Evaluation Approaches of Digital Health Technologies: A Scoping Review

Anneloek Rauwerdink, Pier Spinazze, Harm Gijsbers, Juul Molendijk, Sandra Zwolsman, Marlies P. Schijven, Niels H. Chavannes, Marise J. Kasteleyn

Submitted to: Journal of Medical Internet Research
on: June 24, 2023

Table of Contents

Evaluation Approaches of Digital Health Technologies: A Scoping Review

Anneloek Rauwerdink[1*] MD; Pier Spinazze[2*] MD, MSc, MBA; Harm Gijsbers[3, 4]; Juul Molendijk[5] MD; Sandra Zwolsman[3] PhD; Marlies P. Schijven[3, 4, 6] MD, MSc, MHsc; Niels H. Chavannes[2, 7] MD, PhD; Marise J. Kasteleyn[2, 7] PhD

[1]Dept. of Radiology and Nuclear Medicine Amsterdam UMC Amsterdam NL
[2]Dept. of Public Health and Primary Care Leiden University Medical Centre Leiden NL
[3]Amsterdam Public Health Digital Health Amsterdam NL
[4]Dept. of Surgery Amsterdam UMC Amsterdam NL
[5]Amsterdam UMC Amsterdam NL
[6]Amsterdam Gastroenterology Endocrinology Metabolism Amsterdam NL
[7]National eHealth Living Lab (NeLL) Leiden NL
[*]these authors contributed equally

Corresponding Author:
Anneloek Rauwerdink MD
Dept. of Radiology and Nuclear Medicine
Amsterdam UMC
Meibergdreef 9
Amsterdam
NL

Abstract

Background: Profound scientific evaluation of novel Digital Health Technologies (DHT) is key to enhance successful development and implementation. In such, we developed the eHealth evaluation cycle in previous research. The eHealth evaluation cycle contains five consecutive study phases: conceptual, development, feasibility, effectiveness, and implementation.

Objective: The objective of this study was to create a visual overview of DHTs' evaluation approaches used in original research and to determine to what extent the study phases of the eHealth evaluation cycle have been utilised.

Methods: We conducted a systematic literature search in PubMed including the MeSH term 'telemedicine' in combination with a wide variety of evaluation approaches. Original studies from 2019 (pre-COVID-19 cohort) were included. Data on the following variables were extracted and systematically analysed: journal, country, publication date, medical specialty, primary user, functional classification, evaluation study phases, and evaluation approach.

Results: 824 studies were included after 1583 titles and abstracts were screened. The majority of the evaluation studies focused on the effectiveness (impact) (36.9%) study phase, whereas uptake (implementation) (8.5%) was the least. The RCT (19.0%) was the most commonly used DHT evaluation method. Within the effectiveness (impact) study phase, the RCT was used in half of the studies. In the conceptual and planning phases, survey research (34.6%) and interview studies (34.6%) were most frequently used. The United States published the largest amount of DHT evaluation studies (36.9%). Psychiatry / mental health (10.6%) and cardiology (8.9%) published the majority of the studies within the field.

Conclusions: We found that the study phases of the eHealth evaluation cycle are in equally studied. Also, the majority of the studies in the effectiveness study phase still uses a RCT design. However, in order to successful develop and implement novel DHTs, stimulating equal evaluation of the sequential study phases of DHTs and selecting the right evaluation approach that fits to the iterative nature of technology, might be of utmost importance.

(JMIR Preprints 24/06/2023:50251)
DOI: https://doi.org/10.2196/preprints.50251

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?
✓ **Please make my preprint PDF available to anyone at any time (recommended).**

Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**
Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v
Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Introduction

Background

Healthcare, traditionally a slow adopter of digital innovation, has recently seen an acceleration in the use of digital tools and systems. A transition to more remote patient care and associated services is urgently needed due to an ageing global population [1]. In a highly regulated sector where data privacy and health outcomes are vital, there needs to be a high level of scrutiny and standardised evaluation frameworks in place to limit the inherent risk in rapid innovation and ensure successful outcomes. With an increasing focus on digital health research, it is important to take stock of digital health evaluation methodologies to further guide prospective studies [2].

The World Health Organization (WHO) plays an important role in guiding and accelerating the development of digital health interventions and health innovation globally. In 2018, the WHO developed a *classification scheme of digital health interventions*, aiming to promote a comprehensive and standardised language for health program planners [3]. The scheme organised digital health interventions into the following primary user groups: clients, healthcare providers, health systems or resource managers, and data services (Multimedia Appendix 1).

Another valuable framework was developed by the National Institute for health and Care Excellence (NICE) together with the National Health Service (NHS) England in 2019 [4,5]. They created the *Evidence Standards Framework* (ESF) for Digital Health Technologies (DHTs) in order to promote greater consistency when evaluating or commissioning DHTs and thereby enhance the level of scrutiny, which is generally lower than the level of evidence required for drugs or devices. The subjects of NICE's ESF are: system services (evidence tier 1); Inform, simple monitoring, communicate (evidence tier 2); preventative behaviour change, self-manage (evidence tier 3a); and treatment, active monitoring, calculate and diagnose (evidence tier 3b).

