Shayda M Sabet
Anna C Heard
Scott Neilitz
Annette N Brown

Assessing the evidence base on science, technology, innovation and partnerships for accelerating development outcomes in low- and middle-income countries

March 2017

Scoping Paper 6

Multi-sector

3ie International Initiative for Impact Evaluation

About 3ie

The International Initiative for Impact Evaluation (3ie) is an international grant-making non-governmental organisation promoting evidence-informed development policies and programmes. We are the global leader in funding and producing high-quality evidence of what works, for whom, why and at what cost. We believe that better and policy-relevant evidence will help make development more effective and improve people's lives.

3ie scoping papers

3ie develops its thematic grant programmes in a particular development area or sector through an organised scoping process. This multifaceted exercise helps ensure that new windows address priorities for evidence production where evidence is limited and gaps have been identified. 3ie scopes the demand for evidence through consultation with potential users, including policymakers, programme managers, researchers and civil society. Scoping exercises usually include producing an evidence map of interventions and outcomes of included studies and reviews. This aids identification of where more or better evidence is needed and where evidence is sufficient for producing systematic reviews. Other scoping can include reviewing existing programming and funding. The results include priority questions for which there is demand for answers. Scoping papers consolidate this information into a justification for what a thematic grant programme will address.

About this paper

This paper provides an assessment of the state of the evidence on the effectiveness of promoting and using science, technology, innovation and partnerships to improve development outcomes. It combines a supply analysis using an evidence gap map with a demand analysis using a stakeholder survey, feedback from experts participating in a consultative event and in a roundtable event. The evidence gap map report and the online interactive map are available on the 3ie website.

3ie developed it as part of a project funded by the support of the American People through the United States Agency for International Development (USAID)'s US Global Development Lab. The contents of this report are the sole responsibility of the authors and does not represent the opinions of 3ie, its donors or its Board of Commissioners, or reflect the views of USAID or the United States government. Any errors and omissions are the sole responsibility of the authors.
Direct comments or queries to the corresponding author, Shayda Sabet (ssabet@3ieimpact.org).

Suggested citation**:** Sabet, SM, Heard, AC, Neilitz, S and Brown, AN, 2017. A*ssessing the evidence base on science, technology, innovation and partnerships for accelerating development outcomes*. 3ie Scoping Paper 6. Washington, DC: International Initiative for Impact Evaluation (3ie).

Executive editors: Beryl Leach and Emmanuel Jimenez
Production managers: Angel Kharya and Brigid Monaghan
Assistant production manager: Akarsh Gupta
Copy editor: Jaime L Jarvis
Proof reader: Yvette Charboneau
Cover design: John F McGill
Printer: Via Interactive
Cover photo: Eric Sales/ADB

Assessing the evidence base on science, technology, innovation and partnerships for accelerating development outcomes

Shayda M Sabet
3ie

Anna C Heard
3ie

Scott Neilitz
3ie

Annette N Brown
FHI 360

3ie Scoping Paper 6

March 2017

Acknowledgements

We gratefully acknowledge funding of this paper from the American people through USAID's Global Development Lab.

For this scoping work, 3ie relied on the guidance and feedback from an advisory group, to whom we are thankful: Ashley Bear (National Academy of Sciences), Samuel Bickel (UNICEF), David de Ferranti (Results for Development), Naoto Kanehira (World Bank), Kyriacos Koupparis (USAID) and Saharah Moon Chapotin (USAID).

The authors wish to thank Shannon Griswold, Joshua Kaufman and Jessica Lucas for their constant guidance and support; Kristen Rankin and Ben Wood for their valuable inputs and advice; and the work of consultants Mark Engelbert, Amy Holter, Leah Leavitt, Julie Perng, Nadine Pfeifer, Aryamala Prasad and Ingrid Rojas, who helped with the search and screening process to identify studies for the gap map. Many thanks also go to Edoardo Masset and Beryl Leach for their helpful comments. We would also like to thank Tasmia Rahman, who served as an intern for this project and provided indispensable input and research assistance.

Additionally, 3ie would like to thank the wide range of stakeholders who participated in the consultative workshop and roundtable events in Washington, DC.

Summary

With the decline in official development assistance since the 1990s, the need to scale up innovative approaches that can help address the development challenges in low- and middle-income countries (L&MICs) is greater than ever. Science, technology, innovation and partnerships (STIP) play an important role in accelerating the outcomes of development programmes. Ultimately, L&MICs' ability to initiate and sustain growth will depend on their internal capacities for science, technology and private sector innovation, as well as their access to partnerships with a wide range of organisations and agencies. Employing STIP to enhance and accelerate economic development outcomes should no longer be the unique privilege of high-income countries.

Many bilateral and multilateral assistance agencies have incorporated science, technology and innovation (or science, technology and innovation combined with partnerships) into their development agendas. For example, the UN has established a forum on science, technology and innovation for achieving the Sustainable Development Goals, and the Organisation for Economic Cooperation and Development has a directorate for science, technology and innovation. STIP is becoming a critical component of the theories of change used by organisations such as the United States Agency for International Development (USAID), the World Bank and the UK Department for International Development. The US Global Development Lab at USAID has made a strong push for incorporating STIP into its development efforts. The World Bank focused its 2016 World Development Report on the benefits from digital technology.

As these organisations move to prioritise STIP-related development efforts, it is critical to base policy and programming decisions on solid evidence. Rigorous evaluations of STIP interventions can test unproven approaches, demonstrate impact to government policymakers and reveal cost-effectiveness. Such evidence enables policymakers to maximise the contributions STIP can make to achieving development outcomes. Some good evidence already exists, but more is needed, particularly given the wide variety of STIP approaches.

New investments in the production of evidence about STIP interventions should be targeted to achieve the greatest value added, that is, to answer the questions that are the highest priority for development stakeholders. The purpose of this scoping paper is to identify the priorities for investments in the production of new high-quality evidence for STIP interventions. To identify these priorities, we conducted research on the current supply of evidence and the stakeholder demand for evidence, and we analyse their intersection.

Methods

The first input into the scoping exercise analysis is an assessment of the supply of evidence using an evidence gap map (EGM). 3ie EGMs are thematic collections of evidence on the effects of policies and programmes. They provide an innovative approach for rapid knowledge capture and transfer, combining methods from other

review and mapping approaches with data visualisation using an interactive platform. The STIP EGM[1] catalogues impact evaluations and systematic reviews of STIP-related interventions that met our inclusion criteria. It presents information about this evidence base in a framework by the interventions identified and the outcomes measured.

We group the interventions by each STIP component (science, technology, innovation and partnerships). We include as science interventions those that build the capacities of L&MICs to produce their own scientific and technological research. We include as technology interventions approaches that use mobile devices and the internet to enhance development programming in L&MICs. Innovation interventions include ones that build the innovation 'ecosystem', in other words, interventions that foster the necessary enabling environment to encourage and spur innovation in L&MICs. Finally, partnerships in this mapping framework include the implementation of development programmes through innovative partnerships and financing mechanisms.

Given the broad, cross-sectoral nature of this topic, we chose outcome categories to capture the levels of analysis for the measured effects and to represent the full range of development sectors. The EGM thus catalogues the same studies in two different ways. We also analyse the evidence base according to four crosscutting themes: long-term outcome measurement, cost analysis, sex-disaggregated effects and effects for vulnerable and marginalised populations.

The second input is an assessment of the stakeholder demand for new evidence. We collected data and information from a variety of sources, including expert consultations and a stakeholder survey with 110 respondents, culminating in a roundtable event with 31 participants from STIP fields. We conducted quantitative and qualitative analysis of that information in order to identify the interventions and outcomes for which the demand for more and better evidence is relatively higher.

Findings

The STIP evidence gap map identifies and catalogues a large amount of evidence – 320 completed impact evaluations – on the effectiveness of STIP-related interventions. This existing evidence base is clustered in a subset of intervention categories in the STIP framework. Namely, 134 (nearly 42 per cent) of the studies fall under mobile health (m-health). Science, technology, engineering and mathematics education, grants and subsidies for innovation, digital information systems (other than m-health), innovative financing, technology-assisted learning, digital finance and access to capital for innovation all have more than 10 studies. At the same time, there are four intervention categories with no studies (material resources for scientific

[1] *Science, technology, innovation and partnerships for development: an evidence gap map*. 3ie Evidence Gap Map Report 6 can be accessed at http://www.3ieimpact.org/media/filer_public/2017/03/09/egm6-stip.pdf and the online interactive map is available at http://gapmaps.3ieimpact.org/evidence-maps/science-technology-innovation-and-partnerships-evidence-gap-map

research, technical assistance for scientific research, two-entity partnerships and global multi-stakeholder initiatives), and two with only one impact evaluation (policy and regulation for scientific research and policy and regulation for digital services).

We find that only a small number of studies report outcomes at the community or societal level, and sex disaggregation is largely limited to studies in the health and education sectors. In general, there is a dearth of studies that either focus on or report outcomes separately for any vulnerable or marginalised population. In spite of the sex-disaggregation, the analysis in these studies did not include detailed analysis of gender.

There are only seven completed systematic reviews identified in the map. This is a small number, given the density of completed impact evaluations we found. The seven completed systematic reviews in our EGM are all health related. Six relate to m-health, with one also related to data systems development. The seventh review falls under innovative health financing.

The results from the stakeholder survey reveal that the evidence base for STIP-related interventions is not considered to be strong. If we count the interventions according to those receiving a majority of ‘strong’ plus ‘moderate’ responses versus those receiving a majority of ‘weak’ plus ‘don’t know’ responses, we find that 13 categories have a majority ‘strong’ plus ‘moderate’ and 12 have a majority ‘weak’ plus ‘don’t know’. The survey also asked respondents to select those interventions for which more and better evidence would be useful for their work. We find that the following categories have relatively higher response rates: fellowships and research grants, technical assistance for science research, research collaborations and partnerships, policy and regulation for science, digital infrastructure, digital inclusion, data systems development, digital information services, two-entity partnerships and innovative financing.

In addition, respondents selected more categories than they were supposed to in the innovation ecosystems group, which points to a high demand for evidence across this group.

When we compare the perceived strength of evidence and the demand for more evidence reported in the survey, we do not see what might be an expected correlation, that those categories perceived as having a weak or unknown evidence base are those with a greater demand for more and better evidence. In fact, the correlation may be the other way around. A positive correlation between strength of evidence and demand for more evidence could reflect that stakeholders have a better appreciation for the usefulness of high-quality evidence when they see such evidence, and then desire more of such evidence, perhaps specific to programme context or scale, to inform their work.

Our assessment of demand, combining the other data sources with the survey results, identifies the follow intervention categories as having a relatively higher demand for evidence:

- fellowships and research grants;
- research collaborations;
- digital infrastructure development;
- digital inclusion;
- digital finance;
- digitising identity;
- data systems development;
- digital information services;
- innovation ecosystems;
- two-entity partnerships and global multi-stakeholder initiatives; and
- innovation financing.

Research investment priorities

Given the large number of impact evaluations we found, we consider it surprising that there are so few systematic reviews and that the reviews we did find contain so few of the impact evaluations included in the EGM. By analysing the homogeneity of interventions and outcomes measured amongst the studies for intervention categories with a large number of studies, we can point to several clusters of studies that are promising for systematic review:

- digital finance (in particular, mobile money systems);
- digital information services for agriculture and food security;
- m-health (in particular, antiretroviral therapy adherence, lifestyle changes for chronic diseases and appointment reminders);
- innovation ecosystems programmes in Latin America;
- innovative financing (in particular, results-based financing for health);
- science, technology, engineering and mathematics educational programmes; and
- technology-assisted learning.

By looking at the demand assessment together with the EGM, we identify several intervention categories that are the highest priority for investment in new impact evaluation research:

- research collaborations;
- digital infrastructure development;
- digital inclusion;
- digitising identity;
- data systems development;
- innovation ecosystems programmes in Sub-Saharan Africa; and
- innovative financing (for non-results-based programmes and for results-based education).

Our combined analysis of demand and supply also identifies several outcomes, sectors and crosscutting themes that should be priorities for investments in new STIP-related impact evaluations:

- community- and society-level outcomes in relevant intervention categories;
- technology impact evaluations on democracy, human rights and governance outcomes;
- technology impact evaluations on agriculture and food security outcomes;
- technology impact evaluations on crises and conflict outcomes;
- science impact evaluations that report outcomes disaggregated by sex;
- digital inclusion impact evaluations that report outcomes for vulnerable or marginalised populations;
- science and innovation ecosystems impact evaluations that report long-term impacts; and
- cost analyses for technology, innovation ecosystems and innovative financing interventions.

Contents

List of figures and tables

Abbreviations and acronyms

3ie	International Initiative for Impact Evaluation
ART	Antiretroviral therapy
DfID	UK Department for International Development
e-governance	Electronic governance
EGM	Evidence gap map
FONCYT	Argentinian Scientific and Technological Research Fund
FONTAR	Argentinian Technological Fund
FONTEC	Chilean National Fund for Technological and Productive Development
HH	Household
L&MICs	Low- and middle-income countries
m-health	Mobile health
NGO	Nongovernmental organisation
R&D	Research and development
RBF	Results-based financing
SMS	Short message service
STEM	Science, technology, engineering and mathematics
STIP	Science, technology, innovation and partnerships
UNEP	United Nations Environment Programme
USAID	United States Agency for International Development

1. Introduction

With the decline in official development assistance since the 1990s (World Bank 2016a), the need to scale up innovative approaches that can help address the development challenges in low- and middle-income countries (L&MICs) is greater than ever. Science, technology, innovation and partnerships (STIP) play an important role in accelerating the outcomes of development programmes. Ultimately, L&MICs' ability to initiate and sustain growth will depend on their internal capacities for science, technology and private sector innovation, as well as their access to partnerships with a wide range of organisations and agencies (Juma and Yee-Cheong 2005; World Bank Group 2016b). Employing STIP to enhance and accelerate economic development outcomes should no longer be the unique privilege of high-income countries (UN System Task Team 2013).

Many bilateral and multilateral assistance agencies have incorporated science, technology and innovation (or science, technology and innovation combined with partnerships) into their development agendas. For example, the UN has established a forum on science, technology and innovation for achieving the Sustainable Development Goals and the Organisation for Economic Cooperation and Development has a directorate for science, technology and innovation. STIP is becoming a critical component of the theories of change of the used by organisations such as the United States Agency for International Development (USAID), the UK Department for International Development (DfID) and the World Bank (Runde and Magpile 2014). The US Global Development Lab at USAID has made a strong push for incorporating science, technology, innovation and new partnerships into its development efforts. The World Bank focused its 2016 World Development Report on the benefits of digital technology (World Bank Group 2016b).

As these organisations move to prioritise STIP-related development efforts, it is critical to base policy and programming decisions on solid evidence. Rigorous evaluations of STIP interventions can test unproven approaches, demonstrate impact to government policymakers, and reveal cost effectiveness. Such evidence enables policymakers to maximise the contributions STIP can make to the achievement of development outcomes. Some good evidence already exists. For example, the World Development Report 2016 cites several impact evaluations of digital technology interventions that can inform new programmes and scaling of existing programmes (World Bank Group 2016b). More evidence is needed, however, particularly given the wide variety of STIP approaches.

New investments in the production of evidence about STIP interventions should be targeted to achieve the greatest value added, that is, to answer the questions that are the highest priority for development stakeholders. By stakeholders, we mean the local and international implementers, researchers, funders and policymakers involved in STIP programming in L&MICs. The purpose of this scoping paper is to identify priorities for investments in the production of new high-quality evidence for STIP interventions. To identify these priorities, we conducted research on the current supply of evidence and the stakeholder demand for evidence, and we analyse the

intersection of these. We identify priorities where the current supply does not meet the demand.

Science, technology, innovation and partnership are general terms that may point to different sets of interventions for different people. In section 2, we define the STIP framework that we used for our research in detail by listing all the included intervention categories and providing examples of interventions for each category. Here we provide a brief overview of how we defined the scope of each of these concepts for the framework. Science encompasses interventions that build the capacities of L&MICs to produce their own scientific and technological research. These interventions include efforts by governments and donors to make greater investments in technical and human resources, collaboration and knowledge transfer, and higher-quality education related to science, technology, engineering and mathematics (STEM).

Technology here includes interventions that use mobile devices and/or the internet to enhance development programming in L&MICs. This focused scope for technology allowed us to assess the evidence base in more depth and reflects the strong trends to incorporate mobile and internet into development efforts in a variety of sectors. The emphasis on closing the digital divide, for example, has become a priority for several organisations and is the subject of the World Development Report 2016. It is based on the belief that investments in digital infrastructure and literacy and the provision of digital services can improve information flow, increase access to public services and expand participation in the political process (World Bank Group 2016b).

Innovation interventions here are those that contribute to building an innovation 'ecosystem' in a country, region or market. Innovation ecosystem means the necessary enabling environment to encourage private sector entrepreneurs and firms to innovate, often combined with assistance such as subsidies to jump-start investment in new methods and technologies. The emphasis on the enabling environment for innovation is reflected by the targets in Sustainable Development Goal 9, 'Industry, Innovation and Infrastructure', which encourage L&MIC governments to support the regulatory environment and promote financial incentives to encourage innovation.

Finally, countries and aid agencies are moving towards using partnerships and innovative financing mechanisms to help support the interventions and initiatives that use science, technology and innovation to improve development outcomes. Organisations maintain that using partnerships to increase L&MIC capacity will ultimately lead to better development outcomes. Figuring out which mechanisms and interventions work most effectively and efficiently to do that requires evidence on what works.

The primary focus of this paper is to describe the breadth, depth and features of the existing STIP-related evidence base (the current supply of impact evaluation evidence) and then compare it to the areas where stakeholders think new and higher-quality evidence would be useful. This analysis allows us to identify areas

where investments in new impact evaluation and systematic review research will add the greatest value.

The first input into the analysis is an assessment of the supply of evidence using an evidence gap map (EGM). The EGM catalogues all of the impact evaluations and systematic reviews of STIP-related interventions that met our inclusion criteria and presents information about this evidence base in a matrix format, structured by the interventions analysed and the outcomes measured. In this scoping paper, we analyse the evidence base represented in the map, as well as gaps in this evidence base, to identify areas where additional research could add value. Policymakers and donors can further use the evidence identified in the EGM to set programme priorities and guide their investments.

The second input is an assessment of the stakeholder demand for new evidence. We collected data and information from a variety of sources, including expert consultations and a stakeholder survey, and conducted quantitative and qualitative analysis of that information to identify the interventions and outcomes for which the demand for more and better evidence is relatively higher. Stakeholders expressed a desire for more and better evidence for all of the interventions in our STIP framework, so the demand assessment is an effort to reveal where demand is greater in order to help identify priorities for new research investment.

The remainder of this paper is organised as follows. In section 2 we present the STIP framework, which defines the scope of the project. Section 3 contains the findings of the EGM, which include the search and screening results and an analysis of the characteristics of the impact evaluation and systematic review evidence base. In section 4 we present the findings of our demand assessment, which includes the results of our stakeholder survey. In section 5 we delineate our priorities for future research investments through an analysis of the evidence clusters and gaps. Section 6 notes the limitations to our research and section 7 concludes.

2. Framework

The scope for the analysis in this paper is defined by a framework of STIP-related intervention categories and development outcome categories. For the EGM, we present this framework as a matrix with intervention categories as the rows and outcomes categories as the columns. In section 3.2 we show the evidence map populated into this matrix. We organise both the analysis of the size of the evidence base and the stakeholders' expressed need for more evidence according to the interventions and outcomes included in the framework. We developed the framework using a participatory and iterative process, together with USAID and other stakeholders. Through this process, we identified 25 intervention categories and 16 outcome categories, which we describe in turn.[2] We supplement this framework with

[2] One intervention category was added at a later stage to differentiate mobile health (m-health) interventions from other digital information services, due the vast numbers of m-health studies that were found during search and screening for the gap map. For this reason, the stakeholder survey did not include questions specific to m-health.

four crosscutting themes in order to explore these features of the current evidence base and identify priorities for new evidence production related to these themes. We list these in section 2.3.

2.1 Interventions

We group the interventions by each STIP component (science, technology, innovation and partnerships). We include as science interventions those that build the capacities of L&MICs to produce their own scientific and technological research. We include as technology interventions approaches that use mobile devices and the internet to enhance development programming in L&MICs. Innovation interventions include those that build the innovation 'ecosystem', those that foster the necessary enabling environment to encourage and spur innovation in L&MICs. Finally, partnerships in this framework include the implementation of development programmes through innovative partnerships and financing mechanisms.

Table 1 presents the intervention categories for each group, with brief descriptions and examples.

