHS, WEAKNESSES, AND EXAMPLES OF DATA ENABLING FASTER IMPACT EVALUATION

1. *Geocoded survey data* collected through programs such as USAID’s Demographic and Health Surveys (DHS) Program, Afrobarometer, and national household surveys in many countries has become more readily available. Impact evaluations using geocoded data from comprehensive surveys allow for flexibility in the level of analysis, ranging from the impact of a single intervention, to a specific donor’s projects, to all sector-specific projects across several countries. This flexibility makes these evaluations useful for governments interested in a broader understanding of the development effectiveness of certain implementers or within specific sectors. But since the surveys are not customized for each study, using geocoded data limits the scope of evaluations to the questions covered in the survey instrument at hand. Missing geographic areas and large time gaps between survey rounds can further compound data biases.

2. *Administrative data* refers to information originally collected and stored for operational purposes and record keeping rather than for evaluation (Feeney et al. 2019). Examples include medical records, insurance claims, labor statistics, tax records, sales records, land registry information, and so on. Using administrative data is less expensive than collecting new data; relieves survey fatigue for participants; ensures relatively comprehensive inclusion of study participants; mitigates risk of enumerator bias; allows for observation over long time horizons due to routine collection; and, if quality controls are significantly improved, can boost reliability through biometric and geocoded tagging (as opposed to self-reporting).

But administrative data is not without downsides. The data can be biased if the program being evaluated affects the likelihood of appearing in the data (e.g., crime reports or hospital records); often requires time-intensive (and potentially expensive) cleaning and digitalization; and introduces ethical considerations related to privacy. More broadly, administrative data is often not available for comparison groups, meaning that it may not allow

for assessing the comparative effectiveness of a program. Given that administrative data collection is not designed for specific studies, researchers commonly complement administrative data with survey, remote sensing, and/or other types of data collection to create datasets relevant to their specific research questions (as done by Banerjee, Duflo, et al. 2020 on a financial management reform in India). For a deeper dive, J-PAL’s *Handbook on Using Administrative Data for Research and Evidence-based Policy* (Cole, Dhaliwal, et al. 2020) is a comprehensive resource.

3. *Remotely sensed data* is provided by a growing number of initiatives, such as WorldPop, IPUMS TERRA, the US Geological Survey, and NASA’s Worldview. These offer public access to satellite imagery and other geographical and environmental data. Remotely sensed data can help overcome analytical challenges, including assessing pre-program trends, controlling for covariates, identifying heterogeneous effects, conducting robustness analyses, and assessing external validity for the country or regional context (Rathinam, Goldblatt, et al. 2020).

For example, Burke and Lobell (2017) analyzed smallholder maize yield variation in Kenya using satellite imagery. And Yeh et al. (2020) utilize both satellite imagery and machine learning (ML) techniques to predict local wealth differences across 23 African countries. 3ie’s Big Data Systematic Map also includes 39 impact evaluations that use satellite data (Rathinam, Khatua, et al. 2020). Capacity-strengthening for geospatial impact evaluation could be immensely valuable (especially on the use of nighttime light to assess local economic performance) since the large amount of available satellite data and tools currently outweighs researchers’ capacity to conduct geospatial impact evaluation (Henderson et al. 2012). Capacity is also needed to validate and interpret remotely sensed data, as conducted by IDinsight and Radiant Earth for agriculture data on crop types and field boundaries (IDinsight 2022).

Geospatial Analysis for Development (Geo4Dev), led by CEGA, New Light Technologies, and 3ie, provides public resources and capacity support for use of open-source geospatial tools.

4. *Low-cost remote surveys* have been used expansively during COVID-19. Remote tools include Computer-Assisted Telephone Interviewing (CATI), SMS or text message surveys, Interactive Voice Response (IVR) surveys, and online surveys, each with its own strengths and weaknesses (60 Decibels 2020). Remotely collected survey data risks excluding those without access to phones, computers, and/or internet; requires high literacy rates; and can receive low response rates.

Nonetheless, remote options offer researchers significant flexibility to design their own survey instruments at relatively low cost, and remote surveys can be rapidly conducted by using existing sampling frames. For example, the Cox's Bazar Panel Survey tracks a representative sample of displaced Rohingya households and their host communities and is designed to be a "sandbox" testing environment that streamlines data collection for numerous evaluations. CBPS was leveraged to track the impacts of COVID-19 on labor markets and wages as lockdowns began (World Bank 2020).

5. *Big data* can be drawn from the internet, social media, phone usage, satellite imagery, and so on

(see 3ie's Big Data Systemic Map). As technology generates huge volumes of real-time data on economic activity, infrastructure, migration, human behaviors and preferences, health conditions, and environmental characteristics (Bamberger and York 2020), ML techniques can be used to detect patterns and make inferences (McKenzie 2018; Paul et al. 2018; Rathinam, Thissen, and Gaarder 2021). For example, Leo et al. (2020) use ML techniques to assess vulnerability to climate change in Mali and Malawi. Studying an emergency COVID cash transfer program in Togo, Aiken et al. (2022) show that data from mobile networks can improve the targeting of humanitarian aid by training ML algorithms to recognize patterns of poverty in mobile phone data. Related work in Afghanistan to assess how well ML techniques differentiate ultra-poor households eligible for program benefits from ineligible poor households found ML methods using mobile phone data to be just as accurate as standard surveys in identifying ultra-poor households (Aiken et al. 2020). The study also found that households with phones were wealthier than those without, surfacing a key limitation of ML techniques: those with less access to digital tools will be less represented in the data. ML approaches also introduce concerns related to privacy, transparency, interpretability, and accountability (African Development Bank 2020).

Source: Isaksson 2021.

4. Impact evaluations increasingly involve, or are accompanied by, complementary quantitative and qualitative methods that enhance the ability to derive policy-relevant inferences and implications, such as observational and monitoring data and participant interviews that shed light on underlying reasons why an intervention is (in)effective and other relevant contextual and observational insights.

Cost analysis is a key example of complementary evidence that is crucial to informing policy and deciding how to allocate scarce resources. Despite its importance, cost analysis remains rare. Research suggests that as of 2019 only one in

five impact evaluations integrate cost evidence (Brown and Tanner 2019; see Box 2 for more examples). But efforts like the Costing Community of Practice, consisting of CEGA, 3ie, International Rescue Committee (IRC), J-PAL, IPA, Evidence Action, the World Bank, and others, are underway to address barriers to generating more and better cost evidence produced in conjunction with impact evaluations. Further, since 2016, the IRC has undertaken cost-efficiency and cost-effectiveness analysis of different interventions to make comparisons across programs and identify the best humanitarian approaches for a given cost (IRC 2016).

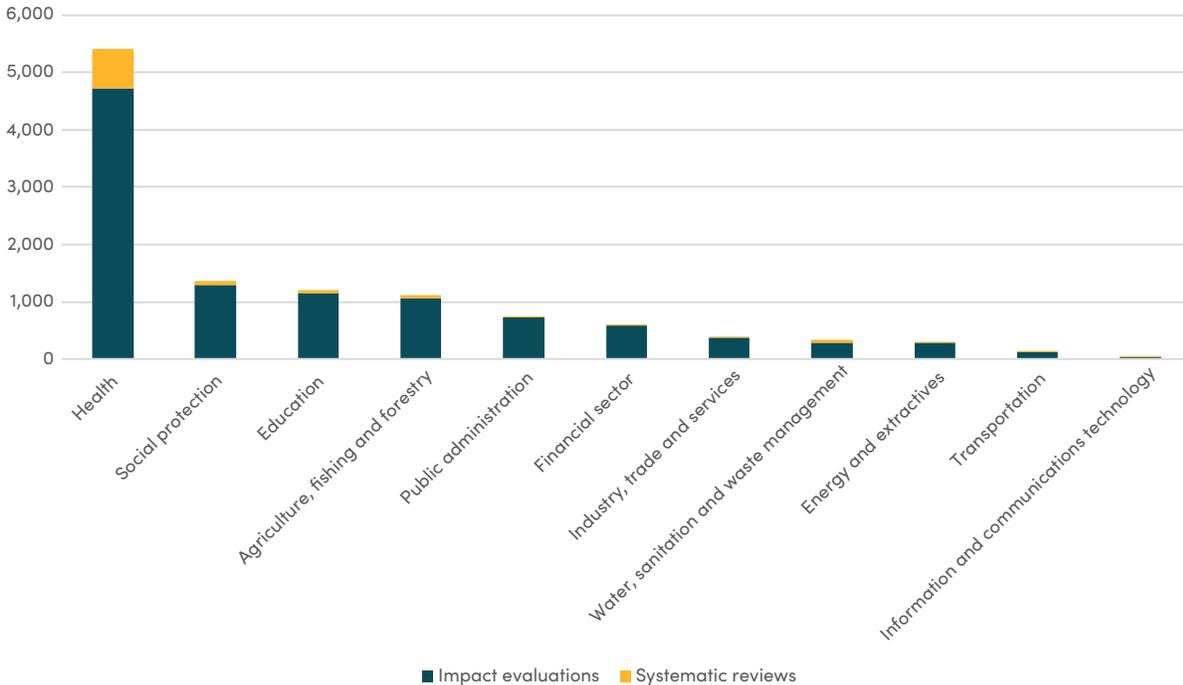
BOX 2. COST ANALYSIS IS OVERLOOKED

- Approximately 19 percent of the World Bank’s impact evaluations include any kind of value-for-money analysis (Brown and Tanner 2019).
- Across all impact evaluations in 3ie’s database, the percentage that include value-for-money analysis has not changed much over time, staying around 15 percent (Brown and Tanner 2019).
- Out of 145 recent empirical studies in education, about one-quarter include detailed costing analysis (Evans and Acosta 2021).
- Out of 50 evaluations of at-scale health programs that used rigorous methods to attribute impact (out of a total of 250 evaluations), only three estimated cost-effectiveness (Glassman and Temin 2016).

5. The application of impact evaluation tools has expanded to new domains, such as deforestation and women’s empowerment, which have historically received minimal attention from the development research and evaluation community. However, evidence in other areas, such as what works best in

fragile settings and the relationship between environmental and poverty reduction interventions, remain understudied (Alpizar and Ferraro 2020). Impact evaluation distribution by sector is concentrated in health, social protection, and education.

FIGURE 4 Distribution of impact evaluations and systematic reviews by sector



Source: 3ie Development Evidence Portal (data as of June 2022).

Note: This figure aggregates totals in 3ie’s portal from 1990 to June 2022. It does not reflect potential changes over time within specific sectors.

Resources on Background and Progress

The Role of Impact Evaluation in Development

- *World Development* special issue (Rodgers et al. 2020) on experimental approaches to development and poverty alleviation
- *Randomized Control Trials in the Field of Development: A Critical Perspective* (Bédécarrats et al. 2020) on the role of randomization in development research
- *The Evidence Commission Report* by the Global Commission on Evidence (2022) on the once-in-a-generation opportunity to systematize evidence use for societal challenges, including by spending at least 1 percent of funding on evidence infrastructure
- *Impact Evaluation in International Development: Theory, Methods and Practice* (Glewwe and Todd 2022) provides a comprehensive overview of how to conduct impact evaluations

Complementary Evidence for Policy Relevance

- Publications, event recordings, and other resources from CEGA's Cost Transparency Initiative
- Washington State Institute for Public Policy's benefit-cost estimates portal on different public programs, based on systematic reviews, return on investment calculations, and risk estimates
- The White House (2022) fact sheet on Recommendations for Advancing Use of Equitable Data, with a focus on disaggregation to understand and improve the impact of federal policies on equity outcomes

- Glandon et al. (2022) on the ten best resources for cost-effectiveness analysis in impact evaluations
- Hirji et al. (2022) on facilitating real-time cost collection and evaluating cost-effectiveness in a multi-armed study with government partners in Ghana
- Godoy et al. (2021) on the effects of savings lockboxes on health, using qualitative and observational methods to understand underlying mechanisms
- Carneiro et al. (2021) on the impacts of a multifaceted prenatal intervention on human capital accumulation in early life and hypotheses raised by qualitative analysis
- Blattman et al. (2017) on assessing cognitive behavioral therapy in Liberia through a randomized experiment alongside qualitative data to understand the context, intervention, and mechanisms

New Domains for Impact Evaluation

- Jayachandran et al. (2017) on the environmental impact of payments for ecosystem services to reduce deforestation
- Story from India Clean Air Markets (2021) on how evaluating an emissions trading pilot program led to Gujarat launching India's first carbon market
- Systematic review by Snilstveit et al. (2019) on the effects of payment for environmental services on environmental and socioeconomic outcomes
- Kochar et al. (2020) on the impact of women's self-help groups as part of India's National Rural Livelihoods Programme

1.3. EXAMINING OUTSTANDING CHALLENGES

Decision makers within governments, bilateral aid agencies, multilateral organizations, and NGOs have not yet fully realized the value of evidence from impact evaluation in improving public policies.

Despite significant progress over the past two decades, impact evaluation has not gained widespread traction for policymaking. While some funders have embraced impact evaluation, the international development community continues to underutilize it as a tool to drive significant value in important policy decisions. Likewise, investments in impact evaluation have not yet resulted in the full potential of evidence use in consequential public policy decision making at both the global and national levels. This inability to make the most of evidence has left social and economic gains on the table. Some posit that limited evidence use is due to shortcomings of the knowledge generation process (Dissanayake, forthcoming). A growing body of research seeks to examine how policymakers use and respond to evidence (Leight 2022a, 2022b), including a deep dive by List (2022) on real-world examples of pitfalls and solutions in scaling.

The working group identified three persistent challenges related to the demand, supply, and funding of impact evaluations:

1. On the demand side, impact evaluations may lack relevance to public policy decisions; and may fail to respond to the priorities, interests, timelines, and questions of decision makers.

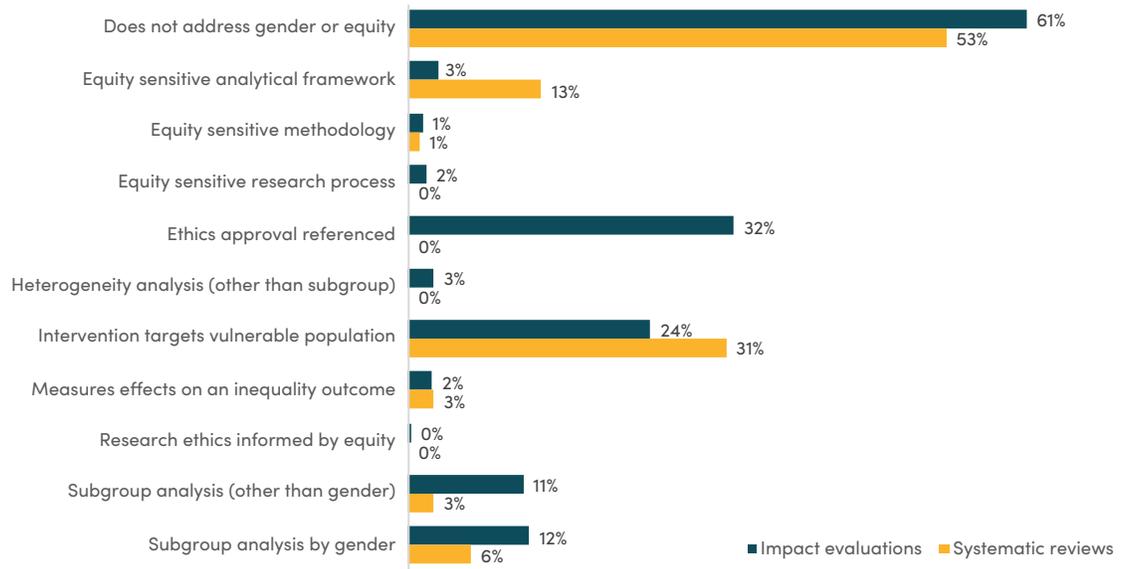
Researchers design impact evaluation studies to isolate and identify the attributable impact of a specific intervention on outcomes of interest. Unless intentional efforts are made,

many impact evaluations overlook the political economy of reform in different settings and contextual factors such as service quality and implementation capacity that influence the relationship between the program and its results (Al-Ubaydli et al. 2019). Whether scaling a pilot intervention, adjusting a widespread program, or introducing a new innovation, complementary analyses on context, cost structure, implementation feasibility, equity, and political economy matter for policy impact.