The WHO's digital health classification scheme and the NICE's ESF could work synergistically, providing a much needed standardised and accepted framework for the varied stakeholders involved in digital health to evaluate and improve the development of evidence-based digital health solutions. To provide further granularity to the focus of digital health research, evaluating the sequential study phases - conceptual to implementation – is important [6,7]. Therefore, we previously developed the *eHealth evaluation cycle* (Figure 1), based on existing eHealth evaluation frameworks [8]. The *eHealth evaluation cycle* contains the following five consecutive study phases: conceptual &

planning, design, development & usability, pilot (feasibility), effectiveness (impact) and uptake (implementation). In the online version of the *eHealth evaluation cycle* one can find a synopsis of evaluation approaches (all the methods, study designs, frameworks, and other philosophical approaches) that could be used to evaluate a particular study phase [9]. For example, a 'concept mapping study design' can be used to gather information in the conceptual & planning study phase and an 'economic evaluation' can be used in the uptake (implementation) study phase [10]. There are also several types of systematic reviews within the *eHealth evaluation cycle*. For instance, a meta-analysis can be found within the effectiveness (impact) phase, and a narrative review can be found in the conceptual & planning phase when one is for example at the start of developing a new DHT and aims to define what is already known.

Figure 1: eHealth evaluation cycle

Although the number of DHT publications has exponentially grown in recent years, not much is known about the use in daily practise of the consecutive study phases [11]. There are indications that there is too much focus on evaluating *effectiveness* instead of, for example, usability testing [12]. Also, large-scale implementation of DHT often fails [13–15]. Therefore, in order to potentially improve DHT evaluation and so implementation of DHTs, the following research question will be addressed in the presented study: What is the actual practice of the consecutive evaluation study phases described in literature?

Aim and Objective

The aim of this study is to develop a better understanding of the daily practice of the *eHealth evaluation cycle*. Therefore, the research objectives are:

- To conduct a structured analysis of literature data to analyse the practice of the evaluation study phases.
- To determine which evaluation approaches are used in which study phase of the eHealth evaluation cycle.

Methods

Overall design

We performed a scoping review subdivided into two phases: 1. systematic literature search to find articles concerning the evaluation of DHT, followed by the extraction of data from the selected articles, and 2. performing a structured data analysis. We followed the PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines

Search strategy

The online database of PubMed was systematically searched using the MeSH term ‘telemedicine’ in combination with a wide variety of ‘evaluation approaches’ (see Multimedia Appendix 2 for a complete list of the search string). Articles written in English published in the year 2019 were included. The year 2019 was chosen since it is the most recent ‘pre-COVID-19’ year cohort. Herewith, we aimed to avoid possible skewing of results by the temporary surge in COVID-19 related DHTs.

Inclusion and exclusion criteria

The following inclusion and exclusion criteria were used:

Inclusion:

- Any original, peer-reviewed study evaluating a specific DHT in one or multiple study phases of the eHealth evaluation cycle.
- All types of literature reviews describing a specific DHT.
- Tele-education studies were included when closely related to the outcomes of health care.

Exclusion:

- Non-English written papers.
- Studies not related to DHTs.
- Non patient-focused studies (e.g. evaluation of technology alone, i.e. sensitivity of sensors) were excluded.
- Poster presentations, protocol studies and opinion papers.

Study selection

Articles were included based on a screening of title and abstract performed by two pairs of researchers. The authors (HG and PS) reviewed articles with titles starting with the letters A through K and (AR and JM) reviewed articles with titles starting with the letters L through Z of the search results. Two test series of twenty cases were discussed between the four researchers to reach a satisfactory level of consensus before the pairs began screening the titles and abstracts. Rayyan QCRI online software was used to be able to collectively screen articles [16].

Once the inclusion screening was complete, each pair discussed conflicting studies. Disagreements between researchers were resolved by the decision of a third researcher (HG or AR). The two sets of included studies were merged and randomly divided into four individual sets. Each of the researchers received one dataset to extract the data from (described in the data extraction section). Most of the data could be extracted from the abstract alone. Therefore, an article's full text was only reviewed in the event that the information in the abstract was insufficient. If the researcher thought the article was not suitable for inclusion in the study after a more in-depth review, it was excluded.

Data extraction

Data extraction was done using a standardised data extraction form (see example in Multimedia Appendix 3) that was developed by AR and HG using MS Office Excel version 16.52. The form was pilot tested by the group of four researchers (AR, HG, JM and PS) on 10 articles and modified afterwards. The data items that were extracted from the articles are described in textbox 1. The variable *functionality* was based upon the NICE Evidence Standards Framework for DHTs and the variable *primary user* was based upon the WHO's digital health classification scheme.

Textbox 1

Generic variables
Journal
Country[i]
Publication date

Medical speciality[II]	
Specific variables	*Categories*
Primary user[III][3]	Clients
	Healthcare providers
	Health system or resource managers
	Data services
Functionality[IV][17]	System services
	Inform
	Simple monitoring
	Communicate
	Preventative behaviour change
	Self-manage
	Treatment
	Active monitoring
	Calculate
	Diagnose
	Multiple functionalities
	Other
	Unclear
Evaluation study phases[V][8]	Conceptual & planning
	Design, development & usability
	Pilot (feasibility)
	Effectiveness (impact)
	Uptake (implementation)
	Multiple phases
Evaluation approach	*Free text field*

[I]Country in which the study was conducted, or if this was not clear the country of the first author. [II]Medical specialty the DHT applies to, multiple options possible. A standardised list was used. [III]Categories of the WHO classification scheme of digital health interventions were extracted, multiple categories may be extracted.

[IV]Catogories of the NICE's evidence standards framework were extracted. [V]Categories of the eHealth evaluation cycles were extracted

The description of the *evaluation approach* was literally extracted by copy pasting from the article. If available, multiple evaluation approaches were extracted.

