2021

RESPONSIVE DECISION-FOCUSED EVIDENCE:

A LANDSCAPE ANALYSIS OF HOW GOVERNMENTS ARE LEARNING BETTER FASTER AND CHEAPER HOW TO IMPROVE LIVES

BY MICHAEL EDDY

FOR THE WILLIAM & FLORA HEWLETT FOUNDATION

MARCH 2021

1. Introduction

This past two decades have seen a dramatic increase in the number of impact evaluations. When provided in a timely way, through trusted relationship, and embedded in a broader understanding of the system, norms, and processes in which the evidence is embedded, governments can and have used impact evaluation evidence to improve people's lives.[1]

Yet the pace of policy learning is still too slow. A policymaker may oversee hundreds or thousands of decisions a year—from big questions (whether to provide cash or food aid in response to COVID-19) to the mundane (whether to include a jingle in an automated phone message going out to farmers). These decisions affect the lives of citizens. Yet the available evidence from impact evaluations is relevant to only a narrow set of these questions—and when the evidence does exist, it may not always be accessible to policymakers; or policymakers may not be motivated to act on that evidence over other inputs into their decision-making.

In addition, some sectors remain systematically understudied—many of which, like infrastructure, are significant sources of public expenditure in LMICs.[2] Research questions that are asked, may or may not reflect policymaker needs and be relevant to them. An impact evaluation might inform a single decision, by comparing a treatment arm to a control arm, and produce results in a few years, when the frequency and nature of choices policymakers face are many orders of magnitude larger.

In this context, investing in the tools that power continuous policy improvement—better, faster, and cheaper evidence—may not only be a good use of philanthropic resources but also an efficient way to improve the use and usefulness of evidence. In science, funders like Chan Zuckerberg Initiative choose to invest in the platforms that power science because they believe that an improvement in the process of scientific discovery even by just a small amount can accelerate the next wave of scientific discovery.

Similarly, investing in the platforms that power continuous policy improvement may be able to improve how governments learn what works for whom in their specific context and informed by the very real practical constraints policymakers face.

Yet important barriers and open questions remain. Will policymakers reasonably use these tools displacing their existing ways of making decisions? How useful are these tools in practice and in what settings are they likely to be most useful? What are the data governance and privacy implications of these tools and how might they be responsibly balanced against the potential social benefits as the field invariably lurches forward?

[1] See, for example, Glassman, Amanda, and Miriam Temin. *Millions saved: new cases of proven success in global health*. Brookings Institution Press, 2016.

[2] Infrastructure investment makes up about 4% of GDP in LMICs. This is about the same as public expenditure on education. Yet even if we take an extraordinarily broad view of what one considers infrastructure impact evaluations (including everything in transport, energy, water & sanitation and ICT), the number of impact evaluations in infrastructure are a third the impact evaluations in education (See Sabet, Brown, 2017, Figure 7).

2. Field Landscape

This section provides an overview of the tools being used that have the potential to enable more responsive decision-focused evidence. I focus on tools that address one or more of the following barriers:

- Relevance of evidence to policymaker needs, in particularly for forward-looking decisions
- Timeliness of evidence
- Cost of evidence

Through interviews and a desk review of relevant papers & reports, I identified several tools that hold potential promise for more responsive, decision-focused evidence. Table 1 lists out the tools and how they aim to address supply-side barriers to evidence use.

This report is unable to do justice to all the ways that evidence is being used in a responsive way to improve decisions. Partly due to a parallel working group hosted by the Center for Global Development, this report takes a particularly close look at innovations that are improving the production and use of causal evidence from impact evaluations. I also look at related tools, like geospatial and administrative data, that have been used in impact evaluations, but also have many more uses outside of impact evaluations as well. Given this initial focus, this report should be considered an initial brief look at the field—it should not be considered a universal census of all tools to improve responsive, decision-focused evidence. We encourage other practitioners to see themselves in the field and share other tools that have helped to encourage more responsive, decision-focused evidence use.

