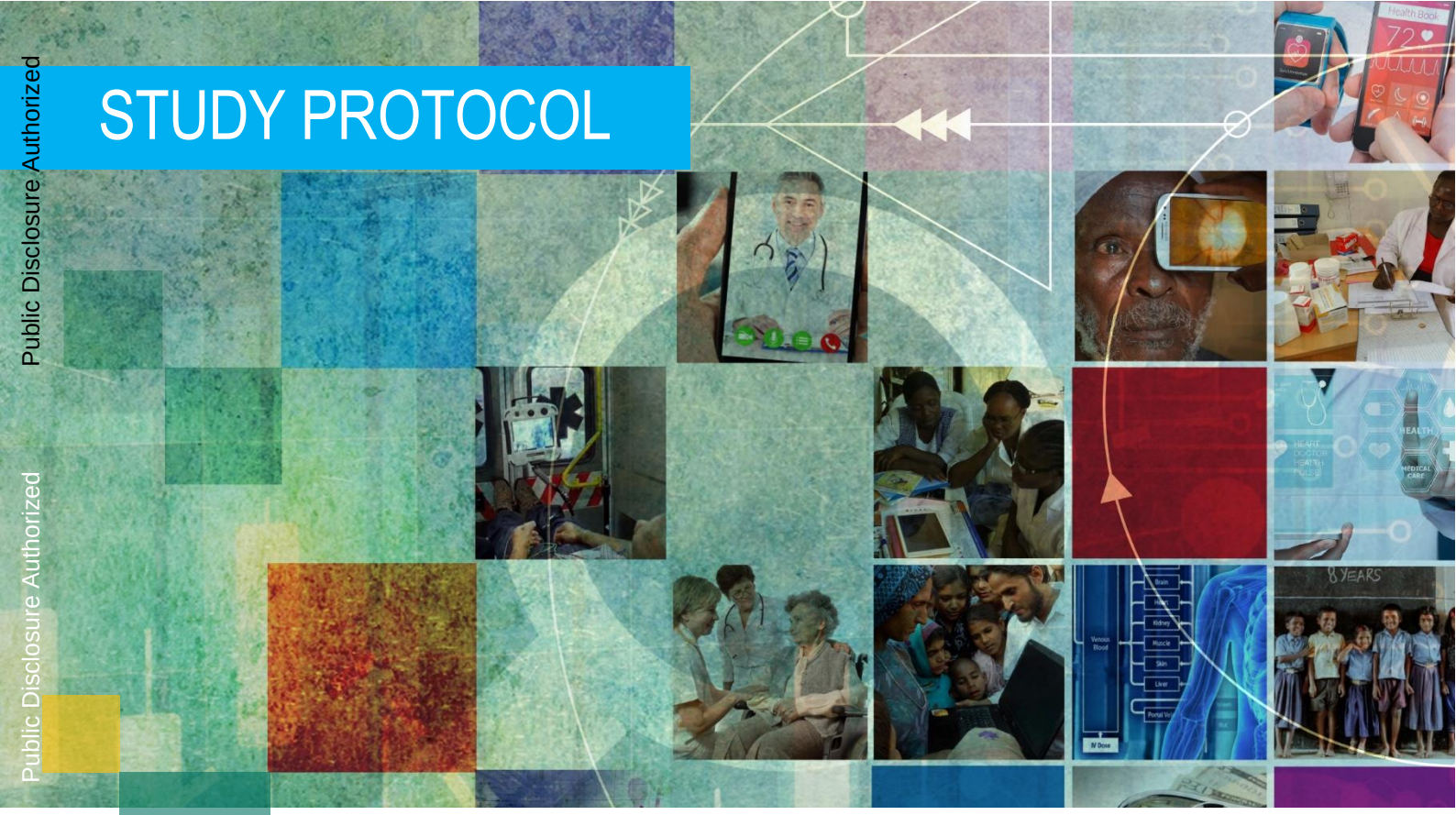
STUDY PROTOCOL



AN EVIDENCE GAP MAP FOR DIGITAL HEALTH INTERVENTIONS

WORLD BANK GROUP

STUDY PROTOCOL :
AN EVIDENCE GAP MAP FOR DIGITAL HEALTH INTERVENTIONS

The World Bank:
Marelize Görgens, Nejma Cheikh
and Thomas Wilkinson

The International Initiative for Impact Evaluation (3ie):
Birte Snilstveit, Collins Zamawe,
Miriam Berretta and Richard Appell

This page is for collation purposes only.

