Research Report

How Much Evidence Is There Really? Mapping the Evidence Base for ICTD Interventions

Annette N. Brown
Hannah J. Skelly

FHI 360, USA

Abstract

To identify impact evaluations on the effectiveness of ICTD interventions, we conducted a systematic search and screening that identified 253 studies. We present our results in an evidence map that shows the amount of evidence for each of 11 intervention categories across nine development sectors. We find large amounts of evidence for some intervention categories and little to no evidence in others. Roughly 80% of the ICTD impact evaluations use randomized assignment, suggesting a relatively low risk of bias across the evidence base. At the same time, roughly 80% evaluate pilot implementations instead of programs, raising questions about how useful the evidence is for informing programs at scale. Less than 20% of studies report costs, limiting our ability to assess cost effectiveness. We make four recommendations to improve the evidence base.

Keywords: information and communication technology, impact evaluation, evidence map, cost effectiveness, developing country, systematic search, pilot

Introduction

In the last two decades international development policy makers and practitioners have increasingly embraced information and communication technology for development (ICTD) products and approaches. Fueled by the rapid spread in Internet access and the diffusion of mobile technologies, the public discourse on ICTD has championed the positive transformation these new tools can bring to development outcomes. The rhetoric on the impact of ICTs has been imbued with overwhelming optimism on the potential of these new tools to accelerate and amplify development outcomes in nearly every sector. In 2015 the development community enshrined faith in the power of ICTs within the new sustainable development agenda. The agenda included individual Sustainable Development Goals (SDGs) focused on ICTs. Several movements have since emerged that advocate for ICT use to achieve each of these goals more broadly.[1]

ICTD enthusiasm remains strong, but there are mixed views on whether there is adequate high-quality evidence that supports the effectiveness of these interventions. As we have seen in our work with implementers and funders, many ICTD practitioners call on a limited but growing number of studies in a specific sector, such as m-health or computer-assisted learning, to inform intervention designs. Indeed, the volume of publications on ICTD has grown, on average, 39% per year from 1999–2008 (WSIS, 2016). However, failure rates of ICTD interventions can be high. The World Bank's Independent Evaluation Group finds that only 30–60% of World Bank ICTD projects achieved their objectives (World Bank, 2011). The United Nations Conference on Trade and Development (UNCTAD) concludes that "many data gaps exist in the area of ICT impacts, particularly with

regard to developing countries" (UNCTAD, 2011, p. 17). UNCTAD (2011) explains that much of the evidence from developing countries comes from local case studies rather than statistical analyses and that developing countries lack data for core ICT indicators, which provide the context for studying impacts. In this article, we address the question at the center of these mixed views: How much good evidence exists about the impacts, or effectiveness, of ICTD interventions?

To answer this question, we present an evidence map of impact evaluations of ICTD interventions in low- and middle-income countries. Evidence maps, also called systematic maps, or evidence and gap maps, have these objectives: "to systematically and transparently describe the extent of research in a field, to identify gaps in the research base, and to provide direct links to the evidence base" (Clapton, Rutter, & Sharif, 2009, p. 11). Systematic mapping methods are derived from systematic reviewing methods (search, screen, assess, and synthesize), but stop short of synthesizing the evidence from the included studies and often do not include a critical assessment for each study mapped. Whereas systematic reviews address a narrow or specific research question, evidence maps catalog thematic collections of research (Snilstveit, Vojtkova, Bhavsar, Stevenson, & Gaarder, 2016), which typically include much larger numbers of studies than a systematic review. For example, our ICTD map includes 253 studies, while the average number of included studies in the ICT systematic reviews in a recent special section of *Information Technologies & International Development* was 11 (Samarajiva, 2018). By revealing clusters and gaps in research, evidence maps inform research agendas for both syntheses and new primary studies and may be used directly for policy briefing (Clapton, Rutter, & Sharif, 2009; James, Randall, & Haddaway, 2016) or for conducting rapid reviews, as we discuss in the Conclusion.

The map in Table 6 builds on prior work, which mapped the impact evaluation evidence base for interventions using science, technology, innovation, and partnerships (STIP) in low- and middle-income countries (Sabet, Heard, & Brown, 2017; Sabet, Heard, Neilitz, & Brown, 2017).[2] The STIP map was built using a framework of intervention categories within the four STIP groups and outcome sectors. The STIP framework catalogs outcomes according to the sector the indicators reflect. Our framework for ICTD uses the same framework that Sabet, Heard, and Brown (2017) used for "technology" in the STIP research, which includes 11 distinct intervention categories that cover digital and data solutions using mobile and Internet technologies as described in Table 1. This framework was designed to focus on interventions using or promoting the use of mobile and Internet technologies such as digital finance and digital literacy, rather than on technologies themselves such as telephones and computers.

The evidence map shows how many studies exist for each of the intervention categories by the types of outcomes measured and the sectors in which outcomes are measured.[3] Appendix A online[4] provides the references for each study on the map by intervention category, so readers can easily access the full studies. We also assess the depth and breadth of the mapped evidence base for ICTD interventions by analyzing data coded for each study, including whether the studies evaluate pilot implementations or programs at scale, whether the evaluations explore questions of equity or provide sex-disaggregated effect estimates, and whether the evaluations provide information related to cost effectiveness. These or similar cross-cutting features of impact evaluations are often coded in evidence maps (see, e.g., Rankin et al., 2015; Rankin et al., 2016) and are useful for understanding the policy relevance of the research cataloged in a map.

Our map focuses on *impact evaluations,* defined as empirical studies that measure the effect of an intervention against a counterfactual, where the counterfactual may be constructed experimentally or statistically (see Gertler, Martinez, Premand, Rawlings, & Vermeersch, 2016, for an overview of impact evaluation practice and explanations of impact evaluation methods). In other words, our map presents studies that examine the outcomes of ICTD policies, programs, or pilots as compared to the outcomes that would have transpired in their

---

2. *Annette Brown was a co-principal investigator for the STIP study.*
3. *We say that we map "evidence," not "studies," because studies can appear in multiple cells based on presenting multiple effect sizes (i.e., multiple pieces of evidence) and because the inclusion criteria for studies are those that provide counterfactual-based evidence of intervention effectiveness.*
4. *All online appendices are available at https://sites.google.com/view/anbrowndc/online-appendixes-and-supplementary-information/how-much-evidence-is-there-really-online-appendices*

Table 1. Intervention Categories in the ICTD Evidence Map.