The four researchers (AR, HG, JM and PS) independently extracted the data into the standardised data extraction form. After the data extraction was completed, AR and HG each randomly checked 15 cases from each investigator to evaluate for inconsistencies amongst the extracted data. When a

data item showed an inter-rater disagreement of more than 10%, a second investigator extracted the data of the specific data item, and a third investigator made the final decision on the inequalities. Considering the extensive workload, this extra step of cross-checking data was performed unblinded. The researchers group discussion was planned ahead of the second researchers' data extraction to increase the level of consensus. Finally, the completed data extraction forms were merged again, checked by AR and altered for consistency. The description of each of the extracted evaluation approaches was checked and substituted by a better general description to allow for statistical analysis, for example 'randomized controlled trial' was changed into 'RCT'

Data synthesis and analysis

RStudio vs. 2022.07.1 software was used to summarise the descriptive data and to perform statistical analyses. The proportions of the categories of each data item were described using percentages. The variable *evaluation approach* was extracted as a free text field in the data extraction form. It was expected that the data extraction for the variable *evaluation approach* would yield a very wide array of different evaluation approaches. Therefore, to design the most clear visual presentation of the data, only the top 8 (cross-tabulation bar chart) and top 10 (bar chart) most frequently used *evaluation approaches* were used for further analysis. Concerning the variable *medical specialty*, the top 5 (cross-tabulation bar chart) and top 10 (bar chart) were used for further analysis. For the variable *country,* only the top 10 (bar chart) were used for analysis.
To examine relationships between the nominal variables, the following cross-tabulations were executed: *primary user* vs *functionality*, *primary user* vs *evaluation study phase*, *primary user* vs top 8 *evaluation approach*, *functionality* vs *evaluation study phase*, *functionality* vs top 8 *evaluation approach*, *evaluation study phase* vs top 5 *medical specialties*, *evaluation study phase* vs top 8 *evaluation approach*, and top 5 *medical specialty* vs top 8 *evaluation approach.*

Results

After the PubMed database search, 1583 studies were considered relevant according to the inclusion criteria. During the screening of the title and abstract, 716 records were excluded. A further 43 articles were excluded after reading the articles more in depth during the data extraction assessment phase. Finally, 824 studies were included. The PRISMA flowchart summarises the article selection process (Figure 2).

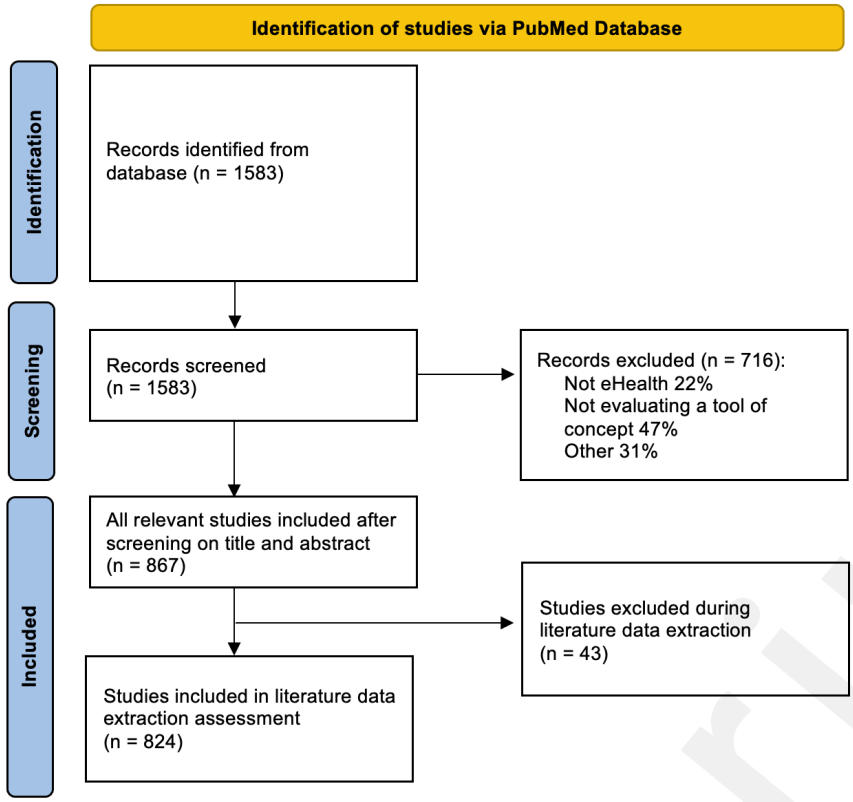


Figure 2: Prisma Flowchart

The independent evaluation of 15 records by each of the researchers AR and HG, revealed unacceptable inconsistencies in the extracted data of the variables *functionality* (29%) and *evaluation study phase*s (21%). Following a discussion among the four researchers, it was decided that the variables *functionality* and *evaluation study phase*s would require a second review. A third reviewer made the final decision on the discrepancies between the two reviewers. All of the researchers participated in the second review, and all made final decisions as a third reviewer.