Table 1: Intervention categories

Intervention	Definition	Example
Science		
Fellowships and research grants	Monetary assistance for postgraduate-level researchers to conduct existing or new research.	Publicly funded grants to professors
Material resources for scientific research	Material resources provided to research institutions for the purposes of conducting research.	Lab equipment, other in-kind donations
Technical assistance for scientific research	Assistance or training for researchers, often provided by an international nongovernmental organisation (NGO) or university from a high-income country.	Training on the use of technology or research equipment
Research exchanges and collaborations	Collaboration between researchers, educational institutions or other research-based entities for the purposes of scientific research or capacity building.	Joint research grants
Policy and regulation for scientific research	Laws and regulations that facilitate research in science and technology.	Patent laws
Education programmes to promote STEM	Educational programmes, scholarships, training and in-kind donations at all non-tertiary educational levels intended to promote the STEM fields.	Pedagogical strategies used to enhance learning in the sciences in secondary schools

Intervention	Definition	Example
Technology		
Digital infrastructure development	Facilitating access to digital technology or improved digital infrastructure.	Rollout of cell phone towers
Policy and regulation for digital services	Laws and regulations that facilitate access to or use of digital technologies.	Reduction of taxes on mobile technologies
Digital literacy	Aims to improve a person's ability to use the internet or mobile devices.	Training on how to use a mobile phone to make financial transactions
Digital inclusion	Facilitating access to digital and data technologies, particularly – though not exclusively – for marginalised groups.	Mobile phone credit to new mothers
Digital finance	Promoting the use of mobile technologies for finance.	Mobile money payment applications
Electronic governance (e-governance)	Facilitating the provision of government services and communication between the public and government agencies using digital technology.	Digitising the process for renewing national ID cards
Digitising identity	Digitising identification systems.	Fingerprinting and biometrics
Data systems development	Using digital technology to improve data collection, management and use.	Use of personal digital assistants for data collection by health workers
Digital information services	Digital technology for information dissemination and the provision of individual services to smooth information asymmetry or to change or 'nudge' behaviour. Services related to finance or health are excluded from this category.	Short message service (SMS) messages to farmers containing information about weather conditions
Technology-assisted learning	Use of the internet or mobile devices to improve learning outcomes.	Web-based computer simulation for teaching science
Mobile health (m-health)	Use of mobile and wireless devices to provide medical care.	SMS messages to patients encouraging medication adherence
Innovation ecosystems		
Access to capital	Facilitating access to capital for small firms and entrepreneurs, intended to spur innovation and improve technology.	Venture capital and seed money; accelerators and incubator programmes

Intervention	Definition	Example
Grants and subsidies	Non-debt instruments provided to firms, intended to spur innovation and improve technology.	Grants, subsidies, prizes and other awards
Policies and regulation that affect innovation	Laws and regulations that affect innovation (positively or negatively).	Regional zones; reductions in trade barriers
Networks and collaboration for innovation	Facilitating the development of networks, partnerships and relationships between individuals or organisations for the purposes of information sharing, technology diffusion, network development or creating credible and recognisable associations.	Managerial associations for production innovation
Capacity building for innovation	Interventions that promote institutional and human capacity building. These interventions foster a culture of innovation or innovation systems, particularly related to promoting science and technology.	Programmes that provide technical assistance, training, mentorship and capacity building to firms to spur innovation; accelerator and incubator programmes
Partnerships		
Two-entity partnerships	Interventions that are created or implemented by a partnership between a public aid agency and a philanthropic or private sector entity. Partners share a vision and values and often may jointly finance a project.	USAID - Rockefeller Foundation Global Resilience Partnership
Global multi-stakeholder initiatives	Collaborations involving three or more entities from the private, public and civil society sectors intended to address complex development challenges in a nontraditional or innovative way.	Child Protection Knowledge and Information Network: an initiative by UNICEF, police, governments and universities in Sierra Leone
Innovative financing	Use of nontraditional, innovative financing instruments to complement traditional development assistance.	Results- or performance-based financing

2.2 Outcomes

Given the broad, cross-sectoral nature of this topic, we chose outcome categories to capture the levels of analysis for the measured effects and to represent the full range of development sectors. This map thus catalogues the same studies in two different ways. The left-most columns of the map (see the map in section 3.2) catalogue the

studies by units of analysis for which they measure outcomes. The middle columns of the map reflect the same information, organised instead by the sectors for which they measure outcomes. For example, a study that measures the impact of SMS reminders on individual health outcomes will be coded in the 'individual and household outcomes' column and in the 'global health' column.

Table 2 presents the outcome categories for each group, with examples.

Table 2: Outcome categories

Outcome	Example
Levels of analysis	
Individual and household outcomes	Patient adherence to medication, researcher's academic output, household consumption
Organisational outcomes	Firm-level profits, health facility's productivity, school-wide average test scores
Community and societal outcomes	Changes in regulation, commodity prices, village-wide disease prevalence
Sectors	
Education and academia	Test scores, numeracy levels, school attendance
Global health	Adherence to treatment, risk of disease, health knowledge, sexual health
Democracy, human rights and governance	Electoral participation, election fairness, government accountability, human rights, and civic engagement
Agriculture and good security	Agricultural production, crop prices, food security
Crises and conflict	Disaster relief, post-conflict reconstruction
Economic growth, finance and trade	Firm profits, employment levels, research and development expenditures
Environment and global climate change	Changes in land regulation, recycling behaviour
Water and sanitation	Access to village water and sanitation resources
Energy	Household energy expenditures, energy/lighting usage

2.3 Crosscutting themes

We also analyse the evidence base according to four crosscutting themes:

- long-term impact;
- cost analysis;
- sex-disaggregated or sex-specific analysis and
- vulnerable or marginalised populations.

We include columns for these on the right side of the map so readers can easily understand the size of the evidence base that is reporting effects specific to these areas. The first column includes studies that include a measurement of long-term outcomes, which includes all studies that provide effect sizes for one or more time periods after endline. The cost analysis column reveals how many studies provide information on programme costs or compare costs to impact. Understanding costs is especially critical for decision-making in resource-limited settings. To be included in this column, a study must have some information about programme cost that can be compared to one or more of the measured net impacts.

The sex-disaggregated or sex-specific analysis includes studies that report effect sizes separately for women and men. These include evaluations of interventions that target only a single sex. Finally, we include a crosscutting theme for vulnerable or marginalised populations. We include here studies that report effect sizes for conflict-afflicted populations, people living with disabilities, rural populations, orphans and vulnerable children and sexual minorities.

3. Supply of evidence: the STIP evidence gap map

3ie EGMs are thematic collections of evidence on the effects of policies and programmes. They provide an innovative approach for rapid knowledge transfer and capture, combining methods from other review and mapping approaches with data visualisation, using an interactive platform. A key feature of an EGM is the use of a framework of interventions and outcomes based on a review of the policy literature and consultation with stakeholders. Snilstveit *et al.* (2013) and Snilstveit *et al.* (2016) provide a detailed description of how the tool was created.

As described in section 2, the rows of the framework represent a list of the key interventions of the sector or thematic area of focus, and the columns cover the most relevant outcomes. We develop each EGM framework to capture the universe of interventions and outcomes within the theme of the map. The evidence that we catalogue in the map includes impact evaluations and systematic reviews. Impact evaluations use counterfactual analysis to measure the net impact of an intervention (3ie 2012). Systematic reviews use transparent and systematic methods to identify, appraise and synthesise findings from studies addressing a specific question or issue (Waddington *et al.* 2012). When we use the term ‘evidence’ in this report, we are speaking of impact evaluations, or primary studies, and systematic reviews, or synthesis studies.

3.1 Methodology

We populate EGMs using systematic search and screening methods to identify impact evaluations and systematic reviews corresponding to the concepts included in the framework. The 3ie report *Science, technology, innovation and partnerships for development: an evidence gap map* (Sabet, Heard and Brown 2017) is the formal supporting documentation for this STIP EGM and presents the detailed methodology used to build this map, along with the complete search and screening protocols in the appendixes. We summarise the methodology here.

In developing the STIP EGM, our first step was to determine the scope of the map, which is represented by the framework described in section 2. We used this framework to develop a comprehensive search strategy and then conducted an extensive search of published and grey literature to find impact evaluations and systematic reviews within the scope. We conducted a search of 12 databases, 32 websites and 3 impact evaluation registries in June and July 2016. We limited our search to studies dated 1990 and later, broadly corresponding with the period when impact evaluations in the sector started to emerge.

We conducted our search in English, but when we happened to capture studies in Spanish, French or Portuguese, we also screened these for inclusion. We searched for STIP and impact evaluation terms, L&MIC country identifiers, and a large set of key terms such as ‘mobile money’, ‘results-based financing’, ‘randomised controlled trial’, and ‘least developed country’.

Our search strategy included three types of searches: publication database searches, targeted searches of specialised websites and databases, and backward and forward snowballing in which references of included studies and the online curricula vitae and websites of authors with at least one included study are checked for additional studies.

We then used the screening protocol to screen results by title, abstract and full text levels. To be included, studies must be impact evaluations or systematic reviews, they must evaluate a STIP-related intervention and the intervention must be conducted in an L&MIC. The screening criteria for identifying a study as an impact evaluation required screeners to look specifically for the following types of methodologies:

- randomised controlled trials;
- regression discontinuity design;
- before and after study using appropriate methods to control for selection bias and confounding variables (propensity score matching or other matching methods), instrumental variable estimation (or other methods using an instrumental variable such as the Heckman Two Step approach), difference-in-differences or a fixed or random effects model (with an interaction term between time and intervention for baseline and follow-up observations);
- cross-sectional or panel studies with an intervention and comparison group using methods to control for selection bias and confounding as described above; and
- studies explicitly described as systematic reviews and reviews that describe methods used for search, data collection and synthesis and that meet the confidence cut-off described below.

The screening criteria for identifying a study as a systematic review requires screeners to look at methodology using the 3ie rating tool that assesses the methodologies used in studies that claim to search, screen and synthesise evidence from individual studies. This tool rates reviews as having low, medium or high confidence in findings based on the methodologies they use. We include as

systematic reviews in this EGM only those reviews that use methodologies resulting in a medium or high confidence in findings. We used this parameter to approximate a similar methodological quality criterion as for the impact evaluations.

To avoid screener bias, at least two reviewers screened each hit. Next, we coded the included studies and populated the map. A second researcher verified the coding for each study. The coding instructions and template can be found in Sabet, Heard and Brown (2017).

3.2 Results

The search and screening resulted in 320 completed impact evaluations,[3] 77 ongoing impact evaluations,[4] and 7 completed systematic reviews. We found no protocols for upcoming systematic reviews related to STIP interventions. Sabet, Heard and Brown (2017) present the bibliographies of all the included impact evaluations, all the ongoing and announced impact evaluations and all the completed systematic reviews. Figure 1 presents the detailed results of the search and screening. Of the 320 completed impact evaluations, 212 are published journal articles (mostly peer-reviewed) and 64 are working papers published on institutional websites. We found 26 draft papers, 14 doctoral or master's dissertations, 3 project reports and 1 book chapter.

We present a picture of the completed EGM as Figure 2. The picture format shows the number of studies that provide evidence for each cell. The darker cells represent those with more evidence. It is important to note that the map shows only where there is evidence, not what the evidence says. Therefore, it is incorrect to interpret a dark cell as meaning that there is a lot of evidence supporting a positive impact of the intervention on the outcome. The evidence may actually show negative effects or null effects or be inconclusive. A dark cell means only that there is a deeper base of evidence for the effect of that intervention on that outcome.

[3] A study is considered complete if it has a published report or is in draft form.

[4] Ongoing impact evaluations were available as pre-registrations, published protocols or pre-analysis plans. Announcements were noted on primary authors' personal websites or in their curricula vitae.

Figure 1: Search and screening results

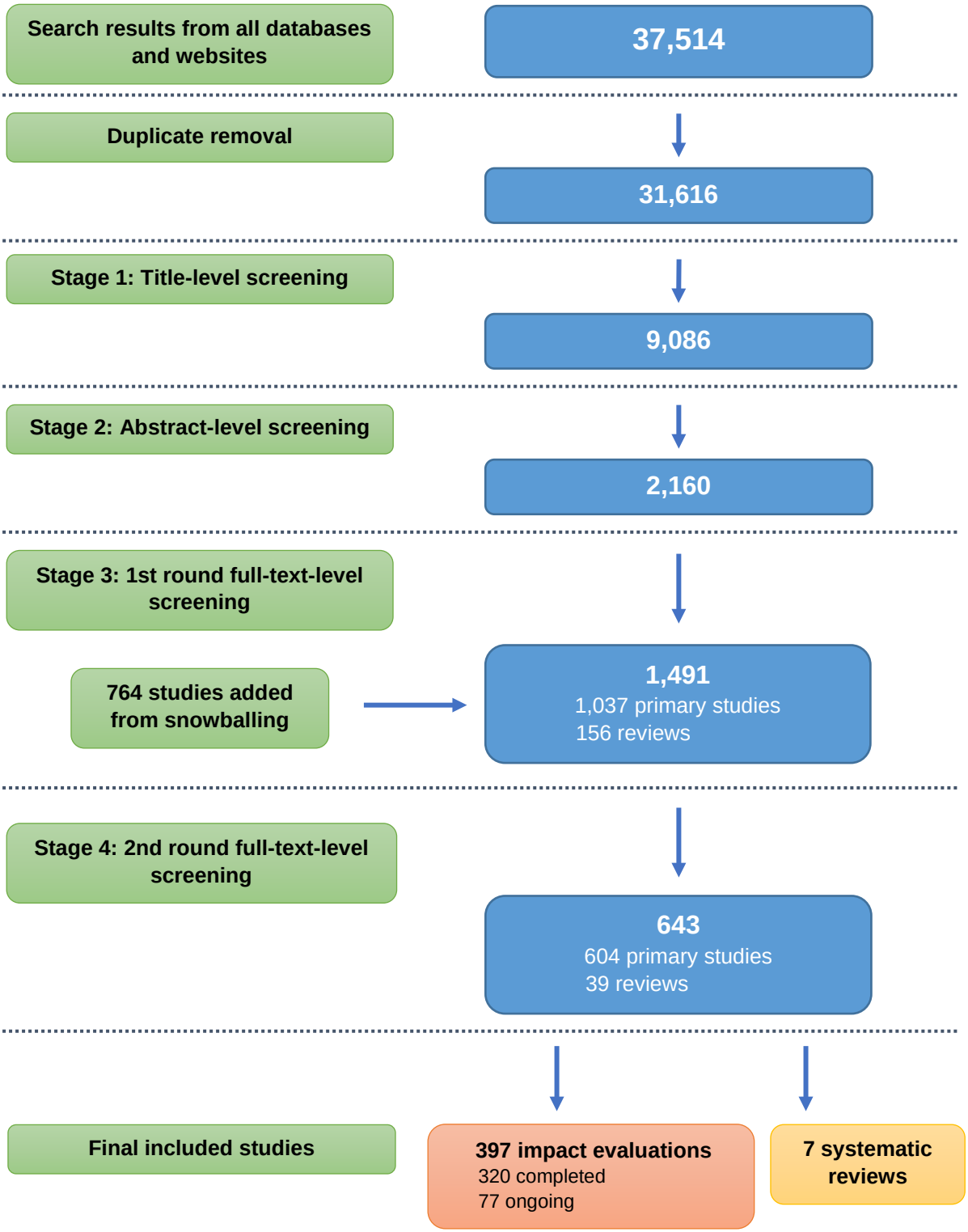


Figure 2: Evidence gap map of completed STIP impact evaluations

	Intervention categories	Levels of analysis			Sectors									Cross-cutting themes			
		Individual & household outcomes	Organisational outcomes	Community & societal outcomes	Education & academia	Global health	Democracy, human rights & governance	Agriculture & food security	Crises & conflict	Economic growth, finance & trade	Environment & global climate change	Water & sanitation	Energy	Long-term impact	Cost-analysis	Sex-disaggregated or sex-specific	Vulnerable or marginalised populations
Science	Fellowships & research grants	8			8									3		1	
	Material resources for scientific research																
	Technical assistance for scientific research																
	Research exchanges & collaborations		2							2							
	Policy & regulation for scientific research	1				1											
	Educational programmes to promote STEM	37			37									1	3	11	1
Technology	Digital infrastructure development	2		3	1			3									
	Policy & regulation for digital services			1						1				1	1		
	Digital literacy	5			2	1	1	1							1	2	
	Digital inclusion	4	1			1	1	2								2	
	Digital finance	16	1					3		13				1	3		3
	e-Governance	6	3				6	1		1					2	1	1
	Digitising identity	6	2			3	1	1		1					2		1
	Data systems development	5	4			8				1				1	5	2	2
	Digital information services	23	3	2	4	6	4	10		3	1			1	9	3	3
	Technology assisted learning	18	2		16	2				1				1	1	3	
	Mobile health	128	6	2		134						1		7	15	50	15
Innovation ecosystems	Access to capital		16							16				3			
	Grants & subsidies		29							29				6	2	1	
	Policies & regulation that affect innovation		7							7				2			
	Networks & collaboration for innovation	1	4	1				1		5	1	1		2			
	Capacity building for innovation		7							7				1	2	1	
Partnerships	Two entity partnerships																
	Global multi-stakeholder initiatives																
	Innovative financing	20	9	2	2	24				1					6	9	2

The 320 completed impact evaluations produce 387 occurrences in the first three columns of the map. These columns categorise the outcomes measured in the studies by the units of analysis. A study can appear in more than one column if it reports effect sizes for outcomes from more than one column. An example is Cole and Fernando (2016), which measures the impact of a mobile advisory service on farmers' agricultural knowledge and the farms' productivity. We think of each cell in which we populate a study as an occurrence of evidence. In the middle section of the map, where the columns categorise the outcomes measured by sector, there are 377 occurrences of evidence. Additionally, a study may evaluate an intervention with multiple components that fall under two or more intervention categories. In this case, it would produce occurrences in more than one row. We can think of this as meaning that it reports two different types of evidence. An example is López, Reynoso and Rossi (2010), which evaluates the impact of a public fund in Argentina that provides credit and matching grants to private firms to improve innovation.

3.3 Features of the impact evaluation evidence base

In this section, we present the features of the evidence base for STIP based on analysis of the numbers of completed and ongoing impact evaluation studies in the gap map. We look at the number of studies by intervention and outcome categories, region, publication year and programme. We present additional analysis of the data from the EGM in Sabet, Heard and Brown (2017).

3.3.1 Impact evaluations by intervention

Figure 3 displays the number of completed impact evaluations by intervention group. The group with the overwhelming majority of studies is technology. Figure 4 presents the number of completed impact evaluation studies by each intervention category.