Intervention category	Description
Digital infrastructure	Facilitating general access to digital technology through improved digital infrastructure
Policy and regulation for digital services	Implementing laws and regulations to facilitate access to or use of digital technologies
Digital literacy	Delivering dedicated training or instruction to improve individual capability to use the Internet or mobile devices
Digital inclusion	Increasing individual access to the Internet and mobile devices, often focusing on marginalized groups
Digital finance	Using mobile technologies for financial transactions and services
e-Governance	Facilitating the provision of government services and communication between government agencies and the public using digital technologies
Digital identity	Digitizing personal identification systems
Data systems	Using digital technology to improve data collection, management, and use
Digital information and individual services	Using digital technology to disseminate information, provide individual services, or change, or "nudge," behavior (interventions in this category that meet the digital finance and m-health definitions not coded here)
Technology-assisted learning	Using the Internet or mobile devices to deliver instruction and learning
Mobile health (m-health)	Using mobile and wireless devices to provide medical care

Note: The definitions here are slightly revised from those used in Sabet, Heard, and Brown (2017). Brown made minor edits for the present study to better reflect the actual screening for STIP and, thus, better guide the screening for the present map.

absence. Impact evaluations measure attributable net outcomes, which are important for establishing the cost effectiveness of interventions, as cost must be compared to a unit of effect to determine cost effectiveness. Other evaluation methods, for example, formative evaluation, performance evaluation, and process evaluation, are useful for answering a wide variety of additional evaluation questions, but are beyond the scope of this map.

To develop this evidence map, we conducted a systematic search-and-screening process covering 14 academic databases and 27 websites. Our inclusion criteria included studies that used a counterfactual method to quantitatively estimate an effect size, evaluated interventions in one or more of our 11 categories, were conducted in a low- or middle-income country, and were published in English in 1990 or later. We identified 253 studies that met our inclusion criteria, 32 of which reported effect sizes across more than one intervention category and seven for more than one sector. Compared to evidence and gap maps for other themes in low- and middle-income countries, this is a large number.

The evidence map is intended as a resource to help readers find the existing evidence most relevant to their needs, as well as a way to identify evidence clusters and gaps for ICTD. In this article we do not synthesize the results from the included studies. That is, we analyze *how much* we know about what works in ICTD, but not *what* we know. Evidence synthesis requires a careful analysis of a set of studies that measure similar outcomes for relatively homogeneous interventions (Samarajiva, 2018). As discussed in the Conclusion, we will conduct such synthesis for subsamples of the evidence map in future articles. Here, we provide the reader with a map and additional information about the full evidence base, covering 253 impact evaluations.

In the next section we present the methods used to develop the map. In the following section we present the results, starting with the flow chart detailing how our screening resulted in the final number of included studies, then the map itself, and then the analysis of the evidence base.

Table 2. ICTD Inclusion Criteria Described According to PICOS.

Participants	Individuals, households, firms, organizations, and governments in low- and middle-income countries
Interventions	11 intervention types defined in Table 1
Comparisons	The intervention may be compared to a situation without the intervention, to a standard of care, or to another intervention, whether ICT or not
Outcomes	Development outcomes in any of nine sectors listed in Table 3; also categorized by level of measurement
Study design	Impact evaluations, defined as using an experimental or quasi-experimental method for establishing a counterfactual against which to measure the effect of the intervention (see Table 3)

Materials and Methods

Evidence mapping, or systematic mapping, uses the first two steps in systematic reviewing methodology: systematic search and screening. The objective is to identify the existing research on the map's theme that meets the inclusion criteria. As with systematic reviews, the criteria can be described using PICOS[5] (participants, interventions, comparisons, outcomes, study design), but the distinctions are typically much broader, as maps are intended to collect large amounts of research, whereas reviews are intended to answer specific research questions. Table 2 presents the PICOS descriptions for this map.

The interventions and outcomes that help define the map also form the framework upon which the map is built, with the interventions as rows and the outcomes as columns following the Snilstveit et al. (2016) methodology. Each study that measures an effect for a particular cell in the map (i.e., that measures an effect for that intervention type in that outcome category) is coded into that cell. Studies can appear in more than one cell if they evaluate more than one intervention type, measure effects in more than one outcome category, or both. Note that this systematic mapping approach differs from mapping exercises that are sometimes also called scoping studies. Those studies define a theme or research question, conduct the search and screening, and then build a framework or classification scheme based on the research identified. See Nascimento and da Silva (2017), as well as Medaglia and Zhang (2017) for examples from ICT research. The objective for scoping study systematic maps is to first view the collection of research and then classify it, whereas the objective of an evidence systematic map is to first establish a framework and then populate it. Both approaches seek to describe a collection of research, but the evidence systematic mapping also helps researchers and policy makers find the individual studies they need.

As noted above, this ICTD evidence mapping work builds on prior work cataloging the impact evaluation evidence base for interventions using science, technology, innovation, and partnerships (STIP) in low- and middle-income countries (Sabet, Heard, & Brown, 2017; Sabet, Heard, Neilitz, & Brown, 2017). The framework uses the 11 intervention categories in Table 1 and the nine sector categories in Table 3. The interventions and outcome sectors were selected in collaboration with the sponsor, United States Global Development Lab, and other stakeholders; the sectors generally reflect the sector distinctions used by the United States Agency for International Development. The team conducting that research identified 220 studies in the technology group of interventions by conducting a search for the period from 1990–June 2016.

Searching

We started our search with the 220 included technology studies from the STIP map. We updated the collection by conducting a new search to capture studies published or indexed since the STIP team's search. Our original

---

5. *Although the acronym PICOS is commonly used, researchers compile it from slightly different sets of words. We have seen: participants, interventions, context, outcomes, and study design; problems, interventions, comparisons, outcomes, and study design; participants, interventions, comparisons, outcomes, and settings, as well as the one we use here. The various configurations generally cover the same principles.*

Table 3. Sector Categories.

Sectors
Education and academe
Global health
Democracy, human rights, and governance
Agriculture and food security
Crises and conflict
Economic growth, finance, and trade
Environment and climate change
Water and sanitation
Energy

plan was to use the same STIP search strategies for technology (see Sabet, Heard, & Brown, 2017, for detailed strategies), except for limiting the publication dates from 2015–present. Although the STIP authors conducted their search in mid-2016, we started the new search from 2015 to ensure that we captured earlier publications that may have not been indexed until a later date.

As we started the search, we realized that the shorter search period allowed us to broaden the search strategy and more carefully tailor the search strings for individual databases. Our core search strategy included strings for four categories of study features and required at least one hit per category. The four categories are: ICTD topical terms; impact evaluation, program evaluation, and systematic review terms; low- and middle-income country identifiers; and publication date. We also restricted our search to English-language publications due to labor constraints, although we recognize that there may be studies that meet our inclusion criteria published in other languages. We present our detailed search strategies in Appendix B online.

We searched 14 academic databases and 27 websites listed in Appendix C online, starting in August 2017. We also conducted a snowball search using the references of all newly included studies and all the included studies in the systematic reviews identified in the search. After the initial search and screening, we learned of an error in the implementation of our search for the databases accessed through EBSCO Host. Because of this error, we could not be sure we correctly captured all the relevant studies conducted in low- and middle-income countries, so we re-searched the EBSCO databases. Figure 1 in the next section combines the hits and screening results from both searches.

Screening

After removing duplicates, we performed title, abstract, and full-text screening, according to the protocols in Appendix B online. For the new studies, our inclusion criteria were these: the study evaluated an intervention in one of the 11 ICTD categories described in Table 1, the intervention was implemented in a low- or middle-income country as classified by the World Bank in 2017, the study used one of the four identification strategies adapted from Waddington et al. (2017) presented in Table 4, and the empirical analysis used a sample size of at least 50 and, if clustered, at least four clusters. Publication types included journal articles, working papers in a series available online, institutional reports available online, books or book chapters, dissertations, and draft papers available online. As part of the screening for working papers, institutional reports, dissertations, and draft papers, we re-searched to ensure we included only the most recent version.