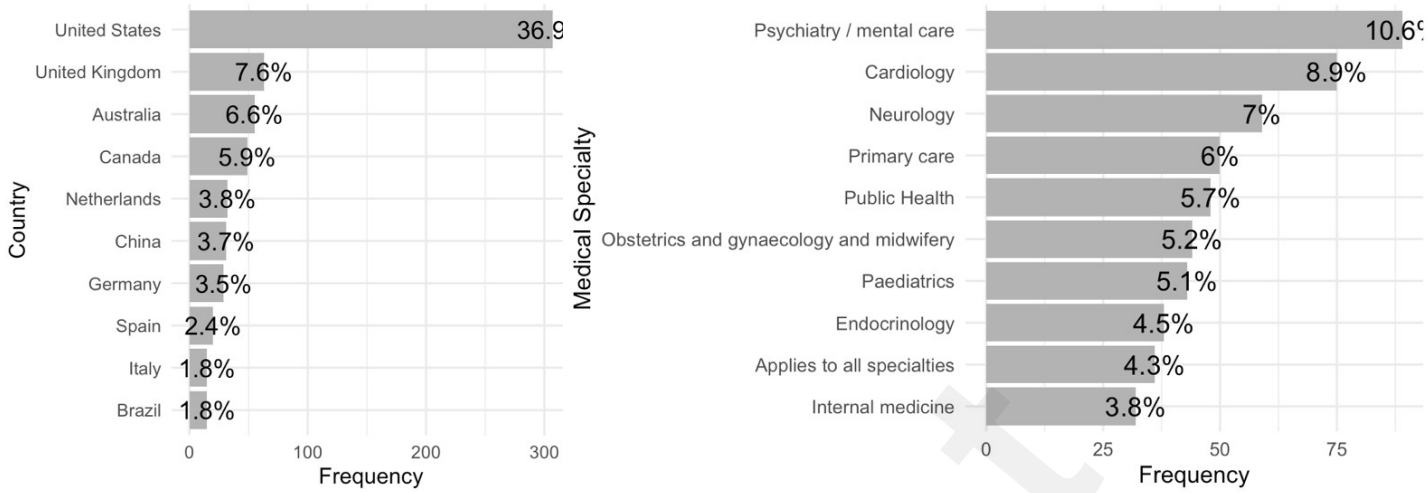


Figure 3: Bar charts country top 10 (total n = 73) and medical specialty top 10 (total n = 44)

The descriptive statistics of the top 10 general variables' *country* and *medical specialty* are illustrated by the bar charts in figure 3. The complete frequency list of the variables *country*, *medical specialty*, and publishing *journals* from 5 studies and up can be found in the Multimedia Appendix 4. Table 1 shows the descriptive statics of the specific variables. The majority of DHT-related research is published in the field of psychiatry/mental health (10.6% (89/840)) and cardiology (8.9% (75/840)). Concerning DHTs *primary user* categories, clients (51.8% (491/947)) are slightly more frequently targeted than healthcare providers (44.9% (425/947)). Communication (19.4% (160/824)) and treatment (18.2% (150/824)) are the largest categories in the *functionality* variable.

Table 1. Descriptive statics of specific variables

Specific variable	**Category**	**Frequency *n* (%)**
Primary user	Clients	491 (51.8)
	Healthcare providers	425 (44.9)
	Health system or resource managers	26 (2.7)
	Data services	5 (0.5)
	Total	947 (100)
Functionality	Communicate	160 (19.4)
	Treatment	150 (18.2)
	Multiple	103 (12.5)
	Diagnose	87 (10.6)
	Self-manage	78 (9.5)
	Active Monitoring	67 (8.1)
	Preventative behaviour change	61 (7.4)
	Inform	49 (5.9)
	Simple monitoring	29 (3.5)
	System services	26 (3.2)
	Calculate	6 (0.7)

	Unclear	5 (0.6)
	Other:	3 (0.4)
	Total	824 (100)
Evaluation study phases	Effectiveness (impact)	304 (36.9)
	Pilot (feasibility)	232 (28.2)
	Design, development & usability	99 (12)
	Conceptual & planning	96 (11.7)
	Uptake (implementation)	70 (8.5)
	Multiple phases	23 (2.8)
	Total	824 (100)
Evaluation approach top 10 *Total distinct evaluation approaches listed = 108*	RCT	170 (18.9)
	Survey Research	91 (10.1)
	Cohort study (prospective)	73 (8.1)
	Interview study	58 (6.5)
	Cohort study (retrospective)	54 (6)
	Mixed method study design	42 (4.7)
	Systematic review	39 (4.3)
	Cross-sectional study	36 (4)
	Feasibility study	31 (3.4)
	Pilot study	25 (2.8)
	Total	899 (100)

The variable *evaluation study phase* shows that nearly 37% (304/824) of the studies were carried out to study the effectiveness of a certain DHT tool. Almost a third of the studies were in the pilot study phase. The RCT was found the most frequently used *evaluation approach*, 19% (171/899) of the studies used a RCT. In total, 108 distinct *evaluation approaches* were encountered (Multimedia Appendix 5). Although the top 10 consists of well-known epidemiologic methods, we did encounter novel evaluation methods such as the "Fit between Individuals, Task and Technology (FITT) framework [18], "Nonadoption, abandonment, scale-up, spread, and sustainability (NASSS) framework [19], "CeHRes Roadmap [20], and "Systems Development Life-cycle (SDLC) methodology [21]. Also, several variants of well-known epidemiological methods, such as a non-randomized group comparison study[22] and a retrospective record review [23] were identified.