Figure 3: Completed impact evaluations by intervention group

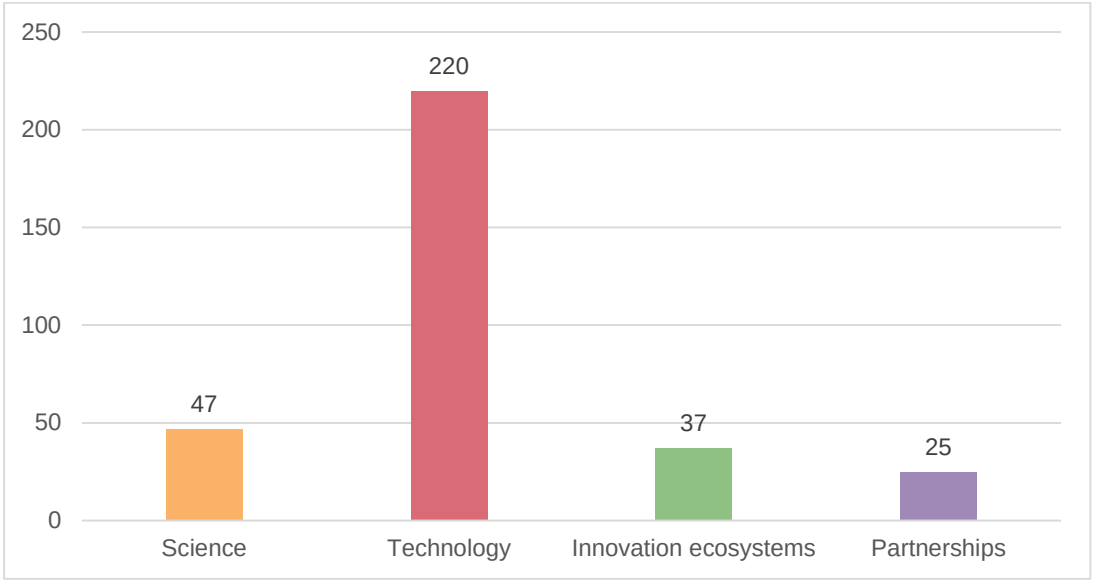
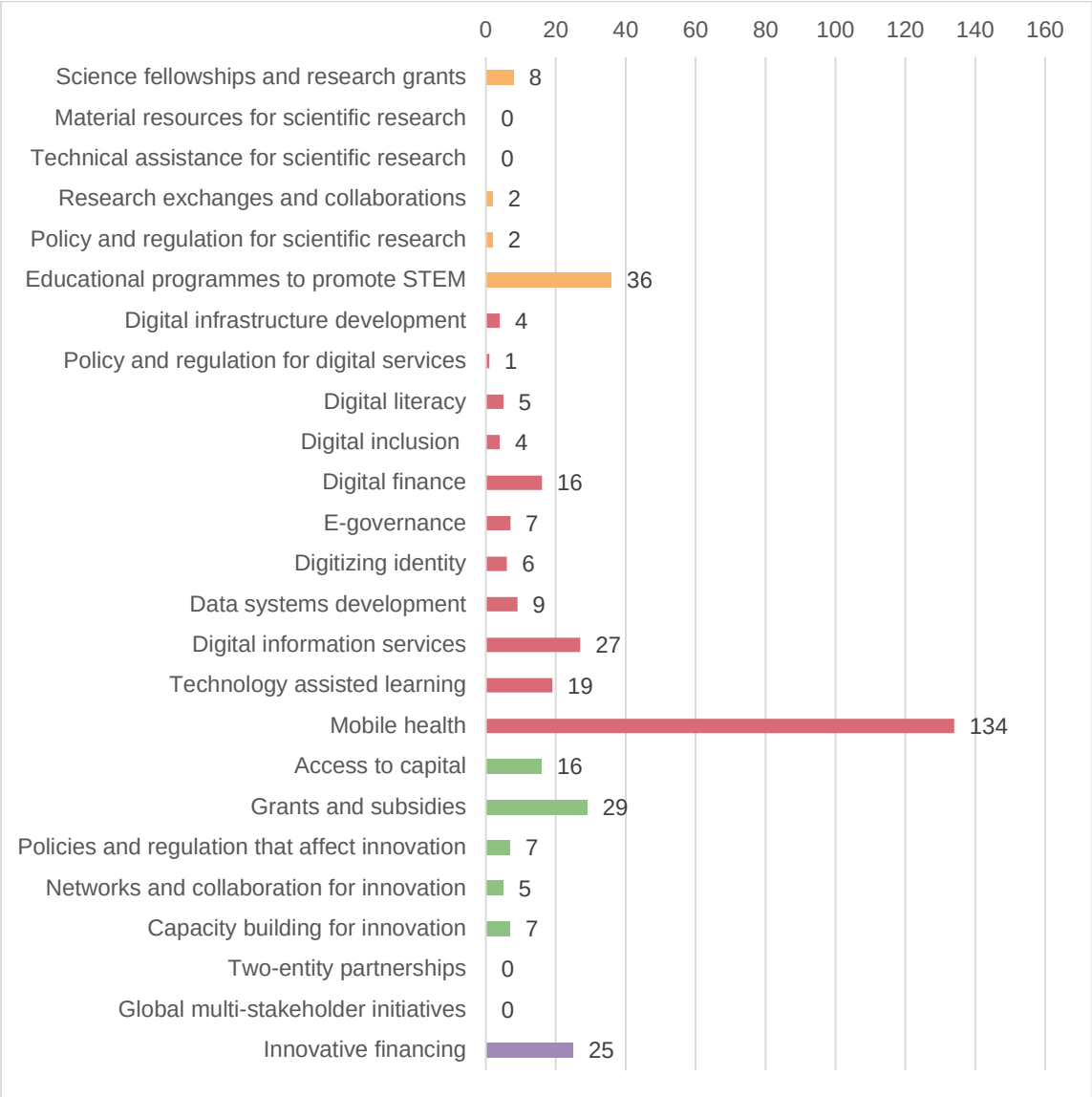


Figure 4: Completed impact evaluations by intervention category



The science intervention category with the most studies is ‘educational programmes to promote STEM’. These interventions include all programmes at the primary, secondary and tertiary level that encourage students to improve their science skills. Different pedagogical methods such as mastery learning and concept mapping intended specifically to encourage learning in the sciences are included here (e.g. Wambugu and Changeiywo 2008; Keraro, Wachanga and Orora 2007). The other prevalent science intervention category is ‘science fellowships and research grants’. These are publicly funded programmes intended to encourage researchers to produce scientific or technological research. We see gaps in the other science categories. In particular, there is little or no evidence on the impacts of material resources and technical assistance for scientific research, exchanges and collaborations amongst researchers, and policies and regulations intended to promote research.

Once we disaggregate the technology group by categories, the source of a large proportion of studies in the group becomes apparent. Of the 220 studies that fall within the technology group, 134 (61 per cent) are m-health related. Most of these studies measure the impacts of mobile phone messages on individual health outcomes. Of these m-health studies, 24 target people living with HIV or are related to HIV prevention. For example, Mbuagbaw *et al.* (2012) evaluate the impact of medication SMS reminders to HIV-positive patients in Cameroon on their adherence to antiretroviral therapy (ART). Other SMS-based studies simply provide health-related information and educational materials to recipients (e.g. Jamison, Karlan and Raffler 2013). Still other m-health studies evaluate the use of mobile devices for improving the quality and efficiency of health facilities. For example, Yu *et al.* (2009) measure the impact of using personal digital assistants to collect patients' health data on data entry error and processing times.

The technology intervention category with second largest concentration is 'digital information services'. The majority of studies coded under this intervention category evaluate SMS interventions intended to improve information asymmetry or nudge behaviour. Common examples include regular SMS messages to farmers with information about weather conditions or regional crop prices (e.g. Cole and Fernando 2016).

Other common technology interventions include financial services that are delivered via mobile phones, which are coded under digital finance. These frequently include SMS messages to individuals encouraging them to repay their loans or reminding them to save (e.g. Karlan, Morten and Zinman 2012) or mobile money applications that facilitate cash transfers and e-payments (e.g. Munyegera and Matsumoto 2016).

Each intervention category under innovation ecosystems is represented by at least five studies. As previously described, these are interventions intended to create the enabling environments in which innovation can thrive. A large number of studies (n=32) evaluate the impacts of publicly funded grants or subsidies to private firms on innovative outputs, profitability, or productivity (e.g. Castillo *et al.* 2014). Other studies examine the effects of policies and regulation, such as tax incentives, on private sector innovation (e.g. Avellar and Alves 2008).

Stakeholders' interest in the effectiveness of business incubators and accelerators motivated us to search specifically for impact evaluations on such interventions. A number of studies evaluate the impacts of seed money (access to capital, grants and subsidies), business networks and business training for new firms, or a combination of them. Three studies (Lopez-Acevedo and Tinajero 2010; Pires *et al.* 2014; McKenzie 2015) evaluate programmes that fit the description of an incubator, though they are not explicitly described as such.

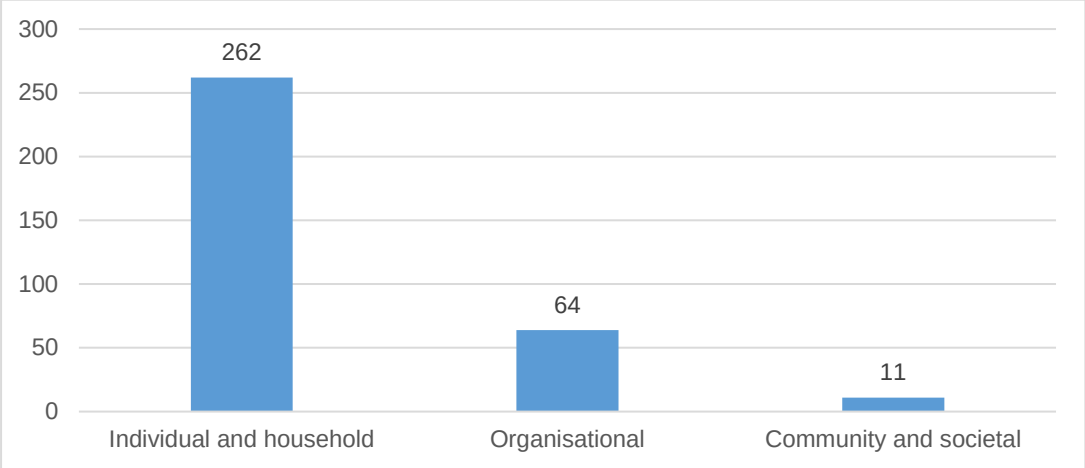
Finally, all of the studies in the partnerships group fall under 'innovative financing'. All 25 studies plotted under 'innovative financing for development assistance' evaluate the effectiveness of results-based financing (RBF), performance-based financing, or pay-for-performance financing mechanisms. These include schemes in which, for example, a health facility receives blocks of payments based on the health of its

patients or the quality of its service (e.g. de Walque *et al.* 2015). Other performance-based financing interventions provide bonus payments to teachers based on student test scores (Muralidharan and Sundararaman 2011) or money to the governing body of an entire village based on aggregate educational and nutritional indicators of the village's children (Olken, Onishi and Wong 2014).

3.3.2 Impact evaluations by outcome

As described earlier, each study appears at least once under the levels of analysis group and at least once under a sector. Figure 5 presents the number of completed impact evaluations by levels of analysis.

Figure 5: Completed impact evaluations by levels of analysis

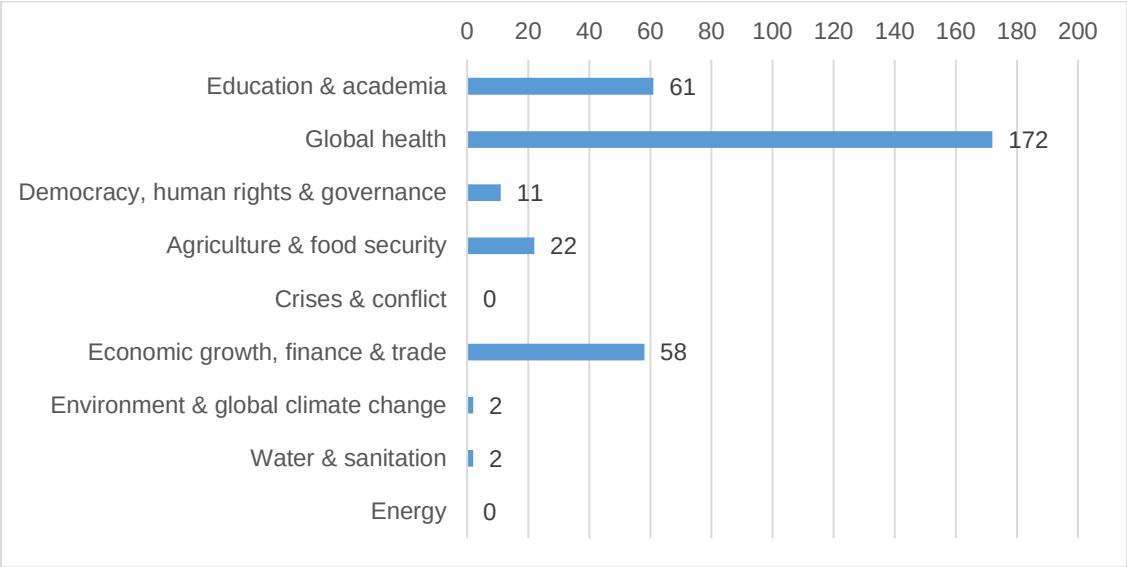


The majority of studies measure individual- or household-level outcomes. These results are not surprising, as most interventions can be easily randomised at the individual level and many researchers rely on household or individual survey data to conduct research. Examples of such outcome types may include individual-level health indicators, student test scores or a researcher's academic output.

Most organisational outcomes pertain to private firms (e.g. profits, productivity and research and development [R&D] input), while some pertain to school-level indicators (e.g. dropout rates) or facility-level indicators (e.g. quality of care). The community and societal outcomes represented in this map include indicators such as agricultural commodity price dispersion and market prices (e.g. Aker and Fafchamps 2014) and village-wide vaccination coverage or prevalence of childhood illness (e.g. Priedeman Skiles *et al.* 2013).

Figure 6 presents the number of completed and ongoing impact evaluations by sector. Most studies (n=172) are coded under 'global health'. This is consistent with our m-health findings, presented above, and with findings from a recent analysis of 3ie's Impact Evaluation Repository (a database of more than 4,000 published impact evaluations), which show that a large proportion of published impact evaluations are health related (Miranda, Sabet and Brown 2016). A large number of studies also fall under 'education and academia' and 'economic growth, finance and trade'. We found no impact evaluations coded under 'crises and conflict' or 'energy'.

Figure 6: Completed impact evaluations by sector



3.3.3 Impact evaluations by crosscutting theme

As Figure 7 illustrates, 79 completed studies report sex-disaggregated or sex-specific effect sizes, though this number is largely driven by the former. In spite of the sex-disaggregation, the analysis in these studies did not include detailed analysis of gender. Twenty-two studies – most of which fall under the innovation ecosystems group – report the long-term impact of an intervention, and 45 studies discuss costs or conduct some form of cost analysis. Only 27 studies report effect sizes for vulnerable or marginalised populations. We code studies under this column if they report effect sizes for conflict-afflicted populations, people living with disabilities, rural populations, orphans and vulnerable children and ethnic or sexual minorities. As Figure 8 indicates, rural populations make up the majority.

Figure 7: Completed impact evaluations by crosscutting themes

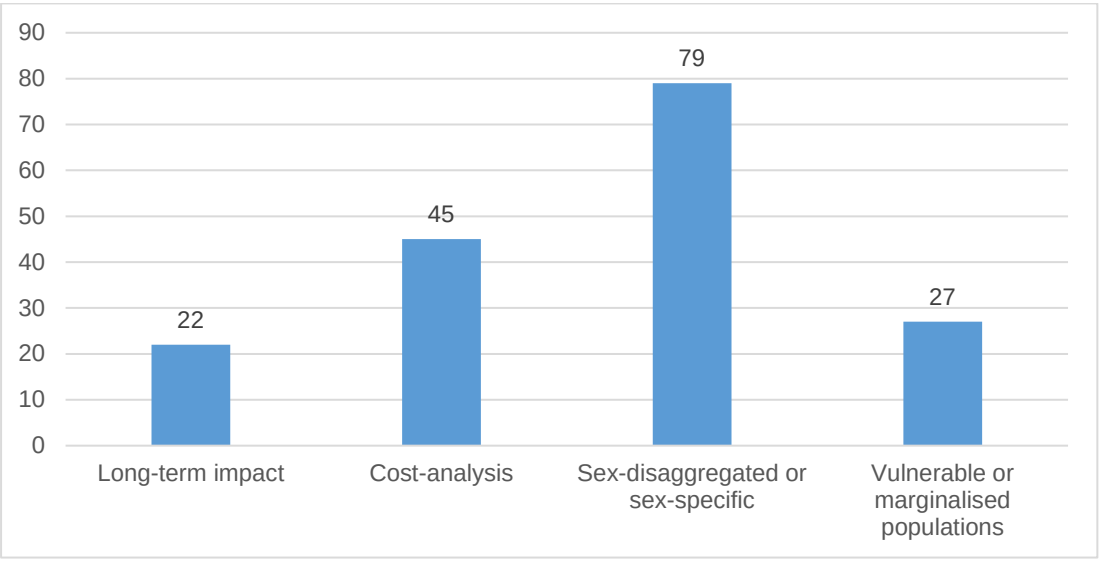
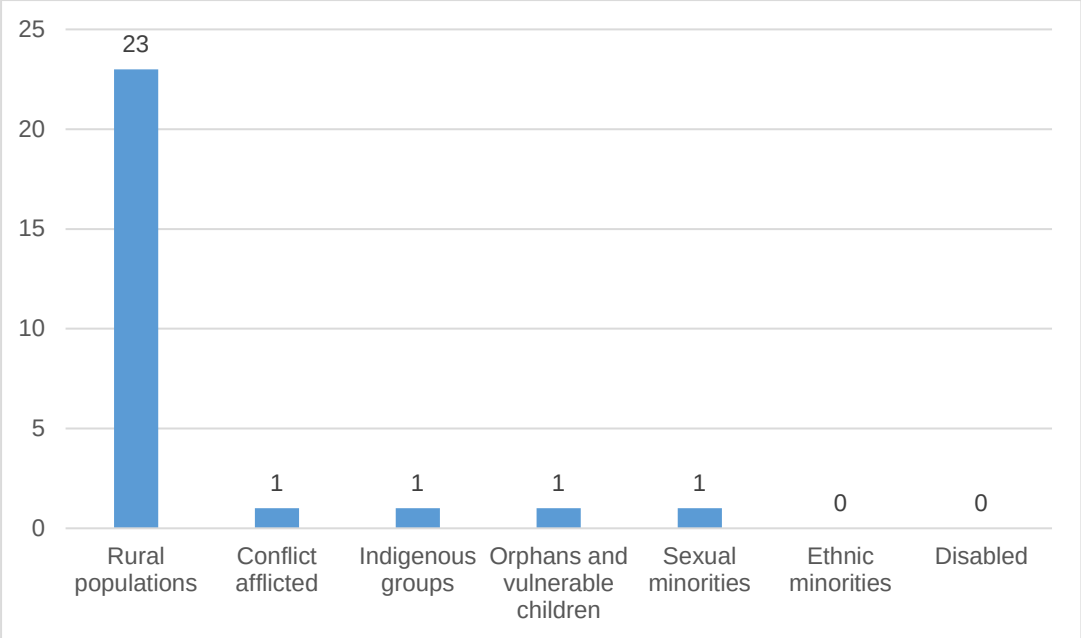


Figure 8: Completed impact evaluations by vulnerable or marginalised populations



3.3.4 Impact evaluations by geography

As Figures 9 and 10 indicate, the majority of studies (n=128) are conducted in Sub-Saharan Africa. Seventy-six studies are conducted in Latin America and the Caribbean, while 48 are conducted in East Asia, South-East Asia and the Pacific. The countries with the most evidence are Kenya and India, followed by China and South Africa. We find a dearth of evidence in French-speaking Africa. This may be due to the language limitations in our search strategy. In south Asia, there are no studies in the science and innovation ecosystems groups, and few studies across all groups are conducted in the Middle East and North Africa.

Of the 33 impact evaluations conducted in Kenya, 14 are education programmes for STEM, 11 are m-health and the remaining 8 are other technology-related intervention categories. Twenty-nine impact evaluations are conducted in India. All but 4 of these are in the technology group, 11 of which are m-health. The four non-technology studies are innovative finance interventions in the partnerships group. Twenty-eight impact evaluation studies are conducted in China. Similar to the Kenyan and Indian studies, many of the studies (n=20) are m-health related. Five of the Chinese studies evaluate education programmes for STEM; one is innovative financing and the rest are either data systems or technology-assisted learning.

Online appendix A contains a breakdown, by intervention category, of each country in which 10 or more impact evaluations are conducted.

