From proxy-indicators to connecting disparate evidence: a multilevel-toolkit for evaluating the impact of digital health implementations on health outcomes

PERRIN FRANCK, Caroline

Abstract

Digital Health is a multi-disciplinary science connecting medicine, technology, and implementation science. A growing evidence base highlights increasing recognition for its potential to strengthen overall health system performance. However, historically, the majority of evidence focuses on improvements of care processes rather than health outcomes. This thesis investigated aspects of evaluating health outcomes of digital health on different levels and proposes approaches to overcome some of the associated challenges. First, on micro level, what to measure by identifying meaningful proxy-indicators for maternal/neonatal outcomes. Second, on meso level, understanding how researchers approach digital health impact evaluations and providing a pedagogical overview on methods. Finally, on macro level, an approach to connecting disparate knowledge, experience, and evidence. Together, these articles provide a multi-level toolkit for evaluating and aggregating outcomes of digital health interventions to establish a comprehensive picture of the digital health landscape and increase the understanding for determinants influencing [...]
From proxy-indicators to connecting disparate evidence: a multilevel-toolkit for evaluating the impact of digital health implementations on health outcomes
DOCTORAT EN SCIENCES BIOMEDICALES
MENTION SANTE GLOBALE

Thèse de:

Caroline PERRIN
originaire de Münich, Allemagne

Intitulée:

From proxy-indicators to connecting disparate evidence: a multilevel-toolkit for evaluating the impact of digital health implementations on health outcomes

La Faculté de médecine, sur préavis du Comité directeur du PhD, autorise l'impression de la présente thèse, sans prétendre par-là émettre d'opinion sur les propositions qui y sont énoncées.

Genève, 30 janvier 2020

Thèse n° 004

Pr Cem Gabay
Doyen

N.B. - La thèse doit porter la déclaration précédente et remplir les conditions énumérées dans les "Informations relatives à la présentation des thèses de doctorat à l'Université de Genève".
Acknowledgement

Undertaking this PhD has been a long and life-changing experience for me and it would not have been possible without the support and guidance that I received from many people.

Foremost, I would like to express my special appreciation and thanks to my supervisor Prof. Antoine Geissbuhler, who has been an exceptional mentor for me. I would like to thank him for his patience, motivation, and encouragement, and for allowing me to grow as a research scientist. His guidance on my research has been invaluable. I admire his vision, his capability to lead and inspire, while being incredibly organized and a great problem solver, all of these qualities were immensely helpful in moving my thesis forward. I would also like to thank my doctoral committee, Dr. Gijs Walraven, Dr. Beat Stoll, and Prof. Jaya Krishnakumar for their support in conducting my research, their encouragement and insightful comments.

I would like to thank my colleagues and co-authors for their work, the many interesting discussions, for sharing their knowledge with me, and for encouraging and motivating me throughout this journey.

I would like to thank the Institute of Global Health at the University of Geneva, particularly Prof. Antoine Flahault, Dr. Nathalie Bot, and Lemlem Girmatsion for their excellent support and guidance throughout this PhD program. I am grateful to have been part of the first cohort of this innovative and interdisciplinary program.

I would like to thank my family and friends for all the support, love and encouragement throughout the years. I am particularly grateful to my parents Barbara, Gerard, and Jürgen, my parents-in-law Odile and Jean-Louis, my aunt Christiane and my uncle Bert, my sister Geraldine and my brother Michael.

My husband Thomas for his support and always listening to my thoughts as well as enriching them with his point of view. My children Elena and Niklas, for always cheering me up and for helping me to see the world from a different perspective.
1. List of publications

1.1. Accepted publications

- Perrin, C.; Hounga, L.; Geissbuhler, A. “Systematic review to identify proxy indicators to quantify the impact of eHealth tools on maternal and neonatal health outcomes in low-income and middle-income countries including Delphi consensus.” BMJ open (2018), 8(8), e022262.


1.2. Submitted publications

- Perrin, C.; Randriambelonoro, M.; Geissbuhler, A. “Navigating through digital health evaluation methodologies: an algorithm based on a scoping review and emerging methodologies.” Submitted to JMIR Medical Informatics (currently under review)
2. Abstract (English)

Digital health, which is the use of digital, mobile and wireless technologies for health, has reached unprecedented presence globally, and is increasingly recognized for its potential to strengthen overall health system performance. Application scope of digital health tools and interventions is broad and evolving, ranging from facilitating access to care, improving continuity of care, risk factor-centric digital health interventions (e.g. lifestyle management, diet, physical activity), empowering consumers to track and manage their own health, to shaping the way to provide personalized and precision medicine. Particularly in low- and middle-income countries, where mobile connectivity has constantly increased over the last decade, digital health can be a catalyst to address some of the most pressing issues in health systems. There is a global shortage and uneven distribution of health workers, which is likely to increase as rising incidence of non-communicable diseases (NCDs) and a globally ageing population generate additional healthcare demand.

The potential of digital health is highlighted by a growing evidence base. However, historically, the majority of evidence focuses on care process improvements rather than improvements in health outcomes. Despite the progress of the last decade in understanding the role of digital health to strengthen the overall health system performance, the availability and use of digital health outcome evidence for informed decision making is still relatively weak. Consequently, there is a strong need to embed implementation science research to identify what works and what doesn’t in different contexts to strengthen digital health from lower facility level to the national level decision makers.

The title of this thesis is: From proxy-indicators to connecting disparate evidence: a multilevel-toolkit for evaluating the impact of digital health implementations on health outcomes. The objectives of this thesis were to investigate at different levels how to evaluate the impact of digital health on outcomes. First, on micro level, what to measure by identifying meaningful proxy indicators for maternal and neonatal outcomes. Second, on meso level, understanding how researchers approach digital health impact evaluations and providing a pedagogical overview on methods. Finally, on macro level, understanding and proposing an approach to connecting disparate knowledge, experience, and evidence.

Part 1 identified a list of 40 potential proxy indicators linked with a positive impact on maternal/neonatal outcomes through a systematic review of the literature. It was completed and prioritized through a Delphi expert consensus, and resulted in a list of 77 potential proxy indicators. These proxy indicators suggest relevant outcome measures to assess whether eHealth interventions directly affect maternal/neonatal outcomes. Some of these indicators require a mapping to the local practices, context and available resources. This mapping allows the utilization of the proxy indicators in different settings while providing the ability to systematically collect data of various implementations, projects or programs.

Part 2 investigated how researchers approach digital health outcome evaluations in different settings through a scoping review. An algorithm was developed on the basis of these results and provides a pedagogical overview
of methods for evaluating health outcomes of digital health interventions. The analysis of the literature, illustrated that digital health outcome evaluations rely on traditional clinical study designs, albeit these interventions are commonly more complex and dependent on the context, culture, and the individual than classical interventions. Desk research identified methodologies, in addition to those extracted from the scoping review, with design features that address some of the shortcomings of traditional clinical methodologies, particularly when applying them to digital health interventions. The studies included in the scoping review were analyzed for reasons why investigators chose them and for identified shortcomings in the context of applying them to a digital health intervention. Together with advantages and disadvantages of the additional methodologies the main decision points were modeled into an algorithm that provides a high-level overview, aiming at enabling the user to navigate through these decision points based on design features and investigator’s priorities. The algorithm aims at facilitating the identification of one or more potential appropriate methodologies.

Part 3 proposes the RAFT annotation model, a semantic description of all elements of various outputs and implementation projects that are part of the RAFT elearning and telemedicine network. This model was developed to annotate various inputs, processes and outputs of the RAFT network in order to facilitate knowledge documentation and sharing. It serves as a basis for the Implementome, which will be an interconnected knowledge system enabling users to navigate on multiple dimensions through metadata of annotated implementations, individuals, or information, and can provide tailored input for consensus building, and the development of best practices and guidelines. The annotation model was developed for the RAFT network; however, it can be extended to enable the general annotation of implementations, outputs, initiatives, or individuals in the digital health domain.

This thesis investigates aspects of evaluating health outcomes of digital health on different levels and proposes approaches to overcome some of the associated challenges. The first article identifies a list of proxy-indicators to measure the impact of digital health on neonatal or maternal health outcomes. The second article provides a pedagogical overview on evaluation methodologies and suggests an algorithm to help identifying approaches on how to measure. The third article outlines the added value of bi-directional learning between the south and north. The forth article proposes a model that enables semantically annotating and connecting information, evidence, knowledge and experience to provide a basis for informed decision-making. Together, these articles provide a multi-level toolkit for evaluating and aggregating the outcomes of digital health interventions.

Aggregation and integration of experience, knowledge and evidence at different levels will help to ensure to get a better, more complete picture of the digital health landscape and will increase the understanding of inputs, processes and outputs and their determinants for success and failure.
3. Abstract (French)

La santé numérique, qui est l'utilisation des technologies numériques et mobiles pour la santé, a atteint une présence mondiale sans précédent et est de plus en plus reconnue pour son potentiel à renforcer la performance globale des systèmes de santé. Le champ d'application des outils et des interventions en santé numérique est vaste et évolutif, allant de la facilitation de l'accès aux soins à l'amélioration de la continuité des soins, en passant par les interventions en santé numérique axées sur les facteurs de risque (p. ex. gestion du mode de vie, alimentation, activité physique), la responsabilisation des consommateurs pour suivre et gérer leur propre santé, et la mise au point d'une médecine de précision. En particulier dans les pays à faible revenu et à revenu intermédiaire, où la connectivité mobile n'a cessé d'augmenter au cours de la dernière décennie, la santé numérique peut servir de catalyseur pour résoudre certains des problèmes les plus urgents des systèmes de santé. Il existe une pénurie mondiale et une répartition inégale du personnel de santé, qui devrait s'aggraver à mesure que l'incidence croissante des maladies non transmissibles (MNT) et le vieillissement de la population dans le monde va générer une demande accrue en soins.

Le potentiel de la santé numérique est mis en évidence par un nombre croissant de preuves. Cependant, historiquement, la majorité des données probantes portent sur l'amélioration des processus de soins plutôt que sur l'amélioration des résultats pour la santé. Malgré les progrès réalisés au cours de la dernière décennie dans la compréhension du rôle de la santé numérique pour renforcer le rendement global du système de santé, la disponibilité et l'utilisation des données probantes sur les résultats de santé pour la prise de décisions éclairées sont encore relativement faibles. Par conséquent, il est absolument nécessaire d'intégrer la recherche en science de l'implémentation pour déterminer ce qui fonctionne et ce qui ne fonctionne pas dans différents contextes afin de renforcer la santé numérique, depuis l'établissement de soins jusqu'au niveau national. Les objectifs de cette thèse sont d'étudier à différents niveaux comment évaluer l'impact de la santé numérique sur les résultats de santé. Tout d'abord, au niveau micro, ce qu'il faut mesurer en identifiant des indicateurs substitutifs significatifs pour les résultats maternels et néonatals. Deuxièmement, au niveau méso, comprendre comment les chercheurs abordent les évaluations numériques de l'impact sur la santé et fournir un aperçu pédagogique des méthodes. Enfin, au niveau macro, comprendre et proposer une approche pour relier des connaissances, des expériences et des preuves disparates.

La Partie 1 dresse une liste de 40 indicateurs indirects potentiels ayant un impact positif sur les résultats maternels et néonataux, notamment la mortalité et la morbidité, grâce à un examen systématique de la littérature, qui a été complété et priorisé par un consensus d'experts Delphi, ce qui a donné une liste de 77 indicateurs indirects potentiels. Ces indicateurs indirects proposent des mesures pour évaluer si les outils de cybersanté ont une incidence directe sur les résultats maternels et néonataux. Certains indicateurs indirects nécessitent une mise en correspondance avec le contexte local, les pratiques et les ressources disponibles, ce qui permet leur utilisation dans divers contextes tout en permettant la collecte systématique de données provenant de différents projets et programmes.

La deuxième partie examine comment les chercheurs abordent l'évaluation des résultats en santé numérique dans différents contextes au moyen d'une revue de la littérature. Un algorithme, basé sur ces résultats, fournit une vue pédagogique des méthodes d'évaluation des résultats pour la santé des interventions en santé...
numérique. L’analyse de la littérature a démontré que les études numériques sur les résultats en matière de santé reposent sur des modèles d’évaluation clinique traditionnels, même si ces interventions sont souvent plus complexes et dépendent du contexte, de la culture et de l’individu. Des méthodologies d’étude supplémentaires, identifiées par recherche documentaire, ont des caractéristiques de conception qui répondent à certaines des lacunes des méthodologies cliniques traditionnelles, lors de leur application aux interventions numériques en santé. Toutes les méthodologies identifiées ont été intégrées dans un algorithme qui fournit une vue d’ensemble de haut niveau et permet à l’utilisateur de naviguer parmi ces méthodologies en fonction des caractéristiques de conception et des priorités de l’enquêteur, et de faciliter l’identification d’une ou plusieurs méthodologies potentiellement appropriées.

La partie 3 propose le modèle d’annotation RAFT, qui permet la description sémantique de tous les éléments des divers résultats et projets de mise en œuvre qui ont été développés, sont utilisés ou font partie du réseau RAFT de télémédecine. Ce modèle a été développé dans un premier temps pour annoter divers intrants, processus et extrants du réseau RAFT afin de faciliter la documentation et le partage des connaissances. Il sert aussi de validation de principe pour la mise en œuvre de l’ « Implémentome ». L’ « Implémentome » sera un système de connaissances interconnecté qui permettra à l’utilisateur de naviguer dans de multiples dimensions grâce aux métadonnées annotées des projets, des personnes et d’autres informations. Il pourra servir de base à l’établissement de consensus, de bonnes pratiques et de lignes directrices. Le modèle d’annotation RAFT peut être développé plus avant pour permettre l’annotation des résultats, des mises en œuvre, des personnes, des initiatives et des projets du domaine de la santé numérique en général.

Cette thèse examine les aspects de l’évaluation des résultats de santé de la santé numérique à différents niveaux et propose des approches pour surmonter certains des défis qui y sont associés. Le premier article dresse une liste d’indicateurs indirects pour mesurer l’impact de la santé numérique sur les résultats de santé néonatale ou maternelle. Le deuxième article donne un aperçu pédagogique des méthodologies d’évaluation et propose un algorithme pour aider à identifier les approches sur la façon de mesurer. Le troisième article souligne la valeur ajoutée de l’apprentissage bidirectionnel entre le Sud et le Nord. Le quatrième article propose un modèle qui permet d’annoter et de relier sémantiquement l’information, les données probantes, les connaissances et l’expérience afin de fournir une base pour une prise de décision éclairée. Ensemble, ces articles constituent une boîte à outils à plusieurs niveaux pour évaluer et agréger les résultats des interventions en santé numérique. L’agrégation et l’intégration de l’expérience, des connaissances et des données probantes à différents niveaux aideront à brosser un tableau plus complet du paysage de la santé numérique. Cela permettra de mieux comprendre les intrants, les processus et les extrants et leurs facteurs clés de succès ou d’échec.
4. Outline

Table of Contents

ACKNOWLEDGEMENT .................................................................................................................................................. 3

1. LIST OF PUBLICATIONS ........................................................................................................................................ 4

1.1. ACCEPTED PUBLICATIONS ............................................................................................................................... 4

1.2. SUBMITTED PUBLICATIONS ............................................................................................................................. 4

2. ABSTRACT (ENGLISH) ............................................................................................................................................ 5

3. ABSTRACT (FRENCH) ............................................................................................................................................. 7

4. OUTLINE ............................................................................................................................................................... 9

5. GENERAL INTRODUCTION .................................................................................................................................... 11

5.1. LOW-RESOURCE SETTINGS HEALTH SECTOR .................................................................................................. 11

5.2. INEQUALITIES IN HEALTH AND ACCESS TO CARE AND SPECIALIZED SERVICES .................................. 11

5.3. MATERNAL, NEWBORN AND CHILD MORTALITY: A PREVENTABLE WORLDWIDE HEALTH CHALLENGE .... 12

5.4. THE POTENTIAL OF DIGITAL HEALTH ........................................................................................................... 13

5.4.1. Interventions for clients .................................................................................................................................. 13

5.4.2. Interventions for healthcare providers ......................................................................................................... 14

5.4.3. Interventions for health system managers .................................................................................................... 15

5.4.4. Interventions for data services .................................................................................................................... 15

5.5. THE NEED TO EVALUATE THE IMPACT OF DIGITAL HEALTH ON OUTCOMES ........................................ 16

5.6. FRAGMENTATION ................................................................................................................................................ 17

5.7. CHALLENGES FOR EVALUATING HEALTH OUTCOMES .................................................................................. 18

5.7.1. Implementation and Evaluation Science ....................................................................................................... 18

5.7.2. Indicators ....................................................................................................................................................... 18

5.7.3. Methodological considerations ................................................................................................................... 18

5.7.3.1. Disparate evidence ................................................................................................................................... 19

5.8. RESEARCH QUESTIONS ..................................................................................................................................... 19

5.8.1. Research question 1 ...................................................................................................................................... 20

5.8.2. Research question 2 ...................................................................................................................................... 20

5.8.3. Research question 3 ...................................................................................................................................... 20

6. ARTICLES ............................................................................................................................................................... 22

6.1. METHODOLOGICAL CONTRIBUTIONS .......................................................................................................... 22

6.1.1. Article 1 – Systematic review to identify proxy indicators to quantify the impact of eHealth tools on maternal and neonatal health outcomes in low-income and middle-income countries including Delphi consensus.................................................................................................................. 22

6.1.2. Article 2 – Navigating through digital health outcome evaluation methodologies: an algorithm based on a scoping review and emerging methodologies. .......................................................................................... 23

6.1.3. Article 3 – Reverse innovation experiences from the RAFT e-learning and telemedicine network. .......................................................... 24

6.1.4. Article 4 – Learning From eHealth Implementations Through “Implementomics”: A Multidimensional Annotation Model Applied to eHealth Projects of the RAFT Network ........................................... 24

6.2. ARTICLE MANUSCRIPTS ................................................................................................................................... 25

6.2.1. Article 1 ......................................................................................................................................................... 25

6.2.2. Article 2 ......................................................................................................................................................... 41

6.2.3. Article 3 ......................................................................................................................................................... 60

6.2.4. Article 4 ......................................................................................................................................................... 65

7. CONCLUSIONS AND PERSPECTIVES .................................................................................................................. 73

7.1. MEASURABLE INDICATORS ............................................................................................................................. 73

7.2. METHODOLOGICAL CONSIDERATIONS ...................................................................................................... 73
7.3. Disparate Experience, Knowledge, and Evidence................................................................. 75
7.4. Implementome...................................................................................................................... 75
7.5. Key Takeaways..................................................................................................................... 77
  7.5.1. The obstacles to measuring health outcomes of digital health are significant........ 77
  7.5.2. Consider innovative evaluation approaches.............................................................. 77
  7.5.3. There is no one-size-fits-all approach in designing digital health evaluations.......... 77
  7.5.4. Experience is the best teacher, even when it’s not our own........................................ 77
  7.5.5. Connect the dots to fight pilotitis................................................................................. 78

8. Bibliography............................................................................................................................. 79
5. General introduction

5.1. Low-resource settings health sector

Health care systems of low-resource settings are challenged by demographic factors, increasing cost of care and suffer from fragmentation, weak information systems, lack of good governance, financial constraints [1] and a deficit of trained human resources for health. In addition, these sparse human resources are unevenly and inequitably distributed [2, 3]. For example, the World Health Organization (WHO) region of the Americas accounts for 10% of the global burden of disease and 37% of the world’s health workforce, whereas the African region has 24% of the disease burden but only 3% of the global health workforce.

These challenges in the health systems have consequences for the population, manifesting in inequalities in health. The importance of eradicating these inequalities has been emphasised in the Alma Ata declaration 40 years ago [4], defining health as a fundamental right and calling for health for all, which is being reinforced by the concept of Universal Health Coverage [5]. The evolution of thinking on Universal Health Coverage has also led to a greater understanding of the functions that health systems should serve. These should be concerned with prevention as well as treatment and should assure: (i) access to essential medicines and health products; (ii) motivated and skilled health workers who are accessible to the people they serve; (iii) integrated, high-quality, patient-centered services at all levels from primary to tertiary care; (iv) a combination of priority programs for health promotion and disease control, including methods for prevention and treatment, which are integrated into health systems; (v) information systems that produce timely and accurate data for decision-making; and (vi) health financing systems that raise sufficient funds for health, provide financial risk protection, and ensure that funds are used equitably and efficiently [6]. However, even high-income countries are struggling to maintain current health services while ensuring that everyone can afford to use them [7, 8]. Globally only half of all countries have the workforce required to deliver quality healthcare services that are critical to achieving Universal Health Coverage [9].

Alarmingly, estimates of health worker density show that the global shortage and unequal distribution of health workers requires urgent attention in order not to undermine the attainment of the Sustainable Development Goals (SDGs) [10]. Especially as this shortage is likely to increase because of rising incidence of non-communicable diseases (NCDs) and a globally ageing population that are estimated to generate a demand for 40 million additional health workers globally, and a shortfall of 18 million health workers by 2030 [9, 11]. Therefore, it is imperative to address the shortage of the healthcare workforce to avoid uneven progress and growing inequalities [10].

5.2. Inequalities in health and access to care and specialized services

Historically, inequalities in health and access to care and specialized services are highest for the poor and rural parts of the population [12-14]. Most difficulties around healthcare delivery in remote or rural areas are common across countries or regions. Often diagnostic capabilities are very limited or mostly absent, being usually only provided in regional referral hospitals or in the capital. Clinical expertise and specialty care is sparse due to the reluctance of clinicians to be based full-time in remote or rural areas, one of the reasons being the lack of access to continuing education [15]. The lack of the provision of clinical services and infrastructure
commonly leads locally to a dis-engagement of the population with the healthcare system endangering higher rates of health problems. Because diagnostic capabilities or specialty care are usually only provided in central referral hospitals, rural populations have to overcome large distances to access the nearest facility. For most of the population living in remote areas it is challenging and expensive to travel to the nearest referral hospital due to financial and logistical constraints. Furthermore, information flows in healthcare facilities are mostly paper-based, resulting in inefficient record-keeping that is prone to human error and damage without the possibility of recovery, and inaccessibility to important information about patients when they are referred. Furthermore, these paper-based electronic systems commonly lead to constraints in reporting, and unavailability of data for decision-making, as the data is not readily available for analytics and manual digitalization frequently increases inaccuracy of the data.

