SUPPLEMENTARY DATA

Appendix 1: Review of the methods used to measure the use of research

A literature review was conducted alongside informal interviews with monitoring and evaluation staff at USAID, other development organizations, and U.S. federal science agencies to understand different research measurement methods.

First, literature was identified through keyword searches including the terms; "research translation international development", "research impact", "assessing evidence use", and "research utilization" in journal databases. Further literature was then identified through a snowball method, reviewing literature cited in the original article to identify related literature.

Program managers and evaluation experts who support similar research for development programs shared assessment methods and directed attention to further tools and literature. These groups included the international program offices of the National Science Foundation, the US Department of Agriculture Forest Service, and the White House Office of Science and Technology Policy. Additionally, feedback sessions were held with USAID Bureaus of Global Health and Food Security and the Lab's Centers for Innovation and Technology programming. Informal conversations also occurred with staff from the International Development Research Centre (IDRC), the Think Tank Initiative, the Wellcome Trust, INASP, and representatives from UK Research and Innovation.

Jacobsson et al. (2014) describe the key outcomes of research as: spurring further research, scientific publishing, commercialization (innovation), education, networking, providing hard and soft infrastructure, and providing explicit guidance (informed recommendations). After gathering resources, the outcomes of interest for the Lab's research for development programming were identified as scientific capacity, pro-poor innovations, and evidence informed policy and practice as. Therefore, focus was placed on three measurement methods related to those outcomes: 1) Metrics of academic impact - peer review, citations, and publication impact factors, 2) Innovation metrics - patents, licensing income, and investment mobilization, and 3) Conceptual frameworks that qualitatively assess research use and the uptake of recommendations.

Academic Outcomes: *Peer Review and Bibliometrics*
Among academic researchers in high income countries, peer-reviewed research papers are seen as the 'gold standard' for measuring research productivity. Metrics include the number of papers published, the impact factor of the journal, and the number of subsequent article citations. Capturing and analyzing this data is referred to as bibliometrics, and is a widely accepted method for research evaluation (Bornmann, 2017; Greenhalgh et al., 2016). However, bibliometrics have limitations, particularly when used as a metric for measuring the outputs and outcomes of development-focused research funding in LMICs.

Bibliometrics often underestimate the productivity of researchers in LMICs and reflect disparities in scientific funding and infrastructure. High impact journals select for cutting edge and highly novel research. As a result, they are dominated by publications from researchers at

large, established labs with substantial budgets and easy access to supplies and equipment - all of which are not common in low-income countries. Publications in high impact journals are more frequently cited as well. As a result, a very small number of scientists account for the vast majority of highly cited papers. Illustrating this trend, a recent analysis using the Scopus database found that less than 1% of scientists accounted for more than 85% of highly cited papers (Ioannidis et al., 2014). Moreover, LMIC researchers are often incentivized to publish in local journals (which may lack impact factors or not indexed in international collections) because it is quicker and fulfills the requirements of local agencies who are often their only source of funding. Likewise, researchers may write policy recommendations, blogs, or editorials that are not captured by traditional bibliometric analysis.

New alternative metrics, or altmetrics, have been developed to capture mentions of peer-reviewed publications across social media, news, and other non-academic platforms such as blogs and editorials as mentioned above. Altmetrics are useful complements to traditional bibliometric methods, especially to measure the current conversation about a study, before it can be cited. There are several difficulties when applying altmetrics to the LMIC context, however. These methods still generally require a persistent identifier (such as a direct object identifier - DOI - or PubMed ID - PMID) for direct tracking in mentions, which many of the aforementioned LMIC publishing methods lack. Additionally, not all non-academic platforms of interest are indexed; for example, national news media may be indexed but more relevant local- or city-level news may not be. Moreover, 'policy mentions' often focus on publications from multinational or prominent scientific advisory organizations like the World Health Organization and the National Academies of Sciences, while neglecting policies from individual countries.

Peer review and bibliometrics are poor proxies for assessing the outcomes of international development research. Peer review provides a widely accepted measure of quality, but it often places greater value on methodological rigor and novelty than applicability and relevance (RAND Europe, 2006). Applicability and relevance is imperative for international development research. In addition, much of international development research is done in social science fields where the dominant bibliometric indices have lower coverage than natural sciences (RAND Europe, 2006) and impacts often depend on linkages with public sector, industry, and civil society rather than academia (Severinson, 2017). Alternative approaches that mix quantitative and qualitative measures, and take biases into account, are needed particularly for assessing the social benefits of research (Agarwal et al., 2016; Severinson, 2017).