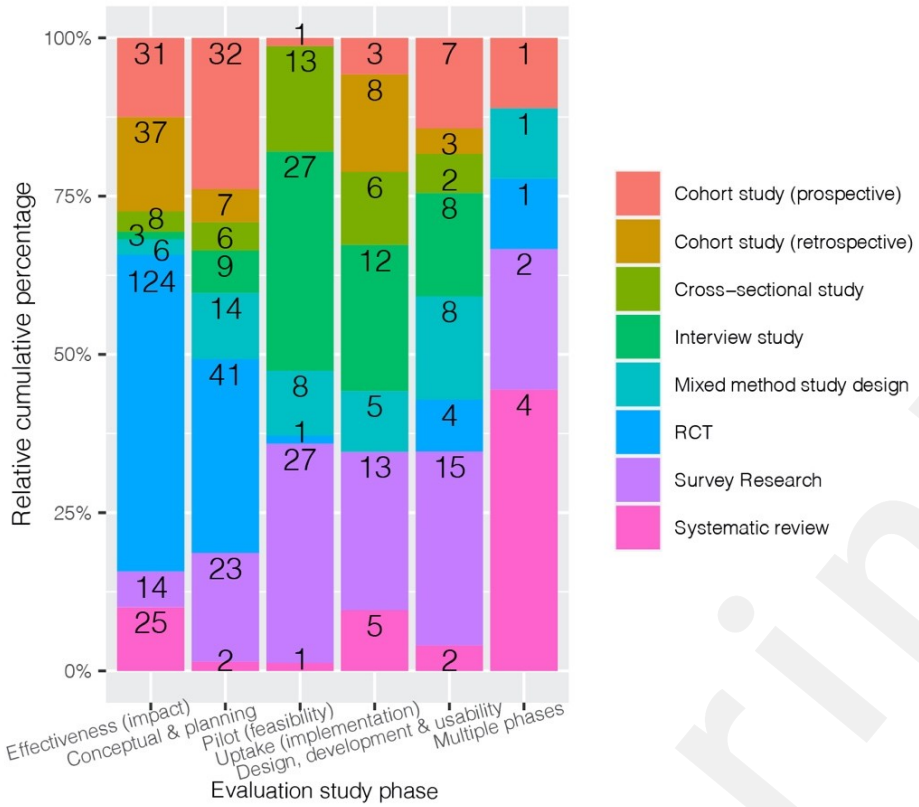


Figure 4: Bar chart evaluation study phase vs. top 8 evaluation approach

The proportional stacked bar chart in Figure 4 illustrates the cross-tabulation *evaluation study phase* vs. *evaluation approach* top 8. In the effectiveness study phase, RCTs were used in half of the studies (50.0% (124/248)). Cohort studies, prospective and retrospective, were used in more than a quarter of the effectiveness studies (27.4% (68/248)). Survey research (34.6% (27/78)), and interview studies (34.6% (27/78)) were more commonly used for the conceptual & planning study phase. In the uptake (implementation) study phase, the evaluation approaches used were generally equally divided.

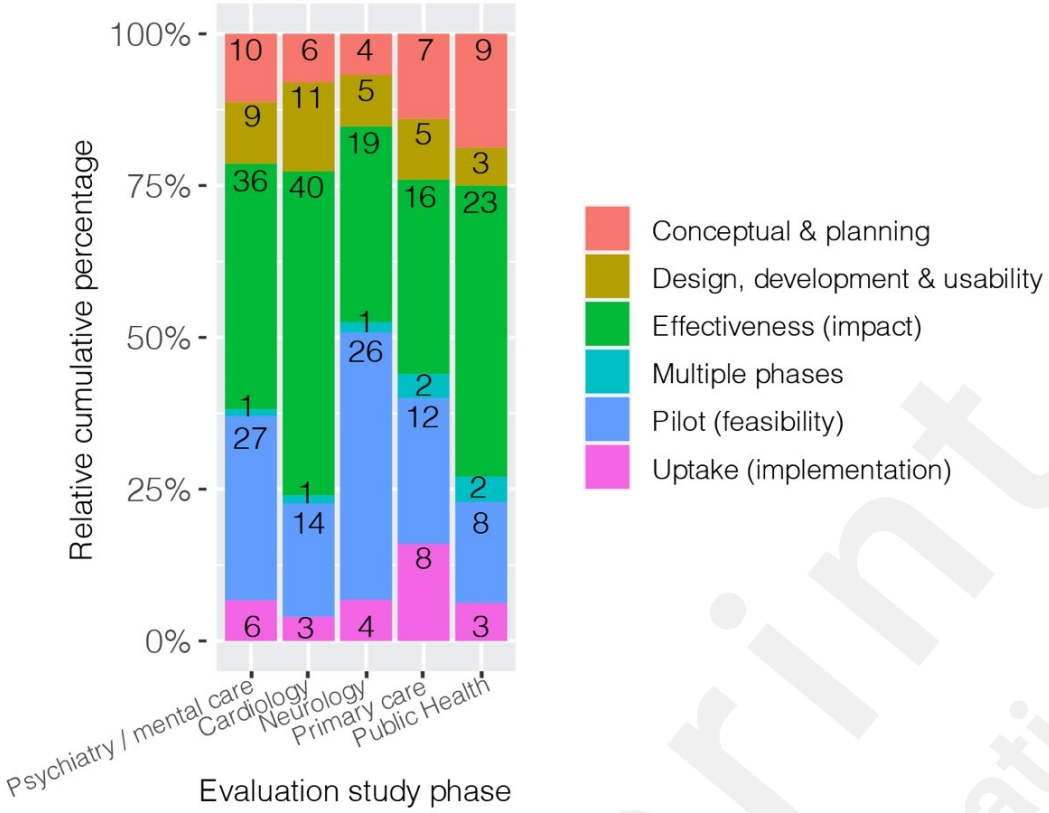


Figure 5: Bar chart evaluation study phase vs. medical specialty

The proportional stacked bar chart of the cross-tabulation *evaluation study phase* vs. the *top 5 medical specialties*, illustrates the differences in emphasis of the selected evaluation study phases between the *medical specialties* (Fig. 5). Cardiology had the biggest number of studies performed in the effectiveness (impact) study phase (53.5% (40/75)) and a lesser share in the pilot study phase (18.6% (14/75)). Within the medical specialty of neurology, the opposite was found; there was more focus on the pilot study phase (44.1% (26/59)) and less on the effectiveness study phase (32.2% (19/59)). Primary care had the most evenly divided chart, with the study phase uptake (implementation) (16.0% (8/50)) appearing to be of greater importance when compared to other specialties.