Figure 9: Completed impact evaluations by region

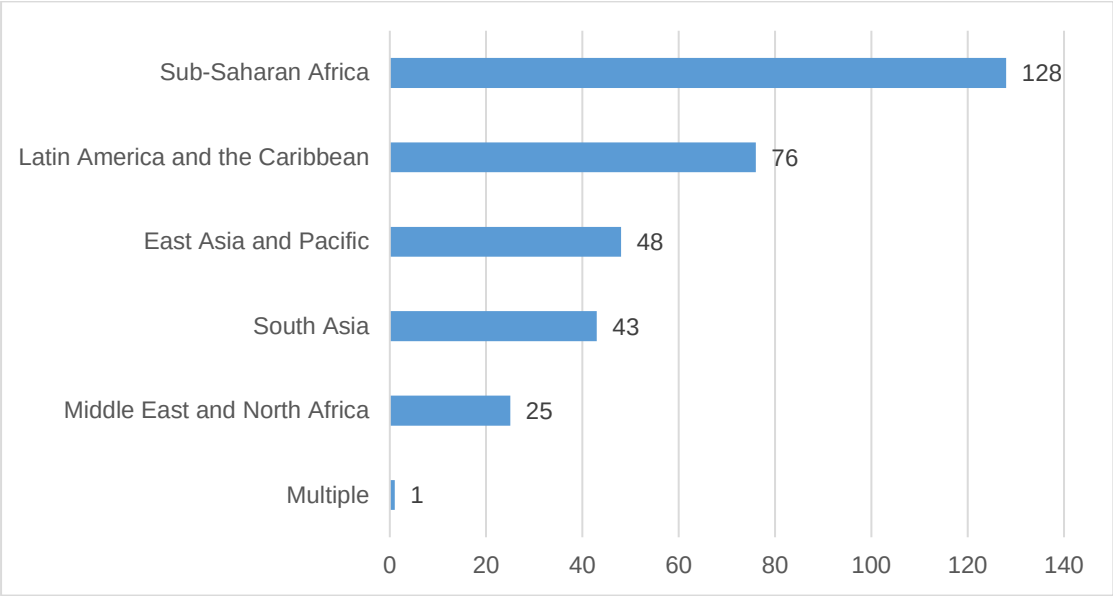


Figure 10: Completed impact evaluations by country

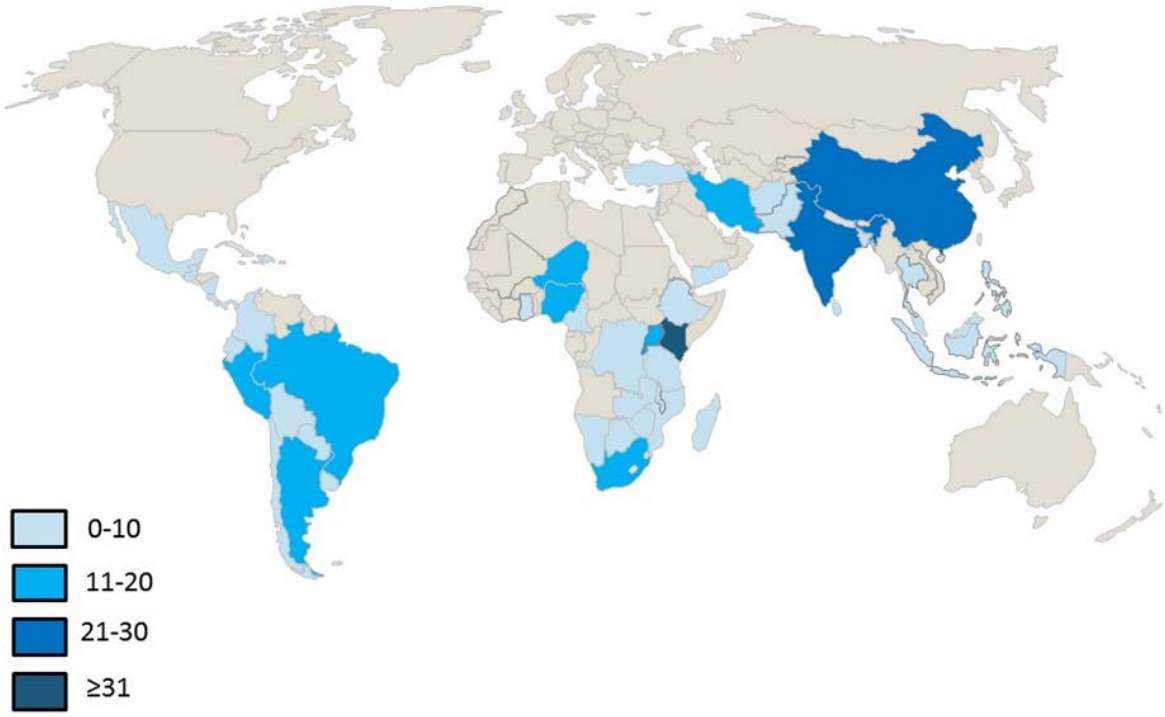
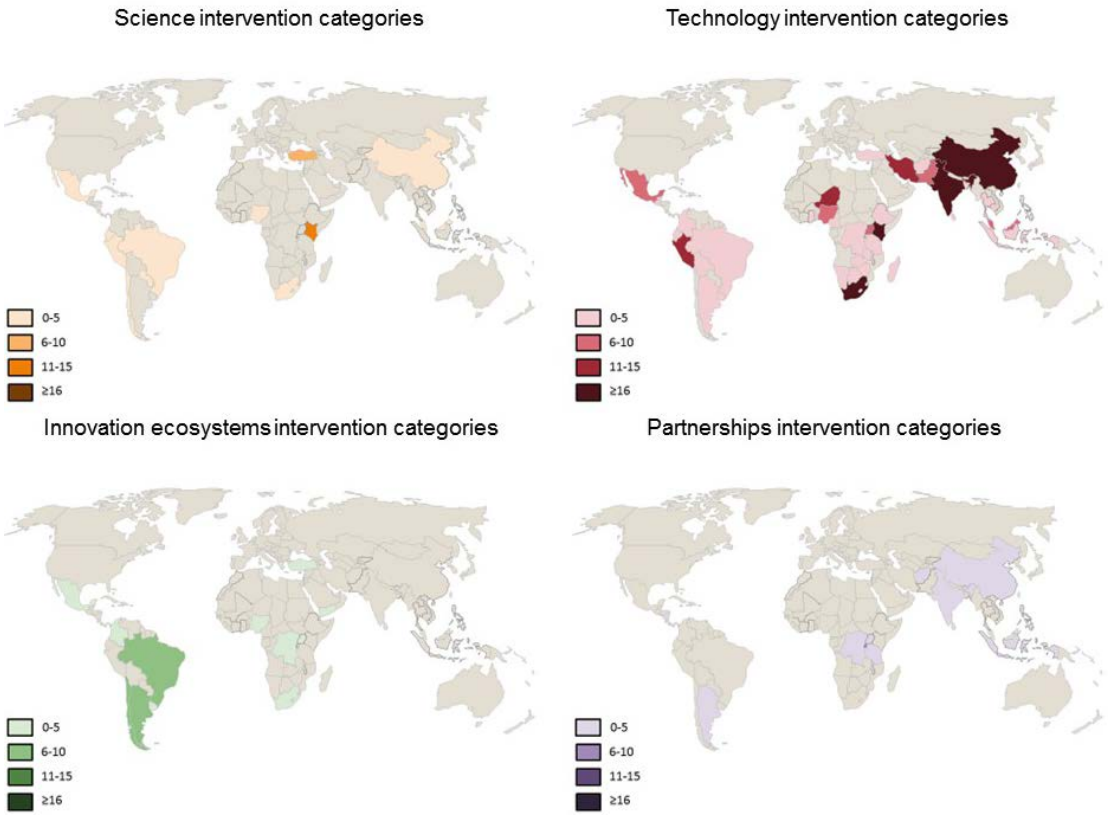


Figure 11 disaggregates the geographical spread of impact evaluations by each STIP group. A large share of innovation ecosystems studies are conducted in Latin America and the Caribbean (n=31). A majority of the partnerships and technology studies are conducted in Sub-Saharan Africa (n=15 and n=92, respectively). As we discuss in later sections, the amount of evidence for each group and region is consistent with the amount of programming in the regions.

Figure 11: Completed impact evaluations by country and STIP group



3.3.5 Impact evaluations by programme

When coding the studies included in the map, we also noted when there were multiple evaluations of the same programme, whether by different authors, in different time frames or reporting on different outcomes. If we found more than one version of the same evaluation with the same reported results (e.g. if a study had a working paper and a journal article), we included only one of them. Two programmes, the Argentinian Technological Fund (FONTAR) and Project ABC in Niger, are each evaluated in five different studies. FONTAR is a national Argentinian fund for private sector 'projects oriented to research and development, pilot scale technologies, applied knowledge generation, innovative products, and process development' (López, Reynoso and Rossi 2010). Project ABC uses mobile phones as a tool to promote adult literacy and numeracy in Niger (Project ABC n.d.). The five studies that evaluate Project ABC are all led by the same author, Jenny Aker, who analyses the same data in several different ways (Aker, Ksoll and Clemens 2011; Aker, Ksoll and Lybbert 2012; Aker and Ksoll 2012; Aker and Ksoll 2015; Aker and Ksoll 2016).

Another programme, the Chilean National Fund for Technological and Productive Development (FONTEC), is evaluated in four different studies (Benavente and Maffioli 2007; Tan 2009; Alvarez, Crespi and Cuevas 2012; Alvarez, Bravo and Zahler 2013). Several programmes have been evaluated two or three times in different studies. Other programmes have multiple analyses, as they also appear in our EGM of ongoing impact evaluation studies.

3.3.6 Ongoing impact evaluation studies

We identified 77 ongoing impact evaluation studies. More than half of these (n=49) are registered as trials on an institutional website (e.g. the American Economic Association or 3ie's Registry or International Development Impact Evaluations), and the remaining are published as protocols in journals. The map of these 77 studies in appendix A of this document provides a glimpse, though not a complete picture, of the direction of future impact evaluation studies of STIP-related interventions.

The pattern of distribution by STIP intervention group remains more or less unchanged in comparison to the map of completed impact evaluations, with the exception of a slightly lower representation of science studies (3 per cent of ongoing studies versus 15 per cent of completed studies) and innovation ecosystems studies (1 per cent versus 19 per cent). Of the six ongoing impact evaluations in the partnerships group, five fall under the innovative financing category (all of which are RBF) and one is an evaluation of a multi-stakeholder consortium that leverages mobile-driven solutions to promote children's welfare (Sheely *forthcoming*).

The pattern by outcome types is also similar to that of the completed impact evaluations. The majority of ongoing impact evaluations measure outcomes at the individual or household level and fall under the global health sector. The pattern of crosscutting themes is also not substantially different on the EGM of ongoing impact evaluations (appendix A). Vulnerable or marginalised populations continue to remain under-represented, and analyses of long-term impacts and costs are scarce. Although a smaller proportion of study proposals specifically state they will disaggregate outcomes by sex than do completed studies, this may not be an accurate representation of what they will actually report. Amongst the studies that provide effect sizes for vulnerable or marginalised populations, the majority (n=5) do so for rural populations, while the others provide effect sizes for conflict-afflicted populations, orphans and vulnerable children and ethnic minorities.

Ongoing impact evaluations continue to be conducted in the same regions. The majority of studies are conducted in Sub-Saharan Africa (n=43), followed by studies conducted in South Asia (n=15). One finding to note is that there are fewer studies being conducted in Latin America and the Caribbean. This is possibly due to the transition of many South American countries to high-income status.

3.4 Features of the systematic review evidence base

There are only seven completed systematic reviews that meet our inclusion criteria. To be included, a systematic review must meet all the relevance and methodology criteria explained above. This is a small number of systematic reviews, given the density of completed impact evaluations we found. The seven completed systematic reviews in our EGM are all health related (Table 3). Six relate to m-health. One of the six also includes studies that evaluate the use of mobile devices to gather health-related data, which falls under data systems development (Lee *et al.* 2016). The seventh systematic review falls under innovative financing.

Table 3: Included systematic reviews

Short citation	Intervention category	Intervention details	Sector	Statistical meta-analysis
Lee *et al.* (2016), m-health interventions for maternal, newborn and child health	Data systems development	Mobile device use to gather data about pregnancies, birth weights, and diagnosis	Global health	Yes
	M-health	SMS reminders and education for maternal and neonatal health		
Sondaal *et al*. (2016), m-health interventions for improving maternal and neonatal care	M-health	SMS for improving maternal and neonatal health	Global health	No
van Velthoven *et al*. (2013), Mobile phone messaging interventions for HIV/AIDS care	M-health	SMS for HIV infection prevention, treatment and care	Global health	No
Arambepola *et al*. (2016), Automated brief messaging interventions to promote lifestyle changes	M-health	SMS to encourage lifestyle changes for diabetics	Global health	Yes
Horvath *et al*. (2012), Text messaging interventions for promoting adherence to antiretroviral therapy	M-health	SMS for ART adherence	Global health	Yes
Beratarrechea *et al*. (2014), Mobile health interventions for treating chronic diseases	M-health	All mobile health interventions targeting chronic diseases	Global health	No
Lagarde and Palmer (2009), The impact of contracting out health services on health outcomes and use of health services in L&MICs	Innovative financing	Contracting out health services	Global health	No

As noted above, EGMs do not report the findings of the impact evaluations catalogued in the framework or attempt to synthesise the findings for any given cell. Rather, the objectives of the map are to analyse the size and scope of the evidence base and provide policymakers easy access to relevant impact evaluations. However, the systematic reviews in the EGM represent cases where a careful synthesis of the findings from individual studies for questions related to the interventions in the framework has already been conducted. Therefore, we do present the results from the seven included systematic reviews, along with an analysis of how well the systematic reviews cover the EGM impact evaluations.

Lee *et al.* (2016) conduct a systematic review of effectiveness studies on m-health interventions for maternal, newborn, and child health in L&MICs. Six of their included studies (eight reports) also appear in the EGM (Flax *et al*. 2014; Jareethum *et al*. 2008; Khorshid, Afshari and Abedi 2014; Lin *et al*. 2012; Lund *et al.* 2012; Lund *et al.* 2014a; Lund *et al.* 2014b; Sharma *et al*. 2011). They report that only one study shows improvements in morbidity or mortality; it shows a decreased risk of perinatal death with SMS support, as compared to routine care. The authors perform a meta-analysis of three studies and find that rates of breastfeeding within one hour after birth are higher amongst mothers who receive an SMS intervention. The meta-analysis also finds that groups that receive an SMS intervention are significantly more likely to exclusively breastfeed for three to six months after birth.

Sondaal *et al.* (2016) review studies that evaluate the effectiveness of m-health interventions on improving maternal and neonatal care in L&MICs. The authors include two studies (five papers) that are included in the EGM (Khorshid, Afshari and Abedi 2014; Lin *et al*. 2012; Lund *et al.* 2012; Lund *et al.* 2014a; Lund *et al.* 2014b); these papers also appear in Lee *et al*. (2016). Given heterogeneity in intervention types, the authors were not able to conduct a meta-analysis. All studies that addressed the impact of m-health on the use of maternal and neonatal service utilisation show significant positive increases. No studies reporting effects on maternal morbidity and mortality are included in their review. A few studies do report on maternal anaemia, gestational age at delivery and mode of delivery, but all find no effect for groups that receive an m-health intervention.

Van Velthoven *et al.* (2013) assess the effectiveness, acceptability and feasibility of using text messaging for HIV prevention, treatment and care. The review includes two studies that are also in the EGM (Lester *et al*. 2010; Pop-Eleches *et al.* 2011). They do not conduct a meta-analysis of the results due to heterogeneity in the intervention types. The authors find that while text messaging is an acceptable way to receive information and communicate with health workers, few studies show a clear benefit. The authors report weak study designs and inadequate reporting resulting in an inability to make conclusions, but caution against using their review as evidence of no effect.

Horvath *et al.* (2012) assess the evidence for promoting adherence to ART with mobile phones. The authors include two randomised controlled trials conducted in Kenya (Lester *et al*. 2010; Pop-Eleches *et al.* 2011), both also included in the EGM

and in van Velthoven et al. (2014). A meta-analysis of the two studies suggests that any weekly text messages is associated with lower risk of non-adherence at 48–52 weeks.

Arambepola *et al.* (2016) assess the use of automated messages for promoting healthy eating and exercises to improve glycaemic control amongst type 2 diabetes patients. The review includes three studies that are also in the EGM (Goodarzi *et al*. 2012; Kamal *et al.* 2015; Tamban, Isip-Tan and Jimeno 2013). The trials included a mix of unidirectional and bidirectional SMS interventions. The systematic review authors report that two studies found no effects of unidirectional SMS on behaviour change, while two others reported significant improvements. The meta-analysis included 13 studies that measure the impacts of unidirectional and bidirectional SMS on haemoglobin levels and finds a significant effect. The systematic review authors do not find a significant impact on body mass index.

Beratarrechea *et al.* (2014) assess the effects of m-health interventions on chronic disease outcomes in developing countries. Five of nine included studies also appear in the EGM (Leong *et al.* 2006; Chen *et al.* 2008; Liew *et al.* 2009; Balsa, Gandelman and Porzecanski 2010; Shetty *et al*. 2011). Some studies target asthma patients, while others target diabetics or people at risk of heart failure. Due to this heterogeneity, the authors could not conduct a meta-analysis. The authors also report that three studies looked at cost effectiveness and conclude that interventions providing SMS appointment reminders are more cost effective and just as efficacious as telephone interventions.

The final systematic review by Lagarde and Palmer (2009) assesses the effect of contracting out health services on health outcomes and utilisation of services in L&MICs. The authors find three studies that meet their inclusion criteria. The results from the studies are mixed. The authors state that drawing conclusions from these studies was difficult and that contracting, as an intervention, was complex, making impact assessment more difficult.

We note that of the six m-health systematic reviews, two relate to maternal and neonatal/child health, two to HIV and two to chronic diseases/diabetes.

4. Demand for evidence

To identify priorities for research investment, it is important to assess the current supply of evidence and the demand for evidence. All of the intervention categories included in the framework were identified by stakeholders as being relevant interventions for using science, technology, innovation ecosystems and partnerships to strengthen development outcomes, but stakeholders may seek more and better evidence for some more than others. Reasons for varying demand include, obviously, the perception of the current state of evidence, but may also include prior beliefs in underlying theories of change; usefulness of other kinds of evidence, such as outcome monitoring, in some contexts; and prevalence and trends in programming. In this section of the scoping paper, we present our assessment of the

demand for evidence across the intervention categories so that we can analyse priorities for new research investments based on both supply and demand.

4.1 Methods

Our methods for assessing the demand for new and better evidence incorporate data from several sources and employ qualitative analysis. We collected information from four sources. First, we draw on the information from the literature review we conducted to support the development of the framework. To include some quantitative data, we conducted a stakeholder survey. We gathered expert inputs from our roundtable event with STIP specialists, which included 31 participants, and from our advisory group members, who are listed in the acknowledgements. We also conducted informal portfolio analysis, primarily using internet searching, to look for the prevalence and location of current programming.

It is important to note that the survey was conducted before the EGM was populated, so the survey responses were in no way informed or influenced by our evidence mapping work. At the roundtable event, however, we did present to the participants a draft of the EGM and the results of the stakeholder survey prior to the in-depth discussions of evidence demands and priorities for investments.

4.1.1 Survey methods

We developed the survey around the framework described in section 2. We uploaded the survey to Survey Monkey and piloted it with the project's advisory group. We incorporated their feedback and modified the survey accordingly. A copy of the survey is available in online appendix B.

We distributed the survey through 3ie's social media channels, the advisory group's networks, 3ie members and the 3ie listserv. USAID also distributed the survey through its networks. To allow respondents to focus on their areas of expertise, 3ie structured the survey so that respondents could self-select into the science, technology, innovation and partnerships intervention groups. Therefore, only a portion of respondents answered each section.

Due to the iterative nature of developing a framework for a gap map, some intervention categories evolved over time, and two categories in the map are slightly different in the survey. First, 'exchanges, collaborations and partnerships' appears as two categories in the stakeholder survey ('exchanges' and 'collaborations and partnerships'). Second, 'digital information services' includes m-health in the survey, whereas in the gap map, m-health is its own category.

After requesting demographic information, the survey allowed respondents to opt in to one or more of the four intervention groups to answer questions about specific intervention categories. The survey asked three questions for the intervention categories in each group. The first question for each category asked how effective the respondent considered the intervention to be. The second question asked respondents to judge the strength of evidence for each category. The final question asked respondents to select up to three categories (for science and technology) and

up to two categories (for innovation and partnerships) in the group for which more and better evidence would be useful to their work. Unfortunately, because the groups are of unequal size in terms of the number of intervention categories, it is not straightforward to compare the responses to this third question across groups. We can compare the relative demand within the groups, however.

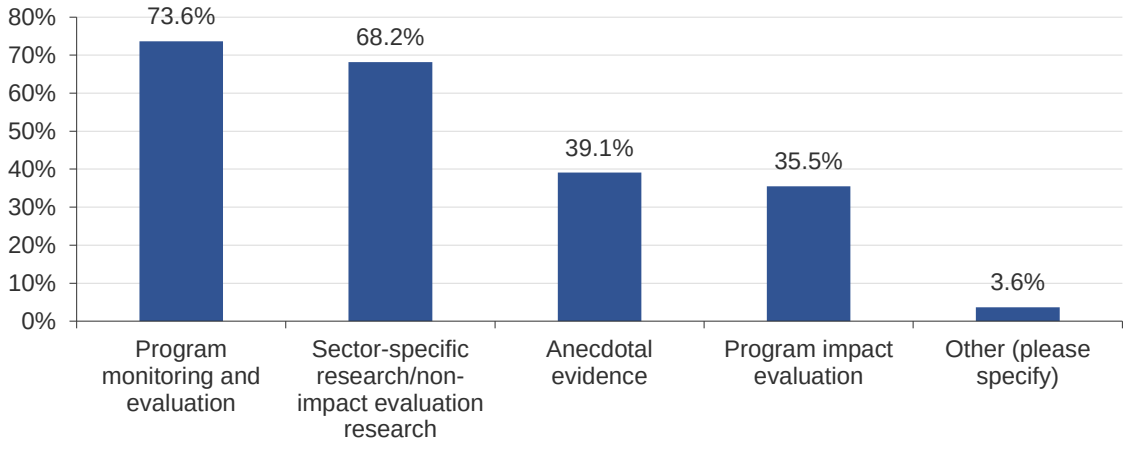
4.1.2 Survey results

The survey was open for three weeks and received 110 respondents working in STIP-related fields. All respondents were required to submit information about their work and experience, including the type of organisation for which they work, their role and their geographic experience. Approximately 28 per cent of respondents were based in L&MICs, and 22 per cent were L&MIC nationals.

A large number of respondents (41 per cent) were mid-level managers, and almost all of the rest were in senior leadership or associate-level positions, divided equally between those two categories. Half of the respondents have been working in STIP fields for more than five years. The majority work for development NGOs (30 per cent) or public aid agencies (35.5 per cent). However, respondents reported that they work on a wide variety of activities in their organisations. Half report working on monitoring and evaluation and technical assistance/capacity building. Fewer respondents report that they work on policy and regulation (23 per cent), higher education (22 per cent) or science programs (15 per cent). The majority (69 per cent) of respondents cite Sub-Saharan Africa as the region in which they have experience, followed in order by east Asia, south Asia, and Latin America and the Caribbean. Online appendix C includes figures for these demographic results.

To get a general sense of the types of evidence that respondents use in their work, we presented them with a list of types of evidence and asked them to choose up to two kinds of evidence that they use most often in their work. The majority use programme monitoring and evaluation (73 per cent) or non-impact evaluation research (68 per cent). Notably, respondents turn to anecdotal evidence slightly more often than they do to impact evaluation evidence.

Figure 12: Which of the following types of evidence do you use in your current work most often? Please select no more than two.



As noted above, the first question in the sections of the survey for each intervention group was whether respondents thought the intervention categories were effective at achieving intended outcomes. We included this question as a way to check whether responses for the questions that followed (about the current state of evidence and where more evidence would be useful) seemed to be moderated by respondents' views of whether an intervention was effective. We present the detailed results from this question in online appendix D.