Impact evaluations often start too late or last too long to influence future policy decisions. In some cases, this results in missing windows of political opportunity. Too often, evaluations follow decisions rather than precede them. Results from past impact evaluations are often not readily available to inform real-time decisions, and impact evaluations are rarely designed and implemented to address known questions and inform expected future decisions. Though some governments choose to scale interventions based on impact evaluations findings (as shown by IPA's embedded labs, DIME's government clients and IDinsight's government partners, for example), evaluation funding and implementation could do more to be decision responsive. And while some evaluations involve ongoing engagement and sharing of preliminary results with program implementers, results are often shared with implementers much later, sometimes over a year after fieldwork ends.

Like all empirical research, policy responsive impact evaluations and related efforts carry risks related to conflicts of interest and other ethical considerations (Evans 2021). While 3ie's (2022) Transparent, Reproducible, and Ethical Evidence (TREE) policy provides tools and principles, more work is needed to translate these policies into consistent research practice beyond 3ie's own studies.

FIGURE 5 Studies with equity-specific analyses in 3ie portal



Source: 3ie Development Evidence Portal (data as of March 2022).
 Note: This figure aggregates totals in 3ie’s portal from 1990 to March 2022.

2. On the supply side, evidence users lack the required institutional incentives and funding to generate and act on relevant evidence.

Relative to other forms of evaluation and research, policy-makers and researchers may not have sufficient funding to generate and act on impact evaluations; the share of public and aid spending that is rigorously evaluated remains small (Manning et al. 2020). And in many sectors, the availability of evidence does not relate to the biggest areas of expenditure. In other words, areas that donors spend most on are not proportionately evaluated (Gaarder 2020).

One key factor in multilateral and bilateral development institutions is the **lack of institutional incentives, consistent signals, and role modeling from leadership** on the importance of learning and evidence use. Professional success is still too often measured by project approval and disbursements, as opposed to learning from, acting on, and sharing evidence. This phenomenon is reflected in the limited interest in, and

capacity for, evidence synthesis and communication to act on existing evidence, despite new synthesis tools and approaches such as Evidence in Governance and Politics’ Metaketa Initiative (which commissions numerous studies on similar questions in different contexts), MCC’s evaluation briefs, VoxDev’s wiki-inspired literature reviews, 3ie’s evidence gap maps and systematic review summaries, and J-PAL’s policy insights.

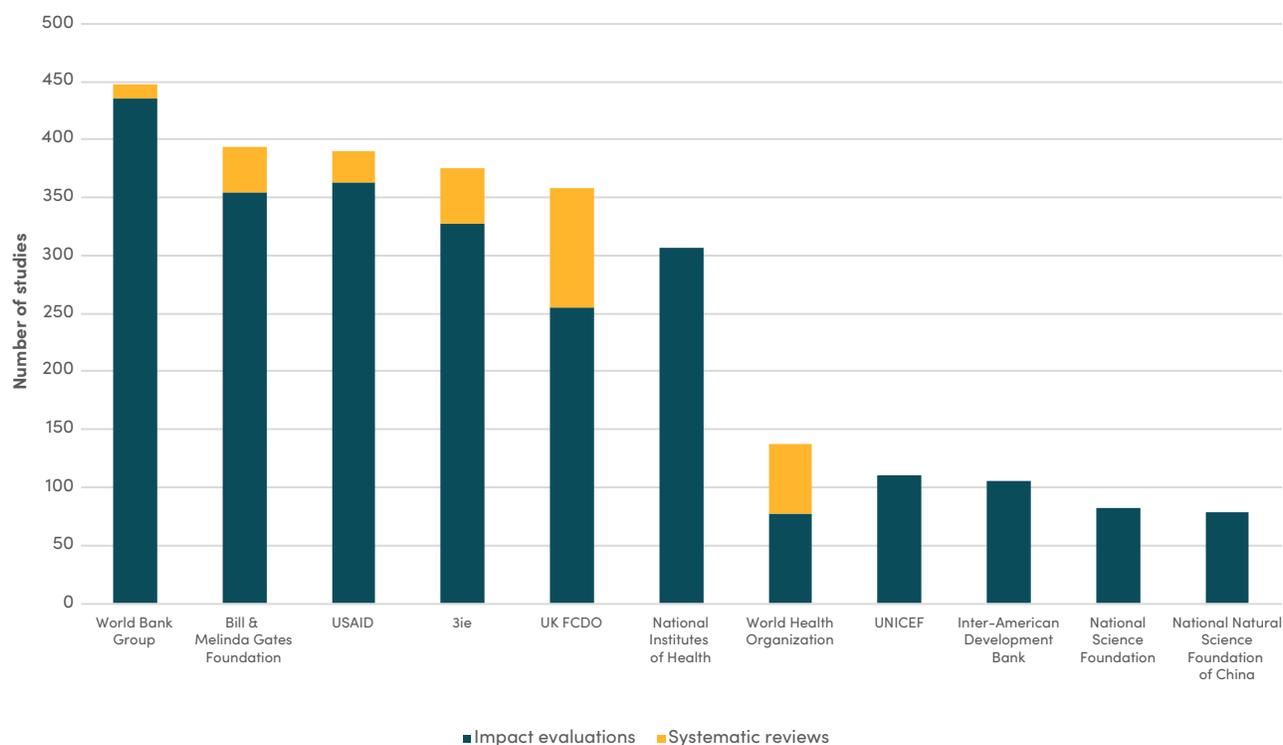
Even when evidence generation is prioritized, decision makers may overlook the methods that are most appropriate and relevant to answering specific policy questions. For instance, some performance evaluations seek to answer questions that are methodologically better suited for an impact evaluation to answer and thus may generate misleading results. Ten percent of USAID’s evaluation portfolio consists of impact evaluations (Steiger et al. 2021), yet many evaluations commissioned or conducted reflect a mismatch between the evaluation questions to answer and the methods used, highlighting the importance of using appropriate analytical methods to address policy questions of interest.

TABLE 1 The role of cost in evidence generation

HOW COST INFLUENCES EVIDENCE PRODUCTION	INTERNATIONAL AGENCIES	MEXICO	COLOMBIA	SOUTH AFRICA	UGANDA	PHILIPPINES
	Donors (official and private) fund, but funding base remains quite narrow	CONEVAL funds, sometimes with support from agencies such as IDB. IE costs a challenge	National budget allocated both to the National Planning Department and to other government agencies	DPME part-funds, departments fund rest. IE costs a challenge	Use basket fund with multiple donors and government. IE costs a challenge	Donors and some agencies (NEDA and PIDS) fund. Other agencies do not have regular funds for IEs

Source: Manning et al. 2020.

FIGURE 6 Reported main funders of impact evaluations in 3ie portal



Source: 3ie Development Evidence Portal (data as of June 2022).

Note: Because some of the included organizations receive funding from each other, data on original funders may be undercounted. Entities that only fund systematic reviews per 3ie's database, such as Australia's Department for Foreign Affairs and Trade, the European Union, and Wellcome Trust, are not included in this figure.

BOX 3. IMPACT EVALUATIONS ARE STILL RELATIVELY RARE

- Of 2,800 evaluations commissioned by CONEVAL in Mexico, only 11 are impact evaluations (Manning et al. 2020).
- In health, less than 10 percent of evaluations conducted directly by major development agencies are impact evaluations (Raifman et al. 2018).
- 10 percent of USAID’s evaluations are impact evaluations (Steiger et al. 2021).

3. Current funding models contribute to misaligned incentives between policymaker needs and academic researchers.

Academic incentives help motivate valuable knowledge production, underpinned by peer review processes, in the public domain. Yet the norms and structures that drive academic research can also limit policy relevance and use. Academic researchers typically have few professional incentives to conduct complementary analyses of costs, equity, implementation capacity, and other contextual factors, in part because peer-reviewed academic journals are generally not designed to assess or reward them. New approaches are needed—not to replace existing rigor and identification standards in academia but to complement them with research that directly responds to near- and medium-term decision-making needs and fills information gaps along the *entire* causal chain, including observational and qualitative data on implementation. Causal chains are context specific, yet ways of designing and conducting impact evaluation continue to lack substantive engagement with local policy processes and rhetoric. Efforts to build equitable, trust-based evidence-to-policy partnerships—a key enabler for policy-relevant analyses and discussions to answer questions that evolve over time—remain a work in progress, in part due to limited institutional funding (Buteau et al. 2020; Taddese 2021).

Resources on Misaligned Incentives and Power Asymmetries between Policymaker Needs and Academic Researchers

- Examples of evidence-to-policy impact from 3ie, classified into seven areas: changes in policies or programs, program closure, improving culture of evidence use, program scale-up, informing discussions on policies or programs, informing global guidelines and policy discussions, and informing the design of other programs
- Consensus statement by Morton et al. (2021) on measures to promote equitable authorship in the publication of research from international partnerships, centered on author reflexivity statements
- Article by Abímbólá et al. (2021) on the imperative to address power asymmetries in global health in the wake of COVID-19
- J-PAL’s framework of pathways to policy impact with case studies on each, reflecting the nonlinear nature of policy influence
- Reflections and funding implications by Keller and Kaufman (2021b) on Abeba Taddese’s background paper on evidence-to-policy partnerships

1.4. WHY THIS CGD WORKING GROUP? A NOTE ON BACKGROUND, SCOPE, AND MANDATE

Working group mandate: To develop a renewed agenda for the next generation of investments in impact evaluation—and evidence and evaluation systems more broadly—that will enhance their value for real-world decision making by government policymakers, multilateral organizations, bilateral aid agencies, and NGOs.

CGD is an independent, nonprofit, nonpartisan research organization that works to reduce global poverty and improve lives by generating independent research and actionable policy proposals on critical global development issues through an economics and financing lens. Since its founding in 2001, CGD's research and policy analysis have focused on the need for better, more rigorous evidence on development impact; what works to improve lives in low- and middle-income countries; and how evidence-informed policies can enhance effectiveness and value-for-money in public and aid spending.

Many major initiatives have sprung from this emphasis. Most notably, CGD's 2006 working group report *When Will We Ever Learn?: Improving Lives Through Impact Evaluation* (Savedoff et al. 2006) led to the creation of 3ie, now a trusted provider and broker of impact evaluation evidence. CGD has collated dozens of examples of how its working groups and other efforts have turned ideas for better development policy and practice into action (CGD 2021).

Amid the significant progress and remaining challenges described in the previous section, CGD drew on its collaborative working group model, emphasis on global public goods, and experience in analyzing, brokering, and encouraging the uptake of evidence to convene a working group to reenergize commitment to, and funding for, the systematic use of evidence—and impact evaluation in particular—in development

policies and programs. To this end, the project specifically aims to highlight how far impact evaluation and the evidence and evaluation field more broadly have come and to elevate the perspectives and experiences of policymakers from low- and middle-income countries.

In December 2020, CGD convened the Working Group on New Evidence Tools for Policy Impact, composed of government policymakers, multilateral organizations, bilateral aid agencies, and NGOs (see the introduction for complete membership and profiles). The group was tasked with reviewing recent progress and remaining challenges in the field and formulating recommendations for how to realize the full potential of impact evaluations and other evidence tools as essential elements of evidence-informed policymaking. The group met four times: December 2020, March 2021, May 2021, and June 2022. To inform the working group's deliberations, CGD also conducted extensive background research and commissioned original research on evidence-to-policy partnerships and rapid evaluations. The project benefited immensely from members' collective knowledge of policy processes and incentives, evaluation studies, and evidence resources and was further aided by consultations with nearly 100 policymakers and evidence experts.

While acknowledging that impact evaluation is an important part—but only a part—of a knowledge system that captures relevant and timely information for public policy decision making, the working group focused on impact evaluation as a form of knowledge-generation that is uniquely suited to estimating the net on *attributable* impact of a policy, program, or intervention social and economic outcomes of interest, which is an essential input for optimal resource allocation. The group focused on key actions to better target and structure the next phase of funding in ways that enhance the value of impact evaluations for public policymaking. The primary audience for the recommendations is evaluation funders—governments, multilateral and bilateral donors, and philanthropies.

Why Now? A Window of Opportunity

Public policy needs are evolving rapidly—as are opportunities to address today’s problems and achieve greater efficiencies in public spending. We see five factors that contribute to an unprecedented opportunity to enact lasting change:

1. As the far-reaching impacts of the COVID-19 pandemic persist, the need for high-quality, timely evidence on the effectiveness of public programs, including pandemic response efforts, deeply resonates with a wide group of stakeholders.
2. The development community has placed a renewed focus on building and advancing more equitable partnerships. This includes pressure to place greater value on understanding local contexts and national priorities. A renewed agenda for research and evaluation that centers on decision-making needs, talent, and expertise in resource-constrained settings fits well within this paradigm shift.
3. As multilateral development banks and other international entities mobilize significant budget support for governments around the world to weather and recover from the COVID-19 crisis, the need to maximize

value-for-money of public spending by national governments is in the spotlight.

4. Five years before the “deadline” for the 2030 Sustainable Development Goals, new negotiations will begin around the next set of global commitments. Bringing impact evaluation evidence to bear on this process can help pave the way for evidence-based goals and priorities.
5. As philanthropies explore how to deliver greater accountability and alignment, funding that genuinely responds to governments’ decision-making needs and strengthens the capacity of public agencies to implement evidence-based programs is a powerful signal. Supporting governments to fulfill their commitments to citizens demonstrates the best of philanthropic engagement in public policy.

Together, these factors offer a compelling opportunity to generate a reinvigorated global conversation about the role of evaluations and to rethink how to fund and conduct impact evaluation as part of evidence-based policymaking to yield lasting benefits for social and economic well-being. Collective action is needed now to change the funding landscape over the coming years.

The Way Forward: A Renewed Agenda for Investments in Impact Evaluation and the Broader Evidence Ecosystem

This section provides an overview of the working group’s understanding of good practice in the design and use of impact evaluation for better public policy and programs. It offers recommendations on “what and how” to fund to deliver on the promise of impact evaluations and bolster the broader

evidence ecosystem. Throughout, we highlight examples of good practice and innovations in the field. The working group directed its recommendations to the development community—government policymakers, funders, researchers, and NGOs.