CONTENTS

TABLES

DIGITAL HEALTH INTERVENTIONS: AN EVIDENCE GAP MAP

Digital health interventions in health systems is a new and fast-growing field of study in public health. Such technologies are vital in meeting new challenges such as the increase of non-communicable diseases, shortage of the health workforce, aging population, unplanned emergencies, and infectious disease outbreaks. Digital Health has been widely recognised as a potential tool to advance the Sustainable Development Goals (SDGs) and to support health systems in all countries, in health promotion and disease prevention. Interventions have targeted all levels of health care delivery including the patient, facility and system levels, aiming to address a range of outcomes, including demand, access, behaviour, quality and integration. However, these interventions can be expensive, so increased importance has been placed on assessing effects, cost-effectiveness and cost-benefit. While some research is available, the evidence base is still relatively nascent. This presents an opportunity to develop a strategic and coordinated research agenda that addresses important evidence gaps. To do so it is important to establish what is already known, identify common outcome measures, establish best practice methods for evaluating effects, cost-effectiveness and cost benefit; and provide access to relevant reference cases which can inform new research. This Evidence Gap Map will contribute to this objective.

1. BACKGROUND

1.1 THE ISSUE

Digital Health Interventions (DHIs) and Artificial Intelligence (AI) have an increasingly important role in the healthcare landscape as advances in technology produce digital health products (analytics and tools). Such tools enable countries to create health systems that deliver personalised, pre-emptive, predictive and participative healthcare.

Currently, there is no conformity regarding how to evaluate the impact of DHIs and AI on healthcare outcomes of interest, or the outcomes of interest themselves. To facilitate governments' decision making with regards to DHI and AI there is a need to develop standardised evaluations of effects, cost-effectiveness and cost-benefit. Such standardisation will help provide comparative evidence to inform future investments in this sector. Therefore, the World Bank is looking to create a reference case on doing economic evaluations in the DHI and AI space, and to conduct selected impact evaluations in this field.

To inform and support this effort the World Bank needs to gain a better understanding of both the characteristics of existing impact evaluations and systematic reviews, and where there are gaps in the available evidence.

1.2 STUDY OBJECTIVES AND QUESTIONS

The broad goal of this EGM is to facilitate decision-making regarding investment in research assessing the effects, cost-effectiveness and cost-benefit of DHIs.

This EGM will do so by reviewing the available evidence, identifying both substantive and methodological characteristics of the literature and describe evidence gaps. By identifying the 'evidence baseline' and best practice for future research it aims to ensure that new studies are as rigorous and useful as possible. A more specific objective is to support the development of a reference case which can be used by stakeholders conducting impact evaluations, cost-effectiveness and cost benefit analyses to address high priority evidence gaps in different contexts. The use of AI in health care can be applied in both digital and non-digital (analog) interventions; this EGM will focus on DHIs with further evidence searches specifically targeted towards AI in a later phase.

To address our objectives, we will address the following research questions:

1. What are the characteristics of the current landscape of evaluations addressing the effects of DHIs (e.g., geographical distribution, interventions covered, outcomes measured, attention to equity, study design, measurement of costs and resource use)?
2. What are the intermediate and final outcomes that are measured in evaluations assessing the impacts of different interventions and at different levels of DHIs?
3. What are the methodological approaches and metrics used to evaluate the effects of DHIs?
4. At what level of the health system is most of the evidence concentrated (patient, facility, system)?