We did not find any strong associations between perceived effectiveness and the other responses. If anything, it appears that people would like more evidence about those intervention categories that they consider to be more effective, which is actually encouraging. Such a correlation would suggest that stakeholders are using evidence in programme design in addition to using it just to identify what works. In general, between 20 and 40 per cent of respondents for each group considered the interventions in the categories to be usually effective. If we combine the 'usually effective' responses with the 'sometimes effective' responses, we find that 70–80 per cent of the respondents for that group consider most intervention categories to be one or the other. Outliers on the low end (with at least 30 per cent responding 'rarely effective' or 'don't know') are digital identity, e-governance, technology-assisted learning and global multi-stakeholder initiatives.

Figures 13 through 16 present the results for each intervention group for both the perceived strength of evidence for each category and the selection of categories where more and better evidence would be useful to stakeholders' work. The bars show the perceived strength of evidence. For example, Figure 13 shows that just under 20 per cent of respondents answering the science questions consider the evidence base for fellowships and research grants to be strong. Another roughly 45 per cent consider it to be moderate; 20 per cent consider it weak; and 15 per cent don't know. The lines in the figures show the relative desire for more and better evidence. Respondents could select up to three for the first two groups; up to two for innovation ecosystems; and were supposed to select only one for partnerships.[5] For example, Figure 13 shows that roughly 50 per cent of the respondents for the science group chose fellowships and research grants as one of the top three categories of interventions for which more and better evidence would be useful for their work.

[5] The Survey Monkey software did not allow us to build in these restrictions, so there are some cases where respondents selected more than the number asked for.

Figure 13: Stakeholder perceptions of strength of evidence and choice of intervention categories where more and better evidence would be useful in the science group

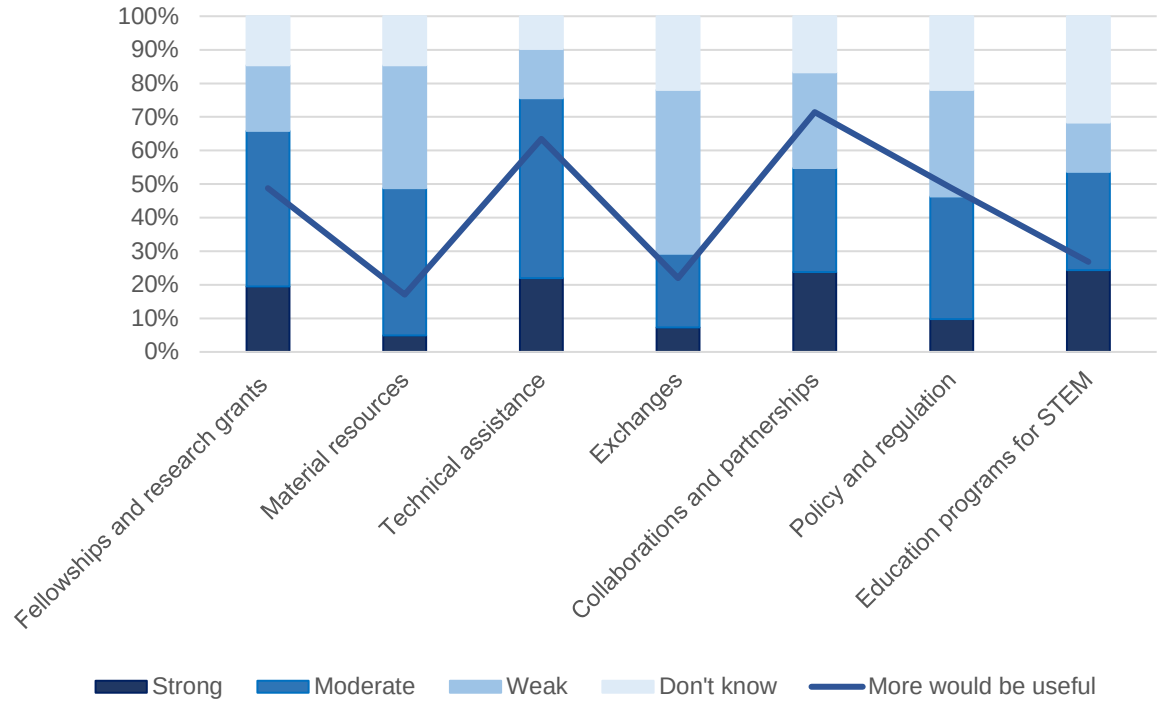


Figure 14: Stakeholder perceptions of strength of evidence and choice of intervention categories where more and better evidence would be useful in the technology group

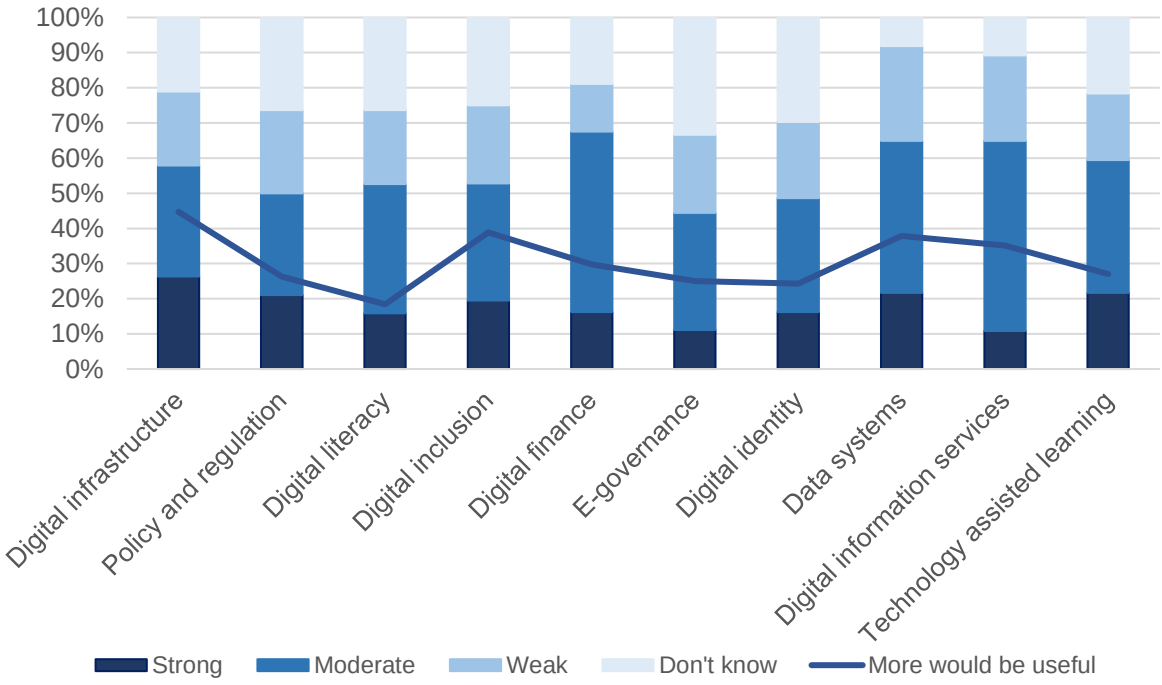


Figure 15: Stakeholder perceptions of strength of evidence and choice of intervention categories where more and better evidence would be useful in the innovation ecosystems group

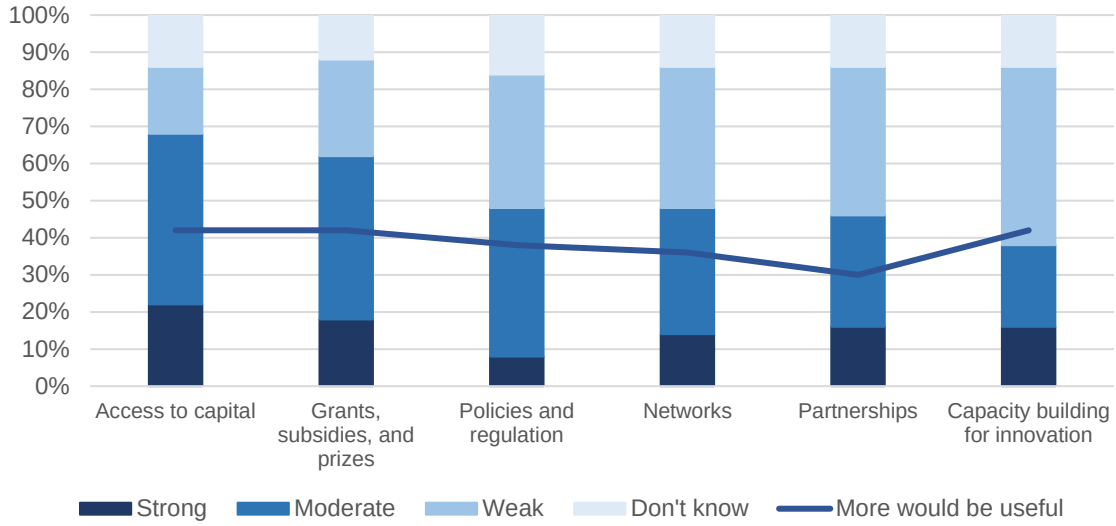
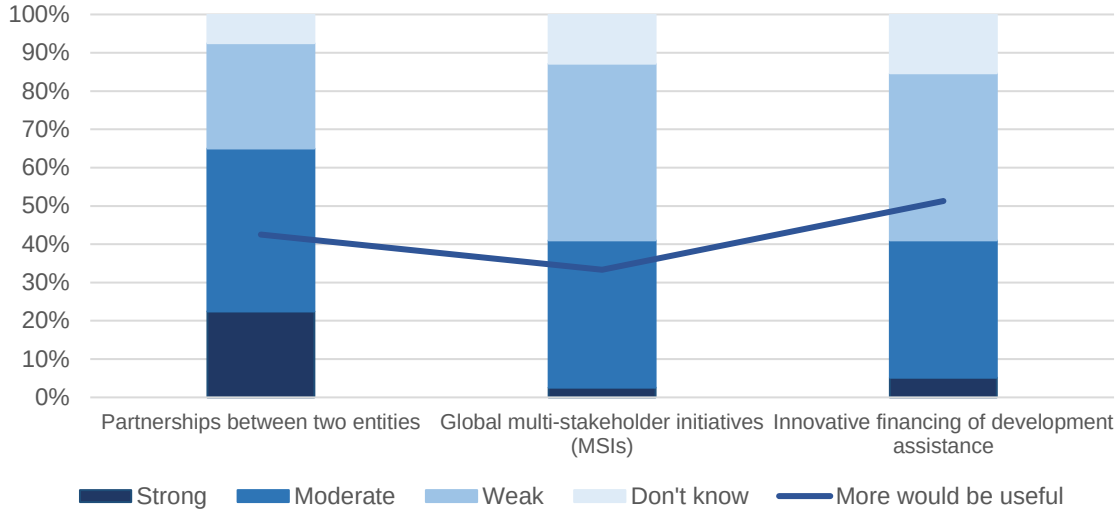


Figure 16: Stakeholder perceptions of strength of evidence and choice of intervention categories where more and better evidence would be useful in the partnerships group



The results from the stakeholder survey reveal that the evidence base for STIP-related interventions is not considered to be strong. The highest rates of ‘strong’ responses are around 20 per cent, and only nine of the 25 intervention categories have ‘strong’ rates at or above 20 per cent. If we group the interventions by those receiving a majority of ‘strong’ plus ‘moderate’ responses versus those receiving a majority of ‘weak’ plus ‘don’t know’ responses, we see a roughly equal split. Thirteen categories have a majority ‘strong’ plus ‘moderate’ and 12 have a majority ‘weak’ plus ‘don’t know’. Some might consider these results to be at odds with the findings, discussed above, that large percentages of respondents considered most

interventions to be usually or sometimes effective. This apparent disconnect suggests that stakeholders make their determinations of effectiveness based on many types of information, as reflected in Figure 12.

Next, we look at the survey question that asks respondents to select interventions for which more and better evidence would be useful for their work. We understand the response as an indication of demand for more evidence. We find that a greater percentage selects the following intervention categories from their group than if each category in the group were demanded equally[6]: fellowships and research grants, technical assistance for science research, research collaborations and partnerships, policy and regulation for science, digital infrastructure, digital inclusion, data systems development, digital information services, two-entity partnerships, and innovative financing. In addition, respondents selected all but the partnerships category under innovation ecosystems more often than for equal demand, meaning that respondents selected more than the allowed two categories. This over-response points to a high demand for evidence across this group. The demand in this group was roughly equal across the categories.

When we compare the perceived strength of evidence and the demand for more evidence, we do not see what might be an expected correlation – that categories perceived as having a weak or unknown evidence base are those with a greater demand for more and better evidence. In fact, as the analysis by groups suggests below, the correlation may be the other way around. A positive correlation between strength of evidence and demand for more evidence could reflect that stakeholders have a better appreciation for the usefulness of high-quality evidence when they see such evidence, and then desire more such evidence, perhaps specific to programme context or scale, to inform their own work.

For the science group, the general findings are that intervention categories with stronger evidence currently are also those for which stakeholders would benefit from more evidence. This correlation is consistent with the idea that interventions with stronger evidence may be preferred interventions, and stakeholders want more evidence to help them design programmes in the preferred intervention categories. An exception is STEM education, for which the evidence base is considered to be relatively strong but the demand for more evidence is relatively low. Here, the low demand for more evidence likely reflects that the survey respondents were stakeholders interested in STIP as approaches to enhance development programmes, whereas the studies on STEM education test pedagogical approaches for teaching STEM, rather than testing whether STEM education improves development outcomes.

---

6 These percentages are not fully comparable across groups, as the number of intervention categories from which respondents could choose was different in each group. If each intervention category were demanded equally in each group and respondents selected only the maximum number allowed, the rates of selection for each would be 43 per cent for science interventions, 30 per cent for technology, 33 per cent for innovation ecosystems and 33 per cent for partnerships.

The categories for which more than 60 per cent of respondents consider the evidence to be strong or moderate are fellowships and research grants and technical assistance. Note that exchanges and collaborations, which were separated in the survey but combined in the EGM, have very different results for the usefulness of more and better evidence, with low demand for exchanges and high demand for collaborations.

For the technology group, the general pattern is similar to science – stronger evidence looks to be correlated with more evidence being useful. The exception is another intervention category tied to education, where the studies test teaching approaches for learning outcomes. The categories for which more than 60 per cent of respondents consider the evidence to be strong or moderate are digital finance, data systems and digital information systems (which, for the survey, included m-health interventions).

Survey responses across the innovation ecosystems categories are relatively equal. Access to capital and grants and subsidies both show that more than 60 per cent of innovation respondents consider the evidence to be strong or moderate. As noted above, a high demand for this group overall is suggested by the fact that respondents selected more than the allowed number of categories. For the partnerships group, the obvious difference between the categories is that more than 20 per cent consider the evidence for partnerships between two entities to be strong and another more than 40 per cent consider it to be moderate, while only roughly 40 per cent consider the evidence for the other two categories, global multi-stakeholder initiatives and innovative financing, to be strong or moderate, with 5 per cent or fewer responding ‘strong’. In this group, some respondents selected more than one category for demand more and better evidence, again suggesting a strong demand for the group generally.

4.2 Assessment of demand

As noted in the methods section (4.1), in order to conduct a more comprehensive assessment of demand, we collected information from a variety of sources. The stakeholder survey focused just on the intervention categories and was conducted before we populated the EGM. Another key source of information on demand for evidence was the roundtable event we held after we had preliminary results from the EGM and the stakeholder survey. At this event, STIP specialists reviewed the early findings and contributed their own assessments of demand for evidence. These discussions included demand by intervention category but also explored the demand for evidence by outcome types, sectors and crosscutting themes. Subsequent reviews by members of our advisory group and by the US Global Development Lab helped refine the assessment.

We identify 11 categories as having a relatively high demand for evidence. We present these, along with the justification and notes about specific needs, in Table 4. Note that we combine some categories for the purpose of the demand assessment. The categories in the innovation ecosystem group all had roughly similar responses in the stakeholder survey, and our discussions with stakeholders revealed a strong

demand for these interventions, particularly when they are combined in comprehensive programmes, such as small business incubators. The two-entity partnerships and global multi-stakeholder initiatives categories are combined, because they follow the same general theory of change. In our discussions with stakeholders, it was often not obvious whether the partnerships they were citing were two-entity or multi-entity partnerships.

There is one category, 'technical assistance for science research', that appears to have high demand in the stakeholder survey but that we did not include in our list. Stakeholders in the consultations did not emphasise this category as needing more evidence. In the informal portfolio review, we did not find programmes centred on technical assistance just for science research. Instead, the technical assistance programmes we found focused on international support to universities and research institutions more generally.

Note that Table 4 mentions some of the findings from our informal portfolio review of programming. We present more details about those findings for the interventions identified as research investment priorities in section 5.

Table 4: Categories with relatively greater demand for more and better evidence

Intervention category	Specific needs	Notes
Science		
Fellowships and research grants	Sex-disaggregated impacts, long-term impacts	Medium response in the survey but strong focus at the roundtable event. Steady trend for programming.
Exchanges and collaborations	Collaboration and partnership interventions	Survey showed low demand for evidence on exchanges but high demand for collaborations and partnerships. Steady trend for programming.
Technology		
Digital infrastructure development	Agriculture, education and economic growth sectors	Strong demand in the survey and strong focus at the roundtable event. Foundations targeting along with public donors.
Digital inclusion	Sex-disaggregated and vulnerable and marginalised populations impacts	Strong demand in the survey and discussion at the roundtable event particularly around programming for specific sub-populations.
Digital finance	Sex-disaggregated and vulnerable and marginalised populations impacts	Low demand in the survey, but discussion at the roundtable suggested a strong demand evidence related to sub-populations. Increasing trend in programming.

Digitising identity	Sub-populations	Low demand in the survey, but strong demand at the roundtable event. Increasing trend in programming
Data systems development	Non-health sectors	Strong demand in survey, increasing trend in programming.
Digital information services	Sex-disaggregated and vulnerable and marginalised populations impacts	Strong demand in survey. Very high prevalence of programming.
Innovation ecosystems		
All categories	Multi-component programmes, especially combining access to capital or grants and subsidies with other categories such as capacity building	Demand in survey similar across categories, strong demand at roundtable, steady trend for programming with many multi-component programmes (e.g. incubators).
Partnerships		
Two-entity partnerships and global multi-stakeholder initiatives	Programming and theory of change very similar for two-entity and multi-stakeholder, so categories can be combined	Increasing trend in programming, strength of current evidence base considered low.
Innovative financing		Increasing trend in programming, strength of current evidence base considered low.

5. Research investment priorities

The EGM allows us to identify clusters of studies that evaluate similar interventions and measure similar outcomes. Ideally, where there is a large enough cluster of impact evaluations addressing the same evaluation question, researchers will conduct a systematic review of that evidence and, if possible, include statistical meta-analysis. We identify clusters of evidence where no systematic review has been conducted as priority questions for systematic reviews. As seen in the discussion of the existing systematic reviews in section 3.4, these studies can provide generalisable evidence that is highly useful for policymakers and programme managers.

The EGM also reveals intervention categories and outcome types for which there are gaps in the evidence. Our analysis of the demand for evidence on the effectiveness of STIP-related interventions, combined with findings from the EGM, allows us to explore potential priority areas for future investments in impact evaluations.

5.1 Priorities for new systematic reviews

The EGM for STIP-related interventions includes several clusters with a large number of studies for which systematic reviews have not yet been conducted or do not cover all of the studies in the cluster. In this section, we examine the studies in these clusters to see whether each cluster is promising for systematic review. Whether synthesis and meta-analysis are possible depends on the homogeneity of the studies in the cluster, particularly whether the studies evaluate programmes that are relatively similar and measure outcomes in ways that can be standardised and aggregated. We analyse the following nine clusters to identify priority areas for evidence synthesis:

- fellowships and research grants;
- educational programmes to promote STEM;
- digital finance;
- e-governance;
- digital information services;
- technology assisted learning;
- m-health;
- all innovation ecosystems categories; and
- innovative financing.

Note that we identify clusters of studies that are promising for systematic review regardless of whether the intervention category is listed as having a relatively higher demand for evidence in Table 5. The fact of a large number of studies in a category suggests that there is demand for evidence about those interventions. The best way to aggregate the learning from those studies is through a systematic review, which is less expensive to conduct than a new impact evaluation. So, we identify all promising clusters of studies as priorities for investment in systematic reviews. Table 5 provides a summary of the analysis, followed by discussion of each of the nine clusters.