BOX 4. UNDERSTANDING THE RETURN ON INVESTMENT

Throughout the report, we consider impact evaluation evidence, when used to inform a decision on the design and/or implementation of a public policy or program, as a development intervention in and of itself. Generating evidence entails a cost and produces a benefit in the form of increased or faster impact on outcomes or cost savings that could be deployed to other policy uses. This value-of-information approach serves as a guiding principle for the working group and the recommendations presented in this report.

An impact evaluation’s tangible value materializes when it informs a resource allocation or implementation decision, such as preventing expenditure on ineffective interventions or informing the scale-up of interventions shown to improve and save lives. The key ratio is the benefit of the evaluation’s information to the cost of the evidence.

There is a growing set of examples where the use of evidence from impact evaluations led to policy changes with tangible impacts on people’s lives. Impact evaluations have shifted global thinking and practice on reducing user fees for preventive health

products like bed nets and deworming medicine; expanded a cash transfer program to reach more poor households in Ghana; and informed the timely design of a “low-tech” remote education intervention during COVID-19 in Botswana and subsequently other countries, among many more examples (J-PAL 2018; 3ie 2020; Youth Impact); see Box 5 for additional details.

Policymakers, researchers, and funders should begin to use a value-of-information approach to proactively identify potential high-value cases of impact evaluation, including in areas such as climate. Programs that have the potential to have the greatest impact on lives and/or receive significant public resources and could easily be evaluated but have not yet been, such as COVID-19 vaccine delivery, should also be prioritized. Prioritizing evaluations of programs and interventions that are most likely to improve lives is sorely needed (Altshuler 2022). By embracing a value-of-information approach to prioritize studies and policymaker engagement efforts, funders can collectively harness greater benefits from impact evaluation.

Going forward, the cost-effectiveness of impact evaluation evidence as an intervention for better outcomes and/or savings should be modelled to quantify the benefits in terms that may resonate with a broad base of funders. The value-of-information literature by Carlson et al. (2013), Claxton and Sculpher (2012), Fenwick et al. (2020), Gratia (2014), Macauley and Laxminarayan (2010), and Myers et al. (2012) provides a comprehensive starting point. The

approach used by the Washington State Institute for Public Policy (2019) to model the potential return on investment of different programs based on existing evidence demonstrates a way to identify potentially high-value interventions, which could then be used to inform future research. IDinsight is also piloting a new value-of-information methodology to assess how much they expect social impact to improve based on additional information (Allier-Gagneur et al. 2022).

BOX 5. EXAMPLES OF EVIDENCE USE AND ITS IMPACT ON POLICY AND LIVES AT SCALE

These examples describe ways in which rigorous evaluation has been used to enhance policy and program impact, and reflect the range of nonlinear pathways to real-world applications. This list is not intended to be exhaustive but illustrative; policy impacts include changes to program design, program scale-up, program drawdown or closure, program adaption and design in other contexts, and influence on policy decisions via wider policy dialogue. Many evaluations also help improve institutional cultures of evidence use by demonstrating benefits, increasing policymaker interest and strengthening capacity (see recommendation #4), but the focus in examples highlighted here is on more direct pathways to real-world impact. While examples are numerous and tracking of use is increasingly recognized as a main component of evidence systems, the social, health and economic benefits of information generated from impact evaluation (in relation to the cost of the study) are only scratching the surface of what is possible.

- Allocating more resources to beneficiaries of a youth livelihood program in Uganda
- Boosting learning outcomes through Teaching at the Right Level in India and Africa
- Combatting COVID-19-induced poverty through community-based cash transfer targeting in Indonesia
- Combining HIV prevention campaigns to address the needs of different populations in Senegal
- Discontinuing rainfall risk insurance to support farmers in India in more effective ways
- Enhancing farmer productivity and livelihoods in India
- Expanding and broadening eligibility of a cash transfer program in Colombia
- Expanding cash transfers across sub-Saharan Africa to reduce poverty
- Expanding the child support grant and youth employment tax incentive in South Africa
- Improving children's health through behavioral nudges in the Philippines
- Improving environmental and human well-being through protected areas and payments for environmental services in Cambodia
- Improving marginalized groups' financial practices and outcomes through new technology in Latin America
- Improving Pakistan's income support program
- Increasing take-up of public benefits programs in the United States
- Influencing major labor law reforms in Mexico
- National adoption of food fortification policy in Uganda
- Preventing COVID transmission through increased mask use in India
- Promoting HIV self-testing among sex workers in Zambia

- Reallocating investments for better early child care in Colombia
- Recommendation of self-testing for HIV by the World Health Organization
- Reducing school dropout rates in Peru
- Safeguarding education progress during COVID-19 through a “low-tech” education intervention
- Saving resources in Peru, Chile, and Mexico based on shortcomings of “One Laptop per Child” program

- Scaling down a financial literacy movie in Nigeria to allocate financial literacy program resources to more effective uses
- Shifting global thinking and practice on free bed nets to fight malaria
- Shifting the focus of a youth program in the Philippines to enhance employability

Sources: Examples collated from Fishman and Christiano 2020, IEG 2012, Manning et al. 2020, OVE 2017, Legovini et al. 2019, Davis et al. 2016, 3ie Evidence Impact Summaries, IPA’s impact case studies, J-PAL’s case studies on pathways to policy change, and IDInsight’s featured projects. Videos available at www.CGDev.org/evidence-to-impact offer additional examples.

RECOMMENDATION 1. DESIGN EVALUATIONS THAT START FROM THE POLICY QUESTION AND DECISION SPACE

Evaluations should be built around policy needs and questions that matter most for development impact, focusing on directly informing policy decisions and/or building the global knowledge base. Many evaluations accomplish both goals through nonlinear, often unpredictable, pathways of influence. But the importance of the former approach has not yet translated into widespread practice of explicitly **supporting decision makers who are interested in using more evidence and have the political space to make related decisions based on that evidence**. (If political space for rigorous evidence to directly inform a decision is not yet available, funders, researchers, and knowledge brokers can provide decision support through a range of appropriate and relevant methods.) To reap the practical benefits of cocreation and successfully identify and act on policy relevant research questions, **evaluations should be conducted by those who understand the operating environment and relevance of the evaluation topic to specific policy priorities**, and can deploy rigorous methods to enhance research credibility and influence, i.e., delivery of both contextual linkages and high-quality research.

As such, impact evaluations must more regularly **integrate a range of complementary analyses** that address decision makers’ information needs and allow them to apply evaluation findings to real-world decisions. From the onset, more research should be designed from the decision makers’ vantage point to **answer both experimental and observational questions** required for successful policy implementation, considering that different types of information will likely be produced by a range of partners and methodological approaches. Policymakers often need to know that the intervention is effective across implementation models. To meet this imperative, researchers should set out—and funders should support—a theory of change that includes the baseline conditions, underlying outputs and outcomes being targeted, and implementation and delivery channels being investigated, giving policymakers **clarity on the intervention’s impact among different settings and groups**.

Through this approach, research proposals would outline a specific method to understanding both generalizability across contexts and, for initial evaluations conducted at a relatively small scale, scalability (i.e., the shape of the benefit and cost curves with respect to intervention size within a specific context). Assessing the scope for scaling up further may or may not involve additional large-scale impact evaluations (as

those conducted by the Yale Research Initiative on Innovation and Scale), but it should at least specify what observational findings would be expected in the case of successful rollout. As part of efforts to consider the implications and inferences related to different implementation approaches and baseline groups, researchers should work with policymakers to ensure that their policy questions are informed by existing evidence drawn from systematic reviews, gap maps, and other sources. If deciding to pilot a new program, interventions found to be generally most effective through systematic reviews should be the starting point, with evaluations then designed to test and monitor those interventions in a given context.

Identifying where causal evidence is needed, where observational or qualitative information is sufficient to inform policy, and which methods will be used to understand scale and context (e.g., a randomized controlled trial, natural experiment, modeling, or observational data) requires a nuanced understanding of local policy processes and questions. For example, Piper et al. (2018) used observational data to successfully

scale a national literacy program in Kenya. And in Colombia, Barrera-Osorio et al. (2022) adjusted an educational program for parents during scale-up based on implementation data.

Investments in impact evaluation should also be paired with embedded technical assistance to support evidence use through the program life cycle and implement **evidence uptake plans**. Like a pre-analysis plan, an evidence uptake plan maps out potential evaluation results alongside related policy responses and pathways to scale. For instance, when an HIV awareness campaign implemented by Youth Impact generated mixed results in Botswana, the government policymakers and other partners involved readily reached a consensus not to scale because they had previously discussed the possibility of negative and ambiguous results (Levy et al. 2018) (Levy et al. 2018). While an evidence uptake plan is not meant to be binding, the process of developing the plan helps secure commitment from policymakers. Platforms that consolidate and communicate insights from different bodies of knowledge, such as 3ie’s synthesis products, are useful for *policy*

BOX 6. INDICATIVE CHECKLIST FOR FUNDERS CONSIDERING IMPACT EVALUATION PROPOSALS WITH THE AIM OF DECISION RESPONSIVENESS

- Does the primary demand arise from policymakers with a commitment and plan to incorporate results into decision making?
- Have the researchers engaged regularly with the relevant policymakers?
- If it is an embedded experiment, does it exclude any personal rewards for government officials who participate in it?
- If evaluating a program at scale, has it been preceded with some sort of safety trial?
- Has responsibility for compensation in the event of any harms been clarified in advance?
- Does the evaluation include an assessment of cost-effectiveness?
- Is bureaucratic feasibility or capacity a factor in the evaluation?
- Does the evaluation precede, rather than follow, decisions to feed into program design and implementation?
- Does the evaluation include parallel data collection activities to assess and improve implementation alongside headline efficacy?
- Does the evaluation include qualitative work with program participants to inform how findings are translated into policy?
- Does the research team include a principal investigator with deep contextual knowledge? (Not solely linked to geography; could be locally based researchers or members of diaspora.)
- Does the evaluation involve capacity building or some form of knowledge transfer to local research institutions and/or policymakers?
- Does the evaluation involve procurement of services from local providers?
- Does the proposal reference and document relevant prior work completed by local researchers, government agencies, and nongovernmental institutions?

translators providing embedded technical assistance—and for interested policymakers themselves. Locally connected experts and policy translators play a key role in garnering interest in evaluation findings among policymakers, which in turn facilitates evidence uptake.

Explicit ethical safeguards and policies are an important part of embedded approaches—and should be across all empirical research. Going forward, researchers should commit to—and funders should require—adherence to improved and more transparent ethical principles and practices (3ie 2022; Evans 2021).

Resources on Developing and Conducting Evaluations to Support Decision Makers (Recommendation 1)

- Muralidharan and Niehaus (2017) on the case for greater use of randomized experiments “at scale” and progress to date
- *The Goldilocks Challenge* by Gugerty and Karlan (2018) on how to create a “right-fit” evidence system that recognizes when (and when not) to measure impact
- Gertler et al. (2016)’s *Impact Evaluation in Practice Handbook* on how to design and implement impact evaluations
- J-PAL’s generalizability framework on how to combine evidence to assess policies in new contexts (Bates and Glennerster 2017)
- Stein et al. (2021) on how to link NGO innovation and testing within NGO operating environments to government scale-up
- Fischer et al. (2021) on the need for alternative analytical approaches when conducting evaluations to inform a specific decision
- Abebe et al. (2021) on a new spatial equilibrium approach to evaluate urban public works at scale
- Bold et al. (2018) on a randomized trial embedded within nationwide education reforms in Kenya
- Implementation handbook by Channon-Wells et al. (2020) to be used across settings for a job-seeker skills certificate program

RECOMMENDATION 2. HARNESS TECHNOLOGY FOR TIMELY, LOWER-COST EVIDENCE

To generate faster, lower-cost, and more policy-relevant evidence, new developments and applications of analytical methods and data sources offer promising prospects. Technological advances in Wi-Fi, cell phones, GPS, and satellite imagery have made gathering and sharing data much easier, and new types of software make this data easier to combine, analyze, and use. As discussed in section 1, five data sources in particular have ample promise for more relevant impact evaluations: geocoded survey data, administrative data, remotely sensed data, low-cost remote surveys, and big data (see Box 1 for more details).

But to **harness the digital transformation** and expand the use of these data sources, development stakeholders must **commit to repurposing existing data** and increase investments in capacity to do so. Administrative data can be used to unlock significant benefits for government functions, such as taxation and public procurement. **Increased investments in the quality, regularity, and granularity of administrative data** are high priority for routine decision making and use for impact evaluation (Glassman et al. 2014). Data quality can also be enhanced through partnerships with the broader “data for development” community, such as through developing a system of checks and balances with trust scores for administrative data.

To translate these data sources into policy action, researchers must **prioritize the role of people, or human capital**. More capacity is needed to validate large datasets through field surveys; provide technical assistance to manage, clean, and analyze big data; incorporate new data privacy and governance policies for legal use of private data; and continue to develop new methodological techniques. CGD’s Working Group on Governing Data for Development discusses how governments and multilateral organizations can strengthen data governance and protection policies for improved access and use, including a common approach to establishing the legality

BOX 7. EXAMPLES OF METHODOLOGICAL INNOVATIONS FOR RAPID IMPACT EVALUATION

1. *Evaluations with multiple treatment arms:*

Many private companies have begun integrating continuous experimentation into their operations through A/B testing and other analyses (Chen 2020). As governments examine how to embed routine evidence into their own decision making for better outcomes, the widespread use of A/B testing to assess variations in program design is especially promising. For example, Banerjee, Alsan, et al. (2020) applied A/B testing to evaluate a COVID-19 prevention campaign in the Indian state of West Bengal, with several treatment arms receiving different messages. Since evaluating multiple treatment arms often requires usable administrative data and large sample sizes to detect impacts of incremental changes, greater support for data collection and infrastructure is critical.

2. *Adaptive/iterative evaluation:* To enable real-time program adaptation, improve targeting, and inform rapid policy responses, evaluations can be set up to include multiple waves of data collection (through low-cost remote surveys, for example) and ongoing engagement with implementers. For example, an evaluation by Angrist et al. (2022) on the extent to which low-tech interventions limit pandemic-related learning loss in Botswana involved multiple rounds of data collection at four- to six-week intervals to facilitate program adaptation. Similarly, Caria et al. (2021) used “adaptive targeted experimentation” to assess the impact of labor market policies on Syrian refugees in Jordan by observing treatment outcomes over time and adaptively optimizing treatment assignment for participants. And Alvarez-Marinelli et

al. (2021) fine-tuned an intervention to improve the reading skills of third-grade students in Colombia for each subsequent cohort and found that program effectiveness increased fourfold over time.

3. *Technology to study heterogeneous treatment effects:*

New ways to understand who benefits most from a given intervention can also be used to inform targeting, a key policy design issue (Vivalt 2015). For example, a study by Islam (2015) on the heterogeneous effects of microcredit programs suggests that effects on consumption vary across different groups of poor household borrowers, though Meager (2019) finds heterogeneity in effects across seven randomized microcredit experiments to be moderate.