To address these questions, we will conduct an EGM, which will include a review of relevant literature and a landscape analysis of outcomes of interest. The EGM will be presented on an interactive platform, and accompanied by a report with analysis of the available evidence.

2. SCOPE OF THE EGM

2.1 INTERVENTIONS

We will use a conceptual framework informed by the World Health Organization classification of DHIs[1] to further define the intervention inclusion criteria. Here DHI are organized into broad categories based on 'targeted primary user', as below:

1. **Clients:** Includes members of the public and caregivers who are potential and current users of health services.
2. **Healthcare providers:** Health workers involved in the delivery of services to clients.
3. **Health system and resource managers:** Administrators of healthcare systems and healthcare providers
4. **Data services:** Cross-cutting activities to support the overall functioning of data collection and digital management of health services.

Within these categories there are a number of overall intervention strategies. Interventions without a clear digital component are outside of the scope of this project.

2.2 OUTCOMES

An extensive range of outcomes will be included in the EGM. The list below is indicative of the types of outcome measures we will include, but is not exhaustive. Part of the objective of this work is to identify what outcomes are commonly measured, and establish meaningful categories of outcome measures. Ultimately this will enable the broader project to propose standardise outcome measures for future research, in consultation with stakeholders. We will therefore not exclude studies on the basis of outcomes, but rather record outcome measures and their definitions and categorise these into broad outcome categories which may include but would not be limited to:

FIRST-ORDER OUTCOMES

- Indication/therapeutic -specific outcomes (eg number of people achieving treatment adherence)
- Quality of care outcomes (eg quality of care survey improvements)
- Client Satisfaction outcomes (eg reported ease of use of technology)
- Process outcomes (eg number of people successfully enrolled in a program)
- Utilisation outcomes (eg number of usages of app functionality)

ECONOMIC OUTCOMES

Any measure of costs, cost effectiveness and cost benefit. Examples of measures that will be included are reduction in treatment costs implementation costs, and cost ratios such as cost-per life year saved, return on investment and net benefit.

GENERALISED HEALTH OUTCOMES

We will include any measure of health behaviour, mortality and morbidity, including measures such as lives saved, Quality-adjusted life-years (QALYs) gained or Disability-adjusted life years (DALYs) averted.

---

[1] World Health Organization *Classification of Digital Health Interventions v1.0. A shared language to describe the uses of digital technology for health.* 2019.

2.3 TYPES OF STUDY

We will include the following study design for impact evaluations:

- Randomised controlled trial (individual or cluster);
- Regression discontinuity design;
- Controlled before-and-after studies using appropriate methods to control for selection bias and confounding, such as:
  - 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;
- Natural experiments;
- Other quasi-experimental studies including studies using synthetic controls;
- Cross-sectional or panel studies with an intervention and comparison group using methods to control for selection bias and confounding as described above.
- Interrupted time series (ITS)

Finally, a criterion that applies to all studies is that they have to be so-called effectiveness studies. These types of studies stand in contrast to efficacy trials which test an intervention under ideal and controlled conditions in order to maximise the likelihood of observing an effect, if one exists. The justification for this approach is that our interest is to identify evidence on the effects of an intervention implemented under circumstances that approach 'real- world' practice.

Although there exists broad agreement on the type of study design characteristics of effectiveness (pragmatic) trials and efficacy (explanatory) trials, there is currently no validated definition of 'effectiveness studies' (Treweek et al., 2009; Gerthlener et al., 2006; Singal et al., 2014). We therefore developed five criteria to help us distinguish more clearly between efficacy trials and effectiveness studies, drawing on two existing tools (Gartlehner et al., 2006; Thorpe et al., 2009). Studies will be considered efficacy trials and be excluded if they fulfil at least one of the criteria outlined below:

- **Research Objective:** Is the study primarily designed to determine to what extent a specific technique, technology, treatment, procedure or service works under ideal condition rather than attempt to answer a question relevant to the roll- out of a large programme?
- **Population:** Are the participants highly selected and therefore unrepresentative of the general population (Are strict inclusion and exclusion criteria used to enrol a homogenous population which may limit the generalizability of the results? e.g., students that truly have a disease of interest or are more likely to adhere to the treatment)?
- **Providers:** Is the intervention primarily delivered by the research study team rather than trained laypersons (parents/ teachers/ community members/ NGOs) who don't have extensive expertise?
- Delivery of intervention: Is the intervention delivered with high degree of assurance of delivery of the treatment? (Is the delivery tightly monitored/ supervised by the researcher following specific protocols; Is adherence to the treatment monitored closely with frequent follow- ups?)
- **Delivery of intervention:** Are concurrent interventions restricted to the study population in order for a witnessed effect to be attributed to the intervention of interest?

For systematic reviews, we will include studies explicitly described as systematic reviews and reviews that describe search, data collection and synthesis methods according to the 3ie database of systematic reviews protocols (Snilstveit et al. 2014).

2.4 POPULATION

Target populations are clients receiving a DHI via digital platform, healthcare providers who prescribe a DHI and monitor outcomes, healthcare systems collecting and reporting information through a DHI, healthcare workers utilizing a DHI to facilitate their work and interact with patients or other healthcare entities. We will include studies regardless of geographical location.

2.5 ADDITIONAL INCLUSION AND EXCLUSION CRITERIA

- There are no restrictions on the population for the geographic region or income status of countries included in the map.
- Given the relatively recent state of the evidence base in this field and the fact that most programs/initiatives have started to flourish in the 2000s, we will conduct the searches for primary studies from 2000 onwards and for systematic reviews from 2014 onwards.
- Purely observational studies (interviews, surveys, correlational studies, design workshops, etc.) and non-systematic literature reviews will be excluded.
- We will not exclude studies on the basis of language, but the search will at least initially focus on English language databases. Where non-English studies are identified in the search results these will be included provided resources are available to do so.

2.6 REPRESENTATION OF EXCLUDED STUDIES

It is expected that there will be a significant number of studies of DHIs, particularly from non-peer review sources, that do not meet the study inclusion criteria. This body of published information is commonly in the form of a simple case-study report on donor/ proprietary/ development partner websites and often provide partial and selective data with no comparative analysis or indication of a structured research design intended to generate evidence to inform decision making. Although these studies would not be incorporated in the EGM, these studies are relevant to the workstream this EGM intends to inform as it is proposed that improved guidance on how to generate evidence would encourage production of higher quality evidence. Therefore, limited metadata will be collected from a sample of these excluded studies to provide an indication of the intervention type and context of this body of published information. Further details about the metadata collected is described in Section 3.5.

3. METHODS

3.1 OVERALL APPROACH

We will follow the standards and methods for EGMs developed by 3ie (Snilstveit et al., 2016; Snilstveit et al., 2017). An evidence gap map aims to establish what we know, and do not know, about the effects of interventions in a thematic area (Snilstveit et al., 2016). The map is populated through systematic searching, screening and data extraction of all relevant completed, and ongoing, impact evaluations and systematic reviews.

The map is populated through systematic searching and screening for all relevant completed, and ongoing, impact evaluations and systematic reviews. The included studies are mapped onto the framework of interventions and outcomes and will be presented on an interactive platform which provides a graphical display of the evidence in a grid-like framework. This provides a visual display of the volume of evidence for intervention-outcome combination, the type of evidence (impact evaluation, systematic reviews, completed or ongoing), and a confidence rating for systematic reviews. The final map will be published on an online interactive platform that provides additional filters so that the users can further explore the available evidence, for example by country or population.

The interactive map will be accompanied by a report addressing the key research questions, including an analysis of the characteristics of the available evidence, key trends and exemplar evaluations.

Evidence gap maps highlight both absolute gaps, which should be filled with new primary studies, and synthesis gaps, which are ready for new systematic reviews and meta-analyses. They are envisioned as a global public good, and this allows them to be used as a tool to democratise access to high quality research.