Further, when looking at the proportional stacked bar chart of the cross-tabulation *medical specialty* vs. the *top 8 evaluation approach*, the chart shows differences concerning medical specialties and the depicted methodology (Fig. 6). The RCT represents the biggest share within all medical specialties, with public health as the front-runner (46.7% (14/30). Again, primary care has the most equally divided chart when compared to other specialties.

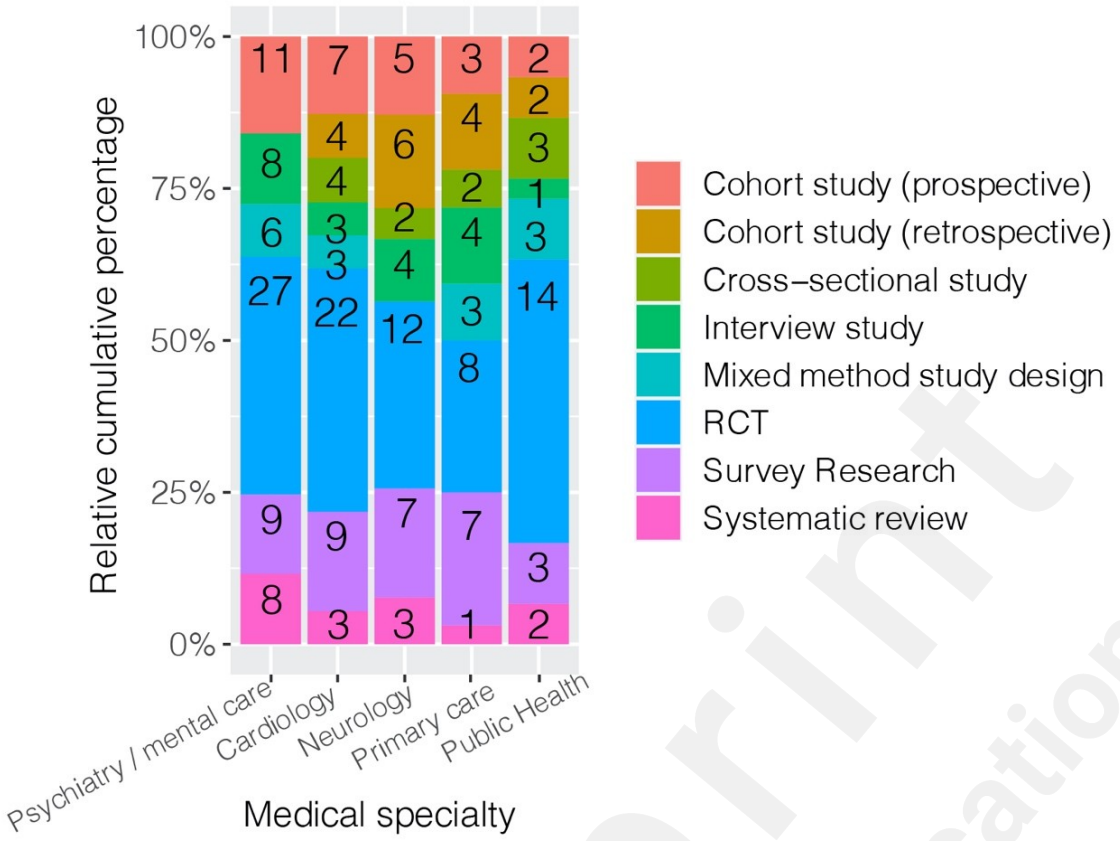


Figure 6: Bar chart medical specialty vs. top 8 evaluation approach

The cross tabulation of *functionality* vs. *evaluation study phase* focuses on the breakdown of the study phase in relation to the assumed DHTs function (Fig. 7). The effectiveness (impact) study phase was most frequently found in the *functionality* category treatment (50% (75/150). The pilot (feasibility) study phase was most dominant in the *functionality* category diagnose (46.0% (40/86)). In almost all categories of the *functionality* variable, significantly less attention was paid to the conceptual & planning, design, development & usability and uptake (implementation) study phases.