4. *Quasi-experimental methods:* While econometric techniques to find “control” groups that are statistically similar to “treated” populations are not new, noteworthy techniques, such as synthetic controls and ML predictions, illuminate the increasingly sophisticated tools at researchers’ disposal (Athey and Imbens 2017; Abadie 2021). These techniques are often used in combination with novel data sources, including granular spatial data and large administrative datasets, to control for potential confounding variables at more specific geographic levels and assess relevant outcomes. “Surrogate” proxies are also a way to define outcomes of interest when using quasi-experimental identification strategies (Athey et al. 2019).

Source: Isaksson 2021.

of cross-border data flows, more resources to strengthen domestic data governance regimes, and better data policy metrics (Pisa et al. 2021). The Development Data Partnership,

a recent collaboration between international organizations and technology companies, provides a useful model for how to facilitate the use of private sector data for development research.

Resources on Data and Methodologies for Timely, Lower-Cost Evidence (Recommendation 2)

- *Handbook on Using Administrative Data for Research and Evidence-based Policy* by Cole, Dhaliwal, et al. (2020)
- 3ie's Big Data Systematic Map on the use of big data from satellites, mobile sensors, cellphone records, businesses, public agencies, citizen reporting, internet searches, and social networks to evaluate development outcomes
- Landscape analysis of how governments are learning better, faster, and cheaper to improve lives by Eddy (2021)
- Reflections from Keller and Kaufman (2021a) on how the development community can increase the speed, decrease the cost, and broaden the applications of rigorous impact evaluation for better evidence that leads to better policy outcomes
- Resources from CGD's Working Group on Governing Data for Development

RECOMMENDATION 3. ADVANCE LOCALLY GROUNDED EVIDENCE-TO-POLICY PARTNERSHIPS

As highlighted throughout the report, locally informed evidence-to-policy partnerships can serve as the foundation of a sustainable evidence ecosystem. The development community should increasingly focus resources on context-specific evidence-to-policy initiatives, guided by **partnership norms that shift power to researchers in resource-constrained settings**. Partnerships involve *knowledge brokers* or *translators*: individuals or entities that facilitate exchanges between research producers and users, discuss how to address policy information needs through research questions, and help determine how to incorporate research into policy and programs. Knowledge brokers may disseminate and synthesize research findings, build networks of knowledge users and

producers, and strengthen capacity to act on evidence. Tadese (2021) identifies five main categories of evidence-to-policy partnerships, each with varying degrees of autonomy and insider knowledge: (1) government agencies, (2) semi-autonomous government units, (3) nongovernmental knowledge brokers and translators, (4) networks and communities, and (5) research-policy maker exchange programs. The map in chapter 1 includes examples of each.

A **community of practice on evidence partnerships** could help develop detailed partnership funding guidelines and envision a center of excellence for continued research and shared learning on best practices for partnerships. The community of practice and eventual center of excellence would also undertake research on how best to assess partnerships, adjust partnership approaches based on the specific policy needs at hand, and regularly apply lessons learned, including lessons on progress toward institutionalizing a culture of evidence use. **These research efforts could inform an investment case on why and how funders should support partnership building as an important intervention in and of itself as part of a commitment to putting power in the hands of those who best understand specific local contexts.**

Assessment of the use of evidence (including through partnerships) is minimal, but some systematic approaches to track policy impact over time are on the rise, as shown by 3ie's evidence use and impact measurement approach (2020) and CEGA's retrospective review (Fishman and Christiano 2020). Building on existing tools, **routine diagnostic processes of evidence use within governments**, conducted either internally or externally, could help countries, donors, and researchers measure their commitment to institutionalizing evidence and adjust which areas and capacities to focus on accordingly. As called for by the Africa Evidence Network (2021b) in their *Manifesto on Capacity Development for Evidence-Informed Decision Making in Africa*, **capacity strengthening should be centered around identifying and building on existing capabilities and local talent and knowledge**, challenging power dynamics that perpetuate transactional and wholly ineffective approaches (Green 2017). In light of unevenly distributed capacity among research institutions, the International Centre for Evaluation

and Development's ALL-IN program calls for researchers at African institutions to take the lead in defining agricultural research priorities and facilitate management of large and complex awards. Other examples of promising partnership initiatives to strengthen government capacity and interest in generating, interpreting, translating, and acting on data include Twende Mbele, a peer learning network of African monitoring and evaluation offices; Utafiti Sera, a program that develops sector-specific "houses" to review and appraise evidence; the International Network for Government Science Advice, a forum to develop approaches to use evidence in policy; the Transfer Project, a multi-country cash transfer research network that invests in long-term relationships with government officials and begins with identifying the most relevant research question and fitting the methods to the question; and the African Institute for Development Policy (AFIDEP)'s parliamentary engagement in Malawi. Capacity building through research cocreation also facilitates buy-in of the evaluation design, results, and recommendations, thereby reinforcing the culture of evidence use.

One institutional function ripe for transformation is procurement. External donors can help governments and other institutions that fund evaluations to better hire high-quality, locally-immersed technical talent by establishing prequalified groups and easy matchmaking programs, among other levers. The partnerships community of practice could systematically solicit and socialize better procurement practices designed to support governments and funders in selecting partners and funding high-quality evidence generation. Program implementers contracting evaluators to assess their own projects may introduce conflicts of interest, but these tradeoffs may be warranted depending on the goals at hand (e.g., iterative learning versus accountability). Multilateral and bilateral funders should make available model procurement documents and contracts for hiring evaluation agencies, templates of evaluation designs, and survey instruments, which could be accessed by local evaluation firms and government agencies committed to undertaking more impact evaluations. The structure could be similar to the World Bank's Public-Private Infrastructure Advisory Facility and build on DIME's existing efforts. Housing

this within a multilateral institution will help give government leaders access to resources.

Last, philanthropic funders and other interested donors should commit to mobilizing and pooling additional resources through a **demand-driven fund** dedicated to supporting governments in articulating evidence agendas that they can then take forward and implement in collaboration with policy-oriented researchers.

Resources on Research-to-Policy Partnerships (Recommendation 3)

- How to build national and regional research capacity in sub-Saharan Africa through innovative funding by Ezeh and Lu (2019)
- Overview of evidence-to-policy partnerships, including evaluation units within governments, knowledge brokers, and networks, by Taddese (2021)
- *New Attitudes, Old Practices: The Provision of Multiyear Support* by Buteau et al. (2020) on the disconnect between the importance of general operating support for nonprofits and the state of practice in philanthropy
- *A Global Procurement Partnership for Sustainable Development: An International Stocktaking of Developments in Public Procurement* by Fagan et al. (2022)
- "How to Keep Good Research from Dying a Bad Death: Strategies for Co-Creating Research with Impact" by Gyamfi and Park (2019)
- Commentary on the importance of collaborative research engagement to build trust, allow for cocreation of research questions, inform operations throughout the evaluation period, and leverage national research expertise by the Transfer Project (2020)
- Reflections from a workshop convened by Results for All, AFIDEP, and IDinsight to facilitate peer-learning between governments on evidence use in policy by Violet Murunga (2018)

RECOMMENDATION 4. STRENGTHEN EVIDENCE USE THROUGH NEW INCENTIVES AND STRUCTURES

National, multilateral, and bilateral policymakers make thousands of consequential decisions everyday about how to invest their resources. In theory, they seek to maximize improvements in living standards and outcomes as part of the moral core of public policy (Levine 2017). In reality, shifting institutional and policy priorities, bandwidth limitations, and a focus on inputs deemed (but not known) to be effective make it difficult to learn about and achieve better outcomes (Gaarder and Bartsch 2015). Organizational leaders rarely express or model sufficient commitment to learning and evidence use, and individual performance is often measured by approval and disbursement as opposed to actual results, which are typically available only after staffing changes between project cycles. At times, speed and confidence are rewarded over efforts to ask and answer questions about actual results. Further, COVID-19 has placed massive stress on officials' abilities to collect, share, and use data and has laid bare just how insufficient existing evidence cultures often can be. **More robust systems and incentives are needed to institutionalize the generation and use of rigorous evidence.**

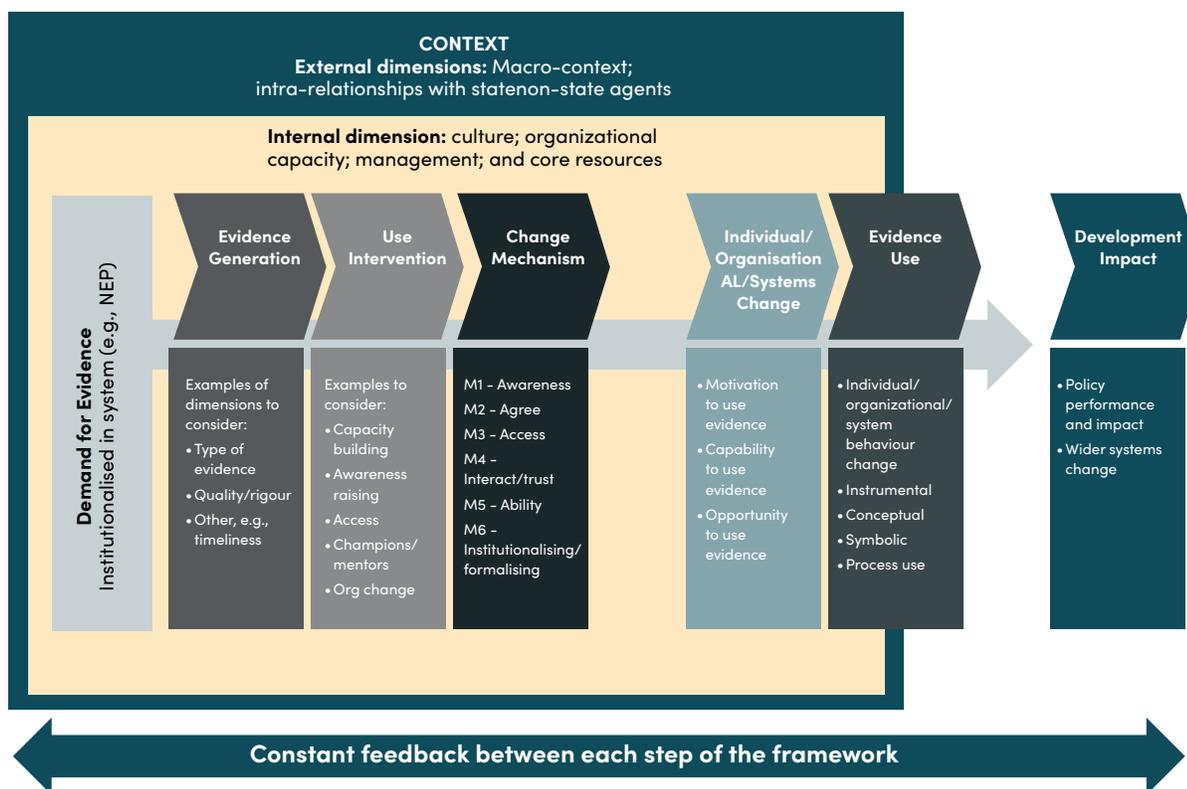
The most successful mechanisms to integrating evidence (including evidence from impact evaluation) into day-to-day operations will necessarily vary between different institutions. Yet across all bureaucracies, aligned incentives are an indispensable ingredient. The lack of institutional incentives and consistent signals from leadership impede evidence use in development institutions. Further, most government policymakers choose not to purchase or use rigorous evidence. Financial and professional rewards could generate demand, motivate capacity strengthening, and increase accountability for measurable impact. Depending on the institution, this might involve new budgetary requirements for evaluation strategies, changes to individual performance assessment criteria, redesigned results agreements, or legal mandates, as in the case of the US Foundations for Evidence-Based Policymaking Act of 2018 (Evidence Act). As policymakers

increasingly put in place environmental and social safeguard policies, the **use of evaluation could be institutionalized as a kind of economic safeguard to ensure the use of appropriate tools to determine whether projects have their intended impacts and whether they should be adjusted or scaled up or down.** The rate of return is immense: a \$1 million impact evaluation could save hundreds of millions in mistargeted or ineffective spending.

External actors can help fortify these functions through supportive partnerships and nudges for accountability. Some funders, such as the Hewlett Foundation, focus on these areas, but current funding levels are inadequate to building—and sustaining—systems and partnerships to withstand the tests of time and turnover. COVID-19 has shown that when organizations help officials use evidence to solve pressing problems, governments are more likely to demand—and in some cases purchase—more evidence. But this requires **upfront demand generation.** Many evidence-to-policy partnership organizations focus on capacity strengthening to build on existing interest and talent. Some organizations first work with governments to quickly respond to short-term evidence needs as an entry point to spurring longer-term interest in commissioning impact evaluation evidence or complementary data sources. For example, 3ie partnered with the Philippines National Economic and Development Authority and Australia's Department of Foreign Affairs and Trade on a multiyear Philippines Evidence Program, in which 3ie and the Philippine government collaboratively commission impact evaluation of major government reforms and service delivery interventions prioritized by select departments.

In addition to external partnership initiatives, in-government or semi-autonomous evidence units—such as the Department of Planning, Monitoring and Evaluation in South Africa, MinEduLAB in Peru, and the Health Intervention and Technology Assessment Program in Thailand—provide value to broader institutional evidence cultures. Government agencies, evaluation units, and “arm’s-length” research and advisory bodies play a central role in commissioning and undertaking research studies and in facilitating engagement with civil society organizations and other external partners. Many of

FIGURE 7 Analytical framework on the elements of evidence use in development decision making



Source: Recreated from Langer and Weyrauch 2020.

these units also provide capacity-strengthening support to build awareness about evaluations as a potential source of information and to help decision makers better understand how to interpret and act on data and research findings.

Resources on Institutional Incentives and Structures for Strengthened Evidence Use (Recommendation 4)

- Case studies by Results for All on incentives for government evidence use in Mexico, Sierra Leone, and South Africa through awards, publication requirements, and staff performance plans
- Manning et al. (2020) on progress to date and continued limitations to embedding impact evaluations in partner governments, looking at Mexico, Columbia, South Africa, and Uganda

- Lucas and Altshuler (2021) on “How NGOs Can Help Institutionalize Government Use of Evidence”
- Carter et al. (2018) on lessons on creating cultures of evidence use from J-PAL’s government partnerships in Latin America
- Examples of embedded government labs supported by Innovations for Poverty Action, which it plans to expand to six new partners in 2022 based on growing government interest
- Interview with Diana Warira Njeri, winner of the Africa Evidence Leadership Award celebrating evidence champions (Africa Evidence Network 2021a)
- Information on the Global Evaluation Initiative, which launched in 2020 with the aim of building country-owned monitoring and evaluation frameworks and capacities

RECOMMENDATION 5. INVEST IN EVIDENCE LEADERS AND COMMUNITIES

A new generation of evidence leaders has emerged further upstream. Young researchers are increasingly interested in applying research to policy, and early-career government officials are increasingly interested in grounding policies in evidence. **Through online teaching resources, civil service institutes, and government training programs, development funders can help build lasting skills and shape meaningful evidence-to-policy communities.** Examples for policymakers include the Network of Impact Evaluation Researchers in Africa (NIERA)'s policymaker and media courses, the Africa Centre for Evidence's new learning program on "Practices of Evidence-Informed Decision-Making," the South Africa National School of Government's course on evaluation management, and causal methods training for policymakers in Pakistan (Mehmood et al. 2021). Examples for scholars include J-PAL's MicroMasters, Rwanda's Impact Evaluation Summer School in partnership with DIME, and coaching on how to prepare grant applications from the BRAC Institute of Governance and Development's Women's Economic Empowerment and Digital Finance initiative. It is particularly valuable to train new cadres of people to use innovative methodologies and data that have until now been relatively underutilized in impact evaluations, including geospatial and remotely sensed data through Geo4Dev and other platforms.