3.2 SEARCH STRATEGY

Due to the broad scope of this EGM we will implement a sensitive search strategy primarily constructed by a combination of intervention and study design terms and developed by an information specialist. The strategy will be translated according to the requirements and functionalities of different databases.

We will search a range of different sources of academic and grey literature, including academic databases (combination of general social science and health focused databases), repositories of impact evaluations and systematic reviews, specialist organizational databases and websites of bilateral and multilateral agencies. The complete list of sources is listed below.

Academic databases:

- Econlit (Ovid): http://www.ovid.com/site/catalog/databases/52.jsp
- Ebsco Discovery (limited to Repec, World Bank E-Library, Econlit, Africa-Wide): https://www.ebscohost.com/discovery
- Scopus: https://www.scopus.com/
- Social Sciences Citation Index (SSCI) (via Web of Science): https://library.maastrichtuniversity.nl/collections/databases/ssci/
- Medline
- Embase
- CAB Global Health
- Popline
- Epistemonikos

- PsycInfo
- CINAHL

Repositories of evaluations and systematic reviews:

- 3ie Repository of Impact Evaluations http://www.3ieimpact.org/en/evidence/impactevaluations/
- 3ie RIDIE (Registry for International Development Impact Evaluations): http://ridie.3ieimpact.org/
- USAID Evaluation Clearing House: https://dec.usaid.gov/dec/content/evaluations.aspx
- Innovations for Poverty Action (IPA) www.poverty-action.org/project-evaluations
- The Abdul Latif Jameel Poverty Action Lab (JPAL) :www.povertyactionlab.org
- AEA RCT Registry : https://www.socialscienceregistry.org/
- Campbell Collaboration, www.campbellcollaboration.org
- Cochrane Collaboration
- African Development Bank (AfDB): https://www.afdb.org/en/documents/publications/
- BREAD: http://ibread.org/bread/papers
- Center for Effective Global Action (CEGA): http://cega.berkeley.edu/evidence/
- DFID Research for Development (R4D): http://r4d.dfid.gov.uk/
- CENTRAL (Cochrane)
- Epistemonikos

We will search specialist organizational databases, which might include evidence on digital interventions in the health sector:

- World Health Organization Database
- Health Technology Assessment Database
- GSMA mWomen
- Mobile Active
- mHealthevidence.org
- mHealth Alliance: https://mhealthknowledge.org
- Knowledge for Health
- International Center for Research on Women
- Women for Women International
- Mobiles for Education Alliance
- mHealth Info, and
- Health Unbound
- Charity Science Health

We will also conduct backwards citation tracking of systematic reviews that meet our inclusion criteria, as well as any relevant non-systematic reviews and guidelines identified to identify additional studies. Once our draft EGM is completed, we will circulate to key experts and stakeholders to identify any studies not already identified.

3.3 SCREENING PROTOCOL

The search results will be imported into EPPI-reviewer, and we will use this platform to manage references, identify and remove duplicate studies, and screen records for inclusion using the procedures outlines below.

At the title and abstract stage, we will combine ‘safety first’ single screening (Shemilt et al., 2016) with EPPI-reviewer’s machine learning functionality to speed up the screening process. We will start by screening a randomly selected set of around 1000 studies to provide a training set for the machine-learning

algorithm and use the prioritisation functionality, to prioritise studies for screening according to their likelihood of inclusion.

In addition, we will explore the potential of using other machine learning strategies for reducing screening workload and speeding up the study selection process, including by building a study classifier based on existing screening data from projects searching for evaluations and systematic reviews. All use of machine learning will be done in consultation with an expert in rapid and systematic literature searching.

We will train all screeners in the implementation of the screening protocol. This will be followed by an independent review by all screeners and senior review team members of the same set of 100 studies. We will then compare the results, discuss any discrepancy in coding decisions and clarify the inclusion criteria as needed.