Table 5: Analysis of evidence clusters

Intervention category	Levels of analysis	Sectors	Number of studies	Existing systematic review	Potential for systemic review	Notes
Science: fellowships and grants	Individual and household (HH)	Education and academia	8	None	No	Limited number of programmes and heterogeneous interventions.
Science: educational programmes to promote STEM	Individual and HH	Education and academia	37	None	Yes	Many studies look at interventions based on the same underlying theories of change, most in secondary school, similar evaluation designs.
Technology: digital finance	Individual and HH, organisational	Ag. and food security, economic growth, finance and trade	16	None	Yes	5 similar studies of mobile money systems.
Technology: e-governance	Individual and HH, organisational	Democracy, human rights and governance, ag. and food security, economic growth, finance and trade	7	None	No	Heterogeneous interventions and outcomes.
Technology: digital information services	Individual and HH, organisational, community and societal	Education and academia, global health, democracy, human rights and governance, ag. and food security, environment and climate change	27	None	Yes	10 studies covering 7 programmes for digital information dissemination to improve agriculture and food security outcomes.
Technology: technology-assisted learning	Individual and HH, organisational	Education and academia, global health, economic growth, finance and trade	19	None	Yes	5 studies of internet or mobile devices in classroom settings.
Technology: m-health	Individual and HH	Global health	134	6	Yes	Existing systematic reviews cover very few of the impact evaluations in the map: • Potential systematic review for ART adherence: 14 studies, only 2 are already in systematic reviews. • Potential systematic review for lifestyle changes for chronic diseases: 20 studies, only 5 are already in systematic reviews • Potential systematic review for appointment reminders: 14 studies.
Innovation ecosystems: all categories	Organisational	Economic growth, finance and trade	37	None	Yes	32 studies on publicly funded monetary assistance to private sector firms to spur innovation; most measure similar outcomes; 27 in Latin America and the Caribbean
Partnerships: innovative financing	Individual and HH, organisational, community and societal	Education and academia, global health, economic growth, finance and trade	25	1	Yes	Existing systematic review on contracting out, no included studies in EGM; 24 studies on RBF for health.

5.1.1 Fellowships and research grants

There are eight impact evaluations of fellowship or research subsidy programmes in the map; however, this cluster is not promising for statistical meta-analysis due to its heterogeneity. All but one of the studies takes place in Latin America, but the programmes differ across countries. The Paraguayan National Programme of Research Support (Aboal and Tacsir 2016) provides ex-ante subsidies to researchers, while other programmes support research projects. The Bin *et al.* (2015) study in Brazil compares two selection mechanisms for awarding scholarships and does not have a pure control, so it provides evidence about mechanism design but does not measure programme impacts against a control, as the other studies do. The effects measured by the Ubfal and Maffioli (2011) study of the Argentinian Scientific and Technological Research Fund (FONCYT) also are not comparable to the other studies, because the study tests whether the programme increases the number of co-authors and not whether the programme increases research products. In addition, three of the eight studies test the outcomes of the same subsidy programme in Argentina, so the overall number of programmes covered by the evidence cluster is lower than the number of studies.

Our analysis of demand for evidence in this area finds that stakeholders consider the evidence base to be relatively strong for this category, yet continue to demand more and higher-quality evidence, particularly on sex-disaggregated impacts. More impact evaluations of these interventions are needed so that synthesis becomes possible.

5.1.2 Educational programmes to promote STEM

There are 37 studies related to STEM education in the map and no existing systematic review that includes L&MICs. This cluster of evidence suggests a few promising questions for systematic review, which would be framed around the effects of pedagogical approaches or enhancements on student outcomes. All but two of the studies evaluate whether some pedagogical approach improves student outcomes. The other two (Sang *et al.* 2012; Sever, Oguz-Unver and Yurumezoghu 2013) evaluate interventions directed at teachers. Unfortunately, none of the studies examine interventions to promote the teaching of STEM in schools or interventions to increase the amount or degree of STEM education in L&MICs.

There are many similarities amongst these studies. They all focus on either math or science, the latter including biology, chemistry and physics. The majority of the tested interventions took place in secondary schools. Many of these studies also use a very similar evaluation design, a Solomon four-group research design, which authors also call non-equivalent control group design. This feature means that the calculated effect sizes will be more comparable, but it is a disadvantage from the standpoint that this is a relatively weak impact evaluation design when cluster effects are accounted for. The typical setup is four classes, two randomly assigned to treatment and the other two control, with a pre-test in one treatment and one control and a post-test in all four. One notable exception is Berlinski and Busso (2015), which presents an active learning experiment run in 85 schools. Another similarity across many studies is that the pilot intervention takes place over a short period, typically a few weeks, and the post-test takes place immediately after. The studies

cover Turkey, China, several countries in Latin America, a few countries in Africa and one study in Malaysia. There are 14 studies in Kenya alone, and another 6 in Turkey and 5 in China.

The primary heterogeneity comes from the interventions. The specific features of the pedagogical intervention design, including the teaching materials, are different in all the studies. These interventions are typically designed by the study authors, who train the treatment teachers to employ the approach. The potential for systematic review comes from commonality in the educational and psychological theories underpinning the various interventions. Many of the studies test interventions that are based on constructivist theory and use experiential or active learning approaches. In this subset are a small group of studies on cooperative approaches. There are also subsets looking at concept mapping type approaches for making the material more relevant to students and mastery learning approaches that break down material into units that students master in turn.

The majority of the studies measure learning outcomes. There is a subset that measure student motivation outcomes. Some interventions are designed just to affect motivation, and only motivation is measured. Others are posited to affect learning and motivation, and the researchers measure both kinds of outcomes.

Seven of the STEM studies test computer-assisted learning interventions, but the contexts are heterogeneous. The studies are spread across China, Ecuador, Kenya, Nigeria and Turkey and span grades from third to secondary school. They include interventions designed for both math and science.

Of the science categories, educational programmes to promote STEM had the highest percentage of respondents who said they did not know what the state of evidence is, however, only 25 per cent of the respondents selected STEM as one of their top three (of seven) intervention categories for which more and better evidence would be useful. A systematic review of the existing impact evaluations in this category would provide useful information to education specialists working in L&MICs, as the studies examine the effectiveness of different pedagogical approaches. This evidence may be less directly relevant for STIP specialists who are interested in whether STEM programmes can enhance development outcomes.

5.1.3 Digital finance

Sixteen studies measure the impacts of digital finance interventions. In this group, we identify one promising area for a systematic review. Nine of the studies evaluate the impacts of mobile money systems, in which applications such as M-Pesa in Kenya and mKesh in Mozambique allow users to deposit and transfer funds and make purchases using mobile phones (Munyegera and Matsumoto 2016). Five of these studies, all conducted in Sub-Saharan Africa, evaluate the impacts of general mobile money use, while four measure the impacts of receiving cash transfers via a mobile money application. The five studies in the former group measure impacts on similar outcomes such as consumption, remittance and savings behaviours. The stakeholder survey results, which show a large percentage of respondents perceiving the strength of evidence for digital finance to be moderate, are consistent with a large

cluster of impact evaluations that have yet to be synthesised. The high demand for evidence, particularly as expressed at the roundtable event, suggests that this is an area where evidence synthesis is promising and would add value for policymakers.

5.1.4 E-governance

There are seven completed impact evaluations of e-governance interventions. Finding a cluster of studies for this category is encouraging, as many e-governance interventions may be harder to evaluate with a counterfactual once they are fully in place. Although the studies in the cluster provide some very useful evidence, we would not recommend that a systematic review be attempted based on the existing evidence base. There are several types of heterogeneity across the studies that make it difficult to perform meta-analysis on them. Two of the interventions address political objectives, primarily political participation. For example, Ferree *et al.* (2015) evaluate an intervention that uses mobile phones and social media in South Africa to encourage citizens to engage politically, including to volunteer at polling stations.

The other five studies evaluate interventions targeted at governance objectives but looking at various services. Chong, Machicado and Yanez-Pagans (2014) look at introducing information and communication technology to improve administrative efficiency at police stations in Bolivia. Rezaee, Hasanain and Khan (2015) test the use of a mobile phone technology to introduce a farmer rating system to improve the productivity of government veterinarians. There are three studies from India, but these examine different interventions: biometric smart cards for payments, computerised service centres and an electronic platform for fund disbursements. The ongoing studies include one political intervention and one governance intervention.

Another feature of these studies that makes them very useful but difficult to synthesise is that they explore a diversity of outcomes, reflecting the complexity of governance reform. Although the studies mostly measure outcomes at the individual level, there are outcome indicators designed to capture uptake, quality, efficiency and corruption.

In the stakeholder survey, respondents were less likely to consider e-governance interventions to be effective, compared to other technology categories, and less likely to report that the evidence base for e-governance is strong or moderate, compared to other technology categories. A relatively small share of respondents selected e-governance as one of the top three intervention categories for which more and better evidence would be useful, but this group had the largest number of intervention categories to choose from.

The Evidence in Governance and Politics network has developed an approach to producing new evidence that greatly facilitates the synthesis of findings from multiple studies. In the approach, called *metaketa*,[7] different research teams undertake concurrent experimental studies of similarly designed interventions in various contexts. The teams agree in advance what outcomes they will measure and how.

[7] A coordinated, multi-site research grant round designed to foster knowledge accumulation (Evidence in Governance and Politics).

Once the studies are complete, the findings can be meta-analysed. *Metaketas* can be designed for most intervention types, but Evidence in Governance and Politics' interest in developing this approach was specifically to address the wide heterogeneity of studies in governance and politics.

In the absence of a *metaketa*, it will likely take some time for a sufficient evidence base to grow around e-governance so that statistical meta-analysis is possible. In the meantime, these individual studies and new impact evaluations of e-governance initiatives can play a significant role in influencing policy and practice in the contexts where they were and are conducted.

5.1.5 Digital information services

The EGM identifies 27 studies that evaluate the use of what we call 'digital information services'. These interventions include the use of mobile devices for information dissemination or SMS messages intended to 'nudge' behaviour that are not included under digital finance or m-health. Of the 27 studies, we identify 10 that are promising for evidence synthesis. These 10 studies evaluate the impacts of agricultural-related information dissemination, such as SMS messages to farmers containing crop price and weather information (e.g. Fafchamps and Minten 2012; Parker, Ramdas and Savva 2016) or a mobile phone hotline service by which farmers can obtain information about agricultural best practices (e.g. Cole and Fernando 2016; Casaburi *et al.* 2014).

Three programmes – Avvaj Utalo and Reuters Market Light in India and Esoko in Ghana – are each evaluated twice, so there are seven different programmes available for synthesis in this cluster. Common outcomes measured include the adoption of new agricultural technologies, crop prices received by farmers, and knowledge of agricultural practices. Despite the impressive number of impact evaluation studies in the EGM, our analysis of the demand finds that stakeholders consider the evidence base in this area to be less strong. These findings, coupled with the expressed demand for higher-quality evidence in this area, suggest that evidence synthesis here is promising and would add value for policymakers.

Amongst the remaining 17 studies in this category, some measure the impacts of using SMS messaging to nudge behaviour. These include SMS messages to encourage voting and civic participation (Marx, Pons and Suri 2016; Aker, Collier and Vicente 2013), SMS messages reminding citizens of their eligibility for social benefits (Blanco and Vargas 2014; Capuno *et al.* 2016) and SMS messages to parents promoting the value of education for their children (López-Vargas *et al.* 2015). We did not identify enough studies in these areas to recommend a synthesis of the evidence.

5.1.6 Technology-assisted learning

Nineteen studies in the EGM evaluate the impacts of technology-assisted learning programmes. We did not identify any systematic reviews of medium or high confidence of L&MICs. Amongst the 19 studies, we identify 8 similar studies, 5 of which form a cluster that is promising for evidence synthesis. Similar to our analysis of the educational programmes for STEM, we recognise that a systematic review of

the existing impact evaluations in this category would provide useful information to education specialists working in L&MICs, rather than STIP specialists.

The eight technology-assisted learning impact evaluations that we identify examine the use of the internet or mobile devices in classroom settings, five of which focus on secondary schools and follow a similar theory of change. The five evaluated programmes use lesson plans that are delivered through either a mobile device or the internet. Three programmes focus on improving mathematics and science proficiencies in particular, but all five studies measure impacts on a variety of academic outcomes, such as literacy and numeracy test scores. The studies are also similar in terms of region. Two are conducted in Nigeria, one in South Africa and two others in Latin America and the Caribbean.

Three other studies (all led by the same two authors) evaluate Project ABC in Niger, which provides mobile phone-based numeracy and literacy education to adults (Aker, Ksoll and Lybbert 2010; Aker, Ksoll and Clemens 2011; Aker, Ksoll and Lybbert 2012). However, the programme's theory of change focuses on improving skills adults can apply to their everyday activities and employment, rather than improving academic performance, which is why we do not suggest including this programme in the cluster for synthesis.

5.1.7 Mobile health

There are 134 studies coded as m-health, an area where we see substantial interest. Our assessment of the demand for evidence in this area is high. Many survey respondents (33 per cent) identified digital information services[8] as an area where more evidence is needed, and participants at the roundtable event specifically identified m-health as an area of interest. In this group of 134 studies, we see multiple promising clusters for evidence synthesis. Synthesis is recommended for the use of SMS for medication adherence, for SMS to promote lifestyle changes (e.g. diabetes, weight loss or exercise, smoking cessation, nutrition and heart health, and oral health) and appointment reminders, and to improve health care providers' knowledge or practices.

We also find two areas where a few more impact evaluations would create a promising cluster for synthesis. These are the use of m-health to improve antenatal care service uptake and to address peri- or postnatal care knowledge and service use. Finally, we identify one area that may warrant new research: the use of m-health for mental health education and support.

Medication adherence

In the m-health intervention category, 14 studies are of interventions using SMS to promote adherence to ART. Two others use phone calls to do the same. Several combine SMS with electronic pill bottles that interact with the SMS system to send messages when a dose is missed. Four other studies use SMS to promote adherence to hypertension or cardiac-related drugs, and one uses SMS for post-

[8] At the time of the survey's distribution, m-health interventions were included under digital information services.

stroke medication. There are other studies for tuberculosis treatment, iron supplementation, malaria medication and birth control. Although not related by disease to ART, the theory of change is similar. We found two high confidence systematic reviews on this topic, published in 2012 and 2013, but they included only two of the studies included in our gap map, and few other impact evaluations, most of which were very small or implemented in a developed country. The additional 10 studies we found could allow for meta-analysis, and we see this cluster of studies as a promising area for synthesis.

Lifestyle behaviour changes

The EGM includes 20 studies that use SMS to promote lifestyle behaviour changes. Many are related to diabetes, with a few to hypertension, weight loss, smoking cessation and oral health for young children. While somewhat heterogeneous, these are all related to risk factors for cardiovascular disease and all have an objective of lifestyle behaviour changes. We found two systematic reviews related to m-health for diabetes, one that includes other chronic diseases. Between the two, they include five of the studies in the EGM. The chronic disease systematic review also includes outcomes related to appointments and interventions targeted towards providers (see the 'Health care providers' section below). Due to the minimal synthesis and the strong interest in m-health, we view additional and new synthesis in this cluster as promising.

Appointment reminders

Fourteen studies assess the effect of SMS, sometimes compared to alternative contacts, for appointment reminders, including several for general clinic appointments and for immunisation. The type of appointment varies, but the theory of change is similar. Two of the studies are included in the chronic care systematic review (Beratarrechea *et al.* 2014). The additional studies specifically focused on appointments, but allowing for a broad range of reasons for the appointments provides a potential opportunity for meta-analysis. The limited number of synthesised studies in the group suggests that this cluster is also a promising area for synthesis.

Health care providers

The EGM contains a number of studies targeting health care providers. Interventions use SMS or hotlines with information on guidelines, treatment help or checklists. Only one appears in an included systematic review. A few use SMS or applications to monitor attendance or encourage reporting of cases, but there are likely sufficient studies to synthesise evidence around SMS or mobile applications for guidance.

Antenatal services

There are also a handful of studies on the use of SMS or phone calls to improve antenatal services uptake or address perinatal or postnatal knowledge and service use. It is interesting that two of these use mobile communication to triage cases and reduce facility-based care. These are areas where a few additional studies could be helpful in building a cluster of studies to synthesise. For example, a few more studies on SMS or mobile phones for promotion of antenatal care service use or for postnatal advice could provide a cluster for synthesis.

5.1.8 Innovation ecosystems

Thirty-seven studies cut across the five categories in the innovation ecosystems group. Several cross intervention categories. Closer inspection reveals that there is a lot of similarity of interventions, even across categories. We analyse all five intervention categories here as a group, and we find one promising area for evidence synthesis.

Amongst the 37 studies in the group, 32 measure the impacts of publicly funded monetary assistance (e.g. grants, loans or subsidies) to private sector firms, intended to spur innovation. Of these 32 studies, 27 are conducted in Latin America. The public funding programmes for innovation in Latin America are similar across different countries, and many of the different studies evaluate the same programmes. Many are grants or lines of credit awarded competitively to private companies with the objectives of increasing R&D expenditures or encouraging investments in innovative activities. These programmes include FONTAR, which funds 'projects oriented to research and development, pilot scale technologies, applied knowledge generation, innovative products, and process development' (López, Reynoso and Rossi 2010, p. 2). Likewise, FONTEC in Chile aims to promote and finance innovation projects support the development of commercial applications of scientific and technological findings (Benavente and Maffioli 2007). In Brazil, the National Technology Development Support Programme provides reimbursable funding for private firms to invest in R&D.

The 27 Latin American studies in this area represent a promising area for a systematic review. Though a number of the studies examine the same programme, more than 20 different monetary assistance programmes in Latin America are represented in the 27 studies. The outcomes measured in each study are also similar. Almost every study measures the impacts of the programme on company sales and profitability, labour productivity, innovative inputs and outputs, as well as technological outputs and R&D expenditures.

5.1.9 Innovative financing

The EGM contains 25 studies in the innovative financing category, all of which evaluate RBF programmes. There are 24 studies on RBF for health, including one that includes both education and health. Although 10 of the studies are on the Rwanda RBF programme, and a few of those evaluate the same outcomes, there are a number of other studies across Sub-Saharan Africa, Afghanistan, Argentina and China. Several evaluate health or service outcomes for maternal and child health. Several others report outcomes for basic primary health care services. We view this cluster as a promising area for synthesis.

5.2 Priority intervention categories for new impact evaluations

In this section, we present our analysis of which intervention categories should be the highest priority for future research. As explained above, this analysis looks at both supply and demand. The EGM provides the supply-side analysis for identifying research questions in this area. The demand assessment is based on information about stakeholders' needs from several sources, along with a rapid portfolio review

of current programming. We identify the following seven priority intervention categories for new impact evaluation research:

- research exchanges and collaboration;
- digital infrastructure development;
- digital inclusion;
- digitising identity;
- data systems development;
- all innovation ecosystems categories; and
- innovative financing.

Of the intervention categories identified as having a higher demand for evidence in Table 4, there is one for which we are not recommending investment in either systematic reviews or new impact evaluations: two-entity partnerships combined with global multi-stakeholder initiatives. We discuss our analysis of these intervention categories at the end of this sub-section.

5.2.1 Research exchanges and collaborations

With only two completed impact evaluations, there is a clear gap in evidence for exchanges and collaborations for scientific research. When the stakeholder survey was written, the category distinctions for these interventions were a bit different than we have in the final map. The survey asked about 'collaborations and partnerships' and 'exchanges' separately. The results for these separated categories are very different. Collaborations and partnerships had the highest response rate, amongst the science categories, as being usually effective, while exchanges had the lowest. Exchanges is also considered to have a much weaker evidence base. At the same time, collaborations and partnerships garnered the most responses amongst science for more and better evidence being useful, while exchanges had one of the lowest.

The strong interest by STIP professionals in evidence for the collaborations sub-category reflects the many efforts of development organisations to strengthen international science collaborations. For example, USAID's Partnerships for Enhanced Engagement in Research Science programme provides grants to developing country scientists to collaborate with US scientists receiving companion National Science Foundation grants. The International Foundation for Science offers grants for scientists living in L&MICs and has adopted a collaborative approach that aims to help researchers find appropriate working partners. The Japan International Cooperation Agency has established the Science and Technology Research Partnership for Sustainable Development, designed to promote joint international research. These are just some examples of how large international organisations are investing in efforts in this area. Given the noticeable gap in supply of evidence in this area, and the high demand, as well as the large amount of programming, we identify this as a priority area for investment in primary research, but note that the interest appears to be limited to collaborations, not exchanges.

5.2.2 Digital infrastructure development

Only four studies evaluate the effectiveness of digital infrastructure development policies or programmes, and three of the four evaluate the same Nigerian policy of

rolling out mobile phone coverage throughout the country. The fourth study measures the impacts of providing internet connectivity for schools in Thailand. Amongst the stakeholders who responded to the technology section of the survey, the largest proportion demanded more and higher-quality evidence in this area (43.6 per cent) and discussions at the roundtable event revealed a strong interest in the impacts of these policies on sectors such as agriculture, education and economic growth. Moreover, the spread of internet and mobile technology is becoming a priority for several international development organisations. For example, through the Alliance for Affordable Internet, USAID, Google.org, the World Wide Web Foundation, the Omidyar Network, DfID and other organisations are encouraging policymakers to facilitate access to broadband connection. This is an area in which policymakers are gaining interest but for which they require more evidence.

5.2.3 Digital inclusion

We identify this category as a priority for future investments in primary research. Only four studies measure the impacts of interventions related to digital inclusion. Two of the four evaluate programmes that provide rural farmers with mobile devices and training on how to use them (Aker and Ksoll 2016; Fu and Akter 2016). The third study evaluates the impacts of facilitating access to internet cafes (Bailard 2012), and the fourth provides pregnant women with mobile phone vouchers to facilitate communication with primary care providers (Lund *et al.* 2012). This intervention category was the second most demanded amongst the technology group, and discussions at the roundtable event centred heavily on the inclusion of vulnerable or marginalised populations, including women.