Among other factors these challenges lead to inequalities in health, access to healthcare, and service coverage, which are highest for the poor and rural parts of the population [12-14], and are particularly high for vulnerable groups such as women and children [16]. For antenatal care for example, there are pro-rich coverage patterns in every region and pro-rich and pro-urban inequalities are observed for all indicators, except for breastfeeding [13]. The widest gaps are seen in skilled birth attendance, and inequalities are highest in Africa [13].

5.3. Maternal, Newborn and Child Mortality: A preventable worldwide health challenge

99% of maternal deaths take place in developing countries, which makes this the most inequitable health indicator in the world [17]. Due to complications during pregnancy and childbirth 1000 women die every day worldwide, of which 570 are in Sub-Saharan Africa. Most maternal deaths could be prevented through the presence of trained healthcare professionals at childbirth, and access to emergency obstetric care [18]. In Sub-Saharan Africa only 46% of the births are attended by a skilled health professional, and only 37% of women in rural areas receive the four recommended visits with trained healthcare professionals during their pregnancy [19].

The situation is similar for newborns and children. Worldwide, 6.9 million children under the age of five died in 2011, and from these nearly three millions passed away within a month of their birth [20]. The highest rates of child mortality are still in Sub-Saharan Africa, where 1 in 9 children dies before age five, more than 16 times the average for developed regions (1 in 152), and slightly more than in Southern Asia (1 in 16) [21]. In these developing regions, child mortality is higher in rural areas, and among poorer and less educated families [20]. Additionally, as under-five mortality rates have fallen more sharply elsewhere, the disparity between these two regions and the rest of the world has grown [21]. About two-thirds of these deaths are preventable through access to practical, low-cost interventions, and effective primary care up to five years of age. Stronger health systems are crucial for improving access to care and prevention [20]. While there was a reduction of under-five mortality rates between 1990 and 2010 in continental Africa, most countries in West Africa made insufficient progress in this period [22]. The highest maternal, neonatal and under-five mortality rates are in sub-Saharan Africa and in Southern Asia [23]. This uneven progress, with pro-rich and pro-urban inequalities, leaves especially women and children in rural areas in high need to access quality health care services.
5.4. The potential of digital health

Continuing education of health-care professionals (HCPs) and access to specialized advice are keys to improving the quality, efficiency and accessibility of health systems. In developing countries, these activities are usually limited to capitals, which constitutes an obstacle for HCPs to practice in remote areas, where they are greatly needed [24], as the majority of the population of Sub-Saharan Africa lives in rural areas [25], where low levels of resources and low infrastructure are common. Providing basic infrastructure, like education and health to the rural population remains a huge challenge [15].

The potential of ICT to alleviate these problems by increasing access to health services, and by improving quality and efficiency of care through de-isolation of healthcare professionals in remote locations is acknowledged by the World Health Organization (WHO) [26]. This is “digital health”, which is an umbrella term to collectively describes a variety of concepts such as telemedicine, eHealth and mHealth, but also more advanced computing sciences as e.g. artificial intelligence. eHealth has been defined as the cost-effective and secure use of information and communication technologies (ICTs) for health and health-related fields, while mHealth (or “mobile health”) is a component of eHealth, which provides information and health services through mobile technologies like mobile phones or tablet computers [27, 28].

Digital health interventions are increasingly used, resulting in a multitude of tools with a wide range of objectives and functionalities. The WHO classification of digital health interventions [29] organizes these into four umbrella groupings based on the targeted primary user: (i) Interventions for clients; (ii) Interventions for health care providers; (iii) Interventions for health system or resource managers; or (iv) Interventions for data services.

5.4.1. Interventions for clients

Interventions for clients group interventions that target members of the public who are potential or current users of health services, and includes health promotion activities and caregivers of clients receiving health services. Interventions for clients can be further mapped into the following categories and subcategories:

- Targeted client communication: includes among others the transmission of targeted health information to patients based on a health status or demographic, as e.g. Alive & Thrive: weekly messages (text and voice) to the mobile phones of the microcredit group leaders to share messages on breastfeeding with the group [30].
- Untargeted client communication: includes the transmission of untargeted health information to an undefined population as e.g. WAHA Maternal mHealth Program, a mass communication campaign to send text messages (SMS) to the inhabitants of the Tambacounda district to inform them of the availability and benefits of the maternal and neonatal health services at their local health-care facilities [31].
- Client to client communication: consists of peer groups for patients as e.g. peer-led online support groups for breast cancer patients [32].
- Personal health tracking: includes the self-monitoring of health or diagnostic data as e.g. wearables and fitness trackers that enable tracking activity, exercise, sleep, weight and food [32].
- Citizen based reporting: includes the reporting of public health events as e.g. Text messages (SMS) feeding into a crowdsourced data collection platform for collecting information from the public [33].
• On-demand information services to clients: consists of client look up of health information as e.g. in the Hesperian Health Guide, which is providing health information freely available to millions of users in digital and multi-media formats [34].

• Client financial transactions: includes among others the transmission or management of vouchers to clients for health services as e.g. the eVoucher scheme in Tanzania, which issues electronic vouchers for long-lasting insecticide-treated nets, that can be redeemed at nearby retailers [35].

5.4.2. Interventions for healthcare providers

Interventions for healthcare providers target members of the health workforce who deliver health services in the following categories and subcategories:

• Client identification and registration: to verify the identity of a client and to enroll clients for health services or a clinical care plan as e.g. MoMConnect which enables the registration of pregnancies at governmental health facilities [36].

• Client health records: includes interventions for the routine data collection and management of health indicators such as DHIS2 which can be used for health program monitoring, evaluation, and national data management and analysis [37].

• Healthcare provider decision support: includes prompts and alerts based on the protocol such as electronic clinical decision algorithms for the integrated primary care management children in low-resource settings [38].

• Telemedicine: includes consultations for case management between healthcare providers such as in the RAFT network where healthcare professionals in remote areas can consult with specialists for treatment or second opinion of clinical cases [39].

• Healthcare provider communication: includes peer groups for healthcare providers such as KevinMD.com which provides a platform for health care organizations and providers to interact with one another, as well a closed groups on Facebook for physicians only [40].

• Referral coordination: includes the management of referrals between points of service within the health sector such as in the RAFT network, where digital health enables determining referral necessity at distance and the transmission of patient relevant data and information [39].

• Scheduling and activity planning for healthcare providers: includes scheduling of clients appointments based on a clinical care plan and scheduling of healthcare provider activities.

• Healthcare provider training: includes the assessment of the knowledge acquisition of healthcare providers and the provision of training content and material, such as those provided by the RAFT network, where interactive courses can be broadcasted or followed in resource constraint settings [39].

• Prescription and medication management: includes the transmission or tracking of prescription orders, such as in an outpatient clinic in Egypt where it reduced medication errors [41].

• Laboratory and diagnostics imaging management: includes among others the capture of diagnostic results from digital devices such as the conversion of a smartphone into a device to screen for cervical cancer, which has been piloted in Madagascar [42].
5.4.3. Interventions for health system managers

Interventions for health system or resource managers group interventions that support or facilitate managerial functions related to supply chain management, health financing, human resource management, which have the following subcategories:

- **Human resource management**: includes the electronic registration of health workforce cadres and related identification management, such as the iHRIS software that supports organizations to track, manage, deploy, and map their health workforce and which is interoperable with DHIS2 [43].
- **Supply chain management**: includes the reporting or verification of drug authenticity and quality, such as the use of electronic solutions combating counterfeit drugs [44] like the m-Pedigree, which authenticates drugs that a consumer buys by dialing a code on their mobile and receiving a message back through short message service (SMS) authenticating the product [45].
- **Public health event notification**: consists of the notification of public health events from the point of diagnosis such as GXAlert, which automatically transmits alerts by SMS or e-mail to officials of the Ministry of Health when a new positive MDR case is detected [46].
- **Civil registration and vital statistics**: include the electronic registration of a birth event such as mobile birth registration in Ghana [47].
- **Health financing**: includes the registration and verification of insurance membership as well as the management of the budget and expenditures.
- **Equipment and asset management**: consists of interventions to monitor status and maintenance of health equipment, as well as the tracking of regulation and licensing of medical equipment.
- **Facility management**: consists of interventions to list health facilities and related information as well as the assessment of health facilities.

5.4.4. Interventions for data services

Interventions for data services group interventions that have crosscutting functionality to support a wide range of activities related to data collection, management, use, and exchange, which have the following subcategories:

- **Data collection, management and use**: includes data synthesis and visualization such as DHIS2, which is enabled through various methods as e.g. pivot tables, spot trends, as well as charting and visualizing geographical data aspects [37].
- **Data coding**: includes interventions to classify diseases codes and cause of mortality such as the coding of algorithms for defining co-morbidities in ICD-10 administrative data [48].
- **Location mapping**: includes interventions the location mapping of health events such as the application of satellite imagery analysis to guide the immunization response to confirmation of previously-undetected, ongoing endemic wild poliovirus transmission in Borno State, Nigeria [49].
- **Data exchange and interoperability**: deal with data exchange across systems.

Digital health applications and interventions are recognized for their potential to address some of the challenges faced by both emerging as well as industrialized countries in providing accessible, cost-effective, high-quality
health care services [24, 39, 50-58], and has brought hope to developing countries and their most remote areas. Digital health technologies such as electronic medical records, decision support systems, diagnostic imaging and bio-signals, mobile computing, artificial intelligence, and robot assisted medical procedures have changed hospitals around the world, as geographic distance has significantly lost impact on service provision. Significant parts of populations live in areas with a lack of health and transport infrastructure, making a trip to the closest hospital or specialized care a time-consuming and expensive journey. In this context, Information and communication technologies (ICTs) offer significant potential and opportunities for global advancement of in health and healthcare by providing innovative approaches for alleviating these problems, as these can help to overcome geographical barriers, increase access to healthcare services, train healthcare professionals through distance education and provide the possibility of collaborative healthcare in remote locations. Bringing technologies like ultrasound imaging, continuous education and second opinion services to remote/rural communities can help with providing better diagnostics and interventions for care problems and reduce fatalities and complications. Digital health applications for clinical, educational, and administrative purposes is a key enabler for supporting health systems and delivery of healthcare. The digital divide is still prevalent between developing and industrialized countries, but connectivity is extending geographically, also into rural areas [59, 60], facilitating implementations of digital health interventions. Their application ranges from civil registration, data interchange, telemedicine, and electronic medical records to public health and diseases surveillance systems, among many other applications. Mobile-phone-based health education and consultations, personalized health tracking devices and mobile diagnostic technologies enable real-time provision of information to improve individual as well as public health. Technologies like smartphones, e-payment systems, or telemedicine improve access to quality care as well as a faster deployment of emergency services. Increased connectivity and innovation in ICTs allow health facilities to enter data through web applications directly into central servers omitting the necessity to install software or database management at the local level [28, 61].

5.5. The need to evaluate the impact of digital health on outcomes

Exploratory evidence points to important socioeconomic benefits to patients, families, health practitioners and health systems, including improved patient-provider communication and educational opportunities [62], however, the majority of the available evidence focuses on care process improvements rather than improvements in health outcomes [63-71]. Even though scientific literature is providing a growing body of literature evaluating the impact of digital health implementations on cost-effectiveness, quality, and safety of health care, there is still a notable gap between the postulated and evidence-based benefits [72-74]. Given the current status of adoption of digital health globally there are relatively few studies measuring the impact on health outcomes [75] which might lead to an ineffective use of resources or investments in potentially inefficient or harmful digital health interventions [76, 77]. Evidence is needed to demonstrate costs and benefits of digital health implementations, and to equip decision makers to choose the most effective, economical and sustainable digital health approaches [78].

There is clearly an urgent need for solid evidence of the impact of digital health interventions [79], especially considering the trend towards an even greater adoption of digital health. Activities should be subject to rigorous,
multidisciplinary, and independent assessment, and impact evaluations should go beyond measuring feasibility, implementation, activity, and attitudes or perceived impact but should also quantify changes caused by digital health interventions in health outcomes [63-71, 78], also to ensure the intervention or implementation does not have adverse or potential harmful effects [80].

Increasing service coverage does not necessarily mean improving the health status. Wagstaff and colleagues investigated differential progress on Health Development Millennium Goals between the poor and the better off within countries [81]. In this study they illustrated that on average overall health service coverage increased and coverage inequalities reduced. However, even though overall health service coverage has increased between 1990 and 2011, health status inequalities have increased in approximately 40% of the countries and in about a quarter of the countries health status declined among the poorest 40% of the population [12, 81]. Therefore, it is essential to not only devote more effort to evaluation, but also to ensure that the adopted methodology has the ability to extricate the commonly complex web of factors potentially influencing the results.

5.6. Fragmentation

The long-term sustainability of digital health services is difficult to achieve, many of these fail to reach critical mass, which is often referred to as pilotitis [82]. The term ‘pilotitis’, referring to the problems due to a plethora of small-scale and short-term pilot projects, is used to express the frustration of professionals in the digital health sector that the vast majority of implementations are a disconnected series of efforts, mostly limited in scale with narrowly focused interventions targeting relatively small populations and conducted in stand-alone, vertical project mode [28, 83]. This leads to disparate evidence from pilot-scale implementations that do not allow decision makers to base investment-decisions on. This is particularly important, as historically, most digital health services or applications in low resource settings have been created using, donations, grants and other funding mechanisms that are a) often not expecting to see returns on their investments, and b) not constant sources of funding, which frequently leads to services that flourish during the initial period of investment but fail to scale-up or establish a sustainable funding mechanism.

Fragmentation has also been associated to donor misalignment. Donor misalignment occurs when donor priorities and objectives do not reflect the recipient ones. These power asymmetries can lead to misaligned investment and objectives, even when stakeholders are theoretically on the same side, and can have a profound impact on the health services available to the population in a country [84]. In October 2018, a group of major global health funders endorsed The Principles of Donor Alignment for Digital Health [85]. The effort was an important step towards reducing historically fragmented approaches to investments in digital health.

Fragmentation, amplified through donor misalignment and pilotitis, results in a landscape where even though low-cost digital health applications have proven to be feasible, clinically useful, sustainable, and scalable in various settings and underserved communities, these are not going to scale, by reason of various barriers [86], and possibly due to the lack of rigorous and general supportive empirical verification of their impact [72-74]. Policy makers, funders and implementers of digital health have a central responsibility and the possibility to ensure that digital health implementations are subjected to, and informed by, rigorous evaluations for example by making them a requirement for continued funding [78, 79].
5.7. Challenges for evaluating Health outcomes

5.7.1. Implementation and Evaluation Science
Implementation science is a relatively new field of study and can be defined as the scientific study of methods promoting the systematic uptake of evidence-based practices and research findings into routine practice to improve the effectiveness and quality of health services [87, 88], and aims at bridging the gap between theory and practice by identifying and addressing the barriers hindering or delaying the uptake of proven health interventions and evidence based practices [89, 90]. Studies evaluating the implementation and integration of digital health interventions, applications, or systems has steadily grown over the last decade, and even though these studies describe important achievements, frequently they explain complex processes with ambiguous outcomes [91], making it difficult to understand the implications of these results, even though facilitating and hindering factors have been systematically reviewed by the research community [92]. One difficulty of understanding the implications for practice and to inform decision making with the digital health implementation literature is that it is fragmented across multiple subspecialty areas [91, 92]. This fragmented literature on the implementation of various e-digital health technologies, may make it difficult for clinicians, managers or policymakers to locate, apply, and integrate an appropriate body of evidence for their specific circumstances or for informing their decision-making [92].

5.7.2. Indicators
An essential component of impact evaluations is the selection or identification of appropriate indicators, which can be generic or very specific [93]. To measure the impact of digital health interventions on health outcomes in a meaningful way requires the consistent use of reliable indicators [93]. Nevertheless, one challenge is the limited identification of measurable and reliable indicators. Scientific literature on digital health impact evaluation is relatively limited, and within the available literature evaluations are mostly based on activity or process indicators, but not on outcomes. The limited availability of outcome indicators has multiple causes. One challenge lies in the long timespan between intervention and potential outcome (e.g. the reduction in maternal/neonatal mortality). In view of this long timespan, outcomes might be affected by a variety of confounding factors, making it difficult to attribute the results to the digital health intervention. Furthermore, these interventions are often embedded in a complex web of factors [94] and depend on the setting, culture, or individual. The relevance of these indicators might be context-dependent which considerably restricts their extrapolation.

5.7.3. Methodological considerations
Besides the need for measurable indicators, another factor influencing the lack of empirical robust health outcome evidence of digital health implementations are methodological considerations. The gold standard among available evaluation strategies is the randomized controlled trial (RCT) [95]. However, while a RCT is a good methodology to test efficacy of specific digital health interventions, e.g. of applications for chronic
conditions like diabetes [96], digital health interventions are often embedded in a complicated system and may depend strongly on the setting, culture, or individual, which is why a traditional clinical evaluation like an RCT might not be the best approach to measure outcomes of these interventions [97]. Conducting RCTs is complex and costly, therefore often not feasible in low-resource settings [98]. Furthermore these have a long lag time [99], and at the pace digital health is evolving the intervention or technology might be outdated by the time the trial is completed. Furthermore, traditional RCTs aim to maximize the internal validity by assuring rigorous control of all variables [100], hindering the generalizability of the results outside of the study setting.

Digital health interventions may also present characteristics, e.g., being flexible or evolving, which are not typical for traditional clinical interventions and require methodologies that acknowledge these characteristics. Digital health interventions may also provide opportunities for evaluation processes that are not supported by traditional clinical evaluations like the automatic collection of process-related data. Understanding the advantages and shortcomings of different methodologies when applied to digital health interventions, and understanding potential new methodologies will contribute to identifying appropriate methodologies for evaluation outcomes.

5.7.3.1. Disparate evidence

Digital health implementations mostly remain in the pilot stage, scaling these implementations to national or international services remains rather the exception. This deployment of digital health technologies produces substantial amounts of information and knowledge, resulting in a paradox: while there are understudied areas in digital health, at the same time there is an exponential increase in published information, guidelines, data, methods, articles, projects and experiences [28, 82]. Additionally, terminology is evolving, and definitions are diverging, making it time-consuming and sometimes impossible to find the appropriate information. Keeping track, connecting, and making sense of data and information, and documenting and using knowledge to develop consensus, best practices and guidelines, or to accumulate evidence of disparate implementations as base for decision-making is challenging, particularly due to the disparate information and varying and evolving terminology. Documenting implementation knowledge and the effectiveness of these tools will help avoid the duplication of efforts and assure that investments are targeted to the most contributive technological innovations.

5.8. Research questions

Digital health has the potential to address many challenges of health systems and is a promising tool to reach vulnerable populations, such as mothers and children, in remote areas that are historically highly disadvantaged in access to quality care and specialized services, as it is one of the major reasons why geographic distance has significantly lost impact on service provision. However, providing access to care does not necessarily mean that health outcomes are improved, therefore it is important to understand if these tools actually improve health outcomes, and have the intended impact. This is certainly important in every context, but particularly in low resource settings and in vulnerable populations, where the duplication of efforts or ineffective interventions imply a stronger impact on population health as the resources could have been invested in other life-saving
interventions. These considerations are reflected in the three research questions that are being investigated in this thesis.

5.8.1. Research question 1

"Which measurements can serve as proxy indicators to evaluate the impact of digital health interventions on maternal and neonatal health outcomes in low and middle-income countries?"

The first research question aims at identifying proxy-indicators that enable to measure actual impact these interventions may have on health outcomes, which can provide a basis to inform decision-making processes. The availability of reliable proxy indicators will facilitate consistent outcome measurements, which would provide signals (positive or negative) about the changes in health status or behavior changes that are assumed to improve outcomes brought about by the digital health interventions [93]. Consistent measurements of reliable proxy indicators or indicators will illustrate how health status and behavior changed over time following implementation of the digital health solution and would be stimulants for change and future investments, identifying digital health solutions that could be implemented when and where matching problems are observed [93]. Availability of indicators (direct and proxy) may facilitate consistent outcome measurements and comparability of studies.

5.8.2. Research question 2

"How do researchers approach impact evaluations of specific digital health interventions?"

The second research question investigates methodologies that are currently used in order to understand how researchers evaluate digital health outcomes in different settings. Exploring how researchers approach measuring the impact of digital health interventions on outcomes in different settings, and understanding the rationale for choosing one methodology provides the foundation for identifying advantages and shortcomings of these methodologies, as well as identifying new emerging methodologies that address some of these shortcomings or specificities digital interventions. Providing a pedagogical overview and inventory of currently used and emerging methods for evaluating health outcomes of digital health interventions, and summarizing their main advantages and shortcomings in an algorithm will help implementers choosing appropriate methods for evaluating the impact of their digital intervention.

5.8.3. Research question 3

"How can implementations or evidence be semantically annotated to describe various outputs of digital health implementations?"

The third research question explores how potentially disparate evidence and implementations can be connected in a meaningful way to facilitate keeping track, connecting, making sense of data and information, and to documenting, and using knowledge to develop consensus, best practices, guidelines, and to support decision-making processes.