Table 1: Barriers to evidence use that the tools potentially address[3]

	Improving use of evidence by addressing....		
	Better (relevance)	Faster (timeliness)	Cheaper (cost)
Innovations in the production of causal evidence			
A/B testing, nimble evaluations, and decision-focused evaluations[4]	✓	✓	✓
Machine learning to identify varied treatment effects (potentially leading to more targeted & tailored treatments)	✓		
Adaptive Experiments[5]	✓	✓	

[3] I assess what I believe to be each tool's *primary* contributions to either better, faster or cheaper policy improvement. Such assessments necessarily involve a question of "compared to what" and therefore can be highly context dependent. For Innovations in the production & use of causal evidence, I compare these approaches to standard impact evaluations and evidence synthesis methods. For the other two categories, I compared to the status quo policy environment. Different analysts may come to different conclusions.

[4] Nimble evaluations use lower-cost administrative data or shorter-term more output-oriented data to lower the costs of evaluation. Often, though not always, this can be combined with faster, more iterative experimentation. See for example, Fabregas, Kremer, and Schilbach (2019), Angrist, et. al., (2020); or SEIF's Nimble Evaluation portfolio.

[5] Adaptive experiments aim to more quickly find optimal treatments among many competing treatment arms by updating likelihood of treatment based on effect sizes in earlier waves. While increasingly used in health, they currently are still on the periphery of development research & policy. I'm aware of only three examples in development: Evaluating Job Search

	Improving use of evidence by addressing....		
	Better (relevance)	Faster (timeliness)	Cheaper (cost)
Novel data in Impact Evaluations[6]	✓	✓	?
Geospatial impact Evaluations	✓	✓	✓
Innovations in the use of causal evidence			
Policy modeling[7]	✓	✓	
Rapid, responsive evidence synthesis	✓	✓	✓
Innovations in the use of data			
Geospatial data (data from satellites; geocoded survey or admin data, highly granular estimates of wealth, mortality, etc.)	✓	✓	
Administrative data	✓	✓	
Private sector data for public policy (e.g. mobile data, social media)	✓		
Citizen-generated data	✓		
Direct client feedback	✓		
Other remote sensor data (e.g. air & water quality sensors)	✓	✓	
Innovations in ways of working			
Performance management / continuous process improvement / iterative feedback loops	✓	✓	
Embedded learning partnerships	✓		
Human centered design	✓		
Adaptive management	✓	✓	
Learning from exemplars / bright spots	✓		

Assistance for Refugees in Jordon; Improving take-up of information with farmers in India with PAD (VoxDev summary) and Increasing uptake of Family Planning in Cameroon.

[6] Novel data includes the use of non-traditional and data from remote sensors in the context of impact evaluations. A few examples include: water or air quality sensors (Rosa et. al. 2014); satellite data like nighttime lights (as in GIEs), financial transaction data (as in Ibarra, McKenzie 2017), mobile phone data (as in Blumenstock 2015). Despite their increasing use of these new data sources by governments, private sector and NGOs, their application in impact evaluations have been limited.

[7] I define policy modeling as tools that integrate rigorous evidence with local data to inform specific decisions in in a target setting. For example, the LiST tool integrates rigorous evidence on effect sizes of health interventions with country-specific data to estimate the impact of coverage change of health products on mortality in LMICs. This is all packaged in a publicly accessible tool, allowing users to easily draw in both descriptive data and effect size estimates (as well as vary the assumptions). There are many examples of COVID-related modeling too. Some academic papers contain one-off modeling or simulation exercises but tend to be retrospective and are rarely accessible to policymakers to localize to their own context. There appear to be even fewer examples of accessible modeling tools outside of health than within health.

Importantly, some of these tools may be more feasible than others. For example, creating a culture of performance management or continuous process improvement can be a challenging task of changing norms and processes. Some innovations, like geospatial data, require specialized software and skills.

Additionally, not all tools are immediately within the reach of policymakers. Based on interviews with stakeholders and desk review of the literature, figure 1 below plots selected tools based on their relative maturity within the field. Tools further to the right of the diagram are more immediately accessible to policymakers in East & West Africa. Appendix B lists key actors & a brief rationale for placement.