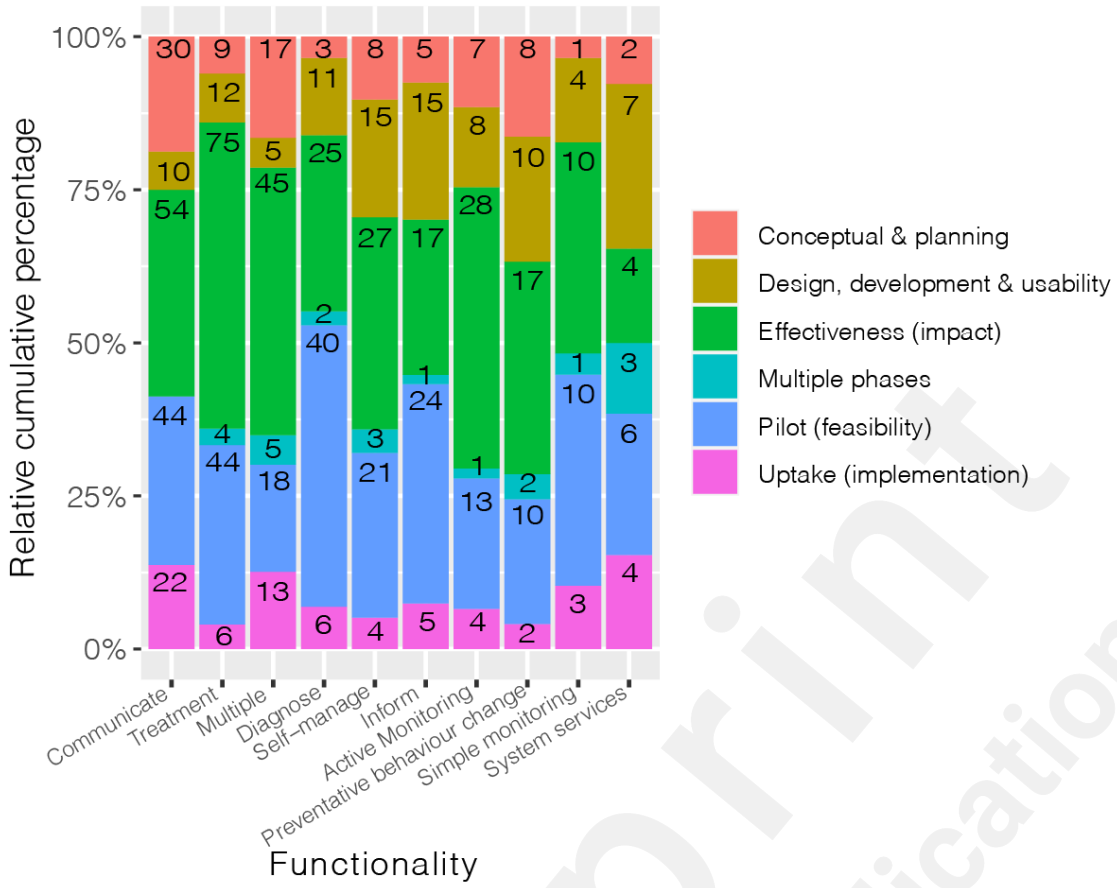


Figure 7: Bar chart functionality vs. evaluation study phase
NB: functionality categories 'unclear', 'calculate' and 'other' were left out due to limited results

The cross tabulations of *primary user* vs. *functionality*, *primary user* vs. *evaluation study phase*, *primary user* vs. the *top 8 evaluation approach*, and *functionality* vs. the *top 8 evaluation approach* resulted in an almost equal distribution of the categories. The frequency tables of all the performed cross tabulation analyses, as described in the methods section, can be found in the Multimedia Appendix 6.

Discussion

The aim of this study was to develop a better understanding of the daily practice of the eHealth evaluation cycle. To our knowledge, we composed the first comprehensive overview of the actually practice of the consecutive DHT evaluation study phases. We summarized our main findings in the infographic in Figure 8.

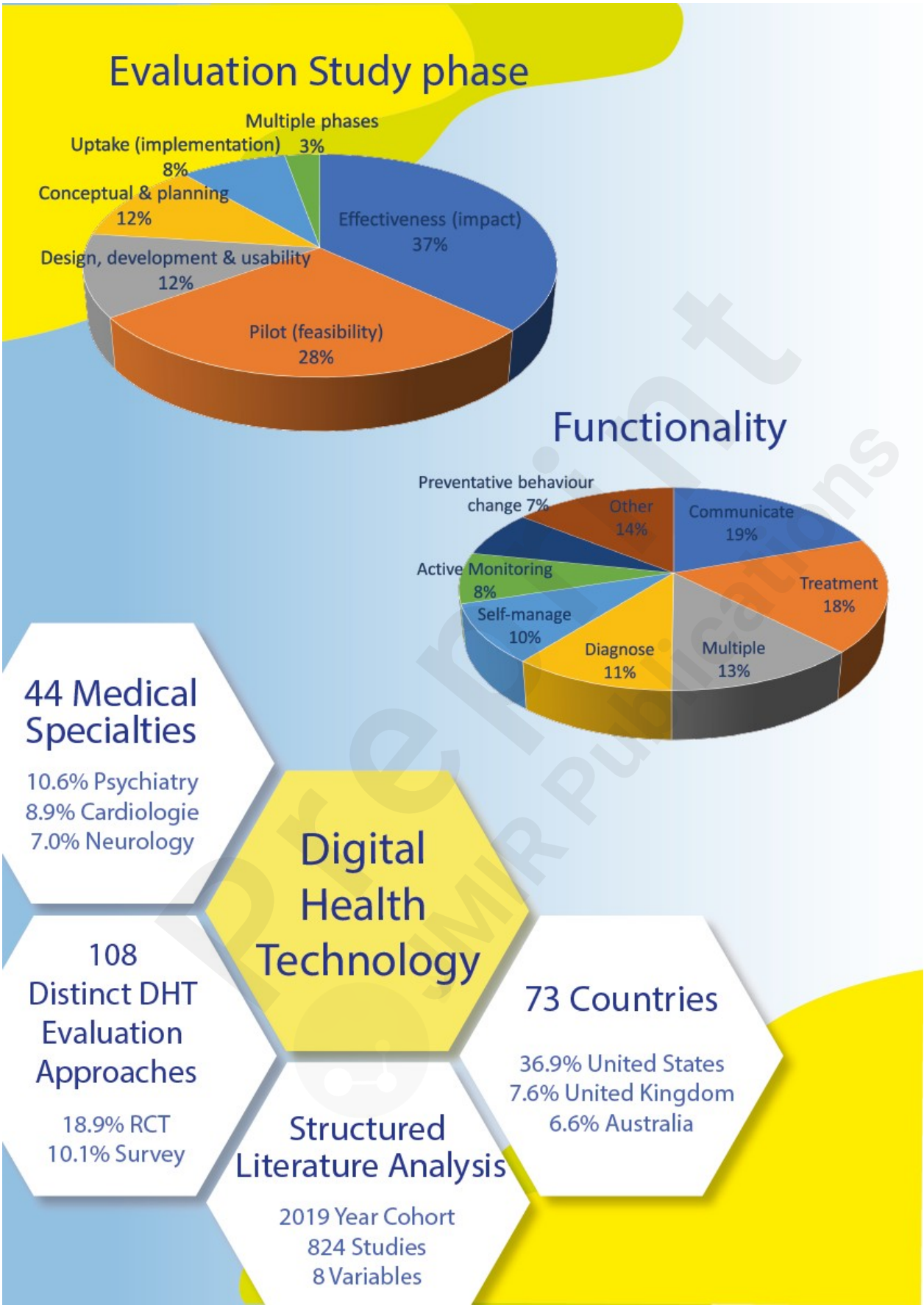


Figure 8: Visual summary of main findings.