Despite the progress of the last decade in understanding the role of digital health to strengthen the overall health system performance, the use of information for evidence-informed decision making is still relatively
weak. Hence, there is a strong need to embed implementation science research. Semantically annotated outputs of digital health will facilitate the documentation and sharing of knowledge, the pooling and connecting of experience and evidence and can help to identify what works and what doesn’t in different contexts to strengthen digital health from lower facility level to the national level decision makers. Documenting implementation knowledge and effectiveness of digital health will help to avoid duplicating efforts and assure that investments target meaningful technological innovations.
6. Articles

6.1. Methodological Contributions

6.1.1. Article 1 – Systematic review to identify proxy indicators to quantify the impact of eHealth tools on maternal and neonatal health outcomes in low-income and middle-income countries including Delphi consensus

Systematic review
Caroline Perrin (CP) developed the protocol for the systematic review with input from the two co-authors, Lothaire Hounga (LH), and Antoine Geissbuhler (AG). This included defining different search strategies for the databases (Pubmed, EMBASE, and the Cochrane Library), which was adapted for the databases using a combination of text words and controlled vocabulary terms, like MeSH (medical subject headings) related to the interventions and potential outcome measures. The protocol was registered and published on the PROSPERO website. CP extracted the results from the different databases into an excel file, and conducted an initial screening aiming at excluding articles with titles that were obviously irrelevant. Afterwards, CP and LH independently rated the remaining titles and abstracts in relation to their relevance to the study objectives. AG resolved discrepancies in the rating. CP extracted from studies that were assessed as eligible study design, setting, population characteristics, description of the intervention, as well as outcomes measured, and effects of the studies. This was thoroughly checked by LH. CP assessed the risk of bias for all included studies and reviews, which was thoroughly checked by LH. Disagreements were resolved by discussion and, when required, by arbitration involving AG. Randomized trials were assessed with the Cochrane risk of bias, non-randomized studies with the Cochrane ROBINS-I (Risk Of Bias In Non-randomized Studies - of Interventions), and systematic reviews with the ROBIS (Tool to assess risk of bias in systematic reviews) tool [101-103]. CP initially summarized by outcome (proxy-indicators) studies and reviews, which met the inclusion criteria and included a quality assessment in a tabular form.

Delphi consensus
CP developed the initial draft for a three-step modified Delphi method. The purpose of the Delphi consensus was to add potential supplementary proxy-indicators and to establish consensus on the potential of the interventions (proxy-indicators) to reduce morbidity and mortality, and if they should be considered an 'essential' intervention, as well as the appropriate level of care. This was reviewed by AG, and updated by CP according to his feedback. CP developed the forms to collect the information from the experts and the documents explaining the process and providing background information.

CP reached out to a list of thirteen international experts, from different regions trained in obstetrics and neonatal care. In the first round the experts were given the possibility to add proxy-indicators to the provisional list. CP integrated the feedback and prepared the form for the second round. In round 2 the completed list of
indicators was transmitted to the experts, which were asked to assess each identified proxy-indicator according to 1) the intervention’s potential in reducing maternal or neonatal morbidity and mortality, 2) whether these should be appraised as an ‘essential’ intervention, and 3) the appropriate level of care (primary, referral or both). CP summarized these rankings using the median and the interquartile range, and included this in a repeat version of the questionnaire for Round 3. In Round 3, the experts re-ranked their agreement to each statement and were given the opportunity to change their score in view of the group’s response. CP summarized the re-rankings and assessed for degree of consensus using interquartile ranges for continuous numerical scales, which were accepted when the interquartile range was 2 or less.

**Writing of the article**

CP drafted the article. AG and critically revised the article. CP updated the article according to the feedback and suggestions of AG and LH and prepared it for submission. After having received the comments from the peer-reviewers CP updated the article accordingly, which was re-reviewed and approved by AG and LH.

6.1.2. Article 2 – Navigating through digital health outcome evaluation methodologies: an algorithm based on a scoping review and emerging methodologies.

**Scoping Review**

Caroline Perrin (CP) developed the protocol for the scoping review with input from the two co-authors, Mirana Randriambelonoro (MR), and Antoine Geissbuhler (AG). This included defining different search strategies for the databases (Pubmed, EMBASE, and the Cochrane Central Register of Controlled Trials (CENTRAL)). For each database the search strategy was adapted. CP extracted the search results and conducted an initial screening aiming at excluding articles with obviously irrelevant titles. Subsequently, CP and MR independently rated titles and abstracts based on relevance to the study objectives. Level of income, system categories, disease addressed by the intervention, evaluation method, and the number of study participants, were abstracted by CP into a standardized spreadsheet and were thoroughly checked by MR. Disagreements were resolved by discussion and, if necessary, by AG. CP analyzed the results and mapped the digital health outcome evaluation methodologies in a Sankey diagram. CP conducted the gap analysis by reviewing the literature to identify additional emerging approaches, that, in the context of digital health address some of the shortcoming of the traditional methodologies, or provide more flexibility, adaptiveness or responsiveness.

**Algorithm**

CP analyzed the various advantages, limitations, and design features of the methodologies, that were identified in the scoping review and the gap analysis. CP further analyzed the studies that were identified in the scoping review to understand why the authors of the included studies chose a methodology and what the
limitations and advantages were. Based on these analysis CP developed a first draft of the algorithm that was reviewed and elaborated by MR and AG.

Writing of the article
CP wrote a first draft of the article. MR and AG made a critical revision of the article. CP updated the article according to the feedback and suggestions of MR and AG and prepared it for submission.

6.1.3. Article 3 – Reverse innovation experiences from the RAFT e-learning and telemedicine network.

Writing of the article
Caroline Perrin (CP), Georges Bediang (GB), Cheick-Oumar Bagayoko (COB), and Antoine Geissbuhler (AG) conceptualized the article outline in discussion. CP reviewed the literature and prepared a first draft of the manuscript that was critically reviewed and completed by GB, COB, and AG. CP integrated the feedback and prepared it for submission. After having received the comments from the peer-reviewers CP updated the article accordingly, which was re-reviewed and approved by GB, COB, and AG

6.1.4. Article 4 – Learning From eHealth Implementations Through “Implementomics”: A Multidimensional Annotation Model Applied to eHealth Projects of the RAFT Network.

Development of the annotation model
Caroline Perrin (CP) reviewed the literature on annotation, annotation models and annotation models for health. CP drafted a first version of the annotation model as a basis for discussions. The model has been developed based on analyzing different case studies of the RAFT network and initial conceptualizing discussions between the co-authors. This first draft was discussed and elaborated between CP, Georges Bediang (GB), Mirana Randriambelono (MR), and Antoine Geissbuhler (AG). After discussion CP integrated the outcomes of the discussion and applied the model to example annotations.

Writing of the article
CP prepared a first draft of the manuscript that was critically reviewed and completed by GB, MR, and AG. CP integrated the feedback and suggestions into the manuscript and prepared it for submission. After having received the comments from the peer-reviewers CP updated the article accordingly, which was re-reviewed and approved by GB, MR, and AG
6.2. Article manuscripts

6.2.1. Article 1

Systematic review to identify proxy indicators to quantify the impact of eHealth tools on maternal and neonatal health outcomes in low-income and middle-income countries including Delphi consensus

Caroline Perrin, Lothaire HOUNGA, Antoine Geissbuhler

**ABSTRACT**

Objective To identify interventions that could serve as reliable proxy indicators to measure eHealth impact on maternal and neonatal outcomes.

Design Systematic review and Delphi study.

Methods We searched PubMed, Embase and Cochrane from January 1990 to May 2016 for studies and reviews that evaluated interventions aimed at improving maternal/neonatal health and reducing mortality. Interventions that are not low-income and middle-income context appropriate and that cannot currently be diagnosed, managed or impacted by eHealth (e.g., telemedicine distance diagnostic or e-learning) were excluded. We used the Cochrane risk of bias, Risk Of Bias In Non-randomised Studies - of Interventions and ROBINS tool to assess the risk of bias. A three-step modified Delphi method was added to identify additional proxy indicators and prioritise the results, involving a panel of 13 experts from different regions, representing obstetricians and neonatologists.

Results We included 44 studies and reviews, identifying 40 potential proxy indicators with a positive impact on maternal/neonatal outcomes. The Delphi experts completed and prioritised these, resulting in a list of 77 potential proxy indicators.

Conclusions The proxy indicators propose relevant outcome measures to evaluate if eHealth tools directly affect maternal/neonatal outcomes. Some proxy indicators require mapping to the local context, practices and available resources. The local mapping facilitates the utilisation of the proxy indicators in various contexts while allowing the systematic collection of data from different projects and programmes. Based on the mapping, the same proxy indicator can be used for different contexts, allowing it to measure what is locally and temporally relevant, making the proxy indicator sustainable.

**Strengths and limitations of this study**

- Limitation: some potential proxy indicators may not have been identified in the systematic review for two possible reasons: (1) due to, for example, a very low GRADE quality, as for some interventions based on ethical reasons, it is not possible to conduct high-quality randomised studies, or (2) no studies have investigated these as they are standard of care. They may also have been overlooked as unforeseen, for example, disruptive uses of eHealth may emerge and offer unexpected ways to improve practices.

- Strengths: to address the limitation of potentially overlooked proxy indicators, the results were assessed and completed in a Delphi consensus process with a group of international experts.

- Strength: a review of this kind, aiming at identifying proxy indicators that could be used to measure the impact of eHealth interventions on maternal and neonatal health outcomes, particularly in low-income and middle-income countries has not yet been conducted.

**INTRODUCTION**

Since 1990, maternal and child mortality have approximately halved; however, most of the remaining deaths are preventable.1 Child mortality decreased disproportionately for older children, and neonatal deaths account for 45% of under-5 mortality.2 Uneven progress between countries and within countries, with pro-rich and pro-poor inequalities, leaves women and children in rural areas with insufficient access to quality health care services.3 Information and communication technologies (ICTs) can provide innovative approaches for alleviating these inequalities, particularly in rural and isolated settings. They do so by overcoming geographical barriers, increasing access to healthcare services, providing continuing education and enabling collaborative healthcare in remote locations.4-5 The WHO
defines electronic health (eHealth) as the cost-effective and secure use of ICTs for health and health-related fields. The potential of eHealth on positive therapeutic and clinical outcomes has been repeatedly postulated, but strong evidence is scarce. Although scientific literature offers an increasing number of publications studying the impact of eHealth tools on the quality, safety and cost-effectiveness of health care, there is still a significant gap between the postulated and empirically demonstrated benefits, including therapeutic and clinical outcomes. It is essential to devote more effort to evaluation and to ensure that the methodology adopted is multidisciplinary and thus capable of disentangling the often complex web of factors that may influence the results. It is equally important that existing activities are subject to rigorous, multidisciplinary and independent assessment. Even though low-cost telemedicine applications have proven to be feasible, clinically useful, sustainable and scalable, they are not being adopted on a significant scale due to a variety of barriers, including the absence of robust and general supportive scientific evidence of their impact.

The need for evaluating eHealth impact on patient outcomes has been strongly emphasised. The main barrier remains in the limited identification of measurable and reliable indicators. The relevance of these indicators may be context dependent and their extrapolation considerably restricted. Availability of outcome indicators (direct and proxy) will facilitate consistent outcome measurements and comparability of studies.

Health outcomes research established as a mean to evaluate the effectiveness of healthcare interventions and an approach to inform resource allocation. Obstacles for the outcomes evaluation of eHealth tools include the absence of methodologies and indicators. The identification of indicators is complex as the time-span between intervention and potential outcome (reduction in maternal/neonatal mortality) is long. Due to this duration, the outcome might be influenced by various confounding factors, and it is difficult to attribute the outcome to the eHealth intervention. The use of proxy indicators helps addressing this issue by measuring changes closer to the intervention.

The objective of this review is to identify proxy indicators that can be used in future studies aiming at measuring the impact of eHealth interventions on maternal/neonatal health outcomes in low-income and middle-income countries (LMICs). The review question is: which interventions that can be impacted by eHealth applications have results that can be clearly linked to maternal and neonatal health outcomes in LMIC countries and could therefore serve as reliable proxy indicators?

METHODS

The review was conducted and reported in line with the standards of the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement. The review protocol is registered in PROSPERO; the detailed description can be accessed on the platform. In short, the review identified interventions, which have an alleged impact on maternal/neonatal health, and are suitable for delivery in LMICs to serve as proxy indicators. In this article, previous reviews are included according to the recommendations for integrating existing systematic reviews into new reviews by Robinson et al.

Searching

To identify studies and reviews that evaluated the effect of interventions on maternal and neonatal health, a comprehensive search of PubMed, Embase and the Cochrane Library was carried out using a combination of text words and controlled vocabulary terms related to the interventions and possible outcome measures. The search strategy was adapted for each database. Studies with an abstract published in English from 1990 to May 2016 were considered for inclusion. The third phase consisted of searching databases of multilateral organisations and Google Scholar.

Inclusion/exclusion criteria

Randomised controlled trials, quasi-experimental studies, observational studies, systematic reviews and intergovernmental and non-governmental agency reports were considered for this review. Population: pregnant women at any gestation age, postpartum women up to 6 weeks after giving birth and neonates (up to 28 days after birth).

Intervention

We included any intervention at health system level aiming at improving maternal/neonatal health and reducing maternal/neonatal mortality.

Type of outcome measures: neonatal outcomes (eg, neonatal mortality, stillbirth, low birth weight and preterm birth) and maternal outcomes (eg, maternal mortality, pre-eclampsia and gestational hypertension).

Studies were excluded if they were not LMIC context appropriate or if the interventions cannot currently be diagnosed, managed or impacted by eHealth interventions, such as telemedicine distance diagnostics or e-learning, as well as qualitative studies and opinion pieces.

Study selection

One author conducted an initial screening to exclude articles whose titles were obviously irrelevant. Subsequently, two reviewers independently rated titles and abstracts based on relevance to the study objectives. The third reviewer resolved discrepancies in the rating. All studies that were rated potentially relevant or definitely relevant underwent full-text review. For each included study, the authors verified that these were not comprised in the included systematic reviews and if so they were excluded.

Figure 1 summarises the study selection.
Figure 1 Flow chart of study selection for inclusion in the systematic review. LMIC, low-income and middle-income country.

Data abstraction, quality assessment and data synthesis and analysis

Study design, setting, study population characteristics, description of the intervention, outcomes measured and effects of studies, which were assessed as eligible, were abstracted by one author into a standardised spreadsheet and were thoroughly checked by the second reviewer. Disagreements were resolved by discussion and, if necessary, by arbitration involving the third reviewer. The risk of bias was assessed for all included studies and reviews. Randomised trials were assessed with the Cochrane risk of bias, non-randomised studies with the Cochrane Risk Of Bias In Non-randomised Studies – of Interventions and systematic review with the ROBIS (tool to assess risk of bias in systematic reviews) tool.26-37 The level of evidence of studies and reviews that met the inclusion criteria were summarised by outcome (proxy indicators) including a quality assessment in a tabular form. For each proxy indicator, the summary of findings (SOFs) table includes the number of studies, a summary of the intervention effect and a measure of the quality of evidence for each outcome according to GRADE.38-49 Existing GRADE assessments of systematic reviews have been included after verification and are marked with an asterisk (*) in the SOF table.

Delphi consensus

A three-step modified Delphi method was used to add additional proxy indicators and to establish consensus on the interventions’ (proxy indicators) potential to reduce morbidity and mortality, if they should be considered an ‘essential’ intervention, and the appropriate level of care. Thirteen international experts, with backgrounds in obstetrics and neonatal care, from different regions were approached. All of them agreed to participate and all completed the three rounds.

In round 1, the experts added potential proxy indicators to the provisional list (table 1). Some proxy indicators may have been missed in the systematic review due to, for example, very low GRADE quality, as some interventions could not be conducted as randomised studies for ethical reasons.

In round 2, the completed list of indicators was circulated to the experts and they were asked to assess each, as proxy indicator identified intervention according to (1) their potential to reduce maternal and neonatal morbidity and mortality, (2) whether they should be considered an ‘essential’ intervention and (3) the appropriate level of care (primary, referral or both). An essential intervention was defined as an essential medical
<table>
<thead>
<tr>
<th>Outcome group</th>
<th>Outcome</th>
<th>Effect</th>
<th>Studies</th>
<th>Quality of the evidence (GRADE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preconception</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birth spacing: interpregnancy interval (IPI) between 6 months and under 60 months</td>
<td>Preterm birth with short IPI (&lt;6 months)</td>
<td>OR 1.40, 95% CI 1.24 to 1.58</td>
<td>8</td>
<td>High*</td>
</tr>
<tr>
<td>Neonatal outcome</td>
<td>Low birth weight with short IPI (&lt;6 months)</td>
<td>OR 1.61, 95% CI 1.39 to 1.86</td>
<td>4</td>
<td>High*</td>
</tr>
<tr>
<td>Neonatal outcome</td>
<td>Birth outcome: preterm birth with long IPI (&gt;60 months)</td>
<td>OR 1.26, 95% CI 1.17 to 1.24</td>
<td>7</td>
<td>High*</td>
</tr>
<tr>
<td>Neonatal outcome</td>
<td>Birth outcome: low birth weight with long IPI (&gt;60 months)</td>
<td>OR 1.43, 95% CI 1.27 to 1.62</td>
<td>4</td>
<td>High*</td>
</tr>
<tr>
<td>Folic acid supplementation and fortification</td>
<td>Primary prevention of neural tube defect</td>
<td>RR 0.38, 95% CI 0.29 to 0.51</td>
<td>4</td>
<td>Moderate*</td>
</tr>
<tr>
<td>Pregnancy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple micronutrient supplementation (with iron and folic acid)</td>
<td>Low birth weight</td>
<td>RR 0.88, 95% CI 0.85 to 0.90</td>
<td>15</td>
<td>High*</td>
</tr>
<tr>
<td>Neonatal outcome</td>
<td>Stilbirth</td>
<td>RR 0.92, 95% CI 0.86 to 0.99</td>
<td>15</td>
<td>High*</td>
</tr>
<tr>
<td>Administration/advice of folic acid to women with history of baby of neural tube defect (NTD)</td>
<td>Secondary neural tube defect reduction</td>
<td>RR 0.30, 95% CI 0.14 to 0.65</td>
<td>3</td>
<td>High</td>
</tr>
<tr>
<td>Neonatal outcome</td>
<td>Preventing neonatal tetanus against neonatal death</td>
<td>RR 0.09, 95% CI 0.02 to 0.20</td>
<td>2</td>
<td>Moderate*</td>
</tr>
<tr>
<td>Diet supplementation (high-energy biscuits) for chronically undernourished women</td>
<td>Mortality within 7 days</td>
<td>OR 0.54, 95% CI 0.35 to 0.85</td>
<td>1</td>
<td>Low</td>
</tr>
<tr>
<td>Neonatal outcome</td>
<td>Stilbirth</td>
<td>OR 0.47, 95% CI 0.23 to 0.99</td>
<td>1</td>
<td>Low</td>
</tr>
<tr>
<td>Neonatal outcome</td>
<td>Mortality within 7 days</td>
<td>OR 0.54, 95% CI 0.35 to 0.85</td>
<td>1</td>
<td>Low</td>
</tr>
<tr>
<td>Tetanus toxoid immunization (at least two vaccinations)</td>
<td>Tetanus-specific neonatal mortality</td>
<td>RR 0.06, 95% CI 0.02 to 0.20</td>
<td>2</td>
<td>Moderate*</td>
</tr>
<tr>
<td>Syphilis screening with treatment</td>
<td>Preventing neonatal tetanus against neonatal death</td>
<td>RR 0.02, 95% CI 0.00 to 0.30</td>
<td>1</td>
<td>Moderate*</td>
</tr>
<tr>
<td>Routine drug administration to prevent malaria and its consequences in pregnant women in areas of moderate to high malaria transmission</td>
<td>Maternal outcome</td>
<td>Severe anaemia (during the third trimester)</td>
<td>RR 0.60, 95% CI 0.47 to 0.75</td>
<td>5</td>
</tr>
<tr>
<td>Maternal outcome</td>
<td>Antenatal parasitesemia</td>
<td>RR 0.39, 95% CI 0.26 to 0.56</td>
<td>8</td>
<td>High*</td>
</tr>
<tr>
<td>Intermittent preventive treatment of malaria in pregnancy</td>
<td>Maternal outcome</td>
<td>Maternal death</td>
<td>RR 0.79, 95% CI 0.29 to 2.20</td>
<td>2</td>
</tr>
<tr>
<td>Maternal outcome</td>
<td>Neonatal mortality</td>
<td>RR 0.69, 95% CI 0.49 to 0.98</td>
<td>6</td>
<td>High*</td>
</tr>
<tr>
<td>Maternal outcome</td>
<td>Low birth weight</td>
<td>RR 0.71, 95% CI 0.57 to 0.89</td>
<td>9</td>
<td>Moderate*</td>
</tr>
<tr>
<td>Smoking cessation during pregnancy (psychosocial interventions)</td>
<td>Preterm birth</td>
<td>RR 0.82, 95% CI 0.70 to 0.96</td>
<td>14</td>
<td>Moderate*</td>
</tr>
<tr>
<td>Maternal outcome</td>
<td>Low birth weight</td>
<td>RR 0.82, 95% CI 0.71 to 0.94</td>
<td>14</td>
<td>Moderate*</td>
</tr>
<tr>
<td>Prevention and management of HIV and prevention of mother-to-child transmission in pregnancy</td>
<td>Maternal outcome</td>
<td>HIV-testing uptake</td>
<td>RR 2.95, 95% CI 1.69 to 5.16</td>
<td>13</td>
</tr>
</tbody>
</table>