5.2.4 Digitising identity

There are only six completed impact evaluation studies and one ongoing study in Malawi for digitising identity. The seven studies are conducted in India, Uganda and Malawi. The four Indian studies evaluate the impacts of biometric identification systems applied in different contexts, related to either health or governance. The Malawian studies examine the use of biometric identification of loan applicants at microfinance institutions, and the Ugandan study uses biometric identification for public health purposes.

Our analysis of the demand for evidence finds that there is an interest in more and higher quality evidence in this area. Digitising identity has become a priority area in development programming. A Center for Global Development working paper (Gelb and Clark 2013) finds more than 160 cases across 70 developing countries where approximately 1 billion people are affected by biometric programmes for social, economic or political purposes. Major development organisations are promoting the use of biometric identification. For example, the World Bank's Identification for Development initiative is devoted to assisting L&MIC governments in the development of national biometric identity systems, and the UN High Commissioner for Refugees is rolling out a biometric identification system to protect refugees across the world.

Given the international development community's rising interest in this area and the resources devoted to establishing biometric identification systems across the developing world, we see this category as a priority. In particular, given the aim of most biometric identification systems to empower populations for whom it is difficult to obtain paper identification cards, we hope to see more studies that provide effect sizes for vulnerable or marginalised populations.

5.2.5 Data systems development

Of the nine studies in the data systems category, four form a potential cluster for evidence synthesis, if new impact evaluations can be added. The four studies – two conducted in China (Zhang *et al.* 2012; Liu, Chen and Win 2013), one conducted in South Africa (Dillon *et al.* 2014) and one in Fiji (Yu *et al.* 2009) – evaluate the impacts of electronic health record systems on data collection accuracy and efficiency. Our assessment of the demand for evidence in this category finds that the need for evidence is relatively high, especially on sex-disaggregated impacts; therefore, we recommend investments in additional primary studies on the effectiveness of electronic health records to form a promising cluster for evidence synthesis.

5.2.6 Innovation ecosystems

Though we find a large number of studies that examine the impacts of innovation ecosystem programmes in Latin America (n=31), there is a dearth of evidence on the impacts of this programming in other parts of the developing world, particularly in Sub-Saharan Africa. Three studies in this group are conducted in Sub-Saharan Africa. Two of these, conducted in Nigeria and South Africa, evaluate the impacts of public grants for innovation (McKenzie 2015; Czarnitzki and Lopes-Bento 2010), and the third evaluates the impacts of 'innovation forums' in Uganda, Rwanda and the Democratic Republic of the Congo (Pamuk *et al.* 2015). The few impact evaluations in this field in Sub-Saharan Africa do not reflect the amount of programming in this area. For instance, the United Nations Environment Programme (UNEP), in partnership with the African Development Bank and other organisations, has established an initiative that provides seed capital to African enterprises to encourage them to invest in innovative clean energy practices.

Other networks and funds, supported by a variety of organisations, are connecting African entrepreneurs with capital, technical assistance and other incubator-like programmes to help them innovate. Examples include Venture Capital for Africa, the African Enterprise Challenge Fund, the Fund for Internet Research and Development, the Savannah Fund and the African Innovation Foundation. Our stakeholder consultations revealed that there is a demand for more evidence on this type of programming in Sub-Saharan Africa, so although there is a cluster of evidence from Latin America that is promising for systematic review, we also recommend investment in new impact evaluations for innovation ecosystem programmes in Sub-Saharan Africa.

5.2.7 Innovative financing

As mentioned earlier, all of the studies in the innovative financing category evaluate RBF programmes. Only two of the RBF studies are related to incentives for education. These two studies reward performance at different levels, one for schools and teachers (Muralidharan and Sundararaman 2011) and the other for villages (Olken, Onishi and Wong 2014). The second also includes a health component. Although we found only two impact evaluations on RBF for education, Cordaid (the Catholic Organization for Relief and Development, headquartered in the Netherlands) is funding RBF programmes for education in five countries. There is an education programme in Malawi, and combined health/education programmes in the Democratic Republic of Congo, Cameroon, the Central African Republic and Burundi. At the same time, other aid agencies and national governments are implementing RBF programmes, and the World Bank, a major funder of RBF for health, has announced it will double RBF for education to $5 billion over the next five years (World Bank Group 2015). The World Bank is also expanding to other sectors, such as energy, municipal solid waste and clean stoves.

There is a generalised interest in more or better evidence on the effectiveness of innovative financing mechanisms, which includes RBF. The fact that there are a growing number of initiatives and limited impact evaluations on RBF outside of health suggests that this is a gap that should be filled, and that funders and implementers of programmes should plan on impact evaluations as they embark on new initiatives. Innovative financing or RBF outside the health sector is a priority area for new research. Given the growing programme base for RBF in education, this sector may be a good place to start focusing research investment, with expansion to other sectors as RBF is implemented.

5.2.8 Two-entity partnerships and multi-stakeholder initiatives

There are clear gaps in the evidence base on the impacts of two-entity partnerships and multi-stakeholder initiatives on all levels of analysis. Interventions implemented through these mechanisms could cut across any sector, but we see only one ongoing study on the impact of a multi-stakeholder consortium that leverages mobile-driven solutions to promote children's welfare (Sheely *forthcoming*). It is worth noting that two-entity partnerships and multi-stakeholder initiatives are not actual programmes or interventions. Rather they are forms of implementing development programming, which were more difficult to capture in our search and screening process. This can potentially explain why our map shows no completed impact evaluations in these two categories. Moreover, where partnerships were used to implement a programme, it is possible that the programme was evaluated, and not the partnership, and thus would not have been included as a partnership intervention in our gap map.

Moreover, the perception of the state of evidence on the effectiveness of delivering development assistance using partnerships is weak, while demand for more and higher quality evidence is high.

In assessing the evidence base for these two categories, it is important to understand that two-entity partnerships and multi-stakeholder initiatives are not programmes or policies themselves but rather are innovative means of implementing development assistance. Thus, the challenge facing any attempt to evaluate the net impact of a programme delivered through such a mechanism is finding a counterfactual, that is, the same programme delivered through 'standard' implementation. Given the challenges inherent in conducting an impact evaluation of a programme delivery mechanism, we do not identify this area as a priority for future investment. Process evaluation and other programme monitoring and evaluation research is a valuable resource when entering into strategic partnerships and making decisions about multi-stakeholder initiative formations. However, it is important for partners, multi-stakeholder initiatives and other consortiums to recognise the need to draw on impact evaluation evidence when selecting and designing the programmes they deliver.

5.3 Other priorities for new impact evaluations

The previous sections have focused on intervention categories. This section identifies priorities for outcomes, sectors and crosscutting themes.

Few studies report outcomes at the community or societal level. In many cases, the reasons for measuring outcomes only for individuals and households or for organisations and firms are obvious and rational. However, there are some few cases, such as e-governance or policies and regulations affecting innovation, in which outcomes observed at the community or societal level are important for understanding development impacts. We consider it a priority for new impact evaluations of programmes with community-level impacts to find ways to measure effects at the community level.

None of the identified impact evaluations measure effects on crisis- and conflict-related outcomes or energy-related outcomes. Other sectors – democracy, human rights and governance; environment and climate change; and water and sanitation – have very few studies. Our engagements with stakeholders reveal that many are interested in the impacts of STIP-related interventions, particularly technology interventions, on democracy, human rights and governance; agriculture and food security; and crises and conflict outcomes. We consider new research on STIP-related interventions in these sectors to be a priority.

Our analysis of the crosscutting themes found 80 studies that are sex-specific or that report outcomes disaggregated by sex; 50 of these fall under m-health, 11 fall under STEM education and 9 fall under innovative financing. The remaining are spread over nine other intervention categories, but with three or fewer studies in any intervention category. Thirteen categories have no studies that report single-sex outcomes. Although we might not expect this in some of the intervention categories, such as policies and regulation for scientific research, it is somewhat surprising that we do not find any studies that report sex-disaggregated outcomes in other categories, such as digital finance.

Similarly, of the five studies in the digital inclusion category, none report outcomes for specific marginalised populations or disaggregate by sex. Further, while 23 studies report outcomes for rural populations, only four report outcomes for other vulnerable or marginalised populations. We feel that producing evidence specific to these sub-populations is a priority, especially since many development programmes target the inclusion of women and vulnerable or marginalised populations, and since gender equality and inequity are specific targets in the Sustainable Development Goals.

Of the 22 studies that measure long-term impacts, 11 are technology interventions, while only 7 are innovation ecosystems interventions and 4 are science interventions. Given the theory of change that investments in scientific capabilities and innovation ecosystems will lead to scientific advancements and private sector innovations that will result in sustained improvements in human development and economic growth, we consider it a priority for new impact evaluations of these interventions to measure long-term impacts. Moreover, embedded in the theories of change of several digital technology programmes is the assumption that the application of new technologies to development efforts will accelerate the intended benefits at a relatively low cost. The same assumption applies to innovative financing mechanisms and to some innovation ecosystems programmes intended to produce high-technology R&D at a low price. With these theories of change in mind, we consider it a priority for new impact evaluations of these interventions to include cost analysis.

6. Limitations

The primary challenge associated with designing the STIP search strategy was that the framework includes four relatively distinct topics, each of which cuts across a variety of sectors. To capture studies across all of the intervention categories in each topic, we chose to focus the search terms on intervention category terms rather than broader thematic search terms for the four topics. It is possible that we missed some studies in which the interventions are described using nonstandard terminology. We tested many terms, including only those that yielded relevant results for the framework. For instance, the ‘partnerships’ search contains the names of certain public aid agencies, but not others, and the acronyms of some are included, whereas the full titles of others are present in the search.

We conducted our search only in English and in primarily English-based databases and websites. Although we screened some foreign-language studies that were captured by the search (in Spanish, French or Portuguese), we invariably missed studies in other languages.

The scope of this gap map was broad, and there was extreme variation in the outcome indicators measured across the studies. Therefore, we were not able to create more distinct categories for the outcomes along a causal chain. Instead, we categorised the outcome indicators by levels of analysis (individual, organisational and community) and by the sectors in which outcomes were measured. An EGM with a narrower focus, such as one that focuses on just one STIP group, would allow for more nuanced outcome categories that better reflect causal chains.

The demand assessment is based on qualitative analysis of largely qualitative information. We were able to triangulate our assessment with experts in the field on multiple occasions. Our assessment is intended to reflect the generally demanded categories, but we expect that there are organisations and agencies that require more and better evidence in other categories, based on their programming. We sent the stakeholder survey to a convenience sample – listservs, groups and individuals with whom 3ie is familiar. It is possible that the stakeholders are not a representative sample of all stakeholders involved in work related to STIP areas. However, the group of respondents were from a broad range of organisations, countries, topical backgrounds, experience levels and research knowledge. We believe that their input provided useful insights into the perception of the state of the evidence and the interest in and demand for new and better evidence. The stakeholder survey focused only on the intervention categories, but we were able to round out our demand assessment using the information from other sources.

7. Conclusion

The STIP EGM identifies and catalogues a large amount of evidence – 320 completed studies – on the effectiveness of STIP-related interventions. This existing evidence base is clustered in a subset of intervention categories in the STIP framework. Namely, 134 of the studies (nearly 42 per cent) fall under m-health. STEM education, grants and subsidies for innovation, digital information systems (other than m-health), innovative financing, technology-assisted learning, digital finance, and access to capital (for innovation) all have more than 10 studies. At the same time, there are four intervention categories (material resources for scientific research, technical assistance for scientific research, two-entity partnerships and global multi-stakeholder initiatives) with no studies, and two categories (policy and regulation for scientific research and policy and regulation for digital services) with only one.

Looking at the columns of the EGM framework, we find that only a small number of studies report outcomes at the community or societal level, and that sex disaggregation is largely limited to studies in the health and education sectors. In general, there is a dearth of studies that focus on, or report outcomes separately for, any vulnerable or marginalised population.

We assess the demand for evidence using a variety of information sources, including a stakeholder survey with 110 respondents, and apply qualitative analysis to identify the follow intervention categories has having a relatively higher demand for evidence:

- fellowships and research grants;
- research collaborations;
- digital infrastructure development;
- digital inclusion;
- digital finance;
- digitising identity;
- data systems development;
- digital information services;

- innovation ecosystems;
- two-entity partnerships and global multi-stakeholder initiatives; and
- innovation financing.

Given the large number of impact evaluations that we found, we consider it somewhat surprising that there are so few systematic reviews, and that the reviews contain so few of the impact evaluations included in the EGM. There are clusters of studies within some intervention categories that are promising for systematic review:

- digital finance (in particular, mobile money systems);
- digital information services for agriculture and food security;
- m-health (in particular, antiretroviral therapy adherence, lifestyle changes for chronic diseases and appointment reminders);
- innovation ecosystems programmes in Latin America;
- innovative financing (in particular, results-based financing for health);
- science, technology, engineering and mathematics educational programmes; and
- technology-assisted learning.

Our analysis also identifies several outcomes, sectors and crosscutting themes that should be priorities for STIP-related impact evaluations:

- community- and society-level outcomes in relevant intervention categories;
- technology impact evaluations on democracy, human rights and governance outcomes;
- technology impact evaluations on agriculture and food security outcomes;
- technology impact evaluations on crises and conflict outcomes;
- science impact evaluations that report outcomes disaggregated by sex;
- digital inclusion impact evaluations that report outcomes for vulnerable or marginalised populations;
- science and innovation ecosystems impact evaluations that report long-term impacts; and
- cost analyses for technology, innovation ecosystems and innovative financing interventions.

Appendix A: Evidence gap map of ongoing STIP impact evaluation studies

	Intervention categories	Levels of analysis			Sectors									Cross-cutting themes			
		Individual & household outcomes	Organisational outcomes	Community & societal outcomes	Education & academia	Global health	Democracy, human rights & governance	Agriculture & food security	Crises & conflict	Economic growth, finance & trade	Environment & global climate change	Water & sanitation	Energy	Long-term impact	Cost-analysis	Sex-disaggregated or sex-specific	Vulnerable or marginalised populations
Science	Fellowships & research grants																
	Material resources for scientific research																
	Technical assistance for scientific research																
	Research exchanges & collaborations																
	Policy & regulation for scientific research																
	Educational programmes to promote STEM	2			2												
Technology	Digital infrastructure development																
	Policy & regulation for digital services																
	Digital literacy	2	1							2					1		
	Digital inclusion	2					1			1						1	1
	Digital finance	11	2		2		1			10				1	1	1	
	e-Governance		1	1			2										
	Digitising identity	1								1							
	Data systems development	5				2	2	2							1		1
	Digital information services	9	1	4			4	3		1	1	1			1		2
	Technology assisted learning	3			1					3	1						1
	Mobile health	36	2	1	2	38								2	11	8	4
Innovation ecosystems	Access to capital																
	Grants & subsidies																
	Policies & regulation that affect innovation																
	Networks & collaboration for innovation		1							1							
	Capacity building for innovation																
Partnerships	Two entity partnerships																
	Global multi-stakeholder initiatives	1	1	1			1										1
	Innovative financing	3	2	4		5									1	3	

Online appendixes

Online appendix A: Countries with 10 or more studies by intervention category

This appendix is only available online and can be accessed from the link below.
http://www.3ieimpact.org/media/filer_public/2017/03/02/sp6-stip-appendix-a.pdf

Online appendix B: Stakeholder survey

This appendix is only available online and can be accessed from the link below.
http://www.3ieimpact.org/media/filer_public/2017/03/02/sp6-stip-appendix-b.pdf

Online appendix C: Stakeholder survey respondents' demographic details

This appendix is only available online and can be accessed from the link below.
http://www.3ieimpact.org/media/filer_public/2017/03/02/sp6-stip-appendix-c.pdf

Online appendix D: Stakeholder survey respondents' views of STIP intervention effectiveness

This appendix is only available online and can be accessed from the link below.
http://www.3ieimpact.org/media/filer_public/2017/03/02/sp6-stip-appendix-d.pdf

References

3ie, 2012. *Impact Evaluation Glossary*. Available at: <http://www.3ieimpact.org/media/filer_public/2012/07/11/impact_evaluation_glossary_-_july_2012_3.pdf> [Accessed 7 November 2016].

Aboal, D and Tacsir E, 2016. The impact of ex-ante subsidies to researchers on researcher's productivity: Evidence from a developing country. *UNU-MERIT Working Paper* 2016-019.

Aker, J, Collier, P and Vicente, PC, 2013. Is information power? Using mobile phones and free newspapers during an election in Mozambique. *Nova Africa Center for Business and Economic Development Working Paper Series*, 1304.

Aker, J and Fafchamps, M, 2014. Mobile phone coverage and producer markets: evidence from West Africa. *The World Bank Economic Review,* 29**,** pp.262–92.

Aker, J and Ksoll, C, 2012. Information Technology and Farm Households in Niger. *United Nations Development Programme Working Paper,* 005.

Aker, J and Ksoll, C, 2015. Call Me Educated: Evidence from a Mobile Monitoring Experiment in Niger. *Center for Global Development Working Paper,* 406.

Aker, J and Ksoll, C, 2016. Can mobile phones improve agricultural outcomes? Evidence from a randomized experiment in Niger. *Food Policy,* 60**,** pp.44–51.

Aker, J, Ksoll, C and Clemens, M, 2011. Mobiles and mobility: The Effect of Mobile Phones on Migration in Niger. Available at: <.https://www.researchgate.net/publication/228409814_Mobiles_and_mobility_The_Effect_of_Mobile_Phones_on_Migration_in_Niger> [Accessed 21 March 2017].

Aker, J, Ksoll, C and Lybbert, TJ, 2010. ABC, 123: The Impact of a Mobile Phone Literacy Program on Educational Outcomes. *Center for Global Development Working Paper,* 223.

Aker, J, Ksoll, C and Lybbert, TJ, 2012. Can Mobile Phones Improve Learning? Evidence from a Field Experiment in Niger. *American Economic Journal: Applied Economics,* 4**,** pp.94–120.

Alvarez, R, Bravo, C and Zahler, A, 2013. Impact Evaluation of Innovation Programs in the Chilean Services Sector. *Centro de Investigaciones Económicas Working Paper,* 2013(SS-IP)-02.

Alvarez, R, Crespi, G and Cuevas, C, 2012. Public Programs, Innovation, and Firm Performance in Chile. *Inter-American Development Bank Technical Notes,* 375.

Arambepola, C, Ricci-Cabello, I, Manikavasagam, P, Roberts, N, French, DP and Farmer, A, 2016. The impact of automated brief messages promoting lifestyle changes delivered via mobile devices to people with type 2 diabetes: a systematic literature review and meta-analysis of controlled trials. *Journal of Medical Internet Research*, 18.

Avellar, APM and Alves, PF, 2008. Avaliação de impacto de programas de incentivos fiscais a inovação – um estudo sobre os efeitos do PDTI no Brasil. *EconomiA,* 9, pp.143–64.

Bailard, CS, 2012. A Field Experiment on the Internet's Effect in an African Election: Savvier Citizens, Disaffected Voters, or Both? *Journal of Communication,* 62**,** pp.330–44.

Balsa, A, Gandelman, N and Porzecanski, R, 2010. The Impact of ICT on Adolescents' Perceptions and Consumption of Substances. *Inter-American Development Bank Working Paper Series,* 219.

Benavente, JM, Crespi, G and Maffioli, A, 2007. Public Support to Firm's Innovation: The Chilean FONTEC Experience. Available at: <http://idbdocs.iadb.org/WSDocs/getdocument.aspx?docnum=1323102&Cache=True> [Accessed 21 March 2017].

Beratarrechea, A, Lee, AG, Willner, JM, Jahangir, E, Ciapponi, A and Rubinstein, A, 2014. The impact of mobile health interventions on chronic disease outcomes in developing countries: a systematic review. *Telemedicine and e-Health,* 20, pp.75–82.

Berlinski, S and Busso, M, 2015. Challenges in Educational Reform: An Experiment on Active Learning in Mathematics. *Inter-American Development Bank Working Paper Series*, 561.

Bin, A, Salles-Filho, S, Capanema, L. M and Colugnati, FAB, 2015. What difference does it make? Impact of peer-reviewed scholarships on scientific production. *Scientometrics,* 102, pp.1,167–88.

Blanco, M and Vargas, JF, 2014. Can SMS Technology Improve Low Take-up of Social Benefits? *Peace Economics, Peace Science and Public Policy,* 20**,** pp.61–81.

Capuno, JJ, Kraft, AD, Quimbo, S, Tan, CR Jr. and Wagstaff, A, 2016. Effects of Price, Information, and Transactions Cost Interventions to Raise Voluntary Enrollment in a Social Health Insurance Scheme: A Randomized Experiment in the Philippines. *Health Economics,* 25**,** pp.650–62.

Casaburi, L, Kremer, M, Mullainathan, S and Ramrattan, R, 2014. Harnessing ICT to Increase Agricultural Production: Evidence from Kenya. Available at: <https://www.gov.uk/dfid-research-outputs/harnessing-ict-to-increase-agricultural-production> [Accessed 21 March 2017].