Innovation metrics

Innovation is often measured in generation of intellectual property (IP), licensing income, patents, and technology transfer. These metrics are commonly used to measure the economic impact of innovation and are employed internationally. They are easy to understand and apply. Innovation measurement frameworks include the university technology transfer reporting standards compiled by the US-based Association of University Technology Managers (AUTM, 2019), the National Science Foundation's Science and Engineering indicators (National Science Board, 2018), and the Organisation for Economic Co-operation and Development's Oslo Manual (OECD/Eurostat, 2018). However, these frameworks have limitations, including comparability across countries, over-reliance on formal tech transfer, and gaps in measurement of innovations produced by small firms, social innovators, and the public sector.

Even among high income countries these metrics may not provide a fair comparison. For example, despite the fact that Europe and the US produce similar high-impact research publications, European universities produce much fewer industry licenses through technology transfer and those that are issued generate significantly less income (Conti and Gaule, 2011). This has been referred to as the 'European Paradox' by the European Commission and suggests that region-specific approaches and capacity for commercializing research vary widely, confusing this measure of impact. Similarly, Canadian universities have argued that technology transfer indicators under-represent their capacity for scientific innovation, conflating differences in the structure and reporting of university tech transfer with actual difference in innovation output (Sigurdson et al., 2015). For LMIC academic institutions, these limitations are magnified. Many do not have technology transfer offices or staff that specialize in developing IP from faculty research.

Reliance on formal tech transfer also limits the usefulness of these metrics - they are capturing only a small portion of innovation activity (Aldridge and Audresch, 2011; Fini et al., 2010; Perkmann et al., 2013; Sigurdson et al., 2015). Estimates show that even in the US, which has a relatively uniform structure for tech transfer across universities, two-thirds of commercialization takes place outside of tech transfer offices (Fini et al., 2010; Sigurdson et al., 2015). Aldridge and Audretsch (2011) demonstrate that these indicators do not capture most new ventures created by university scientists, leading to underreporting of university innovation. Furthermore, public sector innovation and the development of pro-poor products and technologies for LMIC markets largely fall outside the purview of standard innovation metrics (e.g. OECD, 2010).

Development organizations generally measure the impact of social innovations and pro-poor products and technologies in terms of the benefits they provide customers and other end-users. Targeted outcomes include lives saved and improved, enhanced nutritional status, reduced poverty, and improved health (IDIA, 2017; USAID, 2016b; consultations with M&E staff). To track intermediate results, development organizations often use metrics of demand and/or commercial potential such as sales, profits, and investments mobilization (IDIA, 2017; IRIS Metrics, 2019). For early stage research and development (R&D), these metrics present significant measurement challenges, however. Sales and profits, not to mention outcomes such as lives saved, are many years out from the initial research investment.

Conceptual Frameworks and Case-Study-Based Approaches

Other than conceptual frameworks described in 2.2, other structural methods, such as research quality frameworks and contribution analysis, have also emerged as approaches to evaluating research. These methods provide rich information about process, relationships, and results (both expected and unexpected) but they are labor intensive and do not generally provide a quantitative estimate of impact.

Many countries have national-level systems for evaluating the performance of research investments. These frameworks are quite diverse, but many combine quantitative indicators of academic impact with expert panels and case-based assessments (Coryn et al., 2007). The UK uses the Research Excellence Framework, an expert review process that considers a combination of academic and social benefits: research outputs, economic, social, and cultural impacts of

research, and vitality and sustainability of the research environment (REF, 2012). The main benefits of this approach is that it provides a holistic view of the value of research - integrating quantitative metrics with more subjective peer review and case-based approaches. However, some argue that these performance-based, national level systems have become too complex, counting everything, and benchmarking against an exhausting number of criteria (Coryn et al., 2007; Frederiksen et al., 2003).

Case studies are ideal for showcasing impact, for providing context, and when quantitative data are not available. They are labor intensive, however, and not as well-suited to critically assessing impact, which involves subjective perspective that is difficult to compare across cases (Penfield et al., 2014). Bornmann highlights these same weaknesses, but concludes that case studies are still one of the best methods for measuring social impact of research because they can capture complex phenomenon and help highlight the often diffuse path from research to impact (Bornmann, 2013).

Appendix 2: Case Studies

Provided below are additional case studies to illustrate the importance of the PPC framework in supplementing traditional research assessment metrics in research-for-development programming.