Important variation exists which this figure does not capture. Take the example of policy modeling. Policy modeling can take many forms, but generally aims to integrate global evidence with locally relevant data to inform specific decisions. This often happens by modeling the effect of a specific decision (allocating covid tests; scaling-up a specific intervention) on an outcome of interest (COVID rates; budgetary impact, cost-effectiveness, etc.). Policy modeling appears to be comparatively more mature in health than in other sectors. The health sector also has comparatively more African institutions using modeling. A 2020 study of cost-effectiveness analyses found the top 5 African institutions with published health economics CEAs included Makerere University (25 published studies), University of Cape Town (23) & University of Witwatersrand (20), Kenya Medical Research Institute (7), and Tanzania Ministry of Health (7).[8] Within Hewlett's EIP portfolio, IDinsight occasionally conducts policy modeling—for example, in recommending the scale-up of Mama Kits in Zambia.[9]

Outside of health, while many potential use cases exist, a brief search found relatively fewer policy modeling exercises, particularly by African institutions or outside of academic settings. One applied example of policy modeling outside of health is the modeling of the multiplier effect of cash transfer programs by the Transfer Project & UNICEF. This modeling integrates local data with estimated program effect sizes to estimate the general equilibrium effects of cash transfer programs on local economies.[10] Both UNICEF and Transfer Project pointed to the importance of these models in speaking directly to what was most relevant to certain policymakers (like Ministers of Finance) and successful captured the attention of policymakers in ways that impact evaluations alone had not. Another example of modeling I found is that in 2019, UN's Africa Institute for Economic Development and Planning (IDEP) in Senegal conducted a training targeted to African policymakers on modeling the impact of trade policy reforms; though it's not clear if this training was put into practice.[11] Other examples of modeling, not specific to African institutions, includes J-PAL and IPA's cost-effectiveness models; IRC conducting cost-modeling for prospective decision-oriented purposes; and CEDIL supporting a number of initiatives that involve modeling generalizability or transferability of interventions to new contexts.