Continued
<table>
<thead>
<tr>
<th>Table 1</th>
<th>Continued</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pregnancy</strong></td>
<td></td>
</tr>
<tr>
<td>Antiretroviral therapy, for example, Zidovudine given to mothers from 36 weeks' gestation during labour</td>
<td></td>
</tr>
<tr>
<td>Neonatal outcome</td>
<td>Reduced HIV infection at 4–5 weeks</td>
</tr>
<tr>
<td>Adherence to antiretroviral medication; mobile phone messages</td>
<td></td>
</tr>
<tr>
<td>Maternal outcome</td>
<td>Viral load suppression at 52 weeks</td>
</tr>
<tr>
<td>Maternal outcome</td>
<td>ART adherence at 48–52 weeks</td>
</tr>
<tr>
<td>Management of prelabour rupture of membranes and preterm labour</td>
<td></td>
</tr>
<tr>
<td>Calcium channel blockers for women in preterm labour</td>
<td></td>
</tr>
<tr>
<td>Neonatal outcome</td>
<td>Reduction in birth less than 48 hours after trial entry</td>
</tr>
<tr>
<td>Antenatal corticosteroids for accelerating foetal lung maturation for women at risk of preterm birth</td>
<td></td>
</tr>
<tr>
<td>Neonatal outcome</td>
<td>Neonatal mortality</td>
</tr>
<tr>
<td>External cephalic version for breech presentation at term (spinning babies)</td>
<td></td>
</tr>
<tr>
<td>Neonatal outcome</td>
<td>Perinatal death</td>
</tr>
<tr>
<td>Prevention and management of hypertension in pregnancy</td>
<td></td>
</tr>
<tr>
<td>Ultrasound for detection of pre-eclampsia</td>
<td></td>
</tr>
<tr>
<td>Maternal outcome</td>
<td>Abnormal Doppler US developing pre-eclampsia</td>
</tr>
<tr>
<td>Maternal outcome</td>
<td>Increased pulsatility index with notching (low risk patients)</td>
</tr>
<tr>
<td>Maternal outcome</td>
<td>Increased pulsatility index with notching (high risk patients)</td>
</tr>
<tr>
<td>Maternal calcium supplementation</td>
<td></td>
</tr>
<tr>
<td>Maternal outcome</td>
<td>Severe pre-eclampsia</td>
</tr>
<tr>
<td>Maternal outcome</td>
<td>Gestational hypertension</td>
</tr>
<tr>
<td>Maternal outcome</td>
<td>Pre-eclampsia</td>
</tr>
<tr>
<td>Neonatal outcome</td>
<td>Preterm birth</td>
</tr>
<tr>
<td>Antiplatelets for pre-eclampsia (low dose aspirin)</td>
<td></td>
</tr>
<tr>
<td>Maternal outcome</td>
<td>Pre-eclampsia</td>
</tr>
<tr>
<td>Magnesium sulfate</td>
<td></td>
</tr>
<tr>
<td>Maternal outcome</td>
<td>Eclampsia</td>
</tr>
<tr>
<td>Maternal outcome</td>
<td>Case fatality rate of severe pre-eclampsia and eclampsia</td>
</tr>
<tr>
<td>Early administration of magnesium sulfate (at home before referral)</td>
<td></td>
</tr>
<tr>
<td>Maternal outcome</td>
<td>Case fatality rate of severe pre-eclampsia and eclampsia</td>
</tr>
<tr>
<td>Management of unintended pregnancy</td>
<td></td>
</tr>
<tr>
<td>Combination of contraceptive-promoting and educational intervention</td>
<td></td>
</tr>
<tr>
<td>Maternal outcome</td>
<td>Unintended pregnancy among adolescents</td>
</tr>
<tr>
<td>Medications for induced abortion (mifepristone and misoprostol)</td>
<td></td>
</tr>
<tr>
<td>Maternal outcome</td>
<td>No difference in complete abortion rates between medication and clinics group</td>
</tr>
<tr>
<td>Childbirth</td>
<td></td>
</tr>
<tr>
<td>Induction of labour for prolonged pregnancy (uterotonics: oxytocin and misoprostol)</td>
<td></td>
</tr>
<tr>
<td>Neonatal outcome</td>
<td>Perinatal mortality</td>
</tr>
</tbody>
</table>
### Childbirth

| Neonatal outcome | Neonatal mortality or sepsis | RR 0.73, 95% CI 0.64 to 0.76 | Delphi | Low* |
| Neonatal outcome | Neonatal mortality or sepsis | RR 0.85, 95% CI 0.80 to 0.90 | Delphi | Low* |
| Birth attendant hand washing before birth | | | | |
| Neonatal outcome | Cord infection | RR 0.70, 95% CI 0.61 to 0.80 | 2 | Moderate* |

### Management of postpartum haemorrhage

| Active management of third stage of labour (AMTSL) | Maternal outcome | Maternal Hb <9g/dL 24–72 hours postpartum | RR 0.50, 95% CI 0.3 to 0.83 | 2 | Low* |
| Controlled cord traction (as part of AMTSL) | Maternal outcome | Blood loss >500 mL | RR 1.07, 95% CI 1.00 to 1.14 | 2 | High* |
| Preventive uterotonic drugs in the absence of active management of labour | | | | | |
| Oxytocin (when available) | Maternal outcome | Active bleeding controlled within 20 min | RR 0.94, 95% CI 0.91 to 0.98 | 1 | High |
| Oral misoprostol in preventing postpartum haemorrhage (PPH) (when injectable uterotonics not available) | Maternal outcome | Blood loss >1000 mL | RR 0.66, 95% CI 0.45 to 0.98 | 1 | High |
| Uterine balloon tamponade (UBT) (condom catheter) | Maternal outcome | UBT successfully treated PPH | 97% (234 out of 241 cases) | 13 | Low |
| Maternal outcome | All-cause survival | 95% (90 out of 201 cases) | 1 | Low |
| Maternal outcome | Successful treatment of PPH | 97% (223 out of 229 cases) | 1 | Moderate |

### Neonatal care

| Umbilical cord antiseptics in community and primary care settings | Neonatal outcome | Neonatal mortality | RR 0.81, 95% CI 0.71 to 0.92 | 3 | High* |
| Neonatal outcome | Omphalitis/infections | RR 0.77, 95% CI 0.63 to 0.94 | 3 | High* |
| Early skin to skin contact | Neonatal outcome | Breast feeding 0–4 months postbirth | RR 1.27, 95% CI 1.06 to 1.53 | 13 | Moderate |
| Delaying bathing until the second day of life | Neonatal outcome | Hypothermic neonate, rectal temperature | OR 2.90, 95% CI 1.69 to 5.05 | 1 | Moderate |
| Neonatal outcome | Hypothermic neonate, tympanic temperature | OR 4.67, 95% CI 2.62 to 8.38 | 1 | Moderate |
| Early initiation of breast feeding (within the first 24 hours) | Neonatal outcome | Neonatal mortality | RR 0.56, 95% CI 0.40 to 0.79 | 3 | Moderate* |
| Exclusiv breast feeding in the first month of life | Neonatal outcome | Neonatal mortality exclusive versus partial breast feeding | OR 0.27, 95% CI 0.15 to 0.49 | 2 | Moderate* |
| Prophylactic vitamin K for vitamin K deficiency bleeding in neonates | Neonatal outcome | Any moderate to severe bleeding | RR 0.19, 95% CI 0.08 to 0.46 | 1 | Low* |

### Interventions for small and ill babies

| Kangaroo mother care for preterm and for <2000 g babies | Neonatal outcome | Neonatal mortality at discharge | RR 0.60, 95% CI 0.39 to 0.92 | 6 | High |
| Neonatal outcome | Neonatal mortality at latest follow-up | RR 0.67, 95% CI 0.48 to 0.95 | 11 | High |
| Neonatal resuscitation and immediate assessment at facility | Neonatal outcome | Early neonatal deaths | RR 0.62, 95% CI 0.41 to 0.94 | 3 | Moderate* |

---

intervention, or ‘signal function’, that treat the major causes of maternal/neonatal morbidity and mortality and that should be prioritised. Primary level care was defined as care provided by a nurse, family physician or other type of health worker. For example, a rural health centre in Africa would be considered as primary level. Referral level care was defined as care provided in hospitals in general (district or referral); the health care providers at this level are professionals.

The rankings were summarised using the median and the IQR and included in a repeat version of the questionnaire.

In round 3, the experts re-ranked their agreement with each statement, with the opportunity to change their score in view of the group’s response. The re-rankings were summarised and assessed for degree of consensus using IQRs for continuous numerical scales and were accepted when the IQR was 2 or less.

The results of the Delphi consensus are summarised in Table 2 and are rated as low (+) if the median was between 0– and 3, medium (+++) if the median was between 4 and 6 and high (++++) if the median was between 7 and 9.

Patient involvement

Patients were not involved in setting the research question, the outcome measures, the design or the implementation of the study. No patients were asked to advise on interpretation or writing up of results. No patients were advised on dissemination of the present study and its main results.

RESULTS OF THE SYSTEMATIC REVIEW

Our initial search identified 1725 publications, 44 additional records were identified through hand searching. The title and abstract scan resulted in 141 publications that underwent full-text review. Forty-four articles met our selection criteria after the full-text review. The results of the review are 40 potential proxy indicators that are summarised in the SOF table (Table 1).

Preconception

The preconception interventions reviewed included birth spacing and micronutrient supplementation.

Higher risk for preterm birth and low birth weight (LBW) are associated to short interpregnancy intervals (IPIs) (less than 6 months) as well as long IPIs (60 months or more after birth), compared with an IPI of 18–23 months.

Folic acid supplementation and fortification are effective in reducing neonatal mortality.

Pregnancy

The antenatal interventions reviewed included micronutrient and diet supplementation, maternal immunisation, screening and management of infections (syphilis, HIV/AIDS and malaria), prevention and management of pregnancy-induced disorders (notably arterial hypertension), management of prelabour rupture of membranes and preterm labour, drug misuse and management of unintended pregnancy.

Multiple micronutrient supplementation (iron and folic acid) is improving birth outcomes. For women with a history of a baby with neural tube defect, folic acid reduces the recurrence by 70%.

LBW is a major contributor to neonatal mortality and over 95% of LBW babies are born in LMIC countries. While there has been controversy about whether dietary supplementation (e.g., high energy biscuits for chronically undernourished women) in pregnancy can increase birth weight, the 5-year prospective randomised controlled trial in 28 rural Gambian villages by Coxe et al. concludes that supplementation significantly reduces perinatal mortality in at-risk mothers.

Major progress has been achieved for neonatal tetanus, but it remains a significant preventable cause of neonatal mortality globally. Immunisation of pregnant women or women of reproductive age with at least two doses of tetanus toxoid is estimated to reduce mortality from neonatal tetanus by 94%. Infection is a well-acknowledged cause of stillbirth and accounts for an estimated half of all stillbirth, particularly in LMICs. Syphilis screening and treatment with
<table>
<thead>
<tr>
<th>Preconception</th>
<th>Mortality/morbidity</th>
<th>Essential</th>
<th>Primary</th>
<th>Referral</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Family planning</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Birth spacing: interpregnancy interval between 6 months and under 60 months</td>
<td>+ +</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Combination of contraceptive-promoting and educational interventions to avoid unwanted pregnancy*</td>
<td>+++</td>
<td>✓</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Folic acid supplementation and fortification</td>
<td>+ +</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Administration/advice folic acid to women with history of baby of neural tube defects*</td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td><strong>Advice for cessation of alcohol consumption</strong></td>
<td>+++</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education (maternal age, physiology, nutritional status of mother: body mass index (BMI) and so on)*</td>
<td>+++</td>
<td>✓</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Weight reduction in overweight, obese and morbidly obese women*</td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Rubella screening*</td>
<td>+ +</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haemoglobin level/anemia status before pregnancy*</td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td><strong>Pregnancy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Iron and folic acid supplementation (multiple micronutrient)</td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Iron supplementation from second trimester to 3 months postnatal*</td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Nutritional status of mother; BMI*</td>
<td>+++</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Diet supplementation (high energy biscuits) for chronically undernourished women</td>
<td>+ +</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Tetanus toxoid immunisation (at least two vaccinations)</td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Whooping cough immunisation at T2 or T3*</td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Syphilis screening with treatment</td>
<td>+ +</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Intermittent preventive treatment of malaria in pregnancy</td>
<td>+</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Identification of bacteriuria and treatment (urine culture and antibiotic treatment of bacteriuria)*</td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Palpation of uterus and measurement of fundus height (for detecting problems with foetal growth)*</td>
<td>+ +</td>
<td>✓</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td><strong>Advice for cessation of alcohol consumption (adverse effect of alcohol)</strong></td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Smoking cessation during pregnancy (psychosocial interventions)</td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Management of unintended pregnancy: medications for induced abortion (mifepristone and misoprostol)</td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Thyroxine for euthyroid women with positive antithyroid antibodies and recurrent miscarriages*</td>
<td>+ +</td>
<td>✓</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>Kegel exercises to reduce stress incontinence*</td>
<td>+</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Fasting blood sugar checking for high-risk population for gestational diabetes mellitus*</td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Availability of ultrasound</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Foetal echography screening; abnormalities, malformations, growth retardation, macrosomia*</td>
<td>+ +</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Prevention and management of HIV and prevention of mother-to-child transmission in pregnancy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rapid HIV testing</td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Antiretroviral therapy</td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Adherence to antiretroviral medication; mobile phone messages</td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
<td></td>
</tr>
</tbody>
</table>

Continued
### Table 2 Continued

#### Pregnancy

| Management of prelabour rupture of membranes and preterm labour | ++ | +++ | ✓ | ✓ |
| Calcium channel blockers for women in preterm labour | +++ | +++ | ✓ | ✓ |
| Antenatal corticosteroids for accelerating fetal lung maturation for women at risk of preterm birth | +++ | +++ | ✓ | ✓ |
| Antenatal transfer to higher level of neonatal care* | +++ | +++ | ✓ | ✓ |
| Magnesium sulfate in preterm delivery before 34 weeks for neuroprotection* | +++ | +++ | – | ✓ |
| Antibiotics in management of preterm prelabour rupture of membranes* | +++ | +++ | ✓ | ✓ |

#### Prevention and management of hypertension in pregnancy

| Early detection of pre-eclampsia by signs and symptoms* | +++ | +++ | ✓ | ✓ |
| (Better) implementation/adherence to protocols for pregnancy-induced hypertension (PIH)* | +++ | +++ | ✓ | ✓ |
| Antihypertensive drugs to treat PIH* | +++ | +++ | ✓ | ✓ |
| Maternal calcium supplementation (in areas with poor calcium diet) | +++ | +++ | ✓ | ✓ |
| Antiplatelet drugs for pre-eclampsia (low-dose aspirin) | +++ | +++ | ✓ | ✓ |
| Use of magnesium sulfate | +++ | +++ | ✓ | ✓ |
| Early administration of magnesium sulfate (before referral) | +++ | +++ | ✓ | ✓ |

#### Childbirth

| External cephalic version for breech presentation at term | +++ | +++ | – | ✓ |
| Clean birth and postnatal practices at facility | +++ | +++ | ✓ | ✓ |
| Birth attendant hand washing before birth | +++ | +++ | ✓ | ✓ |
| Foetal heart (intermittent) auscultation* | +++ | +++ | ✓ | ✓ |
| Early referral if prolonged labour* | +++ | +++ | ✓ | – |
| Instrumental vaginal delivery (eg, Kiwi vacuum extractor)* | +++ | +++ | ✓ | ✓ |
| Delivery of baby to mother's abdomen* | +++ | +++ | ✓ | ✓ |
| Antibiotic prophylaxis against streptococcus B* | +++ | +++ | ✓ | ✓ |

#### Induction of prolonged pregnancy

| Induction of labour for prolonged pregnancy with uterotonic (oxytocin and misoprostol) | +++ | +++ | – | ✓ |

#### Induction with Foley catheter*

| Management of postpartum haemorrhage | +++ | +++ | ✓ | ✓ |
| Active management of third stage of labour | +++ | +++ | ✓ | ✓ |
| Use of uterotonics for preventing postpartum haemorrhage (PPH) prevention: oxytocin preferred (if available), oral misoprostol second choice (when injectable uterotonics not available) | +++ | +++ | ✓ | ✓ |
| Uterine balloon tamponade (condom catheter) | +++ | +++ | ✓ | ✓ |
| Measurement of blood loss (blood collection bag and blood collection sheets)* | +++ | +++ | ✓ | ✓ |
| Recombinant factor VII in massive PPH* | ++ | ++ | ✓ | ✓ |
| Tranexamic acid in postpartum haemorrhage (PPH)* | +++ | ++ | ✓ | ✓ |
| Uterine massage and emptying the bladder* | +++ | +++ | ✓ | ✓ |

#### Neonatal care

| Umbilical cord antiseptics in community and primary care settings | +++ | +++ | ✓ | ✓ |
| Early skin-to-skin contact | +++ | +++ | ✓ | ✓ |

Contd
Table 2  Continued

<table>
<thead>
<tr>
<th>Neonatal care</th>
<th>+++</th>
<th>+++</th>
<th>✓</th>
<th>✓</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avoidance of hypothermia (delaying bathing until the second day of life, temperature monitoring)</td>
<td>+++</td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Early initiation of breast feeding within 1 hour of life</td>
<td>+++</td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Exclusive breast feeding in the first months of life</td>
<td>+++</td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Prophylactic vitamin K for vitamin K deficiency bleeding in neonates</td>
<td>+++</td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Antibiotic prophylaxis for neonates at risk of bacterial infection*</td>
<td>+++</td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>BCG vaccination before discharge (in areas where tuberculosis is common)*</td>
<td>+++</td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Congenital cardiac disease screening*</td>
<td>++</td>
<td>++</td>
<td>–</td>
<td>✓</td>
</tr>
<tr>
<td>Advise and teach mother to wash hands after change of nappy (infection prevention)*</td>
<td>+++</td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Interventions for small and ill babies</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Parents kangaroo care for preterm and for &lt;2000 g babies</td>
<td>+++</td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Umbilical cord milking for preterm babies*</td>
<td>++</td>
<td>++</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Nasal continuous positive airway pressure for neonates with respiratory distress syndrome*</td>
<td>+++</td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Antibiotics for sepsis*</td>
<td>+++</td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Prevention of hypoglycaemia for small for gestational age and preterm babies (monitor glycaemia and early feeding/glucose)*</td>
<td>+++</td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Neonatal resuscitation and immediate assessment at facility</td>
<td>+++</td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Danger signs predicting severe neonatal illness to be assessed during postnatal contacts (predictive for need for hospitalisation)</td>
<td>+++</td>
<td>+++</td>
<td>✓</td>
<td>✓</td>
</tr>
</tbody>
</table>

Postpartum*

| Precautions to avoid endometritis*                                          | +++   | +++   | ✓     | ✓     |
| Contraception to avoid unwanted pregnancy*                                 | +++   | +++   | ✓     | ✓     |

penicillin reduces syphilis-related stillbirth by 82% and syphilis-specific neonatal death by 80%.59 The effect in all studies was large, and there is a clear biological mechanism, but as only few of the included studies were adjusted for potential confounding factors, quality of the evidence was graded as low.53 54

Intermittent preventive treatment of malaria in pregnancy is a routine drug administration to prevent malaria and its consequences in pregnant women in areas of moderate to high malaria transmission. Routine chemoprevention for malaria and its consequences have been extensively tested in RCTs, with clinically important benefits on anaemia and parasitaemia in the mother39 40 and reduced neonatal mortality.31

The majority of HIV-infected children acquired their infections as a result of mother-to-child transmission during pregnancy, labour or breast feeding. In areas with lower health services infrastructure, infections may stay undetected, which is problematic as early diagnosis and treatment demonstrated improved clinical outcomes.36 37

About 50% of people living with HIV are unaware of their diagnosis.36 37 Reliable point-of-care HIV diagnostic tests, administering antiretroviral drugs to the HIV-infected mother and/or to her child during pregnancy, labour or breast feeding and adherence to antiretroviral medication are essential to prevent vertical transmission.98-100

Preterm birth is a major contributor to perinatal mortality and morbidity. Calcium channel blockers for women in preterm labour have benefits over placebo or no treatment in terms of postponement of birth and were shown to have benefits over beta-mimetics with respect to prolongation of pregnancy, serious neonatal morbidity and maternal adverse effects.98 Corticosteroid therapy used to accelerate foetal lung maturation for women at risk of preterm birth is relatively inexpensive and feasible to implement at primary level in an LMIC context if skilled health care providers are available to identify women at risk of preterm birth and administer intramuscular injections.94 95

Gestational hypertensive diseases, including pregnancy-induced hypertension, pre-eclampsia and eclampsia are leading causes of maternal and infant morbidity and mortality.109 Early detection is crucial for monitoring and prevention. Pre-eclampsia is related to a lack of placental invasion, and its complications on the pregnancy can be detected by ultrasound.107-109 Gestational calcium supplementation is associated with a reduction in hypertensive disorders in pregnancy, especially for women with a low
calcium intake,\textsuperscript{26–28} and reduces gestational hypertension, severe pre-eclampsia and pre-eclampsia.\textsuperscript{70, 73} Administration of antplatelets (e.g., low-dose aspirin) to pregnant women at high risk of pre-eclampsia or those with gestational hypertension prevents pre-eclampsia.\textsuperscript{74} Magnesium sulfate is one of the most effective anticonvulsants to protect women from severe pre-eclampsia and eclampsia and, if administered timely, reduces the risk of seizure repetition and reduces case fatality rate of severe pre-eclampsia and eclampsia.\textsuperscript{75–77} Magnesium sulfate more than halves the risk of eclampsia.\textsuperscript{78, 79} For women who received a magnesium sulfate injection before referral, case fatality rate of severe pre-eclampsia and eclampsia was reduced by 79%.\textsuperscript{80} Even though the effect was strong, due to a small sample size, the evidence was graded low. WHO recommends that magnesium sulfate is administered to women with severe pre-eclampsia before they are transferred to a secondary or tertiary level facility.\textsuperscript{81}

A combination of contraceptive promoting and educational interventions reduce unintended pregnancy, while only contraceptive-promoting interventions showed little or no difference in the risk of unintended first pregnancy (RR 1.01, 95% CI (0.81 to 1.20)).\textsuperscript{82}

Medical abortion uses drugs (mifepristone and misoprostol) to terminate a pregnancy and is an important alternative to surgical methods of pregnancy termination, especially in areas where access to surgical termination is not available.\textsuperscript{83, 84}

**Childbirth**

Interventions during and close to childbirth include clean birth and postnatal practices, the management of postpartum haemorrhage and preventive uterotonics drugs in the absence of active management of labour.