Case 4: Anti-malarials in Mali

Score: Implementation, 3: Influence, 3

This case study highlights how the PPC framework was complementary to bibliometric measures of research quality. The two methods captured different facets of the impact of the research; the PPC framework helped program managers understand the specific interventions and activities that influenced uptake of the research, while traditional publications and news/ blogs disseminated the results and recommendations based on the research.

Background/research project description

Malaria remains the number one killer in Mali (PMI, 2018) despite effective implementation of control measures such as insecticide treated nets and anti-malarial combination therapy. Prior to the PEER award, the researcher and his team showed that Seasonal Malaria Chemoprevention (SMC) reduces malaria infection and disease by more than 80% in Malian children, prompting the WHO to approve SMC as policy for Sahelian countries in March 2012 (Barry et al., 2018; WHO GMP, 2012). As PEER researchers, the team concentrated on finding the optimal delivery method to provide SMC throughout Mali. The team hypothesized that focusing on ease of use and cost-effective deployment were key to widespread adoption and real impact on the rates of morbidity and mortality.

Through the PEER program, the researcher leveraged his NIH partner's five year cohort study to track children exposed to seasonal malaria and given SMC treatment via two interventions. One method was traditional fixed-point delivery (FPD), which is standard and cost-effective but may leave some children unreached since complete coverage may not be achieved. A more robust door-to-door (DTD) strategy, where drugs are left for parents to administer through non-directly observed treatment (NDOT), as well as Directly Observed Treatment (DOT) by community health workers, may guarantee greater coverage, but at greater monetary costs. These delivery modalities were evaluated and compared with respect to coverage, or children reached, in resource-constrained, real-world scenarios.

Description of program and policy change

Coverage, defined as the proportion of children who received all three days of SMC treatment during the three monthly rounds, was based on information collected by interviews of families. The results showed that coverage was significantly higher in children who received SMC using DTD (74%) compared to FPD (60%). Coverage was similar in children who received SMC using DOT (68%) and NDOT (65%). While the difference between DTD and FPD seems small, when scaled across the region, this could amount to hundreds of thousands of children protected.

Armed with the knowledge that DTD delivery of SMC provides better coverage than FPD, and that NDOT had similar coverage to the more time- and resource-intensive DOT, the Government of Mali's Malaria Control Program and several NGOs working to deliver SMC changed their methodologies and policies for the roll-out of SMC during the 2017 season (Barry et al., 2018).

Explanation for score

The research generated from the PEER project impacted policy and programming at the national scale. The research results and associated intervention were scaled across Mali, leading to an implementation score of three and an influence score of three. The PEER researcher and his team worked very closely with Mali's Malaria Control Program and other NGOs working on SMC at the design phase of the project, presenting at conferences and workshops with key stakeholders during implementation. As interventions were designed related to coverage, the PEER researcher worked closely with USAID/Mali and USAID/Washington's Presidential Malaria Initiative office. These important partnerships allowed for a rapid pivot of partners working in this space as the researcher received and verified results.

Bibliometric analysis

Based on annual reports and USAID funding declarations, as of April 2019, nine publications have resulted from this PEER grant. The top journals were Lancet Infectious Diseases with three publications (Dicko et al., 2016; Sagara et al., 2016; Dicko et al., 2018) and the Malaria Journal with two publications (Mahamar et al., 2017; Cairns et al., 2015). Publication in a top journal led to high citation rates; Dicko et al. (2016) received 44 citations and Sagara et al. (2016) received 22 citations based on the journal's PlumX metrics. These papers also received some non-academic discussion, as measured by the free Altmetric bookmarklet. Dicko et al. (2016) was discussed in two news stories, two blogs, and 22 tweets (Altmetric, 2019c). Sagara et al. (2016) was mentioned in one news story, 22 tweets, and one policy document from the World Health Organization titled "The selection and use of essential medicines: report of the WHO Expert Committee, 2017" (Altmetric, 2019d). The most non-academic discussion, however, occurred for Dicko et al. (2018), which was too new to have many academic citations. This paper, titled "Efficacy and safety of primaquine and methylene blue for prevention of Plasmodium falciparum transmission in Mali: a phase 2, single-blind, randomised controlled trial", garnered mentions in 69 news stories from 66 outlets, 3 blogs, and 38 tweets (Altmetric, 2019e).

Case 5: Structured community polling and discussion leads to policy and program change in East Africa

For this project there were two distinct changes as a result of the research, one at the program level and the other at the policy level. Therefore, there are two separate PPC scores.