We also rescreened the 220 technology studies from the STIP map. As shown in Figure 1, we excluded some of those studies for the ICTD map. In a small number of cases, these exclusions represent errors or oversights in the original screening, for example, the publication was a research protocol, not a completed study. Some exclusions arose from elimination of working paper or draft paper versions of studies for which we had a more recent version, either already in the set of 220 or from the new search. And some exclusions arose from the decision to exclude studies that test the feasibility or effectiveness of using mobile devices to collect data, where the outcomes measured are comparisons of different methods of data collection.

Coding

For all the included primary studies after the full-text screening, we coded a set of variables to populate the evidence map and analyze the evidence base. We recorded the country(ies) of study, the publication type of the document(s), and the intervention(s) evaluated, as well as the standard bibliographical information. We coded

Table 4. Empirical Identification Strategies Used for Screening and Coding Studies.

Code	Description
1	Randomized controlled trial: randomized individual assignment (at the unit of analysis or to clusters) to treatment and control by researchers or decision makers
2	As-if randomized assignment, including natural experiments, instrumental variables (based on a natural experiment), encouragement designs, and discontinuity designs (RDD and ITS). May also be clustered
3	Difference in difference or fixed effects, i.e., nonrandomized assignment with control for unobservable time-invariant confounding at the unit of analysis. This code is used if the study uses difference in difference and does not have random or as-if random assignment
4	Single difference study with controls for observables, including matching

Table 5. Variables Coded for Included Studies.

Label	Description
Program name (if provided)	The name of the program evaluated, used to identify multiple studies of the same program
Pilot	*Yes* if the study measures the impact of an intervention implemented as a pilot or trial; *No* if the study evaluates a program otherwise being implemented
Equity focus	*Yes* if the intervention has an equity focus and the study explores outcomes related to the equity focus or if the study measures outcomes for equity-related groups (equity groups include but are not limited to conflict-affected, differently abled, indigenous groups, elderly, refugees, orphans and vulnerable children, ethnic minorities, and sexual minorities)
Sex disaggregated	*Yes* if the study reports counterfactual-based effect estimates separately for men and women
Cost information	*Yes* if the study reports cost information about the intervention
Identification strategy	Coded according to Table 4

the sectors categories in Table 3 as well as three outcome categories based on level of measurement: individual and household level, organization level, and community or societal level. To code a study as reporting an effect for a given outcome category or sector, the study must report an effect size for that outcome and in that sector measured against a counterfactual. We also coded additional study features listed in Table 5. We discuss these features in the Results section.

Note that in addition to rescreening the 220 studies from the STIP map for inclusion in the ICTD map, we recoded the study information from the 193 studies included in the ICTD map. We slightly redefined the categories to emphasize the intervention mechanisms. Impact evaluation evidence is particularly useful for understanding what activities or mechanisms work to achieve different kinds of outcomes. We wanted to ensure that we coded studies accordingly. Thus, for example, if an intervention used mobile technology to teach individuals healthy behaviors, we re-coded it as technology-assisted learning instead of as m-health (the provision of medical care), even if it measures health outcomes. The sector in which the outcomes are measured would still be coded as health.

The digital information and individual services category is the generic intervention category for services provided by short message service (SMS) or interactive voice response. According to their basic definitions, digital finance and m-health would be included as digital information and individual services. To make the map more useful, the STIP team included these two subcategories separately. Thus, what is captured under digital information and individual services are all other such interventions. We maintained this convention.

After completing the evidence map, we began one of the subsequent rapid reviews as described in the Conclusion. For that review, we identified more recent versions of some studies (e.g., published versions of
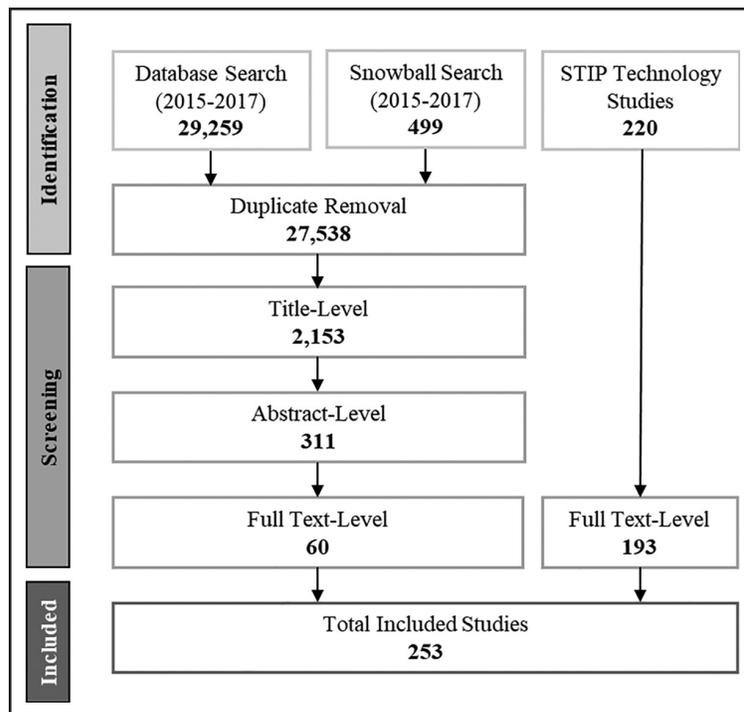
Figure 1. Flow diagram for search and screening results.

working papers, later drafts of draft papers) in that subset. We also corrected a few interventions and outcomes based on the in-depth review. We have included these updates in the current map and analysis.

The title and abstract screening was conducted using EndNote X7.0.2 by one researcher for each search hit, consulting with the second researcher for uncertain cases. Full-text re-screening of the studies included from the STIP map was conducted by the common author for the two teams. The studies included after full-text screening by one researcher were re-screened by a second researcher. One researcher coded each of the included studies, with the second researcher reviewing the full-text coding. For the purpose of the map, we did not conduct risk-of-bias assessments on the included studies beyond using identification strategy as an inclusion criterion. We mirrored our search strategies and screening protocol roughly on the published methods used for the technology group in the referenced STIP evidence map. We did not register or post a separate protocol.

Results

Figure 1 presents the flow diagram for the search and screening results. Our database and website searches (both initial and corrected) yielded 29,259 hits. We also screened 499 additional records from the references in systematic reviews and included studies from our screening. After removing duplicates and screening, we identified 60 new primary studies between 2015 and September 2017. We combined these with the 193 primary studies from the rescreening of the 220 STIP technology studies to create the set of 253 studies for the ICTD evidence map.