Equipped with additional skills, those who are first entering (or pivoting within) the workforce would benefit from **coordinated linkages to career opportunities**, either through fellowships, secondments, matchmaking, or other formal mechanisms that pair researchers and policymakers. For example, Development Impact West Africa is a partnership between CEGA and the Ghana Institute of Management and Public Administration that uses a matchmaking process to bring US and Ghanaian academics and Ghanaian policymakers together for research partnerships (after first determining

that an impact evaluation is feasible and there is scope to inform decision making). J-PAL's African Scholars Program offers a dedicated pool of resources targeted toward African researchers. Many organizations now require that all research projects include a local principal investigator, in line with government regulations in some places. In the United States, the Intergovernmental Personnel Act facilitates the temporary secondment of academic partners to government agencies to develop rigorous studies using administrative data and other sources and then apply insights to government programs (OES 2022). Importantly, scholars from lower-income countries should have opportunities to spend time at US and European institutions in conjunction with US and European scholars traveling to lower-income settings to learn and research.

Resources on Creating and Sustaining Evidence Communities (Recommendation 5)

- "Mixtape Sessions" by Cunningham (2022) to democratize causal inference methods through online and in-person workshops
- Syllabus and slides from "The Practicalities of Running Randomized Trials" taught by Rachel Glennerster (2021)
- Information on the Collaboration for Inclusive Development Research, a new initiative jointly led by CEGA and NIERA to examine the state of inclusion of African scholars in the generation of evidence for policymaking on the continent and identify remaining barriers and promising solutions (see Ranganath and Shipow 2022)
- Information on Plan S, which requires that scientific publications resulting from publicly funded research be published in open access journals or platforms
- Webinar from Pan-African Scientific Research Council (2021) on preparing successful grant applications

From Ideas to Action: Recommendations for a Selection of Development Funders

To illustrate the application of this agenda to specific development funders, the working group developed detailed recommendations for three key audiences with strong existing foundations for evaluation and evidence use: philanthropies, USAID, and the World Bank. Ideas draw from working group discussions, additional consultations with experts from these institutions, and background research. Detailed recommendations for each audience can be found in complementary briefs.

Beyond these three specific audiences, development funders should form an alliance or coalition of development institutions committed to evidence generation and use where it is most in need and demand. Collaboration and collective action across organizations and funders will have a more transformative effect than any single organization can offer.

Key Recommendations for Philanthropies to Invest in a New Era of Evidence-Informed Decision Making to Improve and Save Lives

1. Adjust current grantmaking to better meet high-value decision-making needs of lower-income country governments
2. Identify opportunities for shared learning and aligned funding
3. Develop new funding consortia for policy-responsive evaluation

Key Recommendations for USAID to Mainstream Evidence Use through Locally Led Development

1. Partner with philanthropies on a localization and evidence initiative to systematically support policy-proximate researchers, research organizations, and evidence collaboratives in a set of lower-income countries
2. Structure the new behavioral science and experimental economics unit for maximum impact, including high-value decision-oriented evaluation
3. Bolster agency capacity through new specialized evidence roles
4. Champion cost effectiveness analysis and cash benchmarking
5. Develop accountability policies and performance metrics to incentivize sustained change
6. Build evidence, evaluation, and cost-effectiveness into the agency's routine operational policies and guidance

Key Recommendations for the World Bank to Leverage Knowledge Production for Policy Impact

1. Embed impact evaluation and related evidence resources across the Bank's operational structure and develop sectoral, regional, and country learning agendas
2. Allocate dedicated resources to routinely finance rigorous evaluation
3. Strengthen and centralize tracking and publication systems for data and evidence
4. Develop formal mechanisms to promote evidence-to-policy partnerships, capacity strengthening, and demand generation

A Look Back at Two Decades of Progress in the Impact Evaluation Landscape

1971

RAND's Health Insurance Experiment begins

The RAND Health Insurance Experiment was a 15-year effort funded by the US Department of Health and Human Services on how cost sharing arrangements affect people's use of health services, the quality of care they receive, and their health status. It is the largest health policy study in US history and paved the way for increased cost sharing for medical care in the 1980s and 1990s.

1990–2000

The results agenda takes the stage

The “evidence revolution”—as described by Howard White (2019)—arose as part of New Public Management, which took hold in the 1990s and held government agencies to account for their performance by monitoring trends in high-level outcomes like unemployment and poverty. Examples include the 1993 Government Performance and Results Act in the US and the 1999 Modernizing Government White Paper in the UK. Prior to this period, performance was predominantly assessed by inputs, such as how much money was spent.

1997

PROGRESA kicks off more impact evaluations in development

PROGRESA, a conditional cash transfer program in Mexico, was the first large-scale social policy evaluation implemented in a lower-income country to use a randomized controlled trial (RCT) design. The impact evaluation was designed specifically to inform upcoming policy decisions about whether and how to adapt, expand, and improve PROGRESA across the country. PROGRESA was the first of a series of cash transfer and other social protection programs in Latin America, South Africa, and Asia to be rigorously evaluated, demonstrating the feasibility of evaluating and improving at-scale programs.

2002–2012

Dozens of new evaluation organizations are founded, helping create the evidence-based policymaking field

The number, scope, and funding of evaluation and research partnership organizations grows tremendously. This includes Innovations for Poverty Action (IPA), the Abdul Latif Jameel Poverty Action Lab (J-PAL), the World Bank's Development Impact Evaluation (DIME) group (see next milestone), IDinsight, and the Center for Effective Global Action (CEGA). Prior to this period, there was little academic-level rigor in the evaluation of development interventions. Instead, the focus was on output-level performance measures.

2005–2014

Growth of World Bank trust funds for impact evaluation

The World Bank establishes new multidonor trust funds dedicated to knowledge generation and use through impact evaluation, including DIME (which was created in 2005 and received renewed funding from the Impact Evaluation to Development fund in 2014), the Health Results Innovation Trust Fund in 2007, and the Strategic Impact Evaluation Fund in 2012.

2005–2015

Proliferation of government-embedded evidence entities

Impact evaluation “labs” are embedded in government agencies to test policy reforms and enhance their effectiveness. The establishment of these entities is often due to legislative interest and demand. Examples include CONEVAL in Mexico (prompted by PROGRESA); the Department of Planning, Monitoring, and Evaluation in South Africa; MineduLAB in Peru; the Government Evaluation Facility in Uganda; and the Behavioral Economics Team of Australia (BETA). In the health sector, entities like Thailand's Health Intervention and Technology Assessment Program (HITAP) were created to develop local evidence on cost-effectiveness of health products and services.

2006

Release of CGD's Evaluation Gap Working Group report: When Will We Ever Learn?

CGD publishes the report *When Will We Ever Learn?: Improving Lives Through Impact Evaluation*, highlighting that very few social programs benefit from studies that could determine whether or not they actually make a difference. The report proposed the design of the International Initiative for Impact Evaluation (3ie) and generated support for its creation.

2008

International Initiative for Impact Evaluation established

3ie is founded with technical support from CGD to finance, broker, disseminate, and facilitate use of impact evaluations in collaboration with donors and country governments.

2009–2011

Launch of US agency evaluation policies

Marking a major step forward for impact evaluations of aid programs, the Millennium Challenge Corporation (MCC) and the US Agency for International Development (USAID) developed formal agency-wide evaluation policies in 2009 and 2011, respectively.

2010–2020

New wave of evidence-to-policy organizations

New “knowledge broker” organizations are established to translate evidence into policy and practice. Many have greater proximity to local contexts, an interest in capacity strengthening and demand generation, and a focus on ensuring that research questions align with government decision-making needs. Examples include the African Institute for Development Policy, East Africa Social Science Translation Collaborative, Africa Evidence Network, Knowledge Sector Initiative, Evidence Action, and the Network of Impact Evaluation Researchers in Africa, among others.

2010–present

Rise of evidence synthesis products

Following the rise in the number of impact evaluations, more synthesis products such as systematic reviews are produced in specific policy areas. Some efforts are directly linked to the Campbell Collaboration, an international research network that promotes the production and use of high-quality systematic reviews. The need to create more accessible evidence synthesis products with implications for policy has also been recognized in recent years, as shown by the Metaketa Initiative, MCC’s evaluation briefs, and evidence gap maps like those from 3ie, among many other examples (though many products are not publicly available). Their use is further facilitated by “knowledge broker” organizations that help link evidence to policy and practice (see previous milestone).

2010–present

New impact investment and prize funds established

Starting with Development Innovation Ventures (DIV), based at USAID, impact investment and prize funds for innovations in development based on rigorous impact measurement continued to sprout over the last decade. Examples include the Global Innovation Fund, the Agency Fund, and the Fund for Innovation in Development.

2013

What Works movement picks up in the UK and the US

The UK creates the What Works Network with nine centers, each focused on a different policy area, using a program of impact evaluation and systematic review to better inform the design and delivery of public services at home and abroad. In the US, the What Works Clearinghouse (WWC) synthesizes and distributes knowledge products on education interventions.

2014

First Development Impact Bond launched

Development Impact Bonds (DIBs) finance development programs with money from private investors who earn a return paid by a third-party donor if the program is successful. The outcomes to be measured are agreed upon at the outset and independently verified. A working group jointly led by CGD and Social Finance UK was instrumental in the design, marketing, and adoption of DIBs. The first DIB in education—a partnership between UBS Optimus Foundation, Children’s Investment Fund Foundation, Educate Girls, Instiglio, and IDinsight—improved girls’ education outcomes in Rajasthan, India.

2016

Publication of the third edition of Millions Saved: New Cases of Proven Success in Global Health

Millions Saved is a collection of success stories of large-scale efforts to improve global health. The 2016 edition of the book by Amanda Glassman and Miriam Temin follows on two previous versions published in 2004 and 2007 by Ruth Levine. The three editions showcase more than 30 rigorously evaluated case studies of successful interventions, upending conventional pessimism about public health challenges and drawing broad lessons about what works in global health. *Millions Saved* is a key resource for health policy decision makers, implementers and students worldwide

2018–present

Rise of effective altruism

The rise of “effective altruism” serves as an animating motivation for a new generation of philanthropists, embodied by GiveWell and Open Philanthropy, dedicated to using high-quality evidence to determine how to help others as much as possible. They seek to base recommendations and contributions on the empirically observed cost-effectiveness of different interventions. NGOs like GiveDirectly and New Incentives make use of RCTs to become eligible for funding.

2019

Evidence Act approved in the US

In the United States, the January 2019 bipartisan passage and approval of the Foundations for Evidence-Based Policymaking Act (H.R.4174) promotes data accessibility, responsible data use, and government evidence generation and use for more effective programs. This involves installing chief evaluation officers in every agency, establishing agency-wide learning agendas, and discrete resourcing for evidence generation and use.

2019

Nobel Prize awarded for use of experimental evaluation in development economics

Abhijit Banerjee, Esther Duflo, and Michael Kremer win the Nobel Prize in Economics for their experimental approach to alleviating global poverty.

2019

OECD DAC Network on Development Evaluation adopts new evaluation criteria

The OECD DAC Network on Development Evaluation (EvalNet) publishes six evaluation criteria (relevance, coherence, effectiveness, efficiency, impact, and sustainability), two principles for their use, and guidance to support contextualized application of the criteria, with the aim of helping achieve the Sustainable Development Goals as set out in Agenda 2030 and the Paris Agreement.

2020

Global Evaluation Initiative launched

The Global Evaluation Initiative (GEI) launches with the aim of building country-owned monitoring and evaluation frameworks and evaluation capacities. GEI is supported by a multidonor trust fund managed by the Independent Evaluation Group of the World Bank.

2022 goal

Breakthrough in policy use of impact evaluation funding and practice

CGD launches the final report of its Working Group on New Evidence Tools for Policy Impact, which highlights how far the impact evaluation and broader evidence field has come and proposes a renewed funding agenda for greater value from evidence in public policymaking. The development community elevates the importance of shifting agenda-setting power and resources to those who best understand, and can respond to, decision making needs in different contexts.

2030 vision

Investments transform impact evaluation and evidence use for shared prosperity

Development funders take collective action to unlock the potential of impact evaluation as a powerful tool for better social and economic outcomes, including through support for long-term partnerships for demand generation and decision responsiveness, and investments in data and evidence use systems. In light of these real-world benefits, government policymakers routinely invest in evidence generation and use. Leaders across governments and development institutions commit to institutionalizing evidence production and uptake to improve decision making for improved lives.