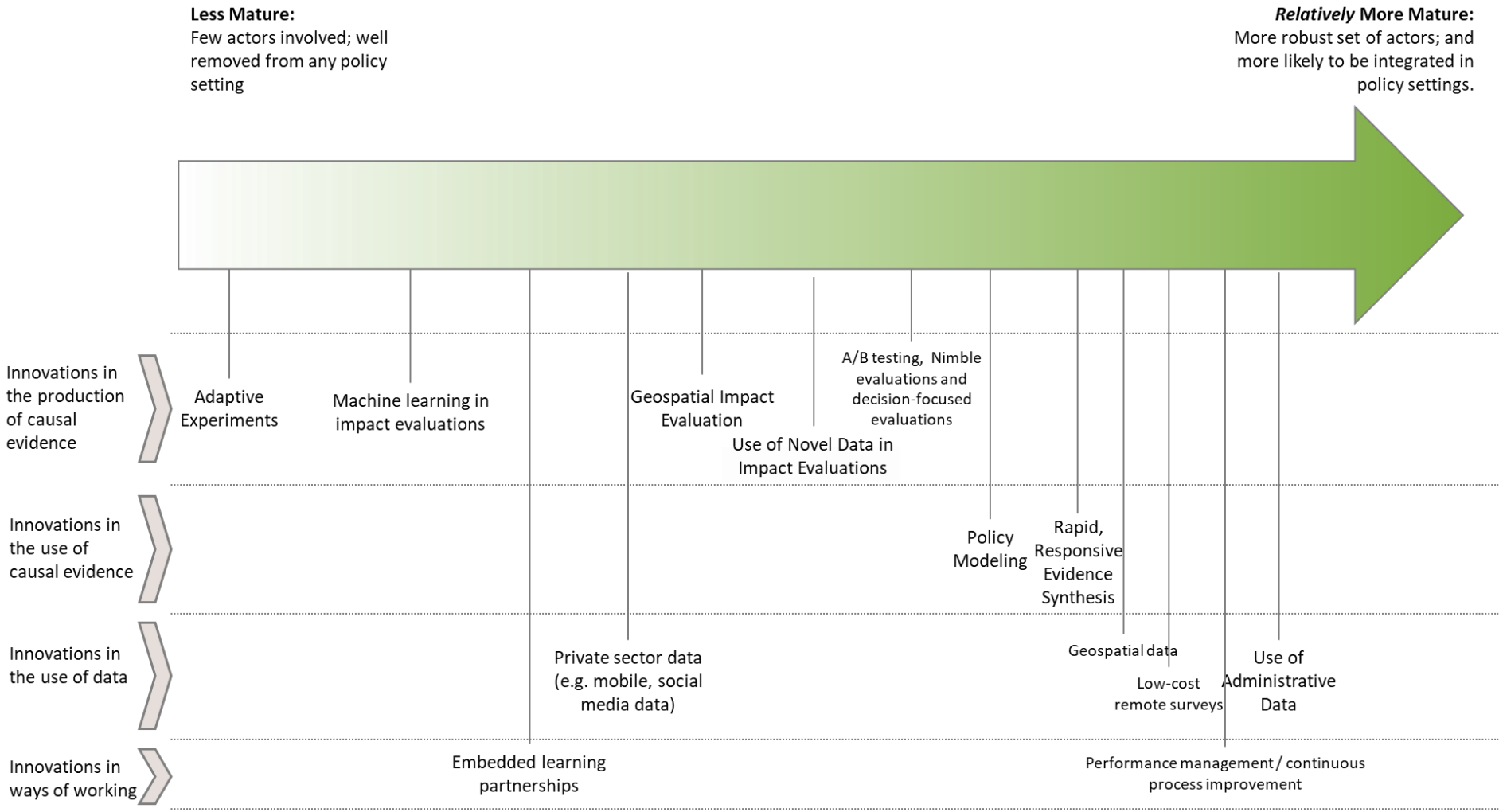
[8] Relying on this Panzer et. al's, study of published cost-effectiveness analyses is limited both because of the scope of the analysis (cost-effectiveness modeling is but one type of policy modeling) and because published CEAs may not always be intended to be informing specific decisions. For example, my understanding is that CHAI regularly conducts health economics modeling (including, for example, budgetary impact analysis) to support specific decisions, but that these analyses may not always be published. Panzer, A., Emerson, J., D'Cruz, B., Patel, A., Dabak, S., & Isaranuwatchai, W. et al. (2020). *Growth and capacity for cost-effectiveness analysis in Africa. Health Economics*, 29(8), 945-954. doi: 10.1002/hec.4029.

[9] See, IDinsight, Mama Kits Policy Brief; and Wang, P., et. Al.. (2016). Measuring the impact of non-monetary incentives on facility delivery in rural Zambia: a clustered randomised controlled trial. Tropical Medicine and International Health, 21(4), 515-524. doi:10.1371/journal. pone.0141455

[10] See, for example, Taylor, J., et. al.. "Local economy-wide impact evaluation of social cash transfer programmes." *From Evidence to Action: The Story of Cash Transfers and Impact Evaluation in Sub Saharan Africa* (2016): 94-116.

[11] See, Economic Modeling applied to Trade Policy Reforms in Africa

Figure 1: Maturity of selected tools in the Responsive Decision-focused Evidence Toolkit



*See **Appendix B** for a description of key actors and brief rationale for placement*

Appendix: Rationale for Figure 1 (Maturity of tools in the Rapid Responsive Toolkit)

Tools are sorted based on category and then by the order they appear within figure 1, starting with the least mature and progressing toward the most mature.