Clean birth practices include: hand washing, clean perineum, clean birth surface, cutting of the umbilical cord using a clean implement and clean cord tying.\textsuperscript{85} Clean postnatal practices include: chlorhexidine, other antimicrobial applications to the cord, avoidance of harmful cord applications, skin applications and emollients and hand washing.\textsuperscript{86} These are estimated to reduce neonatal mortality in a facility and home setting. Even though the evidence quality is low or very low, as there is strong biological plausibility, the GRADE recommendation for these practices is strong.\textsuperscript{83, 85}

Active management of third stage of labour (AMTSL) is a package of three components or steps: (1) administration of an uterotonics, preferably oxytocin, immediately after birth of the baby; (2) controlled cord traction to deliver the placenta, if skilled birth attendants are available;\textsuperscript{89, 91} and (3) massage of the uterine fundus after the placenta is delivered, with administration of an uterotonics as most important part.\textsuperscript{83, 85} In the absence of AMTSL, a preventive uterotonics drug (oxytocin or misoprostol) should be administrated by a health worker trained in its use for prevention of postpartum haemorrhage (PPH).\textsuperscript{87} If both oxytocin and misoprostol are available, oxytocin is the preferred first-line treatment.\textsuperscript{87, 89}

Oral or sublingual misoprostol compared with placebo is effective in reducing severe and is a suitable first-line treatment alternative for PPH in settings where the use of oxytocin is not feasible.\textsuperscript{89, 90}

Uterine balloon tamponade is a relatively simple approach and demonstrated to be an effective technique to treat PPH in developed countries but is underused in LMICs due to the high cost of the balloon. A sterile rubber catheter fitted with a condom was developed as innovative low cost alternative in Bangladesh in 2001.\textsuperscript{84} Three studies suggest that condom catheter uterine balloon tamponade (C-UBT) is simple to use, inexpensive, safe and may be used by any healthcare provider involved in delivery for controlling massive PPH.\textsuperscript{87–90}

**Neonatal care**

Interventions for all neonates include hygienic care, prevention of hypothermia, support for immediate breast feeding and prophylactic vitamin K.

Early skin-to-skin contact benefits breastfeeding outcomes at 0–4 months postbirth,\textsuperscript{94} while early initiation of breast feeding lowers all-cause neonatal mortality among live birth.\textsuperscript{96} Exclusive breast feeding reduces the risk of neonatal mortality compared with partial breast feeding.\textsuperscript{97}

Thermal care (immediate drying, warming, skin to skin and delayed bathing) of neonates prevents hypothermia.\textsuperscript{94} Bathing in warm water 1 hour after delivery is associated with a significant increase in hypothermia in both measurement methods, rectal and tympanic.\textsuperscript{98}

Neonatal chlorhexidine cord care reduces the incidence of omphalitis and neonatal mortality.\textsuperscript{99}

A single dose of 1 mg of intramuscular vitamin K after birth is effective in the prevention of classic haemorrhagic disease of the neonate.\textsuperscript{100}

Interventions for small and ill neonates include neonatal resuscitation and immediate assessment, prevention of hypothermia and danger signs predicting severe neonatal illness to be assessed during postnatal contacts.

Every year, an estimated 10 million babies require assistance to initiate breathing. Basic neonatal care (warming, drying, stimulation and resuscitation including bag-and-mask ventilation) would be sufficient to save most babies in need of resuscitation in low-resource settings.\textsuperscript{101} Training of neonatal resuscitation in facilities could reduce 30% of intrapartum-related mortality RR 0.70, 95% CI (0.59 to 0.84) and 38% of early neonatal mortality.\textsuperscript{102} The coverage of this intervention remains low in countries where most neonatal deaths occur, which presents a missed opportunity to save lives.\textsuperscript{102}

Kangaroo mother care (KMC), among other benefits, reduces neonatal mortality.\textsuperscript{103} KMC in LBW infants is an alternative to conventional neonatal care.

The Young Infants Clinical Signs Study Group developed a simple single algorithm that can identify severe illness in infants aged 0–2 months who are brought to health facilities.\textsuperscript{104} The algorithm was developed from a large prospectively collected dataset and consists of
seventy signs: (1) history of difficulty feeding, (2) history of convulsions, (3) movement only when stimulated, (4) respiratory rate of 60 breaths per minute or more, (5) severe chest in-drawing, (6) temperature of 37.5°C or more and (7) temperature below 35.5°C. Each of these signs is predictive for the need of hospitalization in infants of the age group 0-6 days and 7-90 days and should be used to identify sick infants that need referral faster.

RESULTS OF THE DELPHI CONSENSUS
The Delphi experts completed and prioritised the results of the systematic review, resulting in a table of 77 proxy indicators (table 2). Indicators that were added or modified in the Delphi process are marked with an asterisk (*).

DISCUSSION
Evidence documents the benefits of eHealth tools in terms of increasing satisfaction of health care professionals (HCPs), desilation, acquisition of new knowledge and their potential impact (largely based on observational studies). However, there is little evidence demonstrating that these tools lead to changes in health behaviours, which have a meaningful impact on the patient outcomes. An evaluation of a mobile tool for health workers in India used an approach that is similar to the proposed proxy indicators, measuring the impact of the mobile tools on key health behaviours. On the other hand, this evaluation demonstrated the feasibility of the proposed approach, showing large and statistically significant impacts on many outcomes in the antenatal care domain; on the other hand, it accentuated the need to evaluate the impact of eHealth tools on patient outcomes beyond knowledge acquisition.

The evaluation showed that even though there were significant impacts on mother’s knowledge on exclusive breastfeeding, this did not translate into significant impacts on reported exclusive breastfeeding for 6 months.

The main difficulty of evaluating the impact on patient outcomes can be attributed to the limited identification of measurable and reliable indicators. This systematic review identified a set of proxy indicators (table 1) to evaluate the impact of maternal and neonatal eHealth tools in low-resource settings on health outcomes. Experts completed the results with additional proxy indicators, for example, ‘Whooping cough immunization at T2 or T3’, and reorganised them in a Delphi consensus (table 2). Table 3 provides a summarised view on the identified intervention domains of the proxy indicators, while the granularity of the list of proxy indicators (table 2) is necessary to identify the most appropriate proxy indicators for specific eHealth projects or programmes.

Some of the via the Delphi consensus identified supplementary proxy indicators were not determined in the systematic review, as there were no direct relation to outcomes. They were however added by the experts as they provide essential information for a better case management that may lead to improved outcomes, for example, measurement of blood loss (blood collection bag and blood collection sheets) or nutritional status of mother (BMI). For example, systematically collecting information on blood loss does not prevent PPH, but early detection of excess bleeding may allow for fast and efficient treatment.

The experts also added more general proxy indicators like ‘Antihypertensive drugs to treat pregnancy-induced hypertension (PIH)’ in addition to the more specific ones, for example, antipatelet drugs for pre-eclampsia (low dose aspirin), which were identified in the systematic review. Furthermore, some additional proxy indicators measure whether cases are managed better, which is assumed to improve outcomes, for example, early referral if prolonged labour or antenatal transfer to higher level of neonatal care. In practice, they will need to be mapped to the local context, as the appropriate time for referral in case of, for example, prolonged labour varies depending on the location and context (availability of medication and of the facility).

Moreover the experts identified ‘Tranexamic acid in post-partum hemorrhage’ in the Delphi consensus as an additional proxy indicator. The systematic review did not identify this due to inconclusive literature or poor quality evidence at the time of the systematic review. However, recently, a new randomised, double-blind, placebo-controlled trial was published, concluding that tranexamic acid reduces PPH death of clinically diagnosed women and that early treatment seems to optimise benefits.

Limitations
The proxy indicators are probably more suitable to evaluate maternal and neonatal eHealth programmes or components of a programme. For specific maternal/neonatal eHealth programmes or projects (eg, targeted at HIV infected mothers), additional indicators might be identifiable (eg, vertical transmission of HIV/AIDS).

Some proxy indicators may also have been overlooked as unfeasible and disruptive uses of eHealth may emerge and offer unexpected ways to improve practices.

Application
When applied in future studies, proxy indicators related to the eHealth intervention are identified from table 2. Some of them need to be mapped to the local context, practices and available resources. For example ‘the use of uterotonic for PPH prevention’: oxytocin is the preferred choice when available, while oral misoprostol should be the second choice, when injectable uterotonics are not available for treatment. The proxy indicators can detect and attest changes in behaviour and may explain changes in mortality, even if causality cannot be formally demonstrated.

The local mapping enables the utilisation of the proxy indicators in various contexts, while the ‘high level’ of the indicators allows systematically collecting data from different projects and programmes (collective data/
### Table 3: Categories of proxy indicators

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Education</td>
<td>Education and training of HGPs for interventions that are targeting behaviour changes, knowledge acquisition or awareness of patients or HCps. Examples of proxy indicators are: education on birth spacing, advice for cessation of alcohol, birth attendant hand washing before birth or avoidance of hypothermia (delaying bathing until the second day of life, temperature monitoring).</td>
</tr>
<tr>
<td>Screening for infectious diseases and risk factors</td>
<td>Interventions for a better availability and implementation of screening for infectious diseases and risk factors. Examples of proxy indicators are: nutritional status of mother; body mass index, syphilis screening with treatment, fasting blood sugar checking for high-risk population for gestational diabetes mellitus.</td>
</tr>
<tr>
<td>Availability of ultrasound</td>
<td>The availability of ultrasound allows the detection of abnormalities, malformations, growth retardation and macrosomia but is also assumed to improve the number of prenatal care visits of the pregnant women.</td>
</tr>
<tr>
<td>Management of unintended pregnancy</td>
<td>The better availability and implementation of the management of unintended pregnancy. Examples of a proxy indicator is medications for induced abortion (misoprostol and misoprostol).</td>
</tr>
<tr>
<td>Timely referral</td>
<td>Timely identification and referral of pregnancy-related complications and emergencies are key factors to reduce maternal and new-born mortality. Examples of proxy indicators are: antenatal transfer to higher level of neonatal care, early identification of danger signs predicting severe new-born illness to be assessed during postnatal contacts (predictive for need for hospitalisation).</td>
</tr>
<tr>
<td>Prevention and management of HIV</td>
<td>Interventions for a better availability and implementation of interventions to prevent and manage HIV. Examples of proxy indicators are: rapid HIV testing, adherence to antiretroviral medication and mobile phone messages.</td>
</tr>
<tr>
<td>Management of prelabour rupture of membranes and preterm labour</td>
<td>Interventions for a better availability and implementation of interventions to manage prelabour rupture of membranes and preterm labour. Examples of proxy indicators are: calcium channel blockers for women in preterm labour, antenatal corticosteroids for accelerating foetal lung maturation for women at risk of preterm birth or antibiotics in management of preterm labour rupture of membranes.</td>
</tr>
<tr>
<td>Prevention and management of hypertension in pregnancy</td>
<td>Interventions for a better availability and implementation of interventions to prevent and manage hypertension in pregnancy. Examples of proxy indicators are: better implementation/adherence to protocols for pregnancy-induced hypertension, antithrombotic drugs for pre-eclampsia (low-dose aspirin) and the use of magnesium sulfate.</td>
</tr>
<tr>
<td>Induction of prolonged pregnancy</td>
<td>Interventions for an induction of prolonged pregnancy. Examples of proxy indicators are: induction of labour for prolonged pregnancy with uterotonic (oxytocin and misoprostol) or induction with Foley catheter.</td>
</tr>
<tr>
<td>Management of postpartum haemorrhage</td>
<td>Interventions for a better prevention and management of postpartum haemorrhage. Examples of proxy indicators are: use of uterotonic for PPH prevention: oxytocin preferred (if available), oral misoprostol second choice (when injectable uterotonic not available), the measurement of blood loss (blood collection bag and blood collection sheets) or tranexamic acid in postpartum haemorrhage.</td>
</tr>
<tr>
<td>Interventions for small and ill babies</td>
<td>Interventions for a better availability and implementation of interventions for small and ill babies. Examples of proxy indicators are: parental kangaroo care for preterm and for &lt;2000g babies or neonatal resuscitation and immediate assessment at facility.</td>
</tr>
</tbody>
</table>

Evidence). Because of the mapping, it is the same proxy indicator for different context, measuring what is locally and temporally relevant and therefore sustainable.

Table 2 could also serve as a checklist when implementing a project or as a basis for the baseline questionnaire and for creating the didactic contents.

**CONCLUSION**  
The identified proxy indicators provide a workable approach to measuring the impact of eHealth interventions on maternal and neonatal health. However, their validation and calibration in various settings with different methodologies is still required.

The availability of indicators (direct and proxy) facilitates consistent outcome measurements and comparability of studies, and this methodology could be applied to other domains, for example, chronic diseases.

This implementation research aims at creating evidence to support decision makers to answer questions like ‘why should we invest in eHealth rather than medical staff,”
immunization or medications? and to identify and implement solutions with the greatest potential impact on health. The availability of indicators and the possibility to measure and demonstrate scientific evidence for medical benefits that is based on reliable indicators will accelerate decision makers’ ability to institutionalise eHealth activities and to commit strategically at the regional and national level.

Acknowledgements The authors would like to thank the following persons (in alphabetical order) for their valuable inputs and useful comments: M. Semina, N. Shadf, M. Bosean, D. Desaiwax, O. Iron, J. Krithikaman, S. Masteen, N. Naji, S. Nazneen, R. Plüter, L. Smith, S. Sizensko, B. Stull, M. Temmerman and G. Wolverson.

Contributors GP designed the protocol with input from AG. GP designed the protocol with input from AG. GP and UL drafted the studies. AG solved discrepancies in the ratings. GP extracted the data, which was thoroughly checked and verified by UL. GP drafted the article. AG and UL made a critical review of the article. All authors gave their final approval of the version to be published.

Funding Department of Radiology and Medical Informatics, Faculty of Medicine, University of Geneva.

Competing interests None declared.

Patient consent Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data sharing available.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

REFERENCES


33. Caroline Perimi LH, Gaisibühler A. A systematic review to identify proxy-indicators to quantify the impact of eHealth tools on maternal and neonatal health outcomes in low and middle-income countries - including Delphi consensus. Prospero 2018 https://www.crd.york.ac.uk/PROSPERO/display_record.php?recordID= 27391.


6.2.2. Article 2

Navigating through digital health outcome evaluation methodologies: an algorithm based on a scoping review and emerging methodologies.

Caroline Perrin (M.Sc.), Mirana Randriambelonoro (M.Sc.), Antoine Geissbuhler (MD)
Department of radiology and medical informatics, Faculty of Medicine, University of Geneva, Switzerland

Abstract

Background: Digital health interventions are recognized for their potential and are increasingly implemented globally. The evidence base is growing, but currently there are still relatively few studies evaluating improvements of digital health interventions in health outcomes.

Objective: The objectives of this article were to understand, analyze and map how researchers approach digital health outcome evaluations in different settings through a scoping review and to develop an algorithm, based on these results, to provide a pedagogical overview of methods for evaluating health outcomes of digital health interventions.

Methods: For the scoping review PubMed, EMBASE, and the Cochrane Central Register of Controlled Trials (CENTRAL) were scanned using a predefined search strategy to identify articles measuring the impact of digital health interventions on health outcomes. The algorithm was developed based on analysis and insights from the scoping review.

Results: The database search retrieved 3597 citations of which 208 were included. These articles were reviewed in detail and were classified into different categories: level of income (of the country where the study was conducted), system categories, disease addressed by the intervention, and evaluation method. The gap analysis identified additional emerging approaches that were included in the algorithm.

Conclusion: Through analysis of the literature, we were able to demonstrate that digital health outcome studies rely on traditional clinical evaluation designs, even though these interventions are often more complex and depended on the context, culture, and the patient than classical interventions like efficacy of insulin on diabetes control. In addition to the methodologies extracted from studies through the database search we identified study methodologies through desk research, whose design features address some of the shortcomings of traditional clinical methodologies, when applying them to digital health interventions. We integrated all identified methodologies into an algorithm that provides a high-level overview, enables the user to navigate through these methodologies based on the design features and investigator’s priorities, and to facilitate the identification of one or more potential appropriate methodologies.

Keywords: Digital health; Evaluation; Health Outcomes; Methodologies

Introduction

The potential of digital health is highlighted by a growing evidence base [1, 2]. Historically, the majority of evidence focuses on care process improvements rather than improvements in health outcomes [2-9]. Given the current status of adoption of digital health globally, there are relatively few studies measuring the impact on health outcomes [10]. This dispersed evidence might lead to an ineffective use of resources or investments in potentially harmful digital health interventions [11, 12]. Considering the trend towards an even greater adoption of digital health, impact evaluations should not only
measure feasibility, implementation activity, and attitudes or perceived impact, but quantify changes caused by digital health interventions in health outcomes [2-7, 9, 13-15].

The application of digital health ranges from education, civil registration, data interchange, telemedicine, reimbursement mechanisms and electronic medical records to public health and disease surveillance systems, among many other applications. These interventions are often embedded in a complex web of factors [16] and depend on the setting, culture, or individual, unlike pharmacological interventions like insulin on its receptor, which is why traditional clinical evaluations might not be the best approach to measure outcomes of these interventions [17]. They may also present characteristics, e.g., being flexible or evolving, which are not typical for traditional clinical interventions and require methodologies that acknowledge these characteristics. Digital health interventions may also provide opportunities for evaluation processes that are not supported by traditional clinical evaluations like the automatic collection of process-related data, as for example, in drug development, where data derived from digital health technologies can be employed as proxy outcome measures to reduce random sampling issues, and to potentially increase the statistical power of clinical trials [18, 19].

The objectives of this manuscript were to understand, analyze and map how researchers approach digital health outcome evaluations in different settings through a scoping review and to develop an algorithm, based on these results, to provide a pedagogical overview of methods for evaluating health outcomes of digital health interventions.

Methods

a) Scoping Review

Design
A scoping review aims to map the volume, nature, and characteristics of existing literature in a field of interest [20]. We chose this design as the goal of the review was to map methodologies of studies measuring the impact of digital health interventions on health outcome, and not summing up the best available research, which is the goal of a systematic review [20].

There were no restrictions on the types of study design eligible for inclusion. Studies were excluded when investigating non-health outcomes as for example usability, satisfaction or economic aspects.

Search Method for Identification of Studies
We performed an electronic search of PubMed, EMBASE, and the Cochrane Central Register of Controlled Trials (CENTRAL) to identify all studies published until May 18, 2018, with no restriction on date, or study design, but restricted the search to articles in English language. The search strategy was adapted for each database according to the search capabilities of that database. We used a combination of text words and controlled vocabulary terms (MeSH Terms, Emtree), related to digital health interventions, evaluation and outcomes. MeSH Terms and Emtree both include broader and narrower terms and synonyms. Additional records were identified through grey literature searches.

Selection of Studies
One author conducted an initial screening to exclude articles whose titles were obviously irrelevant, for example a title that was excluded as obviously irrelevant is: What is the definition of sports-related concussio: a systematic review. Subsequently, two reviewers independently rated titles and abstracts based on relevance to the study objectives. The rating scale was: 0 = not relevant, 1 = potentially relevant, 2 = definitely relevant. The third reviewer resolved discrepancies in the rating. All studies that were rated potentially relevant or definitely relevant underwent full-text review.

Categories
For the studies data was extracted for different categories: 1) level of income (of the country where the study was conducted based on the country income classifications for the World Bank[21]), as the setting seems to have an influence on the design [22]; 2) system categories (representing an overarching category of different related interventions based on the WHO classification of digital
health interventions [23]); 3) disease addressed by the intervention; and 4) evaluation method. The authors agreed on these categories by consensus as relevant to understand and analyze the prevalence of different methodologies in relation to these categories.

Data Extraction and Management
Level of income, system categories, disease addressed by the intervention, and evaluation method, were abstracted by one author into a standardized spreadsheet and were thoroughly checked by the second reviewer. Disagreements were resolved by discussion and, if necessary, by arbitration involving the third reviewer.

Gap analysis
The authors searched PubMed and grey literature to identify additional existing or emerging methodologies after the scoping review was completed. The authors searched specifically for innovative methodologies, which included reviewing workshop, conceptual and policy papers. Identified methodologies were evaluated their appropriateness for digital health, based on the limitations of methodologies that were identified in the scoping review.

b) Algorithm
First the authors analyzed the included studies to understand why the investigators in the included studies chose the specific methodology. Then the authors looked at advantages and limitations of all identified methodologies. Based on these analyses the authors developed an algorithm to navigate through the identified methodologies and to provide an overview on different potentially appropriate outcome methodologies.

Results
a) Scoping Review

Literature Search
The flow of the study results is illustrated in Figure 1. Database and grey literature search identified 3597 references. After the removal of duplicates using Endnote (version X8), 2892 were screened by title, after the remaining 1960 references by title and abstract. 1630 articles were excluded, leaving 300 articles to screen for eligibility. Ninety-two articles were further excluded. A total of 208 studies were included [Annex]. The first studies aiming at evaluating the impact of digital health interventions on health outcomes appeared in 2000, with an exponential increase over the years, as illustrated in Figure 2.
Figure 1 - Flow chart

The years in which the included studies were published are represented in Figure 2. The trendline illustrates that there was an increase in studies measuring the outcomes of digital health interventions in recent years.

Figure 2 – Number of digital health outcome studies by year

Almost all studies (95%, 197 studies) were conducted in high-income settings, while only a few were conducted in upper-middle-income (5 studies), lower-middle-income (5 studies), or low-income countries (1). Most of the identified studies are experimental (175 studies), of which 134 have a randomized design (112 superiority design, 5 non-inferiority, 1 equivalency trial, 2 cross-over design, 7 cluster randomized, 5 Pragmatic RCT, 2 step-wedge RCT), and 41 a quasi-experimental design (20 non-randomized studies, 20 controlled before after, 1 interrupted time series). These studies have
predominantly identified mortality and morbidity or proximal outcome measures as e.g. glycemic control of diabetes patients as clinical outcome measures. The other 33 of the identified studies are observational of which 21 are cohort, 7 case control, 4 chart reviews and 1 is cross sectional.