References and Resources

- 3ie (International Initiative for Impact Evaluation). *3ie's Evidence Use and Impact Measurement Approach*. Washington, DC: International Initiative for Impact Evaluation, April 2021. <https://www.3ieimpact.org/sites/default/files/2021-01/3ie-evidence-use-measurement-approach-web.pdf>.
- . “Big Data Systematic Map,” <https://gapmaps.3ieimpact.org/evidence-maps/big-data-systematic-map>.
- . “Development Evidence Portal,” <https://developmentevidence.3ieimpact.org/>.
- . “Evidence Gap Maps,” <https://www.3ieimpact.org/evidence-hub/publications/evidence-gap-maps>.
- . “Evidence Impact Summaries,” <https://www.3ieimpact.org/evidence-hub/evidence-impact-summaries>.
- . “More Poor Households Receive Increased Cash in Hand in Ghana.” 3ie Evidence Impact Summary. Last modified November 13, 2020. <https://www.3ieimpact.org/evidence-hub/Evidence-impact-summaries/more-poor-households-receive-increased-cash-hand-ghana>.
- . “Philippines Evidence Program,” <https://www.3ieimpact.org/our-work/philippines-evidence-program>.
- . *Transparent, Reproducible, and Ethical Evidence (TREE) Policy*, version 3. Washington, DC: 3ie, February 2022. <https://www.3ieimpact.org/sites/default/files/2022-04/3ie-transparent-reproducible-ethical-evidence-policy-2022.pdf>.
- 60 Decibels. “Remote Survey Toolkit: Prepared in Response to COVID-19.” New York: 60 Decibels, March 2020. https://60decibels.com/user/pages/07Work/_remote_survey_toolkit/60_Decibels_Remote_Survey_Toolkit_March_2020.pdf.
- Abadie, Alberto. “Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects.” *Journal of Economic Literature* 59, no. 2 (June 2021): 391–425. <https://doi.org/10.1257/jel.20191450>.
- Abebe, Girum, Simon Franklin, Clément Imbert, and Carolina Mejía-Mantilla. “Urban Public Works in Spatial Equilibrium: Experimental Evidence from Ethiopia.” CEPR Discussion Paper DP16691. London: Centre for Economic Policy Research, November 2021. <https://papers.ssrn.com/abstract=3960295>.
- Abimbólá, Sèye, Sumegha Asthana, Cristian Montenegro, Renzo R. Guinto, Desmond Tanko Jumbam, Lance Louskieter, Kenneth Munge Kabubei, et al. “Addressing Power Asymmetries in Global Health: Imperatives in the Wake of the COVID-19 Pandemic.” *PLOS Medicine* 18, no. 4 (April 22, 2021): e1003604. <https://doi.org/10.1371/journal.pmed.1003604>.
- AFIDEP (African Institute for Development Policy). “Malawi Parliamentary Support Initiative,” <https://www.afidep.org/programme/malawi-parliamentary-support-initiative-mpsi/>.
- Africa Centre for Evidence. “Short Learning Programme: Practices of Evidence-Informed Decision-Making.” Africa Centre for Evidence, <https://africacentreforevidence.org/practices-of-evidence-informed-decision-making/>.
- Africa Evidence Network. “AELA 2021: The Culmination of a Passion That Became a Career.” Interview with Diana Warira Njeri, Africa Evidence Network, April 15, 2021a. <https://aen-website.azurewebsites.net/en/learning-space/article/90/>.
- . *Manifesto on Capacity Development for Evidence-Informed Decision-Making in Africa*. Johannesburg: Africa Evidence Network, 2021b. <https://www.africaevidencenetwork.org/en/learning-space/file/189/>.
- African Development Bank. “Preparing Evaluation of the Future: Big Data, Modern Technologies, and Shifts in Global Development Priorities.” *eVALUation Matters* (Second Quarter 2020). <http://idev.afdb.org/sites/default/files/documents/files/eVALUation%20Matters%20Magazine%20Q2-2020%20%28EN%29.pdf>.
- Aiken, Emily L., Guadalupe Bedoya, Aidan Coville, and Joshua E. Blumenstock. “Targeting Development Aid with Machine Learning and Mobile Phone Data: Evidence from an Anti-Poverty Intervention in Afghanistan.” COMPASS '20: Proceedings of the 3rd ACM SIGCAS Conference on Computing and Sustainable Societies (June 2020): 310–11. <https://doi.org/10.1145/3378393.3402274>.
- Aiken, Emily L., Suzanne Bellue, Dean Karlan, Chris Udry, and Joshua E. Blumenstock. “Machine Learning and Phone Data Can Improve Targeting of Humanitarian Aid.” *Nature* 603 (2022): 864–70. <https://doi.org/10.1038/s41586-022-04484-9>.
- Allier-Gagneur, Zoé, Zack Devlin Foltz, Mallika Sobti, and Marc Shotland. “Improving Decisions about Evidence: IDinsight’s New Value of Information Work.” IDinsight blog, March 28, 2022. <https://www.idinsight.org/article/improving-decisions-about-evidence-idinsights-new-value-of-information-work/>.
- Alpízar, Francisco, and Paul J. Ferraro. “The Environmental Effects of Poverty Programs and the Poverty Effects of Environmental Programs: The Missing RCTs.” *World Development* 127 (March 2020): 104783. <https://doi.org/10.1016/j.worlddev.2019.104783>.
- Altshuler, Norma. “Future Directions for Evidence-Informed Policy.” William & Flora Hewlett Foundation, February 3, 2022. <https://hewlett.org/future-directions-for-evidence-informed-policy/>.
- Altshuler, Norma and Sarah Jane Staats. “A New Look at Impact Evaluation Capacity in Sub-Saharan Africa.” Research Brief, William & Flora Hewlett Foundation, October 2019. <https://hewlett.org/wp-content/uploads/2019/10/A-New-Look-at-Impact-Evaluation-Capacity-in-Sub-Saharan-Africa.pdf>.
- Al-Ubaydli, Omar, John A. List, and Dana Suskind. “The Science of Using Science: Towards an Understanding of the Threats to Scaling Experiments.” NBER Working Paper 25848. Cambridge, MA: National Bureau of Economic Research, May 2019. <https://doi.org/10.3386/w25848>.
- Álvarez Marinelli, Horacio, Samuel Berlinski, and Matias Busso. “Remedial Education: Evidence from a Sequence of Experiments in Colombia.” *Journal of Human Resources* (October 13, 2021): 0320. <https://doi.org/10.3368/jhr.0320-10801R2>.

- Angrist, Noam, Peter Bergman, and Moitshepi Matsheng. "Experimental evidence on learning using low-tech when school is out." *Nature Human Behavior* (2022). <https://doi.org/10.1038/s41562-022-01381-z>.
- Athey, Susan, Raj Chetty, Guido W. Imbens, and Hyunseung Kang. "The Surrogate Index: Combining Short-Term Proxies to Estimate Long-Term Treatment Effects More Rapidly and Precisely." NBER Working Paper 26463. Cambridge, MA: National Bureau of Economic Research, November 2019. <https://doi.org/10.3386/w26463>.
- Athey, Susan, and Guido W. Imbens. "The State of Applied Econometrics: Causality and Policy Evaluation." *Journal of Economic Perspectives* 31, no. 2 (May 2017): 3–32. <https://doi.org/10.1257/jep.31.2.3>.
- Bamberger, Michael, and Pete York. *Measuring Results and Impact in the Age of Big Data: The Nexus of Evaluation, Analytics, and Digital Technology*. New York: The Rockefeller Foundation, March 2020. <https://www.rockefellerfoundation.org/report/measuring-results-and-impact-in-the-age-of-big-data-the-nexus-of-evaluation-analytics-and-digital-technology/>.
- Banerjee, Abhijit, Marcella Alsan, Emily Breza, Arun G. Chandrasekhar, Abhijit Chowdhury, Esther Duflo, Paul Goldsmith-Pinkham, and Benjamin A. Olken. "Messages on COVID-19 Prevention in India Increased Symptoms Reporting and Adherence to Preventive Behaviors Among 25 Million Recipients with Similar Effects on Non-Recipient Members of Their Communities." NBER Working Paper 27496. Cambridge, MA: National Bureau of Economic Research, July 2020. <https://doi.org/10.3386/w27496>.
- Banerjee, Abhijit, Esther Duflo, Clément Imbert, Santhosh Mathew, and Rohini Pande. "E-Governance, Accountability, and Leakage in Public Programs: Experimental Evidence from a Financial Management Reform in India." *American Economic Journal: Applied Economics* 12, no. 4 (October 2020): 39–72. <https://doi.org/10.1257/app.20180302>.
- Barrera-Osorio, Felipe, David Deming, Kathryn Gonzalez, and Francisco Lagos. "Using Information to Foster Parental Engagement in Education: Evidence from Colombia." *VoxDev*, January 17, 2022. <https://voxddev.org/topic/health-education/using-information-foster-parental-engagement-education-evidence-colombia>.
- Bates, Mary Ann, and Rachel Glennerster. "The Generalizability Puzzle (SSIR)." *Stanford Social Innovation Review* 15, no. 3 (2017): 50–54. https://ssir.org/articles/entry/the_generalizability_puzzle.
- Bédécarrats, Florent, Isabelle Guérin, and François Roubaud, eds. *Randomized Control Trials in the Field of Development: A Critical Perspective*. Oxford: Oxford University Press, 2020. <https://doi.org/10.1093/oso/9780198865360.001.0001>.
- Blattman, Christopher, Julian C. Jamison, and Margaret Sheridan. "Reducing Crime and Violence: Experimental Evidence from Cognitive Behavioral Therapy in Liberia." *American Economic Review* 107, no. 4 (April 2017): 1165–1206. <https://doi.org/10.1257/aer.20150503>.
- Bold, Tessa, Mwangi Kimenyi, Germano Mwabu, Alice Ng'ang'a, and Justin Sandefur. "Experimental Evidence on Scaling up Education Reforms in Kenya." *Journal of Public Economics* 168 (December 2018): 1–20. <https://doi.org/10.1016/j.jpubeco.2018.08.007>.
- BRAC Institute of Governance and Development. "Women's Economic Empowerment and Digital Finance Resources," <https://bigd.bracu.ac.bd/all-projects/womens-economic-empowerment-and-digital-finance/resources/>.
- Brown, Elizabeth D., and Jeffery C. Tanner. "Integrating Value for Money and Impact Evaluations: Issues, Institutions, and Opportunities." Policy Research Working Paper 9041. Washington, DC: World Bank, October 2019. <https://doi.org/10.1596/1813-9450-9041>.
- Burke, Marshall, and David B. Lobell. "Satellite-Based Assessment of Yield Variation and Its Determinants in Smallholder African Systems." *Proceedings of the National Academy of Sciences* 114, no. 9 (February 2017): 2189–94. <https://doi.org/10.1073/pnas.1616919114>.
- Buteau, Ellie, Satia Marotta, Hannah Martin, Naomi Orensten, and Kate Gehling. *New Attitudes, Old Practices: The Provision of Multiyear General Operating Support*. San Francisco: Center for Effective Philanthropy, 2020. http://cep.org/wp-content/uploads/2020/11/Ford_MYGOS_FNL.pdf.
- Campbell, Danielle M., Sally Redman, Louisa Jorm, Margaret Cooke, Anthony B. Zwi, and Lucie Rychetnik. "Increasing the Use of Evidence in Health Policy: Practice and Views of Policy Makers and Researchers." *Australia and New Zealand Health Policy* 6, no. 1 (August 2009): 21. <https://doi.org/10.1186/1743-8462-6-21>.
- Caria, A. Stefano, Grant Gordon, Maximilian Kasy, Simon Quinn, Soha Shami, and Alexander Teytelboym. "An Adaptive Targeted Field Experiment: Job Search Assistance for Refugees in Jordan." CEGA Working Paper WPS-172. Berkeley, CA: Center for Effective Global Action, August 3, 2021. <https://doi.org/10.26085/C3C012>.
- Carlson, Josh J., Rahber Thariani, Josh Roth, Julie Gralow, N. Lynn Henry, Laura Esmail, Pat Deverka, et al. "Value-of-Information Analysis within a Stakeholder-Driven Research Prioritization Process in a US Setting: An Application in Cancer Genomics." *Medical Decision Making* 33, no. 4 (May 2013): 463–71. <https://doi.org/10.1177/0272989X13484388>.
- Carneiro, Pedro, Lucy Kraftman, Giacomo Mason, Lucie Moore, Imran Rasul, and Molly Scott. "The Impacts of a Multifaceted Prenatal Intervention on Human Capital Accumulation in Early Life." *American Economic Review* 111, no. 8 (August 2021): 2506–49. <https://doi.org/10.1257/aer.20191726>.
- Carter, Samantha, Iqbal Dhaliwal, Julu Katticaran, Claudia Macías, and Claire Walsh. *Creating a Culture of Evidence Use: Lessons from J-PAL's Government Partnerships in Latin America*. Cambridge, MA: J-PAL, December 2018. <https://www.povertyactionlab.org/page/creating-culture-evidence-use>.
- CEGA (Center for Effective Global Action). "Collaboration for Inclusive Development Research," <https://cega.berkeley.edu/initiative/cidr/>.
- . "Cost Transparency Initiative Resources," <https://cega.berkeley.edu/our-resources/>.
- CGD (Center for Global Development). "Governing Data for Development Working Group," <https://www.cgdev.org/working-group/governing-data-for-development>.
- . "How Is Impact Evaluation Contributing to Evidence Use in Low- and Middle-Income Countries?" Online seminar, July 8, 2020. <https://www.cgdev.org/event/how-impact-evaluation-contributing-evidence-use-low-and-middle-income-countries>.
- Channon-Wells, Helena, Stephanie Palla, and Wim Louw. *Guide to Job-Seeker Skills Certificate Programme*. Washington, DC: World Bank, 2020. <https://mbrg.bsg.ox.ac.uk/sites/default/files/2020-09/Skill%20Cert%20Guide.pdf>.
- Chen, John. "Make Every Marketing Dollar Count with Attribution and Lift Measurement." Ads & Commerce blog, Google, October 12, 2020. <https://blog.google/products/ads-commerce/attribution-lift-measurement/>.
- Claxton, Karl P., and Mark J. Sculpher. "Using Value of Information Analysis to Prioritise Health Research." *Pharmacoeconomics* 24, no. 11 (November 2006): 1055–68. <https://doi.org/10.2165/00019053-200624110-00003>.
- Cohen, Jessica, and William Easterly. *What Works in Development?: Thinking Big and Thinking Small*. Washington, DC: Brookings Institution Press, 2009. <https://www.brookings.edu/book/what-works-in-development/>.
- Cole, Shawn, Iqbal Dhaliwal, Anja Sautmann, and Lars Vilhuber, eds. *Handbook on Using Administrative Data for Research and Evidence-Based Policy*. Cambridge, MA: J-PAL and MIT, 2021. <https://admindatahandbook.mit.edu/book/v1.0/index.html>.
- Cole, Shawn, William Parienté, and Anja Sautmann. "A Revolution in Economics? It's Just Getting Started..." *World Development* 127 (March 2020): 104849. <https://doi.org/10.1016/j.worlddev.2019.104849>.