Table 6 presents the evidence map for the 11 intervention categories and nine sector categories. The number in each cell indicates the number of studies that report an effect size measured against a counterfactual for that intervention on an outcome in that sector. Thus, individual studies are counted in more than one cell in the map if they evaluate more than one intervention or a combined intervention, or if they measure outcomes in more than one sector. For example, a digital finance program could include a digital literacy component, and the evaluation might measure food security outcomes and changes in income or consumption. Such an evaluation would be counted in four cells. In the case of the ICTD map, however, there are limited cases of studies appearing in more than one cell. The map includes 297 occurrences from 253 studies. For comparison, an evidence map for transferable skills programming for youth in low- and middle-income countries covers 90 primary studies and has 609 occurrences in the map's cells (Rankin et al., 2015).

Appendix A online presents the references for all included studies according to the 11 intervention categories. Appendix D online provides the references and hyperlinks for the included studies according to the cells in Table 6.

Table 6. ICTD Evidence Map: Number of Impact Evaluations for Each Intervention Category Measuring Outcomes in Each Sector.

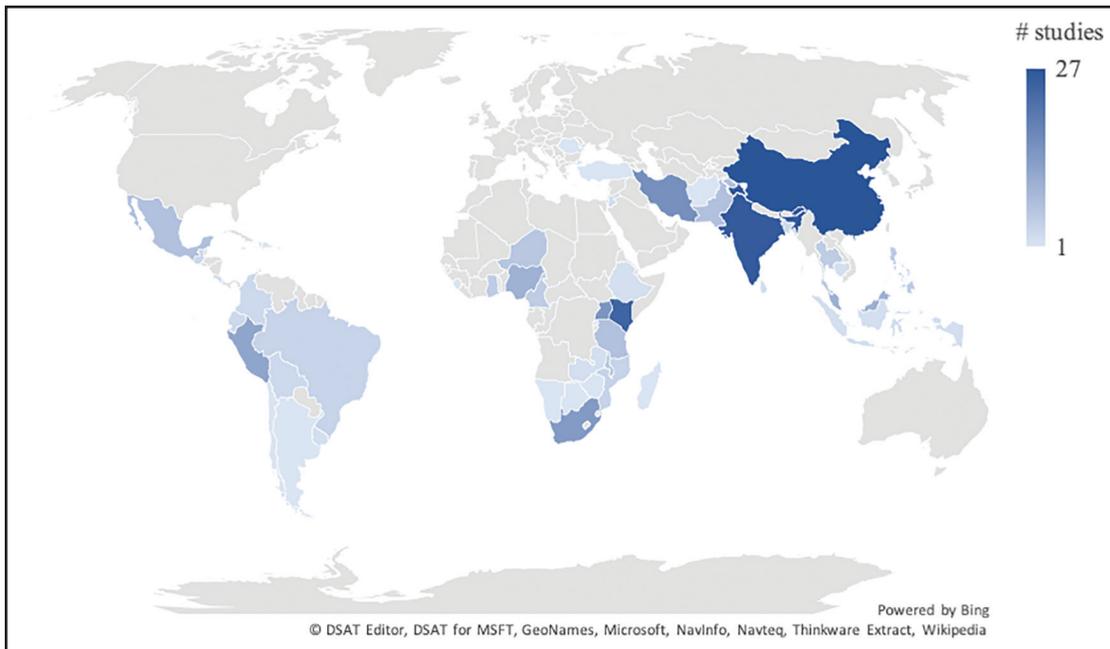| Intervention Categories	Education & Academe	Health	Democracy, Human Rights & Governance	Agriculture & Food Security	Crisis & Conflict	Economic Growth, Finance & Trade	Environment & Climate Change	Water & Sanitation	Energy
Digital infrastructure	1			2		1			
Policy and regulation for digital services									
Digital literacy	3	1	1	2		1			
Digital inclusion	3	4	1	4		1			
Digital finance			1	4		8			
e-Governance	1		5	2		6			
Digital identity		2		1		2			
Data systems		11		1		2			
Digital information services	4	4	5	13	1	6			
Technology-assisted learning	18	24							
Mobile health	1	146		1		1	1	1	

Figure 2. Heat map of number of ICTD impact evaluations by country.

When reading the evidence map, it is important to remember that the numbers in each cell only reflect the amount of evidence, not whether the effect sizes in the studies are positive, negative, or null. Among others, the map shows large clusters of evidence for the effect of digital finance on economic growth, finance, and trade outcomes; digital information services on agriculture and food security outcomes; technology-assisted learning on education and academe outcomes; and m-health interventions on global health outcomes.

Results by Study

The 253 impact evaluations represented in the ICTD evidence map were conducted in 49 low- and middle-income countries. The heat map in Figure 2 shows the prevalence of studies in different countries. The heat map shows that China, India, and Kenya account for many of the ICTD impact evaluations, with many others from Peru, South Africa, Uganda, and Iran. Figure 3 displays the share of impact evaluations by region. Despite the many ICTD impact evaluations from China and India shown in Figure 2, the plurality of the 253 impact evaluations—42% of the total—examine interventions in sub-Saharan Africa. The regions with the least evidence on ICTD are the Middle East and North Africa (8% of the total number of studies) and Europe and Central Asia (one study out of 253).

Figure 4 shows the dramatic growth in the number of impact evaluations of ICTD interventions from 2006–2016, reaching 73 studies published in 2016 (the last full year covered by our search). Our results may be biased toward more studies in 2015 and 2016 because our updated search used broader search strings; however, the trend is clear even up to 2014. Most of the impact evaluations identified through the search and screening are published in journals.

Table 7 presents information about the designs used in the impact evaluations in the evidence map. Roughly 80% of the studies used random assignment as the identification strategy to determine the causal effect of the intervention on the measured outcomes. That is, 80% were randomized controlled trials (RCTs) or clustered RCTs. This share of RCTs is greater than the share of RCTs—roughly 60%—among all development impact evaluations (Cameron, Mishra, & Brown, 2016). Also, roughly 80% of the studies evaluated the impact of a pilot implementation, that is, an intervention implemented as part of the study, while 20% evaluated the
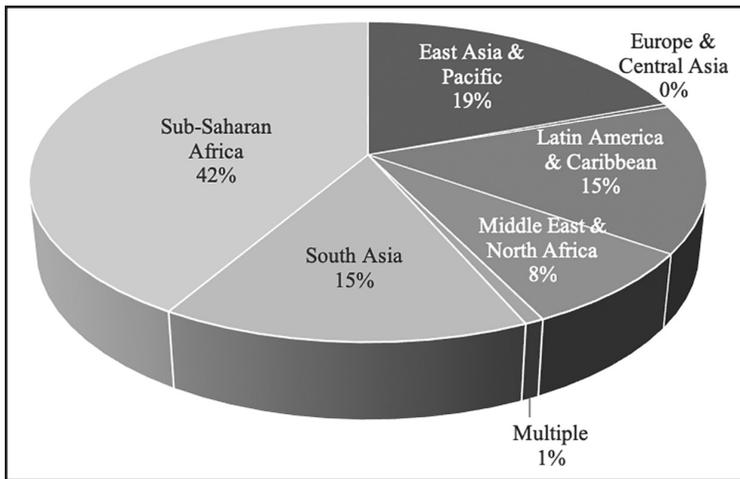
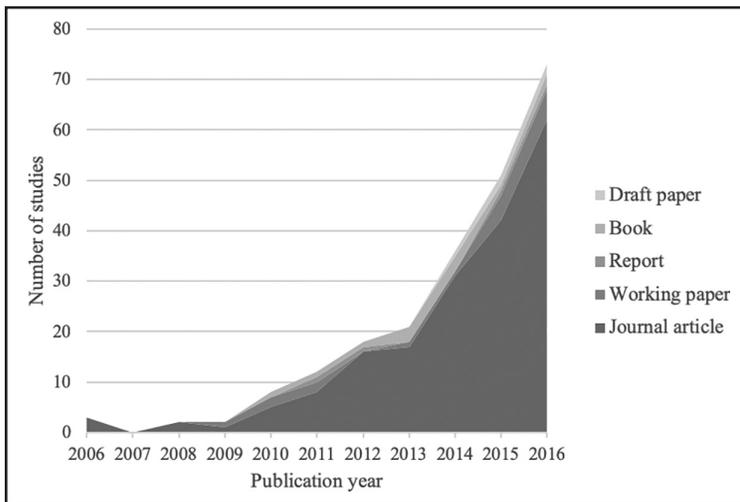*Figure 3. Share of ICTD impact evaluations by region.*