Tool	Key Actors Involved	Rationale for Placement
Innovations in the production of causal evidence		
Adaptive Experiments	Precision Development; WB SIEF; Athey's Social Capital Lab;	I know of only three adaptive experiments in development; only 1 of which is in Africa (a SIEF-supported family planning intervention in Cameroon)
Machine learning to identify varied treatment effects	J-PAL and IPA affiliates, Susan Athey, Guido Imbens, Esther Duflo	The methods (e.g. Wagner Athey 2018 and Chernozhukov, Duflo, Val 2017) are still very recent and are being applied almost exclusively in research settings.
Geospatial impact Evaluations	AidData, WB's Infrastructure Global Practice, Geo4Dev Initiative, Research Groups at Stanford (Burke, Lobell and co-authors), UC Berkeley (Blumenstock); Univ of Southampton/WorldPop (Andrew Tatum);	CEDIL/3ie identify 48 impact evaluations that used data from satellites in their recent Mar 2020 Big Data Systematic Map. Ten of these were in sub-Saharan Africa. GIEs are most prevalent in infrastructure (using nighttime lights) & environmental sustainability (using forest cover and related measures). African governments do not appear to be using GIEs yet and most GIEs appear to be in the context of research. AidData and WB's Transport practice are using GIEs in more applied, decision-oriented settings.
Novel data in Impact Evaluations	World Bank (DIME, Development Data Group), Geo4Dev (CEGA, New Light Technologies), 3ie	CEDIL/3ie's big data systematic map identifies 68 impact evaluations that draw on novel data sources. The majority of these (48) are GIEs.
A/B testing, nimble evaluations, and decision-focused evaluations[12]	IDinsight, Precision Development, Young1ove, IPA's MinEduLab, World Bank SIEF, IDinsight says smaller consultancies like Causal Design and Laterite are doing DFEs.	Multiple actors in the working on this in Africa. Being applied in Ghana, Kenya, etc.
Innovations in the use of causal evidence		
Policy modeling	Within Health: Institute for Disease Modeling, Tufts CEVR, IDSi, a few agencies (often within MoH) doing	Both within, as well as outside of health, policy modeling is far from established. Modeling exercises tend to be one-off or opportunistic/responsive to specific needs

[12] Nimble evaluations use lower-cost administrative data or shorter-term more output-oriented data to lower the costs of evaluation. Often, though not always, this can be combined with faster, more iterative experimentation. See for example, Fabregas, Kremer, and Schilbach (2019), Angrist, et. al., (2020); or SEIF's Nimble Evaluation portfolio.

Tool	Key Actors Involved	Rationale for Placement
	Health Technology Assessments (HTAs) in health. (see this survey), CHAI Outside of health: IDinsight, Transfer Project/UNICEF; J-PAL, IPA, IRC, UN's Africa Institute for Economic Development and Planning	
Rapid, responsive evidence synthesis	Within health: WHO's EVIPNet and its partners. Outside of health: ACRES, 3ie/IDinsight WACIE Initiative; Africa Center for Evidence; IDRC's RREP/PERLSS initiative.	Emerging network of practitioners in Sub-Saharan Africa. Evidence synthesis is, in some ways, a precondition for policy modeling, so therefore is placed relatively more mature on the chart.
Innovations in the use of data		
Private sector data (e.g. mobile data, social media)	Pulse Lab Kampala & UN Global Pulse;, DIAL (mobile data); Flowminder	Relatively few actors; important data governance questions that impede the use of these data and questions around bias in data. For example, DIAL pointed to only one example where they say mobile data was actually used for specific decisions.
Geospatial data (data from satellites; geocoded survey or admin data, highly granular estimates of wealth, mortality, etc.)	Satellite data; Geo4Dev; WB's development data group WorldPop; Grid3, IHME (for granular estimates of measures of health) Josh Blumenstock / Marshal Burke (granular estimates of wealth, farm yields; etc.);	A fair number of actors are either doing the work and supporting training for the work (see Davis' email). Yet, perhaps due to need for more sophisticated tools & specialized knowledge, it doesn't seem to be as mature the use of administrative data.
Administrative data	UNICEF,(and many others), DHIS2 (in health) Within the context of impact evaluations: J-PAL and SIEF	I have low confidence about where this should be placed. In health I'm aware of admin data being used quite regularly for decision-making (e.g. DHIS2, health financing data, etc.). Outside of health, I'm less certain how often admin data is used.
Innovations in ways of working		
Embedded Learning Partnerships	IDinsight, IPA's right-fit evidence unit; Premise runs pulse surveys Zenysis integrates admin data with other data layers to inform decision-making I suspect (but am not confident) that other advisory firms are also supporting governments in somewhat similar ways (e.g. Dalberg, possibly CHAI)	IDinsight says they have only two learning partnerships in Africa; I'm uncertain how prevalent IPA or other firms' work are.