Table 3.1 Screening protocol and data capture at title & abstract level

Metadata	Field values	Detail
Exclusion criteria for studies		
Relevance	Yes/no	Exclude any study that does not describe an intervention or is not describing an intervention in the health sector, or that the intervention is not digital.
Study design	Yes/no	Exclude any study that does not meet the inclusion criteria for study design
Empirical data	Yes/no	Exclude as no real-world data included. Typically, relevant for simulations
Efficacy trials	Yes/no	Exclude if efficacy trials
To inform final decisions about scope and the landscape of excluded studies we will extract the following data from the title and abstracts of all included records and 25% (selected randomly) of studies excluded for reason of study design at this stage.		
Metadata	**Field values**	**Detail**
Country name	Drop down list	If multiple countries then this field can simply return "multiple"
Region	EMRO, AFRO etc.	Using WHO regional classifications. If multiple countries across regions, select predominant region
Country income level	Automatically populated based on country name selection. Where multiple countries, predominant income level should be selected	
Interventions	Client / Healthcare providers / Health systems management / Data services	WHO classification of digital health interventions.
Economic evidence	Yes/No	Does the study include some measurement of costs, cost effectiveness or general economic information

At the full text stage, we will also use safety-first screening, with any study where the first screener is uncertain about inclusion assessed by a second, more senior reviewer. We will follow the same procedure for training screeners as at title/ abstract, using a training sample of 30 full text studies.

3.4 DATA EXTRACTION AND CRITICAL APPRAISAL

For both impact evaluations and systematic reviews, our data extraction templates will be modified from 3ie's repository coding protocol and the coding protocols typically used for EGMs. This includes bibliographic, geographic information and substantive data, as well as standardised methods information and data on how studies address equity. In addition, we will extract data on interventions (type and intervention description), outcomes (type and outcome definition), population, cost effectiveness data and timeframe of intervention.

The full tools for extracting data from impact evaluations and systematic reviews. They will be piloted and then training will be provided to all coders (where they all code the same 3 studies) to ensure consistency in coding and resolve any issues or ambiguities. A single researcher will conduct the data extraction for each study; however, a sample will be double coded to check for consistency. The type of data extracted is listed below.

Variables:

- Coder name
- Study ID
- Title name
- Author name
- Publication information
  - Publication type (journal article, working paper, report, 3ie series report, book/book chapter)
  - Doi
  - Abstract
  - Journal name
  - Journal volume, issue, pages
  - Year of publication
  - Url
  - Open access (yes/no)
- Equity information
  - Study population (healthcare professionals, beneficiaries – children, adult, newborn, adolescents, or other)
  - Equity focus (how this study considers gender and/or equity)
  - Equity dimension (specific vulnerable group addressed)
  - Equity description
  - Keywords
- Geographic information [already collected at screening]
  - WHO region
  - Country name
  - Country income level
- Methodological information
  - Evaluation design (experimental or quasi-experimental)
  - Evaluation method (experimental → RCT; quasi-experimental → regression discontinuity design, difference-in-difference, fixed effects, instrumental variables, statistical matching, synthetic controls, or interrupted time series)
  - Mixed methods (yes/no)
  - Additional methods