**Mapping of digital health outcome evaluation methodologies**

![Sankey diagram](image-url)

**Figure 3 - Sankey diagram – mapping of digital health outcome evaluation methodologies**

Figure 3 maps the design of the outcome studies to the system categories to the diseases addressed, illustrating that most of the identified studies are experimental and fall into the category of telemedicine. The category of telemedicine includes the consultation between remote client and healthcare provider, the remote monitoring of client health or diagnostic data by healthcare providers, the transmission of medical data to the healthcare provider, and the consultation for case management between health care providers.

The diseases that was addressed the most is diabetes with 49 studies followed by cardiac diseases (myocardial infarction, heart failure, ischemic heart disease). In the diagram cardiovascular diseases (cardiac: 27, stroke: 13, Hypertension 12) are grouped into one category, as well as respiratory (COPD: 8, asthma: 6, sleep apnea: 5, smoking: 3).

**Gap analysis**

Reviewing the literature additional emerging approaches were identified, that in the context of digital health address some of the shortcoming of the traditional methodologies, or provide more flexibility, adaptiveness or responsiveness. General additional approaches that were identified are the theory of change and scaling out. Furthermore, the following approaches for multicomponent, evolving or adapting approaches were identified: Single-case experimental designs (SCED), Continuous Evaluation of Evolving Behavioral Intervention Technologies (CEEBIT), and Micro-Randomized Trials (MRT).

**b) Algorithm**

The included articles in the scoping review were carefully reviewed. For each methodology information on advantages, shortcomings, and reasons when it cannot be used was collected. After the gap analysis, literature on the additional identified approaches was reviewed and information on advantages, shortcomings, and reasons when these cannot be used was collected as well. This collected information
was analyzed and main criteria for using or excluding methodologies as well as conditional parameters were defined. These criteria and conditions were then modeled into the algorithm. The algorithm is designed to provide a high-level overview on various methodologies, to navigate through these based on distinctive criteria, conditions, and investigator’s priorities, aiming at facilitating the identification of one or more potential appropriate methodologies. The algorithm was tested by the authors on a set of diverse interventions.
Figure 4 – Digital health intervention outcome evaluation study design algorithm
Discussion
Principal Results of the Scoping Review

Most of the included studies are experimental, randomized designs and were conducted in high-income settings. Analyzing the results based on the categories illustrates that the majority of included studies evaluated telemedicine interventions in a randomized design located in a high-income setting. Other domains, like NCDs, have observed a strong divide in prevalence of RCTs between low and high income settings, which this review also finds for digital health [22]. Historically telemedicine has been established as one of the first applications in the digital health domain with gradually more complex implementations [24, 25]. The relative long use of telemedicine might explain the strong representation in the results. The high representation of studies evaluating interventions addressing diseases that fall into NCDs can be expected considering that almost all studies were conducted in high-income settings where NCDs represent one of the major burden of diseases [26]. All the study methodologies that were retrieved from the included studies represent common clinical methodologies. These results are not surprising; however, considering the vast amount of digital health projects in low-income setting, they illustrate the need to further look at the underlying reasons and to raise the question if methodologies like RCTs are the appropriate standard to measure outcomes of rapidly evolving digital health technologies, also for low-income settings [16, 27, 28].

Experimental Designs
Randomized Controlled Trial

RCTs randomly allocate patients (by chance alone) to receive an interventions or control, which may be a standard practice, a placebo, or no intervention at all [29]. RCTs are considered the gold standard among available evaluation strategies, as confounding factors are balanced across treatment arms due to randomization [29]. However, an RCT tests if an intervention has a beneficial effect in an ideal situation but randomization and strict eligibility criteria often cause delays, impact the sample size, and generalizability of the results [30]. Furthermore, conducting RCTs is complex and costly, therefore often not feasible in low-resource settings [5]. This can be illustrated with one of the included studies, an RCT that was conducted in Egypt [31]. The study assessed the impact of text-message education on glycemic control compared to paper-based education [31]. The trial had 73 participants (34 intervention group, 39 control group) and a duration of 12 weeks. The authors highlight that due to budgetary concerns they restricted the study to 3 months, even though they believe 6 months would have been the optimal timeframe. In addition, they point out difficulties with time delays during recruitment and overlaps between study phases, as well as that the small sample size may limit generalizability.

The literature discusses the appropriateness of RCTs to evaluate the impact of digital with contrary conclusions [5, 32, 33]. RCTs are a good methodology to test efficacy of specific digital health interventions, e.g. of applications for chronic conditions like diabetes [33]. Furthermore these have a long lag time [34], and at the pace digital health is evolving the intervention or technology might be outdated by the time the trial is completed. Traditional RCTs aim to maximize the internal validity by assuring rigorous control of all variables [30], hindering the generalizability of the results outside the study setting. But, measuring and understanding the beneficial effect of these interventions in real clinical practice is relevant to digital health as these are commonly implemented first and evaluated after.

Pragmatic controlled Trial

An alternative when seeking to maximize external validity for the generalization of the results is the pragmatic controlled trial (PCT) [30]. Compared to traditional trials PCTs generally include a wider range of patients, study sites, and outcomes that are aligned with the evidence needs of decision-makers [30, 35], and positive results from pragmatic research can readily inform practice as it provides evidence that the intervention is effective in usual practice [36]. However, to ensure generalizability internal validity may be overly compromised [30], therefore it is a balancing act to define the appropriate level of internal and external validity. An illustrating example is a multi-site, pragmatic, randomized
comparative effectiveness trial comparing the outcomes of patients randomized to practice based collaborative care (PBCC) versus telemedicine based collaborative care (TBCC) [37], which was designed to inform policy. The trial took place in the United States over a duration of 18 months with 179 patients in the TBCC and 185 patients in the PBCC. The authors chose a pragmatic trial design to compare two viable competing strategies to deliver an evidence-based practice to inform policy rather than an explanatory trial to determine why patients randomized to TBCC had better outcomes than those randomized to PBCC. While the results are not conclusive with respect to identifying treatment mechanisms they are valuable for policy-makers as they clearly indicate that contracting with an off-site depression care team yields better depression outcomes than implementing collaborative care with staff available on-site.

Step-wedge and Cross-over Designs

One major limitation of traditional RCTs is that in some cases due to e.g. ethical considerations or constraining situational factors random assignment is not possible. This can be addressed by using a step-wedge design or crossover design, in which all participants are exposed to the intervention. A study comparing enhanced multi-disciplinary teams (EMDTs) with the standard of care for the treatment of pressure ulcers in long term care (LTC) facilities in Canada selected a step-wedge design to retain the power of randomization while offering all facilities enrolled in the trial exposure to what was believed to be a desirable intervention [38].

In crossover designs the patients act as their own controls, enabling comparisons between and within groups and require less patients than a parallel study and increases precision as between-patient variation is removed [39]. However this design is susceptible to carry-over effects (lasting effect of one treatment on outcome in the following period) and cannot be used for interventions with permanent effect [39, 40]. The Step-Wedge RCT randomizes clusters to the start time of the intervention [41], and is useful when treatment and control cannot be simultaneously controlled. But this design can be lengthy and susceptible to temporal confounding, as more clusters are exposed to the intervention later in the study [41]. While the step-wedge design has similar advantages to crossover design and is adapted to small sample sizes, it is particularly useful if the intervention effect is permanent or if there are carry - over effects [41]. A good example for a digital health step-wedge design in a low-resource setting is a study on the Integrated electronic Diagnosis Approach (ieDA) in Burkina Faso[42]. The evaluation used a step-wedge cluster randomized design with health districts (“clusters”) receiving the intervention at different time points in a randomized order because some aspects of the intervention could only be delivered at the district level, and as the implementing agencies needed to roll out the intervention in a phased manner for practical reasons [42]. When due to funding issues the intervention could only be rolled out to 4 instead of 8 districts the evaluation still provided valid results, as the design increases statistical power within and between comparisons. Due to this responsiveness, and other step-wedge design advantages (e.g. that sample sizes can be smaller than in traditional RCTs), in the context of this study this methodology seems more relevant than the traditional RCT for the evaluation of the impact.

Quasi experiments

Quasi experiments are non-randomized intervention studies, which have many different designs that increase or decrease the level of strength of evidence, with some of them being more likely than others to allow causal interpretations of observed associations [43]. As there is no random assignment statistical association does not infer causality, and alternative explanations for the results, which might reduce the evidence of causation, need to be assessed [43]. This design is useful when random assignment is not possible because of ethical considerations or constraining situational factors [44], and has high external validity, as these experiments are conducted in natural settings, which increases applicability of findings [45]. A good example of a digital health quasi-experiment is the TEMPIS study in which the multicenter community setting used may provide better information about effectiveness in routine clinical care than a RCT [46]. The prospective, nonrandomized, intervention study compared 5 community hospitals participating in the TEMPIS (Telemedical Project for Integrative Stroke Care) network with 5 matched control hospitals without specialized stroke facilities or telemedical support. The results show that implementing TEMPIS, which consists of the set-up of specialized local stroke wards, continuous medical education, and telemedical consultation for patients with acute stroke, extends the benefits of acute specialized stroke care to community hospitals. However, due to the
study design it is not possible to determine only the impact of telemedicine on stroke outcome, but the
effects have to be contributed to the complete network concept [46].
This design is also more often used due to the increasing capacity of institutions to collect routine
clinical data [43]. Digital health is often implemented as a response to a problem like e.g. increasing
cost or adverse events [43] which infers regression to the mean as potential alternative explanation.
Regression towards the mean is a statistical phenomenon where extreme outcomes tend to be
followed by moderate ones [47], meaning that the changes in the problem after the implementation
might not be based on the intervention but by the regression towards the mean.

Observational Designs
Observational studies can be viable alternatives to RCTs, e.g. when randomization is not possible.
Populations with multiple chronic conditions and interventions made it difficult to design RCTs with
sufficient sample size and long-term follow-up to account for all the variability this phenomenon entails,
as RCTs are intended to test the efficacy of an intervention in a restricted sample of subjects under
ideal settings [10]. For these cases with co-morbidities and multiple interventions observational studies
as e.g. cohort studies might be appropriate as these can study multiple outcomes in the same exposure
[48]. Cohort studies follow a group of patients with a common condition or exposure over time, the
participants are selected based on the condition/ exposure status and have the ability to control for
multiple confounders [48]. Some observational study designs (e.g. casecontrol, cross sectional) are
relatively fast and inexpensive [49, 50], therefore more feasible for evaluations with limited resources.
In case control designs participants are selected based on the outcome status ("cases"), and compared
to participants who did not experience the outcome of interest but are representative of the same
population as the cases ("controls"), and can may be used for rare outcomes [50]. An illustrating example
is a study evaluating the effect of a heart failure telemonitoring program on outcomes in a
retrospective 1:1 match cohort study [51]. In this study efficiency is increased by matching on potential
confounders [52], which is generally suitable when the investigators have access to large population
data sources [53].
In cross sectional designs the participants are selected based on inclusion and exclusion criteria for the
study, and not on outcome or exposure status. Data is measured at a single point in time, and outcome
and exposure are measured at the same time [49]. A general limitation of a observational studies is
that they are susceptible to bias and cannot establish valid cause and effect relationships as they can
determine association between exposure and outcomes but cannot predict causation [10, 11].

Principal Results of the Gap Analysis
The gap analysis identified additional emerging approaches, that in the context of digital health address
some of the shortcoming of the traditional methodologies, or provide more flexibility, adaptiveness or
responsiveness.

Additional identified approaches
Theory of change (ToC) was developed to evaluate complex public health interventions and is now used
in public health to demonstrate long-term outcomes by measuring proximal outcomes in a logical
sequence using a backward mapping approach [54]. In the mapping process assumptions to reach the
long-term outcome and influencing contextual factors are made explicit [16, 54]. Digital health
innovation is increasingly integrated into health service delivery, often resulting in interventions with
interconnected digital and non-digital components. The ToC approach can be useful to evaluate what
aspects of an intervention have or have not been successful rather than judging the intervention in
their entirety. Aarons et al propose the concept of ‘scaling-out’, which provides a logical framework
for determining what new empirical evidence is required for an intervention to retain its evidence-
based standard in a new context [55]. They propose three types of scaling-out; 1) type I: population
fixed, different delivery system; 2) type II: delivery system fixed, different population, and type III:
different population and delivery system. Aarons et al argument that even if an intervention has shown
to be effective in previous trials, when moved to a new population or delivery system one cannot
assume to keep the same effectiveness, however a part of the effectiveness is "borrowed" from the
previous evidence base. Based on this "borrowed" effectiveness they identified four levels of evidence
that could be applied to predict or measure the expected health impact of implementing an evidence-
based intervention. To identify the required level of evidence they propose a simplified sequential mediational model that provides a schematic view of the major domains to examine regarding whether a clinical/preventive intervention that has been judged to be evidence-based within one setting would be expected to have similar effects when scaled-out [55]. As digital health often enables the delivery or provides a more cost-efficient delivery of evidence-based interventions, the scale-out concept could provide an adequate and practical framework for measuring the impact of these in a new context.

**Approaches for adapting or evolving digital health interventions**

Digital health interventions are often complex interventions consisting of multiple components. Understanding which components impact health outcomes and whether these interact and are affected by their presence or absence is essential to implement or scale-up meaningful solutions. The previously discussed approaches usually investigate single interventions or intervention packages, but are not assessing components or dynamic interventions, which are adapting or evolving like digital interventions for conditions where a change in the status of a patient might require a change in treatment approach [56]. Single case experimental designs (SCEDs) are also referred to as N-of-1 Experiments and include a family of methods in which each participant serves as his or her own control [57]. During the study controls and intervention conditions are systematically introduced and withdrawn, which enables to assess the effect within and across participants [58]. While they are effective in assessing different components they are not able to determine the time or context in which an intervention option is most efficacious [57]. For example, Rosenberg et al conducted 10 sequential N-of-1 (single-case) experimental trials to test the efficacy of inactivity alerts among obese older adults, a highly sedentary population [59]. Mohr et al [60] proposed this research design Continuous Evaluation of Evolving Behavioral Intervention Technologies (CEEBIT) to test continuous evolving technology as an alternative to traditional RCTs. The CEEBIT methodology proposes a statistically powered framework that permits the evaluation of multiple or evolving interventions, and the introduction of additional ones at any time, while eliminating the ones with poorer outcomes [60]. It includes considerations that are also included in an RCT like randomization, inclusion/exclusion criteria and statistical analysis [17], but the required sample size is considerably reduced as it applies a more liberal Type I error rate of 50% [60]. It is an appropriate methodology for rapidly developing mobile technologies with high levels of sign-up and attrition [61]. The Micro-Randomized Trial (MRT) was developed for just-in-time adaptive interventions (JITAs), which aim to provide the right intervention components at the right times and locations to optimally support health behaviors of individuals [57]. An example of a JITAI is the smartphone behavioral intervention FOCUS [62] that provides illness management support to individuals with schizophrenia. Individuals are prompted three times a day via auditory signals and visual notifications to assess their status in five target domains (medication adherence, mood regulation, sleep, social functioning, and coping with hallucinations). The patients can engage or ignore the prompt. If they engage, the system launches a brief assessment, and when this indicates that the individual is experiencing difficulties, FOCUS recommends self-management strategies to improve the difficulties. Otherwise it provides feedback and positive reinforcement. MRTs enable the causal modeling of proximal effects of randomized intervention components and the evaluation of time-varying moderation of those effects [57]. They can be used to evaluate digital health interventions where intervention components or treatments are delivered frequently and which are designed to impact a long-term outcome by impacting a near-term proximal outcomes [63]. For example, HeartSteps is an mHealth intervention that encourages regular walking via activity suggestions tailored to the individuals' current context. Klasnja et al conducted a MRT to evaluate the efficacy of HeartSteps' activity suggestions to optimize the intervention [64]. [2]

**The Algorithm**

The scoping review demonstrated that digital health studies evaluating health outcomes rely on traditional clinical evaluation designs, even though these interventions are frequently more complex and depended on the context, culture, and the individual than classical intervention. Whether a design is good or bad it can be influenced by the characteristics of the digital health intervention, and also depends on the question asked. Experimental designs for example are best for effectiveness questions, but useless if one wants to understand the natural diffusion of an innovation. What study design to choose also depends on if the study aims at locally evaluating a certain intervention, or if the aim is
more general research about the effectiveness of certain features of digital health interventions, for which multi-context trials would be a requirement.

To illustrate the application of the algorithm the authors applied it to a planned study for a digital health intervention named MiranaBot in Figure 5. MiranaBot is a chatbot powered by artificial intelligence, which allows the users to understand their nutritional habits and provides personalized recommendations based on these. The aim of the study is to understand the efficiency of this conversational agent in detecting the user’s nutrition habits and identifying key food items that should to be reduced, and should be compared to a health professional. The first question in the algorithm is: can the investigator assign exposure, determining if the study can be of experimental or observational nature. In this example the investigators will assign the intervention or treatment, leading to the question if it is possible to measure the outcomes of interest in the evaluation period. For MiranaBot this is the case, as the outcome of interest is change in BMI and nutritional habits. For studies that aim at measuring for example changes in maternal mortality rates it might not be possible to measure these within the study duration. For these cases the theory of change, or if the intervention is adaptive, MRT methodology could be appropriate, both using proximal outcome measures. For MiranaBot the next question would be if the intervention has an evidence base and is using a new delivery system, population, or both. This is not the case, as the intervention itself has not been tested yet for its efficacy, but if in another study in the future the same intervention would be tested on another population (e.g. adolescents, instead of adults) it could use the scale out methodology. The next question is if the intervention is evolving or adaptive, which isn’t the case, leading to if it is ethically safe to randomize. For this study it is ethically safe to randomize as standard of care is compared to the intervention. The next question is if internal or external validity is prioritized. Internal validity is measuring a specific constant or variable within the context of the study, while external validity moves to translate the results to a larger population or setting. For MiranaBot internal validity is prioritized, leading to the question if exposure can be assigned randomly at individual level, which is the case. As for this study there is a resource constraint it leads to the question if the intervention effect is permanent. The intervention effect is not necessarily permanent, however as during the study period it is unlikely that the participants immediately unlearn what MiranaBot taught them, and the investigators want to reduce cross-over effects, the algorithm proposes the Step-Wedge Trial. But as the investigators have time constraints the alternative propositions are a quasi-experimental study or a pragmatic RCT. In this example the investigators choose the pragmatic RCT as they prefer a randomized design.
Figure 5 – Application of Example MiranaBot to the algorithm
Limitations
This paper has several limitations. Due to the nature of the scoping review, no quantitative analyses were done, and the selection of the studies was done by interpretation of the researchers. No risk of bias assessment was conducted as this did not affect the interpretative synthesis of the findings. No exclusion criteria were applied with regard to the quality of the studies to ensure broad coverage of the studied methodologies. The algorithm can provide an overview on different methodologies and provide guidance on which methodologies might be appropriate, however due to the complexity of the methodologies and preferences of the investigator the algorithm cannot account for every factor. Two different investigators might receive different results for the same study as their preferences influence the algorithm.

Conclusion
This review illustrates that even though the amount of studies quantifying impact on health outcomes has increased over the years, considering the vast amount of digital health projects and interventions the availability of evidence on the effect on health outcomes of these interventions is still relatively limited, potentially as the standard methodologies might not be appropriate for these complex, intertwined interventions. Through analysis of the literature, we were able to demonstrate that digital health outcome studies rely on standard established methodologies. However, these methodologies were developed for interventions that have a long life-time and are less influenced by rapidly changing technologies there are some shortcomings when applying them to evaluate digital health interventions. For implementers of digital health, it is also worth to look beyond the standard and to consider new or adapted methodologies. It is worth to consider additional approaches, that address some of the shortcoming of the traditional methodologies, or provide more flexibility, adaptiveness or responsiveness. There is no one-fits all solution, and the optimum methodology depends on the context, setting, and the research question but also on the development stage or maturity of the intervention. In such cases, where it is not feasible to measure direct impact as e.g. the distance between intervention and outcome is far, proxy indicators can be used to assess the effectiveness of digital health interventions [65, 66]. Finally, there are many dimensions to why implementations have an impact or success or fail, therefore using different methodologies at different stages in the project or implementation, and looking at the context can help to understand why a digital health intervention did or did not have an impact on health outcomes.

Conflicts of Interest
None declared
REFERENCES

17. Pham Q, Wiljer D, Cafazzo JA. Beyond the randomized controlled trial: a review of alternatives in mHealth clinical trial methods. JMIR mHealth and uHealth. 2016;4(3). PMID: 27613084
22. Fan H, Song F. An assessment of randomized controlled trials (RCTs) for non-communicable diseases (NCDs): more and higher quality research is required in less developed countries. Scientific reports. 2015;5:13221.
23. WHO. Classification of digital health interventions v1.0: a shared language to describe the uses of digital technology for health. 2018.
34. Ioannidis JP. Effect of the statistical significance of results on the time to completion and publication of randomized efficacy trials. Jama. 1998;279(4):281-6. PMID: 9450711.
56. Baker TB, Gustafson DH, Shah D. How can research keep up with eHealth? Ten strategies for increasing the timeliness and usefulness of eHealth research. Journal of medical Internet research. 2014;16(2). PMID: 24554442
Reverse innovation experiences from the RAFT e-learning and telemedicine network

**Background**
Reverse innovation is the term that is used to describe the transfer of innovative processes or technology from developed countries to developing countries. Examples of its application range from business models to global health.

Safaricom and Vodafone for example launched in 2007 in Kenya the mobile phone based money transfer and micro-financing service M-Pesa, that quickly extended to other countries, revolutionized banking in Africa and is now also available in eastern Europe. Orange recently launched a mobile payment service in France, which includes the possibility to send money to African countries.