- Cunningham, Scott. "Mixtape Sessions: Causal Inference Workshops," 2022. <https://www.mixtapesessions.io/>.
- Davis, Benjamin, Sudhanshu Handa, Nicola Hypher, Natalia Winder Rossi, Paul Winters, and Jennifer Yablonski, eds. *From Evidence to Action: The Story of Cash Transfers and Impact Evaluation in Sub-Saharan Africa*. Rome: Food and Agriculture Organization of the UN; New York: UNICEF; Oxford: Oxford University Press. <https://www.unicef.nl/files/From%20Evidence%20to%20Action%20Rapport.pdf>.
- Deaton, Angus. "Randomization in the Tropics Revisited: A Theme and Eleven Variations." NBER Working Paper 27600. Cambridge, MA: National Bureau of Economic Research, July 2020. <https://doi.org/10.3386/w27600>.
- Development Data Partnership. <https://datapartnership.org/>.
- Development Impact West Africa. "About DIWA," <https://ndiwa.gimpa.edu.gh/>.
- Dissanayake, Ranil. "The Elusive Quest for Validity." Center for Global Development, forthcoming.
- DuMont, Kimberly. *Reframing Evidence-Based Policy to Align with the Evidence*. New York: William T. Grant Foundation, January 2019. https://wtgrantfoundation.org/library/uploads/2019/01/Reframing-Evidence-based-Policy_WTG-Digest-2018.pdf.
- Eddy, Michael. *Responsive Decision-Focused Evidence: A Landscape Analysis of How Governments Are Learning Better Faster and Cheaper to Improve Lives*. Menlo Park, CA: William & Flora Hewlett Foundation, March 2021. <https://hewlett.org/wp-content/uploads/2022/02/Responsive-decision-focused-evidence-landscape-analysis.pdf>.
- Erondu, Ngozi A., Ifeyinwa Aniebo, Catherine Kyobutungi, Janet Midega, Emelda Okiro, and Fredros Okumu. "Open Letter to International Funders of Science and Development in Africa." *Nature Medicine* 27, no. 5 (April 15, 2021): 742–44. <https://doi.org/10.1038/s41591-021-01307-8>.
- Evans, David K., and Amina Mendez Acosta. "Education in Africa: What Are We Learning?" *Journal of African Economies* 30, no. 1 (January 2021): 13–54. <https://doi.org/10.1093/jae/ejaa009>.
- Evans, David. "Towards Improved and More Transparent Ethics in Randomised Controlled Trials in Development Social Science." CGD Working Paper. Washington, DC: Center for Global Development, 2021. <https://www.cgdev.org/publication/towards-improved-and-more-transparent-ethics-randomised-controlled-trials-development>
- Evidence in Governance and Politics (EGAP). "Metaketa Initiative," <https://egap.org/our-work/0/the-metaketa-initiative/>.
- Ezeh, Alex, and Jessie Lu. "Transforming the Institutional Landscape in Sub-Saharan Africa: Considerations for Leveraging Africa's Research Capacity to Achieve Socioeconomic Development." CGD Policy Paper. Washington, DC: Center for Global Development, 2019. <https://www.cgdev.org/publication/transforming-institutional-landscape-sub-saharan-africa-considerations-leveraging-africa>.
- Fagan, Joseph, Gustavo Piga, Peter-Armin Trepte, Omar Hayat Tiwana, Vinay Sharma, and Masuid Mozammel. *A Global Procurement Partnership for Sustainable Development: An International Stocktaking of Developments in Public Procurement: Synthesis Report*. Washington, DC: World Bank, 2022. <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/173331642410951798/Synthesis-Report>.
- Feeney, Laura, Jason Bauman, Julia Chabrier, Geeti Mehra, and Michelle Woodford. *Using Administrative Data for Randomized Evaluations*. Cambridge, MA: J-PAL, 2019. <https://www.povertyactionlab.org/resource/using-administrative-data-randomized-evaluations>.
- Fenwick, Elisabeth, Lotte Steuten, Saskia Knies, Salah Ghabri, Anirban Basu, James F. Murray, Hendrik (Erik) Koffijberg, et al. "Value of Information Analysis for Research Decisions—An Introduction: Report 1 of the ISPOR Value of Information Analysis Emerging Good Practices Task Force." *Value in Health* 23, no. 2 (February 2020): 139–50. <https://doi.org/10.1016/j.jval.2020.01.001>.
- Fischer, Torben, Doug Johnson, and Daniel Stein. *Informing Specific Decisions with Data and Evidence: Designing and Analyzing Decision-Focused Evaluations*. Washington, DC: IDInsight, January 2021. <https://www.idinsight.org/publication/informing-specific-decisions-with-data-and-evidence/>.
- Fishman, Sam, and Carson Christiano. "Investigating a History of Impact." CEGA blog, Center for Effective Global Action, December 2, 2020. <https://medium.com/center-for-effective-global-action/investigating-a-history-of-impact-53177e8fbcd5>.
- Gaarder, Marie. "Growth in Impact Evaluations and Systematic Reviews over Time." Presented at CGD Seminar: How Is Impact Evaluation Contributing to Evidence Use in Low- and Middle-Income Countries?, July 8, 2020. <https://www.cgdev.org/sites/default/files/CGD-3ie-ppt-Marie-gaarder.pdf>.
- Gaarder, Marie, and Ulrich Bartsch. "Creating a Market for Outcomes: Shopping for Solutions." *Journal of Development Effectiveness* 7, no. 3 (2015): 304–16. <https://doi.org/10.1080/19439342.2015.1068357>.
- Gerler, Paul J., Sebastian Martinez, Patrick Premand, Laura B. Rawlings, and Christel M. J. Vermeersch. *Impact Evaluation in Practice*, 2nd ed. Washington, DC: World Bank, 2016. <https://openknowledge.worldbank.org/handle/10986/25030>.
- Geospatial Analysis for Development. <https://www.geo4.dev/>.
- Glandon, Douglas, Sam Fishman, Caitlin Tulloch, Radhika Bhula, Grace Morgan, Shahana Hirji, and Liz Brown. "The State of Cost-Effectiveness Guidance: Ten Best Resources for CEA in Impact Evaluations." *Journal of Development Effectiveness*, February 2022. <https://doi.org/10.1080/19439342.2022.2034916>.
- Glassman, Amanda, Alex Ezeh, Jessica Brinton, Kate McQueston, Jenny Ottenhof, and Justin Sandefur. *Delivering on the Data Revolution in Sub-Saharan Africa: Final Report of the Data for African Development Working Group*. Washington, DC: CGD, July 8, 2014. <https://www.cgdev.org/publication/delivering-data-revolution-sub-saharan-africa-0>.
- Glassman, Amanda, and Miriam Temin. *Millions Saved: New Cases of Proven Success in Global Health*. Washington, DC: Center for Global Development, 2016.
- Glennister, Rachel. "The Practicalities of Running Randomized Trials" Course syllabus and lecture slides, Fall 2021. <http://runningres.com/course-details>.
- Glewwe, Paul, and Petra Todd. *Impact Evaluation in International Development: Theory, Methods and Practice*. Washington, DC: World Bank, 2022. <https://openknowledge.worldbank.org/handle/10986/37152>
- Global Commission on Evidence to Address Societal Challenges. "The Evidence Commission Report: A Wake-up Call and Path Forward for Decisionmakers, Evidence Intermediaries, and Impact-Oriented Evidence Producers." Ontario: McMaster Health Forum, January 2022. <https://www.mcmasterforum.org/networks/evidence-commission/report/english>.
- Global Evaluation Initiative. "The GEI Approach," <https://www.globalevaluationinitiative.org/>.
- Godoy, Ricardo, Dean Karlan, and Jonathan Zinman. "Randomization for Causality, Ethnography for Mechanisms: Illiquid Savings for Liquor in an Autarkic Society." NBER Working Paper 29566. Cambridge, MA: National Bureau of Economic Research, December 2021. <https://doi.org/10.3386/w29566>.
- Goldman, Ian, and Mine Pabari, eds. *Using Evidence in Policy and Practice: Lessons from Africa*. London: Routledge, 2020. <https://doi.org/10.4324/9781003007043>.
- Gratia, Zach. *Social Return on Investment*. Saarbrücken, Germany: AV Akademikerverlag, 2014.
- Green, Duncan. "\$15 Billion Is Spent Every Year on Training, with Disappointing Results. Why the Aid Industry Needs to Rethink 'Capacity Building.'" *From Poverty to Power* (blog), July 6, 2017. <https://oxfamapps.org/fp2p/15bn-is-spent-every-year-on-aid-for-training-with-disappointing-results-why-the-aid-industry-needs-to-rethink-its-approach-to-capacity-building/>.

- Gugerty, Mary Kay, and Dean Karlan. "Ten Reasons Not to Measure Impact—and What to Do Instead." *Stanford Social Innovation Review* 16, no. 3 (Summer 2018): 41–47. https://ssir.org/articles/entry/ten_reasons_not_to_measure_impact_and_what_to_do_instead.
- Gyamfi, Bridget Konadu, and Bethany Park. "How to Keep Good Research from Dying a Bad Death: Strategies for Co-Creating Research with Impact." Innovations for Poverty Action blog, January 10, 2019. <https://www.poverty-action.org/blog/how-keep-good-research-dying-bad-death-strategies-co-creating-research-with-impact>.
- Henderson, J. Vernon, Adam Storeygard, and David N. Weil. "Measuring Economic Growth from Outer Space." *American Economic Review* 102, no. 2 (April 2012): 994–1028. <https://doi.org/10.1257/aer.102.2.994>.
- Hirji, Shahana, Bethany Park, Edward Tsinigo, Sabrin Beg, Anne Fitzpatrick, and Adrienne Lucas. "Facilitating Real-Time Cost Collection and Evaluating Cost-Effectiveness in a Multi-Armed Study with Government Partners in Ghana." *Journal of Development Effectiveness*, published online January 25, 2022. <https://doi.org/10.1080/19439342.2021.2024589>.
- HITAP (Health Intervention and Technology Assessment Program). <https://www.hitap.net/en/>.
- IDinsight. "IDinsight Partners with Radiant Earth Foundation to Create High-Quality Ground-Truthed Baseline Geospatial Data for Machine Learning Models in Agriculture." January 18, 2022. <https://www.idinsight.org/article/idinsight-partners-with-radiant-earth-foundation-to-create-high-quality-ground-truthed-baseline-geospatial-data-for-machine-learning-models-in-agriculture/>.
- . "Featured Projects," <https://www.idinsight.org/our-work/projects/>.
- . "Government Agencies," <https://www.idinsight.org/our-work/partners/government-agencies/>.
- IEG (Independent Evaluation Group). *World Bank Group Impact Evaluations: Relevance and Effectiveness*. Washington, DC: World Bank, 2012. <https://documents1.worldbank.org/curated/en/206961468154467890/pdf/726980WP0Box37000impact0eval0report.pdf>.
- India Clean Air Markets. "About Clean Air Markets in India," 2021. <https://www.cleanairmarkets.in/india/about/ets-in-india/>
- IPA (Innovations for Poverty Action). "Case Studies: Impact," <https://www.poverty-action.org/impact/case-studies>.
- . "Embedded Labs," February 26, 2021. <https://www.poverty-action.org/advisory/embedded-labs>.
- . "Ghana's Embedded Evidence Lab in Education," February 7, 2022. <https://www.poverty-action.org/ghana-embedded-evidence-lab-in-education>.
- . "ProcuraLAB - Public Innovation Lab," <https://www.poverty-action.org/procuralab>.
- . "RECOVR Research Hub," <https://www.poverty-action.org/recovr/research-hub>.
- . "Rwanda Embedded Lab," <https://www.poverty-action.org/rwanda-embedded-lab>.
- International Centre for Evaluation and Development. "Feed the Future Advancing Local Leadership, Innovation and Networks (ALL IN)," <https://www.iced-eval.org/programs/all-in>.
- International Network for Government Science Advice. "INGSA Governance," September 17, 2021. <https://ingsa.org/about/ingsa-governance/>.
- IRC (International Rescue Committee). *Cost Analysis Methodology at the IRC*. Washington, DC: International Rescue Committee, September 2016. <https://www.rescue.org/report/cost-analysis-methodology-irc>.
- Isaksson, Ann-Sofie. "Rapid and Rigorous Impact Evaluation: Advances in the Methods and Data Available for Timely and Cost-Efficient Evaluation." CGD Background Paper. Washington, DC: Center for Global Development, August 2021. <https://www.cgdev.org/sites/default/files/Rapid-evaluation-background-paper.pdf>.
- Islam, Asad. "Heterogeneous Effects of Microcredit: Evidence from Large-Scale Programs in Bangladesh." *Journal of Asian Economics* 37 (2015): 48–58. <https://doi.org/10.1016/j.asieco.2015.01.003>.
- Jayachandran, Seema, Joost de Laat, Eric F. Lambin, Charlotte Y. Stanton, Robin Audy, and Nancy E. Thomas. "Cash for Carbon: A Randomized Trial of Payments for Ecosystem Services to Reduce Deforestation." *Science* 357, no. 6348 (July 21, 2017): 267–73. <https://doi.org/10.1126/science.aan0568>.
- J-PAL. "Digital Identification and Finance Initiative in Africa: African Scholars Opportunities," <https://www.povertyactionlab.org/initiative-subpage/african-scholars>.
- . "Evidence to Policy," <https://www.povertyactionlab.org/evidence-policy>.
- . "Free Bednets to Fight Malaria." J-PAL Evidence to Policy Case Study. Last modified October 2018. <https://www.povertyactionlab.org/case-study/free-bednets-fight-malaria>.
- . "MineduLAB," <https://www.povertyactionlab.org/page/minedulab>.
- . "Policy Insights," <https://www.povertyactionlab.org/policy-insights>.
- Keller, Janeen Madan, and Julia Kaufman. "Aligning the Researcher's Toolkit with the Policymaker's Priorities: A Menu of Strategies for Faster, Lower-Cost Impact Evaluations." Center for Global Development blog, August 18, 2021a. <https://www.cgdev.org/blog/aligning-researchers-toolkit-policymakers-priorities-menu-strategies-faster-lower-cost-impact>.
- . "How More and Better Funding Can Unlock the Potential of Evidence-to-Policy Partnerships." Center for Global Development blog, August 18, 2021b. <https://www.cgdev.org/blog/how-more-and-better-funding-can-unlock-potential-evidence-policy-partnerships>.
- Kochar, Anjini, Bidisha Barooah, Chandan Jain, Geeta Singh, Nagabhushana Closepet, Raghunathan Narayanan, Ritwik Sarkar, and Rohan Shah. *Impact Evaluation of the National Rural Livelihoods Project*. Impact Evaluation Report 128. Washington, DC: 3ie, 2020. <https://www.3ieimpact.org/evidence-hub/publications/impact-evaluations/impact-evaluation-national-rural-livelihoods-project>.
- Langer, Laurenz, and Venesa Weyrauch. "Using Evidence In Africa: A Framework to Assess What Works, How, and Why." In *Using Evidence in Policy and Practice: Lessons from Africa*, edited by Ian Goldman and Mine Pabari, 34–53. London: Routledge, 2020. <https://doi.org/10.4324/9781003007043>.
- Legovini, Arianna, Guignon Serge Adjognon, Guadalupe Bedoya, Theophile Bougna, Kayleigh B. Campbell, Paul Christian, Aidan Coville, et al. *Science for Impact: Better Evidence for Better Decisions—The Dime Experience*. Washington, DC: World Bank, 2019. <https://documents.worldbank.org/en/publication/documents-reports/documentdetail/942491550779087507/science-for-impact-better-evidence-for-better-decisions-the-dime-experience>.
- Leight, Jessica (@leightjessica). "The funding mix for published RCTs has re-balanced somewhat." Twitter, December 7, 2021. <https://twitter.com/leightjessica/status/1468297896613318665>.
- . "Wanted to do an #EconTwitter thread on a new + important topic that's growing in the literature. . ." Twitter thread, March 15, 2022a. <https://threadreaderapp.com/thread/1503800956714160132.html>.
- . "Was excited to see #EconTwitter interest in evidence on how policymakers use evidence." Twitter thread, March 17, 2022b. <https://threadreaderapp.com/thread/1504520482745376775.html>.
- Leo, Ben, Shaan Pattni, Catherine Winn, Quinn Lewis, Christina Paton, and Melissa Persaud. "Using Machine Learning for Climate Related Impact Evaluations." *eVALUation Matters* (Second Quarter 2020): 22–33. <https://idev.afdb.org/sites/default/files/documents/files/EM%20Q2-2020-article2-Using%20machine%20learning%20for%20climate%20impact%20evaluation%28En%29.pdf>.