Score, Policy: Implementation, 3: Influence, 3
Score, Program: Implementation, 3: Influence, 3

This project highlights the difficulty of utilizing traditional bibliometrics to evaluate the impact of university-based development implementation projects. The PPC approach allows USAID to measure the development impact of university partners who are using and adapting previous development research in a new context to positively impact the lives of their communities.

Background research/project description
Deliberative Polling is a community engagement tool designed to help policy makers understand what's driving community behaviors and how to influence them for the public good. Deliberative Polling is characterized by an analysis of policy sentiment followed by a facilitated, in-depth, discussion of the problems and possible solutions with the community of interest. These steps are followed by another analysis of policy sentiment. This tool was used by researchers at Makerere University's Resilient Africa Network (RAN) to help understand community feelings around government forced resettlement policies in Uganda established to address the risk from landslides on the community, homes, and businesses.

Description of program and policy change
Following Deliberative Polling on resettlement, RAN drafted policy briefs describing what areas communities preferred for resettlement, then engaged with policy makers, including the office of the Prime Minister and District-level officials. The office of the Prime Minister adopted the policy recommendation of resettlement to areas preferred by the community. Follow-up requests found that over 90% of the people were able to be resettled in a safer area, resulting in a decreased loss of life. Additionally, as communities were resettled, officials realized that education was being affected in mountainous areas unaffected by landslides because schools were far from where people were being resettled. The school dropout rate in young kids was increasing. RAN convened another Deliberative Polling exercise where community support for smaller primary classes in higher altitudes was identified, and specifically for young kids until they were able to walk the long distance over the terrain. In response to this education polling, multiple donors and private organizations are considering funding the construction of new primary schools in the higher altitudes.

As a consequence of this and other resilience research, the United Nations Development program (UNDP) Uganda and the Humanitarian Leadership Academy have partnered with RAN to receive training from them in aspects of resilience measurement and capacity building around resilience programming. With this support, RAN has been able to train over 300 senior and mid-level officers in both Uganda and in East Africa.

Explanation for score
The implementation of the recommended resettlement policy by the office of the Prime Minister scores as both influence and implementation level 3 for a final score of 6. The new leadership training implemented by UNDP Uganda and the Humanitarian Leadership Academy reflects an influence score of 3 (multinational organization, large NGO) and an implementation score of 3 (program change has been implemented).

Bibliometric analysis
The impact of this project is not currently measurable via traditional bibliometrics; as of April 2019, it had produced three publications, but none have citations on Web of Science or discussions visible to Altmetric.

Case 6: The influence of research in understanding and mitigating arsenic contamination in Southeast Asia

For this project there were two distinct program changes as a result of the research, one at the local and the other at the national level. Therefore, there are two separate PPC scores.

PPC Score, Local: Implementation, 3: Influence, 1
PPC Score, National: Implementation, 2: Influence, 3

This project contributed to community-level adoption of good practices and facilitated progress toward national changes in water-use policy. The program and policy change framework was critical to understanding these results as bibliometrics and Altmetrics did not capture them. The research team reported only one research publication during the period of PEER support, which limited bibliometric analysis. Local level dissemination of results via news and blog (as reported by Altmetrics) did give some idea of influence, but it did not capture the breadth of community and policy connections. Of note, the publication does not cite USAID as a funder and illustrates additional difficulties in tracking attribution.

Background/research project description
Arsenic contamination affects 15 million people in Vietnam, second only to Bangladesh. The negative health effects include changes in skin pigmentation, blackfoot disease and cancer. Vietnam's Red River Delta has extremely high levels of arsenic contamination, 20 times the safe limit as set by the WHO (WHO, 2001) and the Vietnamese researchers conducting this project have been documenting the extent and nature of groundwater contamination for many years in collaboration with scientists from around the world. Through the PEER Program, the researchers were able to collaborate with National Science Foundation (NSF)-funded researchers studying hydrological processes at the molecular scale. Together, the research team is documenting how molecular interactions influence large scale geo-chemical dynamics and how practices, like groundwater pumping, affect water availability and quality throughout the region.

Their research showed that assumptions about where groundwater came from were often wrong and the movement of the water through the system was dynamic. Groundwater pumping in Hanoi was pulling arsenic into water sources in the Red River Delta, while carbon infiltration around landfills was pushing arsenic into sediment that fed water treatment facilities. With rapidly changing geochemistry, it was apparent that water needed to be regulated differently.

Description of program and policy change
The research team was able to influence policy and practice at the commune (village) level, plus contributed to a national dialogue that informed water use policy and regulation nationally.