The example of the Narayana Hospital (NH) Group in India illustrates a process innovation and reverse engineering in the medical field: at NH the price of an open heart surgery is just a 10th of what it costs in the United States (US), while the profit margin is slightly higher and mortality within 30 days of bypass surgery is lower. This was achieved by applying industrial concepts of standardization, specialization of labor, economies of scale, and assembly line production to healthcare. NH launched a hospital in proximity to the US on the Cayman Islands offering health services and surgeries at approximately 50% of costs in the US, targeted at...
inhabitants of the US without health insurance.

While there are many great examples of reverse innovation in various industries, this article focuses on reverse innovation experiences related to healthcare and medical education within the RAFT network, and the eHealth and telemedicine group of the University of Geneva and the Geneva University Hospitals.

**Problem-based blended e-learning**

Founded in 1985, the University of Mauritius currently consists of five faculties and their corresponding departments. The Faculty of Science hosts 8 departments, including the Department of Medicine, which has been established in August 2004 and is in charge of implementing the training program for general practitioners (MD Program). The MD program aims to train medical professionals that are able to combine clinical competence, medical knowledge, proficiency in inter-personal and communication skills, and the ability to respond to the psychosocial aspects in health-care and to keep abreast of new developments in the medical field.

To support the implementation of this new training program the Faculty of Medicine of the University of Geneva (UNIGE), the University of Mauritius (UnioM) and the Mauritian Ministry of Health & Quality of Life (MOQH) developed a blended problem-based e-learning program [9] with face-to-face and online components. The teaching and learning model are based on the problem-based learning approach which consists of three phases: 1) analysis of the problem, 2) self-learning, and 3) conclusion. This problem-based approach has been implemented for 20 years at UNIGE through several face-to-face meetings of small groups of eight to ten students with a tutor. This approach was not translatable to UnioM due to shortage of staff, thus the e-learning implementation.

In the MD program five modules that are corresponding to the theoretical second and third year studies were implemented on the e-learning platform. Each of these modules contains a number of clinical cases, for example the module heart and circulation has six clinical cases, e.g. the blue baby or long QT. Each of these clinical cases corresponds to a problem and specific objectives that the student needs to master at the end of the case. The blue baby case represents a newborn with the problem of cyanosis and permits the student to review elements of cardiovascular embryology, physiology of fetal circulation, pathophysiology of congenital heart diseases, etc. The cases are illustrated on the platform with various teaching materials (texts, images, animations, etc.) and the learning progress can be can be self-evaluated.

The blended e-learning MD program started at the University of Mauritius during the academic year 2014 -2015. Based on the positive feedback of the blended learning approach at the University of Mauritius, the medical faculty of the University of Geneva introduces this in the curriculum of the academic year 2016 – 2017.

**Figure 1: Screenshot of the Online Clinical Blue Baby Case**

Source: University of Mauritius Department of Medicine e-Training Portal of the MD program

**Tools initially developed for low-bandwidth environments**

The RAFT network [10] developed tools (Dudus, Bogou) for a low-bandwidth environment that are now also being used in a high-income country context. The RAFT network was launched in 2001 by the University of Geneva and the Geneva University Hospitals, and is recognized as an official WHO collaborating centre for e-health and Telemedicine. It aims at de-isolating healthcare professionals that work in remote settings, by providing them access to adapted, distance continuing medical education, b) the ability to request expert advice when dealing with challenging clinical cases, and c) the use of specialized diagnostic tools (e.g., ultrasonography and electrocardiography) with the remote supervision of experts. The network is currently deployed in 20 African countries and connects hundreds of care professionals. Initially targeted towards French-speaking Africa, it is now deployed in English-speaking and Portuguese-speaking Africa, in Latin America, Nepal and Kyrgyzstan. Most of RAFT activities take place as South-South collaborations, connecting telemedicine facilities in reference (national and regional hospitals), and remote sites (district and rural hospitals).
Dudal, a distance capacity building tool

Dudal is an interactive distance education environment developed by RAFT, specifically designed to function over low-bandwidth connections. Freely available to anyone, 80% of the courses are produced in Africa and are regularly viewed by hundreds of care professionals. Webcast courses are generally followed by 30 min of discussion involving all participants. Once webcasted, these courses are immediately available for playback from the RAFT website.

Initially designed to provide continuous education to healthcare professionals in rural Africa, Dudal is now used by the French Ministry of Culture together with the Université Numérique Francophone Mondiale (UNFM) to provide courses for the e-heritage program. This program is widely attended by experts from around the world, most of them in Europe as the program contains modules on the legislation of subleasing artwork for exhibitions and another module on the international classifications to identify artwork and to fight smuggling.

Bogou, a secure tele-expertise tool

Most Sub-Saharan African countries are experiencing a severe crisis of health professionals. The majority of doctors and specialists are practicing in the capitals, while 60% of the population is living in rural areas. This uneven distribution of health professionals and the lack of human resources are real barriers to equitable access to quality health care and services in developing countries. This has motivated the development of the platform "Bogou", which enables the support of health care professionals in remote areas covering vulnerable populations. This tool has been developed in the context of the RAFT network and enables remote diagnostics and expertise to respond to this severe shortage of qualified health professionals. Bogou enables healthcare professionals (doctors, midwives, nurses, etc.) to request remote secondary advice or a diagnosis of a specialist.

It is a web application that is organized in circles around various topics. The background of members is verified and depending on their specialization and interest they will be added to the circles. One or more moderators who decide to validate or not a user registration application manages each circle. Users can only act within the circle they belong to. When a request is posted, all circle members and experts are automatically alerted by email. The requestor may change the status of a case anytime, for example to "resolved" when he is satisfied with the answers of the experts. In terms of computer security, the data is decrypted on the server and is sent to the client through an encrypted connection (https).

Currently more than 1,000 healthcare professionals across Africa use Bogou. These users from 15 countries are divided into 40 circles on 4 continents. More than 2,000 cases are discussed and response time varies between 13 minutes and 5 days. The tool is used for all health problems ranging from maternal and child health to Non Communicable Diseases (NCDs). It is currently mainly used to request remote expertise in obstetric ultrasound, dermatology, cardiology and reconstructive surgery.

Benefits of Bogou range from economic savings to a de-isolation of care professionals. A study in Mali demonstrates economic benefits on a sample of 215 patients (obstetric ultrasound, and cardiology) over a 10 months period. Distance diagnostics through Bogou helped to save more than $110,000 as unnecessary and costly medical evacuations could be avoided. Additionally, those centers, which used these tools, had an increase in activity between 25% and 36%. Furthermore, connecting isolated healthcare professionals via Bogou is an encouraging factor and has a beneficial effect on the motivation and retention of the care professionals.

World Hospitals and Health Services – Learning from African Innovations in Health Services Vol. 52 No. 3 33

62
Bogou has been developed and is mainly used in low- and middle-income countries, but has also proven to be useful to enable South-North support countries. Additionally, the tool has great potential to be used for various challenges in a developed country context. Even though infrastructure is generally well developed in high-income countries, in specific regions like for example the Alpine Space it can be challenging and time-consuming to get access to specific health care services. The Bogou tool could for example be used to provide distance obstetric consultations and ultrasounds, significantly reducing the travel time for pregnant women in these regions.

**dDopp, a robust, modular and cost-efficient ultrasound imaging device**

The RAFT network utilizes the dDopp ultrasound imaging machine that was designed and developed for the rural African context. Not only is the purchase price a fraction of a standard device, it is also small, mobile, robust, not affected by sand, as it has no ventilator, and it is easy to repair and maintain locally thanks to its modular design. The battery for example can be easily exchanged with a standard lithium battery. The device can be charged and doesn’t require a stable electricity connectivity and can be recharged from various energy sources, e.g. a car battery or a solar panel.

While the dDopp ultrasound machine has proven to be useful in the South it has been discovered in the North and is for example used by an esthetic clinic in Belgium to measure the thickness of adipose tissue and by urologists in Italy.

While in the South the device is used mainly for obstetrics in the North it is applied differently, nonetheless it allows clinics in a high-income context to utilize an ultrasound device for small interventions, which do not justify the purchase price of a standard device. In addition, due to its mobile and robust character paramedics can use it in emergency situations.

**Conclusion**

In many multi-national companies and institutions, innovation is recognized as indispensable for their sustainability, but innovations coming from developing countries have often been neglected, as these tend to involve less technological breakthroughs. But they do present novel and innovative combinations of existing knowledge and technologies addressing local problems using innovative processes and business models. Reverse innovation has the potential to redistribute wealth and to accelerate the development of low-income countries, and applied to health it has the potential to generate effective global solutions.

Reverse innovation has the potential to significantly contribute to the post Millennium Development Goal era in the Sustainable Development Goal number 3 of improving health for all. Non-Communicable Diseases (NCDs) for example are prevalent and posing a major burden in all income-level countries and innovations from a developing country context might help achieving this goal, especially as a NCDs have a disproportionate impact on poor or socially marginalized groups in wealthy nations.

**Biographies**

**Caroline Perrin, M.Sc. (Management of Information Systems and Service Science)**

She is a PhD student in the Geneva University Global Health Program and working in the department of e-Health and telemedicine of the University and the University Hospitals of Geneva. Her research is focusing on impact metrics (and proxy indicators for impact) of e-Health applications in a low- and middle-income countries context.

Prof. Chiek Oumar Bagayoko MD, PhD (Medical Informatics)

He is Professor in Medical Informatics at the Mali University School of Medicine. He is the director of CERTES, the Center of Expertise and Research in Telemedicine and e-Health in Mali and Coordinator of the RAFT Network. Prof. Bagayoko has significant expertise and experience in the implementation of Health Information Systems and e-Health projects in Africa.

**Dr. Georges Bédiang, MD, MAS, MPH**

He is Research and Teaching Assistant at Faculty of Medicine and Biomedical Sciences at Yaoundé University, PhD Student in Medical Informatics at the University of Geneva, and Coordinator of telemedicine activities of the RAFT network. His domains of interest are: eLearning, telemedicine, hospitals/health information systems, mHealth, virtual patients and knowledge engineering.

Prof. Antoine Geissbuhler

He is a Professor of Medicine, Chairman of the Department of Radiology and Medical Informatics at Geneva University, Director of the Division of eHealth and Telemedicine at Geneva University Hospitals, and Director of the RAFT telemedicine network. He is also President of the Health-On-the-Net Foundation, Past-President of the International Medical Informatics Association, and Fellow of the American College of Medical Informaticians.

**References**


Learning From eHealth Implementations Through “Implementomics”: A Multidimensional Annotation Model Applied to eHealth Projects of the RAFT Network

Caroline Perrin¹,², Georges Bediang³, Mirana Randriambelionoro ¹,² and Antoine Geissbuhler ¹,²

¹ Khalifa University of Science and Technology, Al Ain, United Arab Emirates ² Faculty of Medicine and Biomedical Sciences, University of Yaoundé I, Yaoundé, Cameroon

The implementation of digital health technologies has increased globally, producing substantial amounts of information and knowledge. While there are still areas in digital health that are understudied, concurrently there is an exponential increase in published articles, guidelines, methods, projects, and experiences, many of which fail to reach critical mass (pilotitis). Semantically describing and documenting this implementation knowledge and the effectiveness of these tools will help to avoid the duplication of efforts, to reduce preventable implementation obstacles, and to ensure that investments are targeted to the most important technological innovations. The RAFT annotation model, presented in this paper, enables to semantically describe all elements of various outputs and implementation projects that were developed, are used, or are part of the RAFT network. This model was initially developed to annotate various implementations and outputs of the RAFT network to facilitate knowledge documentation and sharing, and to be used as a proof of concept for the Implementome. The Implementome will be an interconnected knowledge system that enables the user to navigate on multiple dimensions through meta-data annotated projects, people, and information, and can serve as base for consensus building, best practices and guidelines. The RAFT annotation model can be further developed to enable the annotation of outputs, implementations, people, initiatives, and projects of the digital health domain in general.

Keywords: eHealth, digital health, implementation, annotation model, implementomics, semantics

INTRODUCTION

Over the last decade internet, connectivity (1), and the implementation of digital health tools, projects and interventions increased globally, addressing challenges faced by both developed and developing countries in providing accessible, cost-effective, high-quality health care services (2–12). Digital health collectively describes the concepts of eHealth, telemedicine and mHealth, as
defined by the World Health Organization (WHO). eHealth is the cost-effective and secure use of information and communication technologies (ICTs) for health and health-related fields, while mHealth is a component of eHealth, and involves the provision of health services and information via mobile technologies (13).

The digital divide between developing and industrialized countries is still prevalent, but connectivity is extending rapidly, including into rural areas (14), facilitating implementations of digital health. In low-resource settings, cellphone-based health education and consultations, personalized health tracking devices, and mobile diagnostic technologies can provide real-time information to improve both individual and public health. Smartphones, e-payment systems, and telemedicine improve access to quality care and more timely deployment of emergency services. ICT innovation, and increased connectivity enables health facilities to enter data directly into central servers through web applications without the need for any software installation or database management at the local level (15).

This deployment of digital health produces voluminous literature on a multiplicity of digital health innovations, and while there are still areas in digital health that are understudied, concurrently there is an increase in published articles, guidelines, methods, projects, and experiences, many of which fail to reach critical mass (piloting) (16, 17). The term "piloting" is used to express the frustration of many of those in the health sector that the vast majority of digital health projects are limited in scale and undertaken in stand-alone, vertical project mode, with predominantly narrowly focused interventions targeting relatively small populations (18).

Furthermore, terminology is evolving, but definitions are diverging, making it time-consuming and occasionally impossible to find appropriate information. Documenting digital health implementation knowledge and help will help to avoid duplicating efforts and ensure that investments target meaningful technological innovations. We refer to these developments as "Implementomics," the ability to capture, organize, and exploit the multidimensional knowledge related to implementation issues. As for other "omics" domains, a key challenge is to master the variety and volume of information. Knowledge models can improve this.

The variety and volume of information is challenging in the domain of digital health, but also on a smaller scale as within the Réseau en Afrique Francophone pour la Télémédecine (RAFT) network. Established in 2003, RAFT is a telemedicine and elearning network that is currently operational on four continents. It supports isolated healthcare professionals by providing telemedicine and elearning services using affordable, low-bandwidth technologies. It is not only a platform to share and exchange knowledge, but has implemented a solid local infrastructure to ensure sustainability and maintenance of the network. From experience developed with South-South collaboration to top-down and bottom-up approaches and various certification models, much knowledge was produced and various perspectives for improvement were proposed (11, 12, 15, 20).

The model, presented in this paper, enables to semantically describe elements (people, results, etc.) of implementations, articles, courses, tools etc. that were developed, or are part of the RAFT network. It serves as a proof of concept for the Implementome, which will be an interconnected knowledge system. The basic building block for the Implementome is to associate machine-readable metadata to content.

An example of an annotation model or metadata schema is one for digital photos, which enables to describe, among other properties, the camera used to take the picture, shutter speed, date, and location (21). A useful side-effect of this model is that the same piece of metadata can describe the content, as well as to organize and classify it, therefore setting up other properties that were initially not considered, e.g., the possibility to search for photos of a location, taken at a given time (17).

The example of an annotation model designed for a ward of an oncological hospital illustrates how annotation models can facilitate group decision making in a complex environment, and depicts the crucial role of annotations to address organizational complexity and manage heterogeneous flows of essential data and information (22).

RAFT ANNOTATION MODEL

Annotations are not only a way of explaining and enriching a resource with observations, but are also a means of transmitting and sharing ideas to improve collaborative work practices. The RAFT annotation model was developed to initially enable the annotation of implementations, activities, and outputs of the network. Figure 1 provides an overview of the RAFT annotation model with its eight super-dimensions, and sub-dimensions.

Description of the Super-Dimensions

MeSH (Medical Subjects Headings)

MeSH is a hierarchical controlled vocabulary used for indexing, cataloging, and searching for biomedical and health-related information and documents (23), and in particular the MEDLINE database of biomedical scientific publications. MeSH terminology annotates the entry, ranging from specific diseases, population characteristics, or information science to geographic location.

DCMI (Dublin Core Metadata Initiative)

The DCMI was developed to describe web resources and consists of a set of fifteen elements (24). The RAFT annotation model uses seven of these:
- Title: given to the resource by the creator or publisher.
- Date: indicates a date associated with the creation or availability of the resource.
- Identifier: number or string to uniquely identify the resource (e.g., URL or ISBN).
- Language: language of the content.
- Publisher: entity responsible for publishing.
- Relation: reference to a related entry.
- Type: the genre or nature of the resource e.g., homepage or journal paper.
Stakeholders
This may be either an organization or a person, or both. There may also be several organizations or persons.

- Person: affiliated person(s) to a project, or authors of a publication. When available, the Open Researcher and Contributor ID (ORCID) will be included, which aims
at uniquely identifying and connecting persons to their contributions across disciplines, borders, and time (25).

- **Organization**: stakeholder organizations in a project, or affiliations of authors for publications.

**Interventions**

Interventions are annotated using the Classification of digital health interventions that was developed by WHO (26).

- **Health System Challenges**: high-level description of needs and addressed challenges in the implementation context.
- **System Category**: describes the types of ICT applications and information systems designed to deliver one or more digital health intervention.
- **Digital health interventions**: organized into four umbrella groupings based on the targeted primary user: (1) Clients; (2) Health care providers; (3) Health system or resource managers; or (4) Data services.

**Process**

The process section annotates implementation aspects. As opposed to outcome evaluations, process evaluation focuses on inputs, activities, and outputs, and evaluating how they work together. Evaluating the process may explain why implementations did or did not work.

- **Challenge**: annotates specific implementation challenges, e.g., resistance to change or change in leadership, but also overarching challenges like political or ethical.
- **Barriers**: annotates barriers like adoption, technical illiteracy, or missing legislation.
- **Risks**: annotates risks like political instability, but also funding continuity.
- **Facilitators**: annotates factors contributing to a successful implementation, e.g., governmental support.
- **Best practices**: annotates a procedure or process that produced optimal results and is established or proposed as a standard suitable for widespread adoption.
- **Lessons learnt**: annotates the learning gained from the process of performing the project or service, e.g., institutional anchoring, or identification of champions.
- **Key performance indicators (KPIs)**: annotates KPIs, like the number of telemedicine-cases and their status: measuring the number and status of cases is an important KPI and helps identifying if a connected site is having technical or organizational challenges and might need additional support.

**Evidence/Evaluation Outcomes**

Based on a classification from Zanaboni et al. (27), entries are annotated with the following five categories:

- **Health and Clinical**: can be specified with general indicators, disease-specific indicators, or patient-reported outcomes (e.g., improvement in health status, quality of life, medication management, mortality, physical activity, or related to diabetes or hypertension);
- **Psychological and Behavioral**: e.g., patients changing their behavior toward the way they manage their health or a specific disease;
- **Health care utilization**: impact of digital health interventions on the resources involved, including economic effects and time used by patients and providers, and use of the health care system;
- **System adoption and use**: e.g., how patients use a digital health intervention in practice, or the organizational change for health care professionals;
- **System attributes**: other effects focusing on the evaluation of systems themselves, e.g., usability for patients and providers;

**Project/Service**

Can be annotated with

- **Maturity levels**: (Informal: early adoption in the absence of formal processes and policies; Pilot: Testing and evaluating in a given situation; Scale-up: beyond the initial pilot, extension to other populations or centers; Established operation: ongoing, since at least 1 year, with funding to continue).
- **Status of the project/service**: (ongoing, completed and ended, completed and continued, discontinued).
- **System level of the implementation**: (local, regional, national, international).
- **Business model**:.
- **Funding**: (Public funding; Private funding).
- **Adaptability**: the project/service may have broad contextual adaptability, where it broadly applies to a range of settings and usage scenarios, or it might have specific adaptability, where it is only suited to specific needs, users, or geographical localities.

- **Technology**: Hardware, Software (open source, publicly available, proprietary).
- **Interoperability**: according to four interoperability levels defined by the Healthcare Information and Management Systems Society (HIMSS) (29).
- **(1) Foundational**: basic level of technical interoperability. Data from one IT system can be received by another, but the receiving system does not need to be able to interpret it.
- **(2) Structural**: intermediate level of technical interoperability, where the data exchanges between IT systems can be interpreted at the data field level, and clinical or operational purpose and meaning of the data is preserved.
- **(3) Semantic**: highest level of technical interoperability, where two or more systems can exchange information, and the exchanged information can be used. It allows the electronic exchange of patient summary information among caregivers and other authorized parties via potentially disparate electronic health record (EHR) systems.
- **Organizational**: includes non-technical considerations and enables interoperability that is integrated into end-user processes and workflows in a manner that supports efficiencies, relationships and overall health and wellness through cooperative use of shared data both across and within organizational boundaries.

**Context Elements**

Enables the annotation of context elements that influenced, enabled, challenged, or facilitated the implementation.
• Strategy: the entry corresponds, implemented or contributed to a digital health strategy (national or organizational) in the context.
• Infrastructure: annotates the need for infrastructure, or leveraging on nationally or private governed infrastructure, or existing services.
• Legislation/Policy/Compliance: annotates e.g., compliance with national guidelines or regulatory statuses, or needs for legislation or policy. If there is e.g., no legislation regulating various aspects of telemedicine, questions around medical liability might complicate the implementation significantly.
• Ethics: annotates ethical challenges and frameworks.
• Governance: annotates national, regional or organizational governance, and Governance challenges, models, and decision-making in acute care.
• Standards: annotates need or use of standards.
• Cybersecurity: annotates e.g., cybersecurity guidance, mitigation for cybersecurity risks, or to report issues.

For illustrative purposes the annotation model was applied to two examples (Figure 2). The first one is a journal article reporting on a randomized controlled trial that investigated the effect of SMS reminders on the adherence and cure of tuberculosis patients in Cameroon. The second example is on the RAFT network in general.