Castillo, V, Maffioli, A, Rojo, S and Stucchi, R, 2014. Knowledge Spillovers of Innovation Policy through Labor Mobility: An Impact Evaluation of the FONTAR Program in Argentina. *Inter-American Development Bank Working Paper Series,* 488.

Chen, Z, Fang, L, Chen, L and Dai, H, 2008. Comparison of an SMS text messaging and phone reminder to improve attendance at a health promotion center: A randomized controlled trial. *Journal of Zhejiang University Science B,* 9, pp.34–38.

Chong, A, Machicado, CG and Yanez-Pagans, M, 2014. Information Technologies and Provision of National Identification Cards by the Bolivian Police: Evidence from Two Randomized Natural Field Experiments. *Institute for the Study of Labor Discussion Series Paper*, 7975.

Cole, SA and Fernando, AN, 2016. 'Mobile'izing Agricultural Advice: Technology Adoption, Diffusion and Sustainability. *Harvard Business School Working Paper Series*, 13-047.

Czarnitcki, D and Lopes-Bento, C, 2010. Evaluation of Public R&D Policies: A Cross-Country Comparison. *Center for European Economics Research Discussion Paper,* 10-073.

de Walque, D, Gertler, PJ, Bautista-Arredondo, S, Kwan, A, Vermeersh, C, Bizimana, JDD, Binagwaho, A and Condo, J, 2015. Using provider performance incentives to increase HIV testing and counseling services in Rwanda. *Journal of Health Economics,* 40**,** pp.1–9.

Dillon, DG, Pirie, F, Rice, S, Pomilla, C, Sandhu, M. S, Motala, AA and Young, EH, 2014. Open-source electronic data capture system offered increased accuracy and cost-effectiveness compared with paper methods in Africa. *Journal of Clinical Epidemiology*, 67, pp.1358–63.

Evidence in Governance and Politics (n.d.). Available at: www.egap.org [Accessed 13 March 2017].

Fafchamps, M and Minten, B, 2012. Impact of SMS-Based Agricultural Information on Indian Farmers. *The World Bank Economic Review*, pp.1–32.

Ferree, K, Gibson, C, Jung, DF, Long, JD and McIntosh, C, 2015. Using Technology to Promote Participation in Emerging Democracies: VIP: Voice and the 2014 South African Election. Available at: <http://scholar.princeton.edu/sites/default/files/ictandgov/files/paper_panel_2_mcintosh.pdf> [Accessed 21 March 2017].

Flax, VL, Negerie, M, Ibrahim, AU, Leatherman, S, Daza, EJ and Bentley, ME, 2014. Integrating Group Counseling, Cell Phone Messaging, and Participant-Generated Songs and Dramas into a Microcredit Program Increases Nigerian Women's Adherence to International Breastfeeding Recommendations. *The Journal of Nutrition*, 144, pp.1120–24.

Fu, X and Akter, S, 2016. The Impact of Mobile Phone Technology on Agricultural Extension Services Delivery: Evidence from India. *The Journal of Development Studies*, 52, pp.1561–76.

Gelb, A and Clark, J, 2013. Identification for Development: The Biometrics Revolution. *CGD Working Paper,* 315.

Goodarzi, M, Ebrahimzadeh, I, Rabi, A, Saedipoor, B and Jafarbadi, MA, 2012. Impact of distance education via mobile phone text messaging on knowledge, attitude, practice and self-efficacy of patients with type 2 diabetes mellitus in Iran. *Journal of Diabetes and Metabolic Disorders*, 11.

Horvath, T, Azman, H, Kennedy, GE and Rutherford, GW, 2012. Mobile phone text messaging for promoting adherence to antiretroviral therapy in patients with HIV infection (Review). *Cochrane Database of Systematic Reviews,* 3.

Jamison, JC, Karlan, D and Raffler, P, 2013. Mixed Method Evaluation of a Passive m-health Sexual Information Texting Service in Uganda. *National Bureau of Economic Research Working Paper Series*, 19107.

Jareethum, R, Titapant, V, Chantra, T, Sommai, V, Chuenwattana, P and Jirawan, C, 2008. Satisfaction of healthy pregnant women receiving short message service via mobile phone for prenatal support: A randomized controlled trial. *Journal of the Medical Association of Thailand*, 91, pp.458–63.

Juma, C and Yee-Cheong, L, 2005. Innovation: applying knowledge in development. (ed.), Task Force on Science, Technology and Innovation. New York: UN Millennium Project.

Kamal, AK, Shaikh, Q, Pasha, O, Azam, I, Islam, M, Memon, AA, Rehman, H, Akram, MA, Affan, M, Nazir, S, Aziz, S, Jan, M, Andani, A, Muqeet, A, Ahmed, B and 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.

Karlan, D, Morten, M and Zinman, J, 2012. A Personal Touch: Text messaging for Loan Repayment. *National Bureau of Economic Research Working Paper Series,* 17952.

Keraro, FN, Wachanga, SW and Orora, W, 2007. Effects of Cooperative Concept Mapping Teaching Approach on Secondary School Students' Motivation in Biology in Gucha District, Kenya. *International Journal of Science and Mathematics Education*, 5, pp.111–24.

Khorshid, MR, Afshari, P and Abedi, P, 2014. The Effect of SMS Messaging on the Compliance with Iron Supplementation among Pregnant Women In Iran: A Randomized Controlled Trial. *Journal of Telemedicine and Telecare,* 20, pp.201–06.

Lagarde, M and Palmer, N, 2009. The impact of contracting out on health outcomes and use of health services in low and middle-income countries. *Cochrane Database of Systematic Reviews*, 4.

Lee, SH, Nurmatov, UB, Nwaru, BI, Grant, L and Pagliari, C, 2016. Effectiveness of m-health interventions for maternal, newborn and child health in low– and middle–income countries: Systematic review and meta–analysis. *Journal of Global Health,* 6, pp. 1-17.

Leong, KC, Chen, W. S, Leong, K. W, Mastura, I, Mimi, O, Sheikh, M. A, Zailinawati, A. H, Nh, C. J, Phua, KL and Teng, CL, 2006. The use of text messaging to improve attendance in primary care: a randomized controlled trial. *Family Practice,* 23, pp.699–705.

Lester, RT, Ritvo, P, Mills, EJ, Kariri, A, Karanja, S, Chung, MH, Jack, W, Habyarimana, JP, Sadatsafavi, M, Najafzadeh, M, Marra, CA, Estambale, B, Ngugi, E, Ball, TB, Thabane, L, Gelmon, LJ, Kimani, J, Ackers, M and Plummer, FA, 2010. Effects of a mobile phone short message service on antiretroviral treatment adherence in Kenya (WelTel Kenya1): a randomised trial. *The Lancet,* 376, pp.1838–45.

Liew, S, Tong, SF, Lee, VKM, Ng, CJ, Leong, WC and Teng, CL, 2009. Text messaging reminders to reduce non-attendance in chronic disease follow-up: a clinical trial. *British Journal of General Practice,* 59, pp.916–20.

Lin, H, Chen, W, Luo, L, Zhang, NCX, Zhong, X, Liu, Z, Chen, W, Wu, C, Zheng, D, Deng, D, Ye, S, Lin, Z, Zou, X and Liu, Y, 2012. Effectiveness of a Short Message Reminder in Increasing Compliance with Pediatric Cataract Treatment: A Randomized Trial. *Opthalmology,* 119, pp.2463–70.

Liu, GG, Chen, Y and Win, X, 2013. Transforming rural health care through information technology: an interventional study in China. *Health Policy and Planning,* 29, pp.975–85.

López, A, Reynoso, AM and Rossi, M, 2010. Impact evaluation of a program of public funding of private innovation activities: an econometric study of FONTAR in Argentina. *Inter-American Development Bank Office of Evaluation and Oversight Working Paper,* OVE/WP-03/10.

Lopez-Acevedo, G and Tinajero, M, 2010. Mexico: Impact Evaluation of SME Programs Using Panel Firm Data. *World Bank Policy Research Working Paper,* 5186.

López-Vargas, K, Parodi, S, Pineda, A and Tristao, I, 2015. Promoting Pro-School Behavior through Reinforcement Messages. Available at: <http://lacer.lacea.org/bitstream/handle/123456789/53165/lacea2015_promoting_pro_school_behavior.pdf?sequence=1> [Accessed 2017].

Lund, S, Hemed, M, Nielsen, BB, Said, A, Said, K, Makungu, MH and Rasch, V, 2012. Mobile phones as a health communication tool to improve skilled attendance at delivery in Zanzibar: a cluster-randomised controlled trial. *British Journal of Obstetrics and Gynaecology,* 119, pp. 1256-64.

Lund, S, Nielsen, BB, Hemed, M, Boas, IM, Said, A, Said, K, Makungu, MH and Rasch, V, 2014a. Mobile phones improve antenatal care attendance in Zanzibar: a cluster randomized controlled trial. *BMC Pregnancy and Childbirth,* 14:29.

Lund, S, Rasch, V, Hemed, M, Boas, IM, Said, A, Said, K, Makungu, MH and Nielsen, BB, 2014b. Mobile phone intervention reduces perinatal mortality in Zanzibar: secondary outcomes of a cluster randomized controlled trial. *JMIR m-health and uhealth,* 2(1):e15.

Marx, B, Pons, V and Suri, T, 2016. The Perils of Building Democracy in Africa. <https://www.povertyactionlab.org/sites/default/files/publications/SMS%20Experiment_NBER.pdf> [Accessed 13 March 2017].

Mbuagbaw, L, Thabane, L, Ongolo-Zogo, P, Lester, RT, Mills, EJ, Smieja, M, Dolvich, L and Kouanfack, C, 2012. The Cameroon mobile phone SMS (CAMPS) trial: a randomized trial of text messaging versus usual care for adherence to antiretroviral therapy. *PLoS One,* 7.

McKenzie, D, 2015. Identifying and Spurring High-Growth Entrepreneurship: Experimental Evidence from a Business Plan Competition. *World Bank Policy Research Working Paper,* 7391.

Miranda, J, Sabet, S and Brown, AN, 2016. Is impact evaluation still on the rise? Available at: <http://blogs.3ieimpact.org/is-impact-evaluation-still-on-the-rise/>

Munyegera, GK and Matsumoto, T, 2016. Mobile money, remittances, and household welfare: panel evidence from rural Uganda. *World Development,* 79**,** pp.127–37.

Muralidharan, K and Sundararaman, V, 2011. Teacher performance pay: experimental evidence from India. *The Journal of Political Economy,* 119**,** pp.39–77.

Olken, BA, Onishi, J and Wong, S, 2014. Should aid reward performance? Evidence from a field experiment on health and education in Indonesia. *American Economic Journal: Applied Economics,* 6(4), pp.1–34.

Pamuk, H, Bulte, E, Adekunle, A and Diagne, A, 2015. Decentralised innovation systems and poverty reduction: experimental evidence from Central Africa. *European Review of Agricultural Economics,* 42, pp.99–127.

Parker, C, Ramdas, K and Savva, N, 2016. Is IT enough? Evidence from a natural experiment in India's agriculture markets. *Management Science,* 62**,** pp.2,481–503.

Pires, JCL, Lodato, S, Cravo, T and Vellani, S, 2014. A Comparative Analysis of IDB Approaches Supporting SMEs: Assessing Results in the Brazilian Manufacturing Sector. Washington, DC: Inter-American Development Bank.

Pop-Eleches, C, Thirumurthy, H, Habyarimana, JP, Zivin, JG, Goldstein, MP, de Walque, D, Mackeen, L, Haberer, J, Kimaiyo, S, Sidle, J, Ngare, D and Bangsberg, DR, 2011. Mobile phone technologies improve adherence to antiretroviral treatment

in a resource-limited setting: a randomized controlled trial of text message reminders. *AIDS,* 25, pp.825–34.

Priedeman Skiles, M, Curtis, SL, Basinga, P and Angeles, G, 2013. An equity analysis of performance-based financing in Rwanda: are services reaching the poorest women? *Health Policy and Planning,* 28, pp.825–37.

Project ABC (n.d.). 'Project ABC Mobiles 4 Literacy'. Tufts University. Available at: <https://sites.tufts.edu/projectabc/> [Accessed 13 November 2016].

Rezaee, A, Hasanain, A and Khan, Y, 2015. Crowdsourcing government accountability: Experimental evidence from Pakistan. Available at: <http://econweb.ucsd.edu/~arezaee/pdfs/arezaee_jmp_11nov2015.pdf> [Accessed 21 March 2017].

Runde, DF and Magpile, J, 2014. Science, technology, and innovation as drivers of development. *CSIS Analysis, Commentary*. Available at: <https://www.csis.org/analysis?&field_contributor=220> [Accessed 2016].

Sabet, S, Heard, A and Brown, AN. 2017. *Science, technology, innovation and partnerships for development: an evidence gap map*. 3ie Evidence Gap Map Report 6. New Delhi: International Initiative for Impact Evaluation (3ie).

Sang, G, Valcke, M, Braak, JV, Zhu, C, Tondeur, J and Yu, K, 2012. Challenging science teachers' beliefs and practices through a video-case-based intervention in China's primary schools. *Asia-Pacific Journal of Teacher Education,* 40, pp.363–78.

Sever, S, Oguz-Unver, A and Yurumezoghu, K, 2013. The Effective Presentation of Inquiry-Based Classroom Experiments Using Teaching Strategies that Employ Video and Demonstration Methods. *Australasian Journal of Education Technology,* 29, pp.450–63.

Sharma, R, Hebbal, M, Ankola, AV and Murugabupathy, V, 2011. Mobile-phone text messaging (SMS) for providing oral health education to mothers of preschool children in Belgaum City. *Journal of Telemedicine and Telecare,* 17, pp.432–36.

Sheely, R (forthcoming). *Child Protection Knowledge and Information Network (CPKIN)* [Online]. Innovation for Poverty Action. Available at: <http://www.poverty-action.org/study/child-protection-knowledge-and-information-network-cpkin> [Accessed 13 March 2017].

Shetty, AS, Chamukuttan, S, Nanditha, A, Raj, RK and Ramachandran, A, 2011. Reinforcement of adherence to prescription recommendations in Asian Indian diabetes patients using short message service (SMS) – a pilot study. *The Journal of the Association of Physicians of India,* 59, pp.711–14.

Snilstveit, B, Vojtkova, M, Bhavsar, A and Gaarder, M, 2013. Evidence gap maps: a tool for promoting evidence-informed policy and prioritizing future research. *Policy Research Working Paper 6725*, The World Bank. Available at: <http://www-

wds.worldbank.org/external/default/WDSContentServer/IW3P/IB/2013/12/13/000158349_20131213135609/Rendered/PDF/WPS6725.pdf> [Accessed 13 March 2017].

Snilstveit, B, Vojtkova, M, Bhavsar, A and Gaarder, M, 2016. Evidence & gap maps: a tool for promoting evidence informed policy and strategic research agendas. *Journal of Clinical Epidemiology*, 79, pp. 120-29.

Sondaal, SFV, Browne, JL, Amoakoh-Coleman, M, Borgstein, A, Miltenburg, AS, Verwijs, M and Klipstein-Grobusch, KM, 2016. Assessing the Effect of m-health Interventions in Improving Maternal and Neonatal Care in Low- and Middle-Income Countries: A Systematic Review. *PLoS One,* 11(5): e0154664.

Tamban, C, Isip-Tan, IT and Jimeno, C, 2013. Use of Short Message Services (SMS) for the Management of Type 2 Diabetes Mellitus: A Randomized Controlled Trial. *Journal of the ASEAN Federation of Endocrine Societies,* 28.

Tan, HW, 2009. Evaluating SME Support Programs in Chile using Panel Firm Data. *World Bank Policy Research Working Paper,* 5082.

Ubfal, D and Maffioli, A, 2011. The Impact of Funding on Research Collaboration: Evidence from a Developing Country. *Research Policy,* 40, pp.1269–79.

UN System Task Team, 2013. Science, technology and innovation for sustainable development in the global partnership for development beyond 2015. *In:* ITU, O., UNCTAD, UNEP, UNESCO, UNFCCC, UNIDO, WIPO, WMO (ed.).

United Nations, 2014. The Road to Dignity by 2030: Ending Poverty, Transforming All Lives and Protecting the Planet. *Synthesis Report of the Secretary-General.*

United Nations, 2015. Addis Ababa Action Agenda of the Third International Conference on Financing for Development. In: *Report of the Third International Conference on Financing for Development, Addis Ababa, Ethiopia.* New York: United Nations.

United Nations, 2017. *Sustainable Development Goals* [Online]. Available: https://sustainabledevelopment.un.org/?menu=1300 [Accessed 2017].

UN Secretary-General Scientific Advisory Board, 2014. The Crucial Role of Science for Sustainable Development and the Post-2015 Development Agenda.

UNESCO (n.d.). Building Capacity in Science and Engineering. Available at: http://en.unesco.org/themes/building-capacity-science-and-engineering [Accessed 2016].

van Velthoven, MHMMT, Brusamento, S, Majeed, A and Car, J, 2013. Scope and effectiveness of mobile phone messaging for HIV/AIDS care: a systematic review. *Psychology, Health and Medicine,* 18, pp.182–202.

Waddington, H, White, H, Snilstveit, B, Hombrados, JG, Vojtkova, M, Davies, P, Bhavsar, A, Eyers, J, Koehlmoos, TP, Petticrew, M, Valentine, JC and Tugwell, P, 2012. How to do a good systematic review of effects in international development: a tool kit. *Journal of Development Effectiveness,* 4.

Wambugu, PW and Changeiywo, JM, 2008. Effects of Mastery Learning Approach on Secondary School Students' Physics Achievement. *EURASIA Journal of Mathematics, Science & Technology Education,* 4, pp.293–302.

Watson, R, Crawford, M and Farley, S, 2003. Strategic Approaches to Science and Technology in Development. *World Bank Policy Research Working Paper,* 3026.

World Bank Group, 2015. *World Bank Group Doubles Results-Based Financing for Education to US$5 Billion over Next 5 Years*. Available at: < http://www.worldbank.org/en/news/press-release/2015/05/18/world-bank-group-doubles-results-based-financing-for-education-to-us5-billion-over-next-5-years> [Accessed 17 December 2016].

World Bank Group, 2016a. *Net ODA received (% of GNI)*. Available at: <http://data.worldbank.org/indicator/DT.ODA.ODAT.GN.ZS> [Accessed 11 August 2016].

World Bank Group, 2016b. World development report 2016: digital dividends. Washington, DC: World Bank Group.

Yu, P, Courten, MD, Galea, G and Pryor, J, 2009. The development and evaluation of a PDA-based method for public health surveillance data collection in developing countries. *International Journal of Medical Informatics,* 78, pp.532–42.

Zhang, S, Wu, Q, van Velthoven, MHMMTV, Chen, L, Car, J, Rudan, I, Zhang, Y, Li, T and Scherpbier, RW, 2012. Smartphone Versus Pen-and-Paper Data Collection of Infant Feeding Practices in Rural China. *Journal of Medical Internet Research,* 14(5):e119.

Other publications in the 3ie Scoping Paper Series

The following papers are available from http://www.3ieimpact.org/en/publications/3ie-scoping-paper-series/

Adolescent sexual and reproductive health: the state of evidence on the impact of programming in low- and middle-income countries. 3ie Scoping Paper 5. Rankin, K, Heard, AC and Diaz, N (2016)

The state of evidence on the impact of transferable skills programming on youth in low- and middle-income countries, 3ie Scoping Paper 4. Brown AN, Rankin, K, Picon, M and Cameron, DB (2015)

Engaging communities for increasing immunisation coverage: What do we know?. 3ie Scoping Paper 3. Sabarwal, S, Bhatia, R, Dhody, B, Perumal, S, White, H and Puri, J (2015)

The current state of peacebuilding programming and evidence. 3ie Scoping paper 2. Brown, AN, McCollister, F, Cameron, DB and Ludwig, J (2015)

What evidence is available and what is required, in humanitarian assistance?, 3ie Scoping Paper 1. Clarke, M, Allen, C, Archer, F, Wong, D, Eriksson, A and Puri, J (2014)

This scoping paper summarises a range of activities 3ie undertook to assess stakeholder demands and priorities for generating new impact evaluation evidence on how science, technology, innovation and partnerships accelerate development outcomes in low- and middle-income countries.

We identify several important gaps in the evidence, such as the impacts of technology interventions that use biometrics and non-health-related innovative financing interventions. We also identify several sectors and themes that should be prioritised for new impact evaluations. They include impacts on crises and conflict and agriculture and food security outcomes. We also recommend that new impact evaluations conduct cost analysis. There are several clusters of studies that are promising for synthesis, including studies on mobile money systems, digital information services for agriculture and food security, and m-health interventions.

Scoping Paper Series

International Initiative for Impact Evaluation
203-202, 2nd Floor, Rectangle One
D-4, Saket District Center
New Delhi – 110017
India

3ie@3ieimpact.org
Tel: +91 11 4989 4444

www.3ieimpact.org

USAID
FROM THE AMERICAN PEOPLE