At the local level, the main behavior change was that villages along the Red River in the area of Van Phuc no longer dig tube wells, with many villages prohibiting this type of well due to susceptibility to arsenic contamination.

On the national level, the government has closed some water treatment facilities in areas prone to arsenic contamination and is moving from extracting groundwater to treating surface water in certain areas, particularly around Hanoi.

The years of research on the dynamics of arsenic contamination in the region was essential to change in policy and practice. PEER research alone did not change policy. The research was also conducted at a time when the political context facilitated relatively swift change. Water was not seen as a controversial issue, and control was centralized at the level of the national government. At both the local and national levels, the PEER research team played an important role as a broker. They had strong connections with other academics, village and municipal authorities, and relevant national ministries. Many of these relationships pre-dated the PEER research project and were long term relationships. During the project, the local research team had continuous interactions with local officials and hosted workshops and disseminated water quality information to villagers. At the national level, the research team organized stakeholder convenings, including a large, international conference around arsenic.

Explanation for Score
In scoring this project, there were program and policy changes at both the local and national levels. At the local level, this project was rated a 3 on implementation and a 1 on the influence axis reflecting the change in behavior and policy at the commune level. At the national level, the project was assigned a 2 on implementation and a 3 on the influence axis, demonstrating national level commitment to research-based water use policy. The score of 2 for implementation reflected that the government had demonstrated commitment by adopting research-informed policy in certain sectors and regions, but had not fully adopted this approach nationally or integrated all the needed infrastructure to do so. The scoring framework enabled program managers to track the pathway and actors who translated the research into action, helping document the underlying drivers of research utilization and how the PEER program contributed.

Bibliometric analysis

The lead PI reported one academic article connected to the PEER research, although USAID funding is not acknowledged in the paper itself (Stahl, 2016). As of April 2019, this paper was discussed in 9 news articles and 1 blog post, which indicates moderate conversation via Altmetric (Altmetric, 2019f), and has been cited 13 times with a percentile of 10% (cited more than 90% of papers in its subject area and year) via Web of Science.

Case 7: Collaborative research & development for designing interventions in African food markets

Score: Implementation, 3: Influence, 1 (market improvements)
Implementation, 2: Influence, 1 (tax pol

This project highlights student-driven projects that use research to drive an implementation-based approach to understanding and solving discrete development challenges. As such, the subsequent development interventions have a direct impact on the community that isn't publicized in a manner that is easily quantified by traditional bibliometrics. This project has led to important programmatic changes at the community level.

Background research/project description
Throughout Africa, and particularly in Malawi, urban food retail operates through both formal markets such as grocery stores, and informal markets such as open-air spaces where individual vendors sell their wares. Because these informal markets operate on a small scale with limited oversight, appropriate food and occupational safety infrastructure is often lacking. Likewise, little is known about how the informal markets interact with other, more formal, food systems. HESN's Global Center for Food Systems Innovation (GCSFI) at Michigan State University (MSU) has been working to address informal market safety, operation, and interaction with the formal food systems.

The Frugal Innovation Practicum (FIP) trains students to engage in what they call "action research". Rather than serving as traditional research assistants, FIP students engage in international development through a collaborative, student-centered service learning program that connects MSU students with Malawian students at Lilongwe University of Agriculture & Natural Resources (LUANAR). Students from both universities work together with the community to help identify safety and operations issues in the informal food markets. Then, they work together to raise funds for quick turnaround interventions through a crowdfunding platform that draws on donations from both university communities. These interventions are complemented by more traditional research analyses of how informal urban markets function.

Description of program and policy change
The FIP students partnered with the Lilongwe city council and the Director of Commerce to identify and implement interventions to improve the market safety and function. These include co-funded (crowd-sourced and public) installation of improved water delivery, sanitation,

parking infrastructure, and lighting. Additionally, the City of Lilongwe has drafted a policy that would reinvest 25% of the retailer taxes back into the markets, as a means to further support market safety and effective operation.

Explanation for score
The infrastructure and safety improvements that occurred as a result of the student's research have been implemented at a local level, resulting with implementation score of three and influence score of 1. The proposed change in retailer tax policy scored a two for adopter commitment and a one for influence. USAID was not able to verify whether the tax policy has been implemented.

Bibliometric analysis
This project did not result in any trackable publications, thus the impact of the project is not currently measurable via traditional bibliometrics. The larger GCSFI award did produce a substantial body of publications (21 journal articles and two conference proceedings as of April 2019) including several highly cited publications that generated policy, news, and social media mentions. However, the results of community research projects, such as the one described above, were not detected in this analysis because they were not directly connected to publications.