Figure 4. Growth in the number of impact evaluations published each year by publication type.

impact of a program that was otherwise in place or to be implemented. These design choices overlap, with more than 90% of the RCTs testing the impact of pilot interventions. We provide some possible explanations and implications for a large share of RCTs and pilot studies in the Discussion section.

Results by Intervention Category

The 253 studies appear 289 times when counting studies by intervention type. Thirty-two studies present evidence for two intervention categories, and two studies present evidence for three intervention categories. In a few cases, the studies present separate evidence for the different intervention types. In most of these cases, however, the two intervention types are combined in a single program so that the evidence produced is only for the interventions when combined. The study by Aker, Boumnijel, McClelland, and Tierney (2016) is an example where two intervention categories are evaluated, but one is combined and the other not. Aker et al. examine a cash transfer program in Niger using an RCT, where there was one study group (called a treatment arm in RCTs) receiving the transfers through mobile phones provided by the program and another treatment arm receiving the cash transfers, but where the program still provided recipients with mobile phones. The first arm combines digital finance with digital inclusion, whereas the second arm tests only digital inclusion.

Figure 5 shows the number of studies for each intervention category arranged in order of prevalence. M-health has the most evidence. Both technology-assisted learning and digital information services have evidence from more than 30 impact evaluations. As explained above, many of the specific interventions in the m-health and technology-assisted learning categories are the same types of interventions—that is, they use similar mechanisms—as in the digital information services category. For example, Bobrow et al. (2016) test an intervention using SMS text messaging to support medication adherence for patients with high blood pressure. We code it as an m-health intervention. Bruxvoort et al. (2014) test an intervention using SMS text messages to retail staff to reinforce their training about what advice to give customers on malaria medication. This study is coded under technology-assisted learning, as the SMS text messages are used for training purposes rather than providing health services directly. Finally, Cadena and Schoar (2011) test an intervention using SMS text

Table 7. Design Characteristics of ICTD Impact Evaluations.

Identification Strategy	Implementation Type		Total
	Pilot	**Program**	**Total**
Random assignment	181	17	198
As-if random assignment	8	8	16
Difference in difference	10	18	28
Statistical controls	3	8	11
Totals	202	51	253
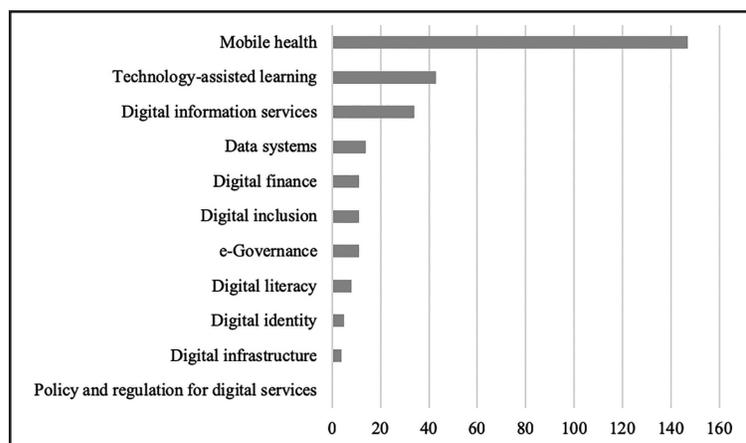
Figure 5. Number of impact evaluations providing evidence for each intervention category.

messaging to remind borrowers to make their loan payments. This study is included under digital information services.

The last row of Figure 5 shows there is no evidence from impact evaluations for the impacts of policy and regulation for digital services. This finding is unsurprising as it can be challenging to design impact evaluations for policies and regulations, which are typically enacted across all beneficiaries at the same time, thus making it difficult to construct a valid counterfactual.[6]

The evidence for the intervention categories varies in terms of the share of total evidence that comes from evaluations of actual programs and what share of the studies provide cost information about the intervention evaluated. Table 8 presents the shares for each. For digital infrastructure, digital finance, e-governance, and digital identity, more than 50% of the studies measure the impact of a program implemented at scale or independent of the research as opposed to testing a pilot, or trial, intervention. This result is consistent with the intervention types; for example, it is difficult to build infrastructure on a pilot basis. This result is also consistent with the lower number of studies in these categories, as programs may be more costly to evaluate with a counterfactual than pilots. On the other hand, only 7% of the impact evaluations of m-health interventions look at actual programs.

Overall, the share of impact evaluations that present cost information about the interventions evaluated is 18.6%. Those categories with notably higher shares of cost information are e-governance, digital identity, and digital information services. Impact evaluations that measure the impact of actual programs and also report cost information may provide evidence that is more relevant for determining which interventions to scale up.

Results by Type of Outcome
The 253 studies appear 260 times when distinguishing sector-of-outcome measurement. Seven studies report outcomes in two sectors. An example of a study reporting outcomes in two sectors is Beuermann (2015), who measures the impact on agricultural productivity and school enrollment by providing villages with satellite pay

---

6. *It is possible to conduct impact evaluations of policies and regulations, but possible approaches to doing so depend on the circumstances of each situation.*

Table 8. Share of Evidence from Programs and Share of Evidence with Cost Information for Each Intervention Category.