Application of the Model to Examples
RAFT Article
• MeSH Headings: Tuberculosis, Pulmonary (Diseases); Cameroon (Geographic locations); Adult (Persons); Clinical Trial (Study Characteristics); Randomized Controlled Trial (Epidemiologic study characteristics); Urban Population (Population characteristics); Multicenter Studies as Topic (Environment and Public Health); Journal Article (Publication Format);
• DCMI: Title: SMS reminders to improve adherence and cure of tuberculosis patients in Cameroon (TB-SMS Cameroon):
RAFT Article
• Stakeholders: Person: Georges Bediang (https://orcid.org/0000-0001-9177-8798); Beat Stoll; Nadia Elia; Jean-Louis Abena; Antoine Geissbuhler (https://orcid.org/0000-0001-5039-3373); Organization: Faculty of Medicine and Biomedical Sciences, University of Yaoundé I, Yaoundé, Cameroon; Geneva Tumor Registry, Institute of Global Health, Faculty of Medicine, University of Geneva, Geneva Switzerland; National Tuberculosis Control Program, Ministry of Public Health, Yaoundé, Cameroon; Department of Radiology and Medical Informatics, Faculty of Medicine, University of Geneva, Geneva, Switzerland;
• Intervention: Health System Challenge: 5.3 Low adherence to treatments; Digital Health Intervention: 1.1.3 Transmit targeted alerts and reminders to client(s); System Category: D Client Communication System;
• Process Evaluation: Challenge; Barrier; Lessons learnt;
• Evidence Evaluation Outcomes: Health and Clinical, Psychological and Behavioral;
• Project/Service: Maturity: informal; Status: completed and ended; System Level: local;

RAFT Network
• MeSH Headings: Online Social Networking (Information Science); Teaching (Education); Remote Consultation (Health Services Administration);
• DCMI: Title: Réseau en Afrique Francophone pour la Télémédecine; Identifier: http://raft.network; PMCID: PMC5932834; Language: English, French, Spanish, Portuguese;
• Type: Website
• Stakeholder: Person: Antoine Geissbuhler (https://orcid.org/0000-0001-5039-3373); Cheick-Oumar Bagayoko; Organization: Department of eHealth and Telemedicine, University Hospital of Geneva, Geneva, Switzerland; Department of Radiology and Medical Informatics, Faculty of Medicine, University of Geneva, Geneva, Switzerland; CERTES Expertise Center and telemedicine Research in e-Health, Ramako, Mali;
• Intervention: Health System Challenge: 2.2 Insufficient supply of services, 2.4 Insufficient supply of qualified health workers, 3.2 Insufficient health worker competence, 3.4 Low health worker motivation, 5.2 Geographical inaccessibility, 6.2 Lack of or inappropriate referrals, 6.4 Delayed provision of care, 6.5 Inadequate access to transportation; Digital Health Intervention: 2.4.4 Consultation for case management between healthcare provider(s), 2.6.1 Coordinate emergency response and transport, 2.8.1 Provide training content to healthcare provider(s); System Category: 5 Learning and training system, Y Telemedicine;
• Project/Service: Maturity: established operation; Status: completed and ongoing System level: international, Funding: Public funding, Donor/non-public funding;
• Elements Infrastructure, Governance.

DISCUSSION
The presented model was developed to annotate activities and outputs of the RAFT network. The purpose of this model is not to replace existing annotations like MeSH, but to connect and extend these to enable a detailed annotation of digital health activities. It will present opportunities for extension when applying it to the digital health domain in general, e.g., annotating health care organizations electronic medical record implementations with the stages of HIMSS Analytics Electronic Medical Record Adoption Model (EMRAM), which measures adoption and utilization of electronic medical record (EMR) functions (29); or the inclusion of technologies like blockchain in the model.

The value of the Implementation will be determined by the dependability, and the number and quality of annotated entries. Different strategies exist to populate the Implementation. These range from manual annotation by trained experts, crowdsourcing, and hybrid strategies to auto-harvesting of entries.
Crowdsourcing, or “citizen science,” is a strategy for the collection, analysis and sharing of large amounts of data, generally via the Internet. For researchers it represents an opportunity to overcome common barriers to data collection, such as ensuring extensive geographic coverage and maintaining long-term projects. For example, crowdsourcing has been used for decades to harness the power of citizen bird-watchers to better document the distribution and migratory patterns of a wide range of bird species (30). Millions of people around the globe help professional scientists with tasks that range from
monitoring changes in local biodiversity to providing innovation and computing power for new drug development.

Manual indexing is precise and trustworthy, but can be time-consuming and costly. Automatic annotation is widely used in genomics (31) and proteomics (32), considering the enormous amount of data shared and classified in these domains. Uniprot, a protein knowledge-based platform, uses two prediction systems the Unified Rule System (Unirule) and the statistical Automatic Annotation System (SAAS) to automatically annotate unreviewed entries in an efficient and scalable manner, in addition to their manual annotation.

MeSH terms in Pubmed are currently manually assigned by human indexers. However, there are consistent efforts in improving automatic indexing (33–35). Probabilistic models and other machine learning algorithms (33, 35) are tested with different sets of data to predict correct MeSH terms for documents. MESH Now (34) is an integrated approach using multiple strategies to generate a combined list of candidate MeSH terms for a target article. While automatic classification remains challenging and is undergoing research, the advance in artificial intelligence, especially in Natural Language Processing and Deep Learning, could lead to the way to more precise annotation.

Manual annotation by trained experts produces high-quality records but is expensive and time consuming, while entirely automatic strategies are fast and low-cost, but have higher probability for poor quality. As a perspective for the future, the Implementone could use a hybrid annotation system with an automatic indexing followed by manual curation.

The main limitation of the proposed RAFT annotation model is that it might not capture all the complexity and detail of the annotated items, as the model uses ontologies to simplify the complicated web of inter-related and resembling terms, and might assign a broader more general term instead of a more detailed concept.

Another limitation comes from the annotations themselves. The model is based on experience of the RAFT network, and is therefore limited to describe well-known processes within the network.

Establishing the possibility to propose additional annotations that will be reviewed and potentially added when implementing the model addresses these limitations. These additional annotations might describe more details, use more accurate terms, or add new emerging concepts.

When implementing the model further limitations or issues might appear like duplicate subjects. To minimize this, the model proposes to use unique identifiers like ORCID ID, or PMCID when possible, however this is not possible for all subjects and needs to be considered in the implementation strategy.

CONCLUSION

The presented annotation model enables the collection, annotation and connection of information to encourage the exchange of knowledge and learning, and to facilitate joint knowledge generation to address knowledge gaps, between and across digital health projects, programs and initiatives. This model has been presented and the application was demonstrated. The next step will be to develop the Implementone, based on the annotation model and to potentially refine and evolve the model by extending it to the domain of digital health in general.

The vision of the Implementone is: (1) to create a central multidimensional digital health implementation hub to facilitate knowledge documentation and sharing, (2) to pool and connect knowledge resources produced by various projects, initiatives and organizations, (3) to enable joint knowledge creation, and (4) to link organizations, academia, people, policymakers, civil society, and other users to digital health knowledge.

To develop the proof-of-concept for the Implementone the annotation model will be applied to describe a variety of digital health projects, organizations, tools, and experts, which are identified in a mapping study. These annotated contents will be added to an interconnected knowledge system that will enable the user to navigate on multiple dimensions through metadata annotated contents. The metadata will cover a broad set of elements of relevance for the understanding and processing of the information at different levels of granularity. For example, for an annotated study, it can include information on the regions of the data collection, the study methodology, the intervention, the authors, outcomes, or pointers to related studies.

DATA AVAILABILITY

No datasets were generated or analyzed for this study.

AUTHOR CONTRIBUTIONS

CP, MR, GR, and AG contributed to the conception and development of the annotation model, which is presented and discussed in this paper. CP wrote the first draft of the manuscript and integrated proposed revisions. MR and AG wrote sections of the manuscript. All authors contributed to manuscript revision, read and approved the submitted version.

FUNDING

Department of Radiology and Medical Informatics, Faculty of Medicine, University of Geneva.

REFERENCES


Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
7. Conclusions and perspectives

This thesis investigated aspects of evaluating health outcomes of digital health on different levels and proposes approaches to overcome some of the associated challenges. The first article identified proxy-indicators to shed light on what could be measured [104]. The second article suggests an algorithm to help identifying approaches on how to measure. The third article outlines the added value of bi-directional learning between the south and north. The forth article proposes a model that enables semantically annotating and connecting information, evidence, knowledge and experience to provide a basis for informed decision-making.

7.1. Measurable indicators

One difficulty of evaluating the impact of digital health remains in the limited identification of measurable and reliable indicators. Proxy indicators address this issue by measuring changes that have a well-established evidence base for improving health or decreasing mortality closer to the intervention.

Article 1 identified a set of 77 proxy-indicators to measure the impact of maternal and neonatal digital health tools in low resource settings through a systematic review, which was completed by a Delphi consensus of international experts. Depending on the context (e.g. availability of medication in a setting) different interventions will be recommended. Some of the proxy-indicators address the same issue, therefore the full indicator list won’t be relevant to an implementation. This list will need to be mapped to the implementation as well as the local context, practices, and available resources. An illustrating example is ‘the use of uterotonics for PPH prevention’: where oxytocin is the preferred choice if locally available, while oral misoprostol is the second choice, in settings where injectable uterotonics are not available for treatment [105, 106]. Which proxy-indicator should be measured depends on the availability of these medications in the implemented setting.

The proxy-indicators are summarized in intervention domain, like e.g. Education, Management of unintended pregnancy, or Prevention and Management of Hypertension in Pregnancy. This ‘high level’ of the indicators allows systemically collecting data from different projects and programs (collective data/evidence), while the local mapping enables the utilisation of the proxy-indicators in various contexts. For example, The Prevention and Management of Hypertension in Pregnancy includes interventions to improve availability and implementation of interventions aiming at preventing and managing hypertension in pregnancy. Examples of proxy-indicators in this intervention domain are: (better) implementation/adherence to protocols for pregnancy-induced hypertension (PIH), antiplatelet drugs for preeclampsia (low dose aspirin), as well as the use of magnesium sulfate. The proxy indicators identified though the systematic review and Delphi consensus provide a workable approach to measure the impact of digital health interventions on maternal and neonatal health. However, these have not been validated yet and limitations might arise when implementing them. These will need to be measured in various settings with different methodologies, and potentially be calibrated to establish their reliability.

7.2. Methodological considerations

Article 1 answered the question of what to measure, and raises questions on a different level of granularity around how outcomes of digital health interventions can be measured once indicators are identified. Article 2
reflects on this question by first conducting a scoping review to understand, analyze and map how researchers approach digital health outcome evaluations in different settings. The results of the scoping review illustrate that even though the amount of studies quantifying impact on health outcomes has increased over the years, considering the vast amount of digital health projects and interventions the availability of evidence on the effect on health outcomes of these interventions is still relatively limited, potentially as the standard methodologies might not be appropriate for these complex, inter-tangled interventions. The analysis of the results of the scoping review demonstrated that digital health outcome studies rely on traditional clinical evaluation designs, even though these interventions are often more complex and depended on the context, culture, and the individual. 95% of the included studies in the scoping review were conducted in high-income countries, and most of the identified studies are experimental with a randomized design. All the study methodologies that were retrieved from the included studies represent common clinical methodologies, which demonstrates that digital health outcome studies rely on standard established methodologies. But, as these methodologies were developed for interventions that have a long life-time and are less influenced by rapidly changing technologies there are some shortcomings when applying them to evaluate digital health interventions. When reviewing the literature, in the gap analysis, additional emerging approaches were identified, that in the context of digital health address some of the shortcoming of the traditional methodologies, or provide more flexibility, adaptiveness or responsiveness, like for example the theory of change, scaling out, or single-case experimental designs. However, most of the emerging approaches are relatively novel, therefore they have not yet been used for evaluations, or just for very few. One study for example, conducted 10 sequential N-of-1 (single-case) experimental trials to test the efficacy of inactivity alerts among obese older adults, a highly sedentary population, where patients were monitored for a baseline phase followed by an intervention phase, which was then removed and participants were monitored to test an experimental effect [107]. The pedagogical overview is summarized by the algorithm; however, this algorithm has some limitations. Due to the complexity of various factors influencing digital health implementations and the dependency on the preferences of the investigator there is no one-size-fits-all solution, and the algorithm might provide different guidance to users that apply it to the same intervention. This is due to the fact that preferences of the investigator can influence the navigation within the algorithm. The optimum methodology depends on the context, setting, and the research question but also on the development stage or maturity of the intervention. Furthermore, there are many dimensions to why implementations have an impact or success or fail, therefore using mixed methods, or looking at the context and other factors can help to understand why a digital health intervention did or did not have an impact on health outcomes. Finding studies that are evaluating health outcomes of digital health interventions and that are using methodologies that differ from the standard established methodologies are rather the exception. However, for implementers of digital health, it may also be worth to look beyond the standard and to consider new or adapted methodologies, which is the rationale for the development of the algorithm in Article 2. The algorithm provides guidance as a snapshot on the current digital health evaluation landscape. New technologies, interventions, or methodologies will provide potential for updating the algorithm to incorporate developments.
7.3. Disparate Experience, Knowledge, and Evidence

Article 3 emphasizes the importance of global knowledge transfer, not only from the north to the south, but also vice versa, and, more generally, in every direction. Reverse innovation describes the transfer of innovative processes or technology from developing countries to developed countries. Examples of its application range from business models to global health. In many multi-national companies and institutions, innovation is recognized as indispensable for their sustainability, but innovations, including innovations in digital health, coming from developing countries have often been neglected, as these tend to involve less technological breakthroughs. But they do present novel and innovative combinations of existing knowledge and technologies addressing local problems using innovative processes and business models. Reverse innovation has the potential to redistribute wealth and to accelerate the development of low-income countries, and applied to health it has the potential to generate effective global solutions. Reverse innovation has the potential to significantly contribute to the post Millennium Development Goal era in the Sustainable Development Goal number three of improving health for all. Non-Communicable Diseases (NCDs) for example are prevalent and posing a major burden in all income-level countries and innovations from a developing country context might help achieving this goal, especially as a NCDs have a disproportionate impact on poor or socially marginalized groups in wealthy nations. It is encouraging to see health leaders turning to technology for innovative solutions that will not only drive efficiencies from a cost perspective, but also create a positive outcome on patient care [108]. These positive examples emphasize the need for global exchange and connection, of information, knowledge, experience and evidence, particularly as the variety and exponential increase in volume of information and data in the domain of digital health aggravate finding relevant information or data. Article 4 refers to these developments as “Implementomics”, which describes the ability to encapsulate, organize and exploit the multidimensional knowledge that is related to implementation issues. Similar to other “omics” domains, a crucial challenge is to grasp the variety and volume of information. Knowledge models such as the RAFT annotation model presented in Article 4 can improve this. Annotations do not only explain and enrich a resource with observations, but are also a mean of transmitting and sharing ideas to improve collaborative work practices. One limitation of this model is that it might not be capable to capture the comprehensive complexity and detail of the annotated items, because the model is based on ontologies to simplify the complicated web of inter-related and resembling terms. As a result, a broader more general term might be assigned instead of a more detailed concept. Another limitation is rooted in the model itself, as it has been developed based on the experience and well-known processes within the network it might need to be extended in order to be applicable to the digital health domain in general.

7.4. Implementome

The perspective to advance this research further is to implement the tools, which have been developed in this thesis in the Implementome. The vision of the Implementome is: 1) to construct a multidimensional digital health implementation hub to expedite knowledge documentation and sharing, 2) to agglomerate and connect knowledge resources by various projects, initiatives and organizations, 3) to facilitate joint knowledge creation, and 4) to link various organizations, individuals, academia, policymakers, civil society, and other users to digital
health knowledge. Metadata annotated contents in the Implementome will create an interconnected knowledge system that enabling the user to navigate on multiple dimensions. This metadata will cover a broad set of elements, which is necessary to understand and process information at different levels of granularity. For example, an annotated study could contain information on the geography of the data collection, the study design, the intervention, the authors, outcomes, outputs or provide pointers to related information or studies. One challenge will be the number and quality of annotated entries and the dependability. Different strategies could be applied to add content to the Implementome, ranging from manual annotation by trained experts, to crowdsourcing, or hybrid strategies that include auto-harvesting of entries.

Crowdsourcing describes the delegation of tasks to a crowd of non-specialists and is a strategy to collect, analyze or share vast amounts of data, commonly via the Internet. It provides an opportunity for researchers to overcome typical barriers to data collection, such as ensuring wide-ranging geographic coverage and the maintenance of long-term projects. Citizen science is the involvement of citizen in scientific projects and is a subset of crowdsourcing. Even though these concepts are nowadays generally used via the internet these are not new as they have been used for decades as for example by citizen bird-watchers to better document the migratory patterns and distribution of a wide range of bird species (30). Millions of individuals around the globe support professional scientists in tasks that range from monitoring changes in local biodiversity to providing innovation and computing power for new drug development to identification of patterns or annotations.

While manual indexing by experts is more precise and trustworthy, it is also usually time-consuming and costly. An additional annotation strategy that could be used is automatic annotation. This strategy is widely used in genomics (31) and proteomics (32), because of the enormous amount of data that is shared and classified in these domains. Uniprot, for example, is a protein knowledge-based platform that uses two prediction systems the Unified Rule System (UniRule) and the statistical Automatic Annotation System (SAAS). These systems automatically annotate entries that were not reviewed in an efficient and scalable manner, in addition to their manual annotation. This strategy is also used to create semantically annotated corpus of clinical texts, which are drawn from the free text component of patient records, and which capture clinically significant information documented in these texts [109].

Pubmed is another well-known example of annotation, where MeSH terms are currently manually assigned by human indexers. However, there are efforts in extending the annotation strategies to include automatic indexing (33-35). Probabilistic models and other machine learning algorithms (33, 35) are run and tested on various sets of data to predict accurate MESH terms for documents. For example, MESH Now (34) is an integrated approach, which uses multiple strategies to generate a combined list of potential MeSH terms for a specific manuscript. While automatic classification currently still remains challenging and needs more research, the advance in artificial intelligence, particularly in Natural Language Processing and Deep Learning, could pave the way to more precise annotation.

In the future, the implementome could use a combination of these strategies with a hybrid annotation system that uses automatic indexing followed by manual curation.
7.5. Key Takeaways

7.5.1. The obstacles to measuring health outcomes of digital health are significant
The challenge is to demonstrate why a digital intervention leads to better health outcomes than the status quo or than other interventions. Proxy indicators presented in Article 1 may help overcome some of the challenges associated to measuring the impact of digital health on health outcomes.

7.5.2. Consider innovative evaluation approaches
Even though there is an increasing amount of literature on outcome measures of digital health, as outlined in Article 2, these studies rely on traditional clinical evaluation approaches. Some traditional clinical evaluation approaches inhibit rather than support measuring the full value of innovative digital health solutions. Therefore, implementers and researchers should be encouraged to consider innovative and emerging evaluation methodologies besides the traditional clinical evaluation methodologies. Article 2 proposes for example an algorithm that can help implementers to gain insight on new evaluation approaches to achieve more efficient evaluation processes. Removing such barriers and allowing for controlled and evaluated experiments with new technologies can also facilitate the identification and up-scaling of decentralized innovation, where some of the best services are being developed and implemented to strengthen health systems based on local needs as outlined in Article 3.

7.5.3. There is no one-size-fits-all approach in designing digital health evaluations
Digital health is an interdisciplinary, dynamic field bringing various scientific domains and implementation science together. Furthermore, application range is large and interventions are often integrated in a complex system. Therefore, these efforts require an evaluation design that is fit-for-purpose, striking a balance between what can be demonstrated in the project period and long-term goals, while leaving enough flexibility to address potential challenges, developments, and evolving technology in the long run. Usually the development phase of the digital health intervention as well as the implementation phase are crucial components, which might require a combinations of different evaluation types to collect relevant information for decision-makers at different moments as discussed in Article 2.

7.5.4. Experience is the best teacher, even when it’s not our own
Collaboration, partnerships and team-work are a fundamental part of digital health due to its inter-disciplinary nature. To develop sustainable digital health services there is significant value in learning from each other’s experiences, which applies to multilateral implementations, but also across disconnected implementations. These collaborative and bi-directional learning processes must be sought for and attended to with diligence. Knowledge transfer and exchange of expertise are key elements. However, typically, individuals, institutions or organizations might have bias or imprecise perspectives about other institutions, organizations or projects and connecting these meaningfully as e.g. described in Article 4 can help to break down these silos and accelerate impact across implementing organizations and sectors.
7.5.5. Connect the dots to fight pilotitis

In the field of digital health, hundreds of projects aim to tackle either a particular disease or the health challenges of a specific population. Because these siloed projects often struggle to reach critical mass, they have created a state of “pilotitis”, a disconnected series of implementations that fail to provide a holistic view of an individual’s health or the strength of the health system as a whole. In order to decrease this fragmentation decision makers and implementers need to move from a solution mindset to systems thinking. As discussed in Article 4, connecting the dots by annotating and linking experts, projects, data and information of these digital health implementations on a global scale in the Implementome (Article 4), an interconnected knowledge system, could facilitate the navigation across multiple information dimensions and could promote greater transparency and accountability, reduce redundancies, optimize resource allocation, and ultimately accelerate impact for organizations with shared objectives and advance systemic change to improve outcomes.
8. Bibliography


43. iHRIS. iHRIS Software: Track, Manage, and Plan Your Health Workforce [Available from: https://www.ihris.org/.


60. Union IT. Measuring the information society 2017: International Telecommunication Union; 2017. 9261138012
61. ITU. ICT for improving information and accountability for women’s and children’s health International Telecommunication Union; 2013.
71. Khoja S, Durrani H. Introduction to mHealth-focused issue on evidence-based eHealth adoption and application. AME Publications; 2017.
104. Perrin C, Hounga L, Geissbuhler A. Systematic review to identify proxy indicators to quantify the impact of eHealth tools on maternal and neonatal health outcomes in low-income and middle-income countries including Delphi consensus. BMJ open. 2018;8(8):e022262. PMID: 30121608.