- Program, funding and implementation
  - Project/program name
  - Program implementation agency category and name
  - Program funding agency category and name
  - Research funding agency category and name
  - Study of intervention implemented at national or provincial or local level
- EGM specific information
  - DHI category First Tier eg Clients (1.0)
  - DHI category Second Tier eg Targeted Client Communication (1.1),
  - DHI category Third Tier eg Transmit health event alerts to specific population groups (1.1.1).
  - Intervention name (as described in study)
  - Intervention description
  - Health system application (Client education and behaviour change communication, Workforce development and support, Service delivery, Information systems, Financial transactions and incentives, Supply chain management, Governance)
  - Broad outcome category (quality of care outcomes, client satisfaction outcomes, process outcomes, utilization outcomes, behaviour change, health status, lives saved, or generalised health utility-outcome (quality-adjusted life years or disability adjusted life years)
  - Outcome label
  - Outcome description
  - Cost data (cost only, cost-benefit analysis, cost-effectiveness analysis, return on investment or costs not specified above or none)
  - Cost description
  - Health domain (family planning and reproductive health, maternal health, child health, Infectious, parasitic and vector-borne disease, Non-communicable disease, nutrition or not applicable (for DHIs not specific to particular domain eg general electronic patient records)
  - Whether the intervention involves the application of artificial intelligence component
  - What is the category of the artificial intelligence component (data, processing, action or other)
  - What is the sub-category of the artificial intelligence component?
  - Description of the AI component
  - Whether adverse findings or unintended consequences from the study
  - Type of adverse finding (health related or non-health related)
  - Description of the adverse event or unintended consequence.

Additionally, all included systematic reviews will be critically appraised following the 3ie systematic review database protocol (3ie, n.d.), which draws on Lewin et al. (2009). The tool assesses the review on how the search, screening, data extraction, and synthesis was conducted, and covers all the most common areas where biases are introduced. Overall, each systematic review will be rated as low, medium, or high confidence. One reviewer will conduct the initial critical appraisal, and a systematic review methods expert will conduct a final review of all appraisals.

It is important to note that we will not be critically appraising primary studies, as this is beyond the scope of the project.

3.5 METHODS FOR REPRESENTATION OF EXCLUDED STUDIES

As described in Section 2.6, it is necessary to also represent the landscape of studies that although are excluded from the EGM are relevant to the boarder workstream. To represent this body of information, we will record the reasons for exclusion (where appropriate) and limited metadata about the type of intervention, country context of the study, health topic and whether there is costing or cost effectiveness data (detailed in table 3.1).

This will be carried out for the random sample of studies required to build the machine learning classifier (see section 1.2) and will provide an indicative view of the literature beyond our inclusion criteria. Based on this metadata, we will provide a summary of the evidence landscape using the sample. We will also identify example studies that provide an illustrative view of the state of evidence for the metadata we have collected.

When completed, the EGM will be accompanied by a report summarising the findings of our searches. We will describe the size and characteristics of the existing evidence base and identify major evidence gaps. Such analysis will be supported by descriptive statistics, presented in tables and figures as appropriate. We will provide data on interventions and outcomes studied in impact evaluations or systematic reviews, types of studies (and designs), geographical location of studies, sex disaggregation, and how studies address cross cutting issues like gender and equity. We will also provide details of the methods used to assess costs, cost-effectiveness and cost-benefit.

The EGM report will also include a systematic analysis of evidence gaps, focusing on (a) absolute gaps (no or few primary studies); (b) synthesis gaps (clusters of impact evaluations but no high-quality systematic reviews; and (c) other gaps (any gaps or in-balance in the evidence base, including for example type of studies, or geographical locations).

4. REFERENCES

Hugh Waddington, Edoardo Masset & Emmanuel Jimenez (2018) What have we learned after ten years of systematic reviews in international development? *Journal of Development Effectiveness,* 10:1, 1-16, DOI: 10.1080/19439342.2018.1441166

Snilstveit, B. 2012. "Systematic Reviews: From 'Bare Bones' to Policy Relevance." *Journal of Development Effectiveness* 4 (3): 388–408. doi:10.1080/19439342.2012.709875.

Snilstveit, B., M. Vojtkova, A. Bhavsar, and M. Gaarder, 2013. Evidence Gap Maps: A Tool for Promoting Evidence-Informed Policy and Prioritizing Future Research. Policy Research Working Paper WPS6725. Washington, D.C.: World Bank.

Snilsveit, B., Vojtkova, M., Bhavsar, A., Stevenson, J., and Gaarder, M. (2016). Evidence and gap maps: A tool for promoting evidence informed policy and strategic research agendas. *Journal of clinical epidemiology*, 79, pp. 120–129.