Intervention category	Program Evaluation Share	Cost Information Share	Total Number Studies
Digital infrastructure	100%	0%	4
Policy and regulation for digital services	n/a	n/a	0
Digital literacy	13%	13%	8
Digital inclusion	27%	9%	11
Digital finance	82%	27%	11
e-Governance	75%	33%	12
Digital identity	80%	60%	5
Data systems	43%	21%	14
Digital information services	18%	24%	34
Technology-assisted learning	14%	16%	43
Mobile health	7%	16%	147

Note: Impact evaluations are counted for each intervention category in which they appear in the map, so the total number represented in the table is 289.
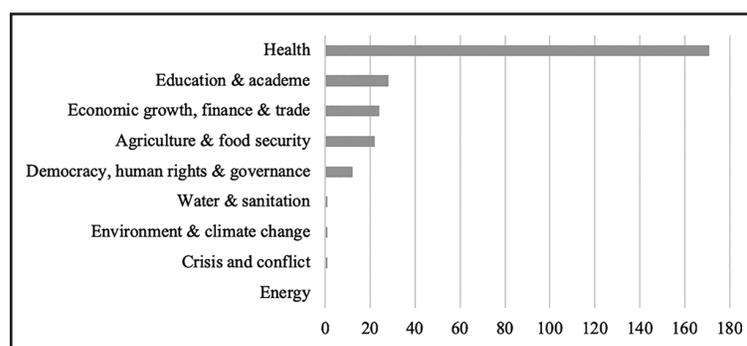

Figure 6. Number of impact evaluations providing evidence for each sector.

phones. Figure 6 presents the prevalence of evidence by sector of outcome measurement in order of prevalence. By far the most evidence is for health outcomes. As seen in Table 6, much of this evidence comes from impact evaluations of m-health interventions, but there are also health outcomes measured for six other intervention categories. There is a dearth of evidence on the impacts of ICTD interventions on outcomes in water and sanitation, environment and climate change, crisis and conflict, and energy. We provide possible explanations and implications in the Discussion section.

Table 9 shows the distribution of outcome measurement by the level of measurement for each intervention category. Of the 253 studies, eight studies measure outcomes for more than one level. One example is Marx, Pons, and Suri (2016), who evaluate the impact of an SMS campaign to mobilize voters in Kenya and measure outcomes for voters (at the individual level) and for polling stations (at the organizational level). The results show that the vast majority of outcomes measured in the impact evaluations at the individual level are consistent with impact evaluations in international development generally. Data systems is a category with relatively more measurement at the organizational level, although there are only four studies. Note the total numbers of impact evaluations by intervention category are not identical to Table 7 because the studies that measure outcomes in multiple sectors are generally not the same as the studies that measure outcomes at multiple levels.

Other Results

The final two features we coded for each study indicated whether the intervention had an equity focus, which we define broadly (see Table 5), and whether the study measured outcomes against a counterfactual

Table 9. Number of Impact Evaluations for Each Level of Outcome by Intervention Category.

Intervention Category	Individual	Organizational	Societal
Digital infrastructure	2	1	2
Policy and regulation for digital services	0	0	0
Digital literacy	8	0	0
Digital inclusion	11	0	0
Digital finance	11	0	0
e-Governance	10	1	2
Digital identity	5	1	0
Data systems	9	4	1
Digital information services	33	3	2
Technology-assisted learning	42	1	0
Mobile health	144	1	3

Note: Impact evaluations are counted across the levels of measurement for each intervention category they cover, so the total number represented in the table is 297.

Table 10. Share of Evidence for Interventions Targeting Disadvantaged Groups and Share of Studies Measuring Outcomes Specific to Females or Males for Each Intervention Category.

Intervention Category	Equity Focus	Sex Disaggregated	Total Number Studies
Digital infrastructure	0%	0%	4
Policy and regulation for digital services	n/a	n/a	0
Digital literacy	0%	38%	8
Digital inclusion	18%	64%	11
Digital finance	36%	9%	11
e-Governance	25%	8%	12
Digital identity	20%	0%	5
Data systems	14%	29%	14
Digital information services	15%	6%	34
Technology-assisted learning	5%	19%	43
Mobile health	0%	37%	147

Note: Impact evaluations are counted across the levels of measurement for each intervention category covered, so the total number represented in the table is 289.

for women or men separately. Table 10 reports the shares of the impact evaluations for each intervention category with these features. To help identify studies that might include useful information about programs benefiting disadvantaged groups, we code studies as having an equity focus if the program targeted a disadvantaged group, even if the study does not provide a separate effect size for that group. We find few ICTD interventions studied in impact evaluations have an equity focus. The highest shares are for digital finance and e-governance, with 36% and 25%, respectively. An example of an equity-focused intervention is the delivery of maternal health messages and animated films by mobile phones to rural women in India, which is evaluated by Joshi, Patil, and Hegde (2015). Although we did not code the specific disadvantaged groups targeted for the interventions with equity focus, we can say anecdotally that the vast majority are interventions targeted at rural populations. Sabet et al. (2017) document this result for the larger STIP sample.

The intervention categories with the most evidence measured for women and men separately are digital literacy, digital inclusion, and data systems. An example of a study with sex disaggregated measurement is Jamison, Karlan, and Raffler (2013), who evaluate an m-health intervention in Uganda designed to provide sex

health information using an interactive text messaging platform, and the study reports the effect sizes separately for women and men. While this example includes outcomes measured for both women and men, many of the studies coded as measuring sex-disaggregated outcomes only measure outcomes for women. They are often interventions focused on maternal and child health. One hypothesis for why more studies do not measure outcomes for the sexes separately is that most are evaluating pilot implementations and, likely due to research cost considerations, are not powered for measuring effect sizes for subsamples.

Discussion

Using systematic search-and-screening methods, we found 253 impact evaluations of ICTD interventions. Especially given that the first study in the set was published only 11 years prior to the search, this is an impressive number of studies for one theme within international development, although there are noticeable gaps in evidence. There is a dearth of evidence on the effectiveness of ICTD interventions in the energy, water and sanitation, crisis and conflict, and environment and climate change sectors.[7] This finding is consistent with the development impact evaluation base overall, which has far fewer studies in energy, water and sanitation, and environment and disaster management than for sectors such as health, nutrition, and population and education (Sabet & Brown, 2018). It is true that impact evaluations can be harder to design for programs in these sectors, as programs are often implemented across all beneficiaries at the same time, but it is not impossible, as demonstrated by the development impact evaluations that do exist. ICTs can provide valuable innovations in these sectors, so it is important that we start to produce evidence on the effectiveness and cost effectiveness of these innovations. For example, Youngman (2012) presents case studies for several ICT interventions to increase energy efficiency, but he notes that even the case study information comes primarily from the developed world. Many of the innovations he presents, such as smart building innovations designed to help buildings use less energy, could feasibly be evaluated using a comparison group.

We have little evidence about the effects of ICTD interventions on disadvantaged groups as revealed by our focus on coding for equity. Less than 10% of the studies evaluate interventions targeted to disadvantaged groups or measure outcomes specifically for disadvantaged groups. We also have limited evidence specific to the recipients' sex. Of the studies, 27% measure outcomes separately for women or for men, but a large share of these are studies of interventions that only apply to women, such as m-health for maternal care, which means for ICTD interventions that apply to both men and women, we know little about whether men and women have different outcomes.

Where there is a lot of evidence, one question to ask is: How useful is the evidence for making policy and designing programs? As mentioned above, one feature that affects the relevance of evidence is whether it comes from an evaluation of a program versus a pilot. As shown in Table 7, the ICTD evidence base is skewed toward evaluations of pilot interventions. Pilot studies can serve different purposes.[8] One is to test theories, in which case the specific design of the pilot intervention typically is not intended to inform program design or scaling. Another is to test mechanisms, in which case the design of the pilot intervention is meant to inform programming, but only one program component or element. A third is to test a program on a limited basis before taking it to scale. Even in the third case, careful assessments of site selection bias (Alcott, 2015) and the ecological validity of the pilot intervention (Brown, 2016) are necessary to determine whether these studies can be used to predict the success of scaled-up programs over time. In addition, pilot studies often only measure short-term outcomes or may produce outcomes specific to a partial equilibrium, meaning the nature or size of the benefit depends on only some people receiving the intervention and may change when everyone receives it. Thus, the evidence from pilot studies is valuable, but often only part of what is needed to inform policy and programming.