- Levine, Ruth. "The Moral Case for Evidence in Policymaking." William & Flora Hewlett Foundation, September 29, 2017. <https://hewlett.org/moral-case-evidence-policy-making/>.
- Levy, Karen, Julie Wang'ombe, and Diksha Radhakrishnan. "An Evaluation of a Relative-Risk HIV Awareness Campaign Generated Mixed Results . . . Here's What We Learned from It." Evidence Action blog, June 27, 2018. <https://www.evidenceaction.org/no-sugar-not-scaling-ii/>.
- List, John A. *The Voltage Effect: How to Make Good Ideas Great and Great Ideas Scale*. New York: Currency, 2022.
- Lucas, Sarah, and Norma Altshuler. "How Non-Governmental Organizations Can Help Institutionalize Government Use of Evidence: Four Possible Paths." William & Flora Hewlett Foundation, September 22, 2021. <https://hewlett.org/how-non-governmental-organizations-can-help-institutionalize-government-use-of-evidence-four-possible-paths/>.
- Macaulay, Molly, and Ramanan Laxminarayan. "The Value of Information: 'Methodological Frontiers and New Applications for Realizing Social Benefit' Workshop." *Space Policy* 26, no. 4 (November 2010): 249–51. <https://doi.org/10.1016/j.spacepol.2010.08.007>.
- Manning, Richard, Ian Goldman, and Gonzalo Hernández Licona. "The Impact of Impact Evaluation: Are Impact Evaluation and Impact Evaluation Synthesis Contributing to Evidence Generation and Use in Low- and Middle-Income Countries?" WIDER Working Paper 2020/20. Helsinki: UNU-WIDER, 2020. <https://doi.org/10.35188/UNU-WIDER/2020/777-4>.
- MCC (Millennium Challenge Corporation). "Evaluation Briefs," <https://www.mcc.gov/our-impact/evaluation-briefs>.
- McKenzie, David. "How Can Machine Learning and Artificial Intelligence Be Used in Development Interventions and Impact Evaluations?" *Development Impact* (blog), World Bank, March 5, 2018. <https://blogs.worldbank.org/impactevaluations/how-can-machine-learning-and-artificial-intelligence-be-used-development-interventions-and-impact>.
- Meager, Rachael. "Understanding the Average Impact of Microcredit Expansions: A Bayesian Hierarchical Analysis of Seven Randomized Experiments." *American Economic Journal: Applied Economics* 11, no. 1 (January 2019): 57–91. <https://doi.org/10.1257/app.20170299>.
- Mehmood, Sultan, Shaheen Naseer, and Daniel Li Chen. "Training Policymakers in Econometrics: Evidence from Pakistan." *VoxDev*, August 29, 2021. <https://voxdev.org/topic/methods-measurement/training-policymakers-econometrics-evidence-pakistan>.
- MITx MicroMasters Programs. "Data, Economics, and Development Policy MicroMasters," <https://micromasters.mit.edu/dedp/>.
- Morton, Ben, Andre Verceuil, Refiloe Masekela, Eva Heinz, Lisa Reimer, Sepeedeh Saleh, Chisomo Kalinga, et al. "Consensus Statement on Measures to Promote Equitable Authorship in the Publication of Research from International Partnerships." *Anaesthesia* 77, no. 3 (2022): 264–76. <https://doi.org/10.1111/anae.15597>.
- Muralidharan, Karthik, and Paul Niehaus. "Experimentation at Scale." *Journal of Economic Perspectives* 31, no. 4 (November 2017): 103–24. <https://doi.org/10.1257/jep.31.4.103>.
- Murunga, Violet. "Reflections from a Peer-Learning Workshop on Using Evidence to Improve Policy Implementation." African Institute for Development Policy blog, July 25, 2018. <https://www.afidep.org/reflections-from-a-peer-learning-workshop-on-using-evidence-to-improve-policy-implementation/>.
- Myers, Evan, Amanda J. McBroom, Lan Shen, Rachael E. Posey, Rebecca Gray, and Gillian D. Sanders. *Value-of-Information Analysis for Patient-Centered Outcomes Research Prioritization*. Washington, DC: Patient-Centered Outcomes Research Institute, March 9, 2012. <https://www.pcori.org/assets/Value-of-Information-Analysis-for-Patient-Centered-Outcomes-Research-Prioritization.pdf>.
- National School of Government. "Introduction to Monitoring and Evaluation in the Public Sector: Analysing and Using M&E Evidence to Improve Government Performance." *Course Directory*, 53–54. Pretoria, South Africa: National School of Government. https://www.thensg.gov.za/wp-content/uploads/2016/08/NSG_Course_Directory_Final-002.pdf#page=55.
- NIERA (Network of Impact Evaluation Researchers in Africa). "Decision Focused Evaluation Trainings for Media Practitioners in East Africa," June 2–4, 2021. <https://nieraglobal.org/event/view/decision-focused-evaluation-trainings-for-media-practitioners-in-east-africa>.
- . "Decision Focused Evaluation Trainings for Policymakers," February 2–4, 2021. <https://nieraglobal.org/event/view/decision-focused-evaluation-trainings-for-policymakers>.
- OES (Office of Evaluation Sciences). "Evidence Act Toolkits: Learning Agenda, Annual Evaluation Plan, Capacity Assessment," <https://oes.gsa.gov/toolkits/>.
- . "Intergovernmental Personnel Act Toolkit: Internal Guide for IPAs," February 2022. <https://oes.gsa.gov/assets/files/ipa-toolkit-oes.pdf>.
- OVE (Office of Evaluation and Oversight). *IDB's Impact Evaluations: Production, Use, and Influence*. Washington, DC: Inter-American Development Bank, 2017. <https://publications.iadb.org/publications/english/document/IDB-Impact-Evaluations-Production-Use-and-Influence.pdf>.
- Ogden, Timothy N., ed. *Experimental Conversations: Perspectives on Randomized Trials in Development Economics*. Cambridge, MA: MIT Press, 2017. <https://mitpress.mit.edu/books/experimental-conversations>.
- Oliver, Kathryn, Simon Innvar, Theo Lorenc, Jenny Woodman, and James Thomas. "A Systematic Review of Barriers to and Facilitators of the Use of Evidence by Policymakers." *BMC Health Services Research* 14, no. 1 (January 2014): 2. <https://doi.org/10.1186/1472-6963-14-2>.
- Pan-African Scientific Research Council. "PASRC Career Development Webinar: Funding, Grant Applications and New Funding Opportunities," 2021. <https://pasrc.princeton.edu/news-events/meetings-events/2021/pasrc-career-development-webinar-funding-grant-applications-and>.
- Partnership for African Social and Governance Research. "Utafiti Sera." <https://www.pasgr.org/what-we-do/research/utafiti-sera/>.
- . "Utafiti Sera: Institutionalizing Evidence Use: What We Mean and What We're Learning." Hewlett Foundation. YouTube video, 2021. <https://www.youtube.com/watch?v=9UNIJQLyaOs>.
- Paul, Amy, Craig Jolley, and Aubra Anthony. *Reflecting the Past, Shaping the Future: Making AI Work for International Development*. Washington, DC: USAID Center for Digital Development, September 2018. <https://www.usaid.gov/digital-development/machine-learning/AI-ML-in-development>.
- Piper, Benjamin, Joseph Destefano, Esther M. Kinyanjui, and Salome Ong'ele. "Scaling up Successfully: Lessons from Kenya's Tusome National Literacy Program." *Journal of Educational Change* 19, no. 3 (August 2018): 293–321. <https://doi.org/10.1007/s10833-018-9325-4>.
- Pisa, Michael, Pam Dixon, and Ugonma Nwankwo. "Why Data Protection Matters for Development: The Case for Strengthening Inclusion and Regulatory Capacity." CGD Note. Washington, DC: Center for Global Development, December 2021. <https://www.cgdev.org/publication/why-data-protection-matters-development-case-strengthening-inclusion-and>.
- Plan S. "What Is COAlition S?," <https://www.coalition-s.org/about/>.
- PPIAF (Public-Private Infrastructure Advisory Facility). "PPIAF Knowledge Resources," <https://ppiaf.org/knowledge>.

- Raifman, Julia, Felix Lam, Janeen Madan Keller, Alexander Radunsky, and William Savedoff. "How Well Are Aid Agencies Evaluating Programs? An Assessment of the Quality of Global Health Evaluations." *Journal of Development Effectiveness* 10, no. 2 (2018): 277–89. <https://doi.org/10.1080/19439342.2018.1452779>.
- Ranganath, Maya, and Amy Shipow. "Shaping the Future of Inclusion in Global Development Research." Center for Effective Global Action blog, April 11, 2022. <https://medium.com/center-for-effective-global-action/shaping-the-future-of-inclusion-in-global-development-research-76f9763c208a>.
- Rathinam, Francis, Ran Goldblatt, Douglas Glandon, and Ghermay Araya. "Remotely Sensed Data for Efficient Data Collection." GIM International, July 17, 2020. <https://www.gim-international.com/content/article/remotely-sensed-data-for-efficient-data-collection>.
- Rathinam, Francis, Sayak Khatua, Zeba Siddiqui, Manya Malik, Pallavi Duggal, Samantha Watson, and Xavier Vollenweider. "Using Big Data for Evaluating Development Outcomes: A Systematic Map." CEDIL Methods Working Paper 2. Oxford: Centre of Excellence for Development Impact and Learning, 2020. <https://doi.org/10.51744/CMWP2>.
- Rathinam, Francis, Paul Thissen, and Marie Gaarder. "Using Big Data for Impact Evaluations." CEDIL Methods Working Brief 2. Oxford: Centre of Excellence for Development Impact and Learning, February 2021. <https://doi.org/10.51744/CMB2>.
- Results for All. "Incentivizing Evidence-Informed Decision Making in Government: A Series of Case Studies," <https://results4america.org/incentives-for-evidence-informed-policy-making/>.
- Rodgers, Yana van der Meulen, Anthony Bebbington, Catherine Boone, Jampel Dell'Angelo, Jean-Philippe Platteau, and Arun Agrawal, eds. "Symposium on Development and Poverty Alleviation." Special Issue, *World Development* 127 (March 2020). <http://www.sciencedirect.com/journal/world-development/vol/127/suppl/C>.
- Savedoff, William D., Ruth Levine, and Nancy Birdsall. "When Will We Ever Learn?: Improving Lives through Impact Evaluation." Washington, DC: Center for Global Development, 2006. <https://www.cgdev.org/publication/when-will-we-ever-learn-improving-lives-through-impact-evaluation>.
- Shah, Neil Buddy, Paul Wang, Andrew Fraker, and Daniel Gastfriend. "Evaluations with Impact: Decision-Focused Impact Evaluation as a Practical Policymaking Tool." Working Paper 25. 3ie, September 2015. <https://www.idinsight.org/publication/decision-focused-impact-evaluation-as-a-practical-policy-making-tool/>.
- Snilstveit, Birte, Jennifer Stevenson, Laurenz Langer, Natalie da Silva, Zafeer Rabath, Promise Nduku, Joshua Polanin, Ian Shemilt, John Eyers, and Paul J Ferraro. "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." 3ie Systematic Review 44. Washington, DC: 3ie, 2019. <https://www.3ieimpact.org/evidence-hub/publications/systematic-reviews/incentives-climate-mitigation-land-use-sector-effects>.
- South Africa Department of Planning, Monitoring and Evaluation. "DPME: About Us," <https://www.dpme.gov.za/Pages/default.aspx>.
- South Africa National School of Government. "Introduction to M&E," <https://www.thensg.gov.za/training-course/introduction-to-me-copy/>.
- Steiger, William, Christopher Maloney, and Daniel Handel. "Passing the Baton on Data and Evidence." Wilson Center, August 3, 2021. <https://www.wilsoncenter.org/article/passing-baton-data-and-evidence>.
- Stein, Daniel, Rupika Singh, and Madhav Seth. "Making the Government Adoption of Social Innovations Work." *Stanford Social Innovation Review* (November 2021). https://ssir.org/articles/entry/making_the_government_adoption_of_social_innovations_work.
- Taddese, Abeba. "Meeting Policymakers Where They Are: Evidence-to-Policy and Practice Partnership Models." CGD Background Paper. Washington, DC: Center for Global Development, August 2021. <https://www.cgdev.org/sites/default/files/meeting-policy-makers-where-they-are-background-paper.pdf>.
- Teele, Dawn Langan. *Field Experiments and Their Critics: Essays on the Uses and Abuses of Experimentation in the Social Sciences*. New Haven, CT: Yale University Press, 2014. <https://yalebooks.yale.edu/book/9780300169409/field-experiments-and-their-critics/>.
- The Transfer Project. "About," <https://transfer.cpc.unc.edu/about/>.
- . "Beyond Internal Validity: Towards a Broader Understanding of Credibility in Development Policy Research." *World Development* 127 (March 2020): 104802. <https://doi.org/10.1016/j.worlddev.2019.104802>.
- Twende Mbele. "About Us," <https://twendembele.org/about-us/>.
- VoxDev. "VoxDevLit: Wiki-inspired literature reviews." <https://voxdev.org/voxdevlit>
- Vivalt, Eva. "Heterogeneous Treatment Effects in Impact Evaluation." *American Economic Review* 105, no. 5 (May 2015): 467–70. <https://doi.org/10.1257/aer.p20151015>.
- . "How Much Can We Generalize From Impact Evaluations?" *Journal of the European Economic Association* 18, no. 6 (December 2020): 3045–89. <https://doi.org/10.1093/jeea/jvaa019>.
- Washington State Institute for Public Policy. "Benefit-Cost Results," <https://www.wsipp.wa.gov/BenefitCost>.
- . *Benefit-Cost Technical Documentation*. Pullman: Washington State Institute for Public Policy, December 2019. <https://www.wsipp.wa.gov/TechnicalDocumentation/WsippBenefitCostTechnicalDocumentation.pdf>.
- White, Howard. "The Twenty-First Century Experimenting Society: The Four Waves of the Evidence Revolution." *Palgrave Communications* 5, no. 1 (May 2019): 1–7. <https://doi.org/10.1057/s41599-019-0253-6>.
- The White House. "Biden-Harris Administration Releases Recommendations for Advancing Use of Equitable Data," Briefing Room Fact Sheet, April 22, 2022. <https://www.whitehouse.gov/briefing-room/statements-releases/2022/04/22/fact-sheet-biden-harris-administration-releases-recommendations-for-advancing-use-of-equitable-data/>.
- World Bank. *Cox's Bazar Panel Survey: Rapid Follow-up Round 1: Impacts of COVID-19 on Work and Wages in Cox's Bazar*. Washington, DC: World Bank, July 2020. https://fscluster.org/sites/default/files/documents/labor_cxb_covid.pdf.
- . "DIME Overview," <https://www.worldbank.org/en/research/dime/overview>.
- . "Rwanda Impact Evaluation Summer School," <https://www.worldbank.org/en/research/dime/brief/rwanda-impact-evaluation-summer-school>.
- Yale Research Initiative on Innovation and Scale (Y-RISE). "Challenges and Implications of Scale," <https://yrise.yale.edu/>.
- Yale MacMillan Center Program on Refugees, Forced Displacement, and Humanitarian Responses. "Cox's Bazar Panel Survey," <https://refugee.macmillan.yale.edu/research-outputs/coxs-bazar-panel-survey>.
- Yeh, Christopher, Anthony Perez, Anne Driscoll, George Azzari, Zhongyi Tang, David Lobell, Stefano Ermon, and Marshall Burke. "Using Publicly Available Satellite Imagery and Deep Learning to Understand Economic Well-Being in Africa." *Nature Communications* 11 (May 2020). <https://doi.org/10.1038/s41467-020-16185-w>.
- Youth Impact. "Adapting to the Times, Committed to the Cause: COVID-19 Response." <https://www.youth-impact.org/covid-19response>.

PUBLICATION PHOTO CREDIT:
Cover image: Yagazie Emezi/Getty Images/
Images of Empowerment



CENTER
FOR
GLOBAL
DEVELOPMENT

Center for Global Development

2055 L Street NW
Floor 5
Washington DC
20036

CGD Europe

1 Abbey Gardens
Great College Street
London
SW1P 3SE

WWW.CGDEV.ORG