Another important feature for usefulness is whether the study measures cost effectiveness as well as

---

7. While only one study measures crisis- and conflict-related outcomes, our dataset includes six studies conducted in countries in situations of fragility, crisis, and violence (as labeled by the World Bank for the publication year of the study).
8. Feasibility studies, or formative evaluations, without a comparison group are not discussed here as they are not included in the map.

effectiveness. Although we coded the cost-effectiveness variable generously by requiring only that the study provide information on the cost of the intervention, but not that it present cost-effectiveness estimates, we find that only 18.6% of the impact evaluations provide cost information. This finding suggests that even where we have evidence that an intervention is effective, we may still need research to determine whether it is cost effective.

We can look at the example of m-health, for which we have by far the most evidence from impact evaluations. For m-health, more than 90% of the studies measure the impacts of pilot interventions, and only 16% include information about intervention costs. The irony is that many of these studies use the cost effectiveness of m-health interventions as a motivation for evaluating m-health interventions, but then do not include information on cost in the study. Kamal et al. (2015) is one example. They even include "cost effectiveness" as one of the article's keywords, but conclude by saying that cost effectiveness is an area for future research. The cost of pilot implementations should not be overly difficult to calculate, as the implementations are of limited duration and applied to only a sample of recipients. While pilot implementation costs may not be proportional to the costs at scale, they are at least good information to start with.

The prevalence of pilot studies and studies using RCT design for m-health, as well as for other intervention categories where health outcomes are measured, is unsurprising. Public health researchers have a long tradition coming from medical trials of looking to RCTs to test interventions. Because the researchers must assign participants to the intervention (treatment) and the status quo (control) groups before implementation, and they must carefully separate and monitor the treatment and control groups, it is often easiest to have the implementation be a pilot, or trial. Similarly, researchers in other fields who want to conduct RCTs often look to pilot studies to ensure they can randomly assign the intervention and maintain some control over the comparison. What this means, however, is that as governments, the private sector, and civil society take ICTD interventions to scale, they must continue to rigorously evaluate for effectiveness (Brown, 2016).

Where there are clusters of evidence, the next step is to carefully synthesize that evidence to identify whether and when there are generalizable findings. We screened for systematic reviews when conducting our search and screening, and we found a handful of reviews on m-health, but few others. It is encouraging that this journal recently published three systematic reviews on ICTD (Alampay & Moshi, 2018; Ilavarasan & Otieno, 2018; Stork, Kapugama, & Samarajiva, 2018). In follow-up research from this evidence map, we have begun two rapid reviews, one to look at which ICTD interventions work for achieving economic growth outcomes and one to look at which ICTD interventions work for achieving governance outcomes. We call them rapid because we are starting with sets of studies from the map rather than conducting new searches.

Limitations

The search-and-screening process for this evidence map was consistent with a limited-resource review. We did not conduct parallel searches or screening by two researchers, and we limited our search to English-language documents.

As noted above, we found an error in our initial search strategy for the EBSCO databases. We re-ran those searches and screened an additional 7,914 records. We spent a lot of time refining the search strategy for EconLit as our strings returned a small number of records. We expanded the terms to increase the number of hits, but ultimately concluded there is a challenge in conducting systematic searches of social science indexes arising from the way in which social scientists write their paper abstracts, that is, not following a standard template and often intended to pique interest rather than give away the results. We hypothesize that we missed more studies outside the health sector than in it.

For the updated search, which focused on a limited number of years, we broadened the topical search terms from those in the published strategies for the STIP work. Our resulting sample of studies from 1990–2017 therefore is biased toward including more studies from the 2015–2017 period. Nonetheless, the dramatic increase in the number of impact evaluations of ICTD intervention in the current decade means there are likely to be several studies published before our search cut-off, but not yet indexed, and several more published since our search cut-off.

Conclusion

In the last decade there has been huge growth in impact evaluations of ICTD interventions in low- and middle-income countries, from three studies in 2006 to 73 in 2016. The evidence comes from all regions except Europe and Central Asia. The largest clusters of evidence are for m-health, technology-assisted learning, and digital information services. The map reveals evidence for ICTD interventions across many sectors, but some sectors have little to no evidence. Roughly 80% of the ICTD impact evaluations use randomized assignment to identify causal effects, which suggests a relatively low risk of bias across the evidence base. At the same time, roughly 80% evaluate pilot implementations instead of programs, raising questions about how useful the evidence is for informing programs at scale. Less than 20% of studies report costing data, so our ability to assess cost effectiveness is limited. Less than 10% of the evidence base comes from interventions targeted to disadvantaged groups.

Our findings support four recommendations. First, we who want evidence-informed ICTD need to start evaluating the impact of ICTD interventions on outcomes in sectors with little evidence. These include conflict and crisis, environment and climate change, water and sanitation, and energy. Second, we must continue to rigorously evaluate the impact of ICTD interventions at scale or when implemented as programs and not just pilots. Pilot studies are useful, even for understanding outcomes, but are not enough. Third, we need to collect cost information and perform a cost-effectiveness analysis. It is not enough to assert that ICTs are low cost and therefore presume these interventions are cost effective. And finally, where there are clusters of evidence, we must produce more systematic reviews to identify generalizable results and better understand heterogeneous outcomes. As noted above, the authors have started two such reviews based on the studies catalogued in this map. ∎

Acknowledgments

Annette N. Brown, Principal Economist, Chief Science Office, FHI 360, USA. ABrown@fhi360.org

Hannah J. Skelly, Project Director, Global Education, Employment, and Engagement, FHI 360, USA. HSkelly@fhi360.org

References

Aker, J. C., Boumnijel, R., McClelland, A., & Tierney, N. (2016). Payment mechanisms and anti-poverty programs: Evidence from a mobile money cash transfer experiment in Niger. *Economic Development and Cultural Change, 65*(1), 1–37. doi:https://doi.org/10.1086/687578

Alampay, E., & Moshi, G. (2018). Impact of mobile financial services in low- and lower-middle-income countries: A systematic review. *Information Technologies & International Development* (Special Section), *14,* 164–181.

Alcott, H. (2015, August 1). Site selection bias in program evaluation. *The Quarterly Journal of Economics, 130*(3), 1117–1165. https://doi.org/10.1093/qje/qjv015

Beuermann, D. W. (2015). Information and communications technology, agricultural profitability and child labor in rural Peru. *Review of Development Economics, 19*(4), 988–1005. doi:10.1111/rode.12180

Bobrow, K., Farmer, A. J., Springer, D., Shanyinde, M., Yu, L. M., Brennan, T., . . . Levitt, N. (2016). Mobile phone text messages to support treatment adherence in adults with high blood pressure (SMS-text adherence support [StAR]). *Circulation, 133*(6), 592–600. doi:10.1161/CIRCULATIONAHA.115.017530

Brown, A. N. (2016, April 4). The pitfalls of going from pilot to scale [Web log post]. Retrieved from http://blogs.3ieimpact.org/the-pitfalls-of-going-from-pilot-to-scale-or-why-ecological-validity-matters/

Bruxvoort, K., Festo, C., Kalolella, A., Cairns, M., Lyaruu, P., Kenani, M., . . . Schellenberg, D. (2014). Cluster randomized trial of text message reminders to retail staff in Tanzanian drug shops dispensing artemether-lumefantrine: Effect on dispenser knowledge and patient adherence. *The American Journal of Tropical Medicine and Hygiene, 91*(4), 844–853. https://doi.org/10.4269/ajtmh.14-0126

Cadena, X., & Schoar, A. (2011). *Remembering to pay? Reminders vs. financial incentives for loan payments* (Working Paper 17020). Retrieved from https://www.nber.org/papers/w17020

Cameron, D. B., Mishra, A., & Brown, A. N. (2016). The growth of impact evaluation for international development: How much have we learned? *Journal of Development Effectiveness, 8*(1), 1–21. doi:10.1080/19439342.2015.1034156

Clapton, J., Rutter, D., & Sharif, N. (2009). *SCIE Systematic mapping guidance; April 2009.* Retrieved from https://www.scie.org.uk/publications/researchresources/rr03.pdf

Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. J. (2016). *Impact evaluation in practice.* Washington, DC: Inter-American Development Bank & World Bank. doi:10.1596/978-1-4648-0779-4

Ilavarasan, P. V., & Otieno, A. (2018). Tiny impact of ICTs and paucity of rigorous causal studies: A systematic review of urban MSMEs in the developing world. *Information Technologies & International Development* (Special Section), *14,* 134–150.

James, K. L., Randall, N. P., & Haddaway, N. R. (2016). A methodology for systematic mapping in environmental sciences. *Environmental Evidence, 5*(7), 1–13. doi:10.1186/s13750-016-0059-6

Jamison, J. C., Karlan, D., & Raffler, P. (2013). Mixed-method evaluation of a passive mHealth sexual information texting service in Uganda. *Information Technologies & International Development, 9*(3), 1–28.

Joshi, S., Patil, N., & Hegde, A. (2015). Impact of mhealth initiative on utilization of ante natal care services in rural Maharashtra, India. *Indian Journal of Maternal and Child Health, 17*(2), 1–7.

Kamal, A. K., Shaikh, Q., Pasha, O., Azam, I., Islam, M., Memon, A. A., . . . Khoja, S. (2015). A randomized controlled behavioral intervention trial to improve medication adherence in adult stroke patients with prescription tailored short messaging service (SMS)-SMS4Stroke study. *BMC Neurology, 15*(212), 1–11. https://doi.org/10.1186/s12883-015-0471-5

Marx, B., Pons, V., & Suri, T. (2016). *Voter mobilization can backfire: Evidence from Kenya.* Unpublished paper. Retrieved from https://pdfs.semanticscholar.org/bff0/f8274caa2cf1067b1e3547a77a19bb0a01e2.pdf

Medaglia, R., & Zheng, L. (2017). Mapping government social media research and moving it forward: A framework and research agenda. *Government Information Quarterly, 34,* 496–510. doi:10.1016/j.giq.2017.06.001

Nascimento, A. M., & da Silva, D. S. (2017). A systematic mapping study on using social media for business process improvement. *Computers in Human Behavior, 73,* 670–675. doi:10.1016/j.chb.2016.10.016

Rankin, K., Cameron, D. B., Ingraham, K., Mishra, A., Burke, J., Picon, M, . . . Brown, A. N. (2015). *Youth and transferable skills: An evidence gap map.* Retrieved from http://www.3ieimpact.org/media/filer_public/2015/09/01/egm2-youth_and_transferable_skills.pdf

Rankin, K., Jarvis-Thiébault, J., Pfeifer, N., Engelbert, M., Perng, J., Yoon, S., & Brown, A. N. (2016). *Adolescent sexual and reproductive health: An evidence gap map.* Retrieved from http://www.3ieimpact.org/media/filer_public/2016/12/29/egm5-asrh.pdf

Sabet, S., & Brown, A. N. (2018). Is impact evaluation still on the rise: The new trends for 2010–2015. *Journal of Development Effectiveness, 10*(3), 291–304. doi:10.1080/19439342.2018.1483414

Sabet, S., Heard, A. C., & Brown, A. N. (2017). *Science, technology, innovation and partnerships for development: An evidence gap map.* Retrieved from http://www.3ieimpact.org/media/filer_public/2017/03/09/egm6-stip.pdf

Sabet, S., Heard, A. C., Neilitz, S., & Brown, A. N. (2017). *Assessing the evidence base on science, technology, innovation and partnerships for accelerating development outcomes in low- and middle-income countries.* New Delhi, India: International Initiative for Impact Evaluation.

Samarajiva, R. (2018). Introduction: What do we know about ICT impact and how best can that knowledge be communicated? *Information Technologies & International Development* (Special Section), *14,* 182–190.

Snilstveit, B., Vojtkova, M., Bhavsar, A., Stevenson, J., & Gaarder, M. (2016) Evidence & gap maps: A tool for promoting evidence informed policy and strategic research agendas. *Journal of Clinical Epidemiology, 79,* 120–129. http://dx.doi.org/10.1016/j.jclinepi.2016.05.015

Stork, C., Kapugama, N., & Samarajiva, R. (2018). Economic impacts of mobile telecom in rural areas in low- and middle-income countries: Findings of a systematic review. *Information Technologies & International Development* (Special Section), *14,* 191–208.

United Nations Conference on Trade and Development (UNCTAD). (2011). *Measuring the impacts of information and communication technology for development* (UNCTAD Current Studies on Science, Technology and Innovation, No. 3). Retrieved from http://unctad.org/en/docs/dtlstict2011d1_en.pdf

Waddington, H., Aloe, A., Becker, B. J., Djimeu, E. W., Hombrados, J. G., Tugwell, P., . . . Reeves, B. (2017). Quasi-experimental design series—paper 6: risk of bias assessment. *Journal of Clinical Epidemiology, 89,* 43–52. https://doi.org/10.1016/j.jclinepi.2017.02.015

World Bank. (2011). Capturing technology for development: An evaluation of World Bank Group activities in information and communication technologies. Retrieved from http://ieg.worldbankgroup.org/sites/default/files/Data/Evaluation/files/ict_evaluation.pdf

World Summit on the Information Society (WSIS). (2016). *Final WSIS targets review: Achievements, challenges and the way forward.* Retrieved from http://www.oecd-ilibrary.org/science-and-technology/final-wsis-targets-review_pub/80a85799-3001f517-en

Youngman, R. (2012). *ICT solutions for energy efficiency.* Washington, DC: World Bank. Retrieved from https://openknowledge.worldbank.org/handle/10986/12685