Evaluation study phase and evaluation approach

Our study highlighted disparities in the attention given to the consecutive evaluation study phases. A predominant focus on effectiveness and pilot study phases was found. The uptake phase, crucial for successful implementation and scale-up, received the least emphasis. Also, less focus has been placed on the planning and development study phases. This fundamental mismatch between the context and the technology is the main reason recognised by the World Health Organization (WHO) that up to 75% of new medical devices fail [24]. A recently published original study from Royle, JK et al. also aimed to created awareness for lack of evaluation of the methodological steps necessary for developing and testing new clinical care pathways. They proposed 'the technology clinical trial' [25], describing: 1. co-creation of care pathways and addressing information governance, data protection, regulatory, and ethical questions by design; 2. delivery of the technology clinical trial and 3. supporting future research and uptake in practice. The three-step methodology is comparable to the *eHealth evaluation cycle*; however, the relation of the study phase to DHT evaluation approaches is not described.

The RCT emerged as the most frequently used evaluation approach, posing challenges in the context of rapid and iterative DHT development [26]. We did, however, encounter studies using RCT 'sub-designs' such as Sequential Multiple Assignment Randomized Trial (SMART) and micro-randomized trials (Multimedia Appendix 5). These designs allow for limited interventional modification within pre-set boundaries and therewith might bridge the methodological gap. Recently, Hrynyschyn, R et al. conducted a scoping review aiming to provide an overview of existing evaluation methods of DHTs beyond the RCT [27]. They described in detail the micro randomization trial, (fractional) factorial randomised controlled trial, Sequential Multiple Assignment Randomisation Trial (SMART), and stepped-wedge cluster randomized trial as promising alternatives.

In the conceptual & planning study phase, survey research and interview studies were each used in one third of the cases. Although these findings seem obvious, as formative research approaches apply well to the earlier study phases of a development cycle, the importance of thourough 'background' research in relation to successful implementation cannot be emphasized enough.

Functionality *(NICE)*

In this study, the *functionality* categories communication and treatment of the NICE's ESF were most frequently studied [17,28]. In the *functionality* category treatment, half of the studies were focused on the effectiveness (impact) study phase. Concerning the functionality category 'diagnose', the

majority of the studies were found in the pilot (feasibility) study phase. The DHTs in the 'treatment' and 'diagnose' classes are both considered 'Tier 3b' according to the ESF. This implies that the designed DHT must demonstrate effectiveness through a high-quality intervention study (experimental or quasi-experimental design) showing improvements in relevant outcomes. Therefore, it seems to make sense that the majority of the studies focus on the pilot (feasibility) and effectiveness (impact) study phases. Unfortunately, there seems to be too little attention paid to the pre-development and post-deployment phases, which are also not included in the NICE's ESF. DHTs are often developed by commercial entities, which then need validation through research to allow for regulatory clearance by governing bodies (e.g., the FDA) and subsequent adoption and commercialisation. Hence, they may only pursue research following the development of a DHT, allowing for speed and rapid prototyping unhindered by slow academic rigour. Often, the development phase involves some form of user testing and prototyping with feedback; however, the value in publishing this is less pronounced and perhaps even avoided to protect IP and avoid competition before commercialisation. Despite the time limitations and stakeholder involvement required, perhaps greater attention should be placed on evaluating DHTs focused on treatment in the earlier design and conceptual phases to improve success and uptake at later phases (in terms of costs, usefulness, adoption, etc). The CeHRes roadmap is one proposed approach to the development of eHealth interventions. It incorporates both a human-centred design and a business model focus, to create potentially value-adding and sustainable eHealth technologies [29].

Primary User (WHO)

The target end-users of studies, as described by the WHO, were primarily focused on clients and healthcare providers. Health systems, or resource managers, and data services accounted for less than 4% of studies. This might be due to the limitation of the search to the clinical database PubMed. For example, the IEEE database perhaps would have yielded more relevant results. However, it also might suggest that the majority of technological innovation and implementation is focused on the providers and clients, i.e., at the point of care rather than at the system and infrastructure level. This is understandable, considering that health organisations have a level of complexity and fragmented structures that constrain their ability to adopt organisation wide digitisation [30,31]. In such, a recent study describes how to improve and manage sustainable hybrid (eHealth and face-to-face) health care on an organisational level through the use of the Hybrid Health Care Quality Assessment [32]. The other consideration regarding target-end users is the rate of adoption, which is a big contributor to successful implementation and uptake of health technology. Providers remain, to a large extent, the gatekeepers of this, as healthcare professionals' acceptance is reported as an important need to

the success of the clinical systems [33]. This could explain why the majority of DHTs and their evaluation studies are focused on the providers and clients.

Countries and medical specialties

The United States published the largest number of DHT evaluation studies. The number of studies per country aligns with the overall ranking of health publications in general per country being the United States, United Kingdom, Australia, and Canada. However, China, which usually features in third place, featured slightly lower in sixth, and the Netherlands significantly higher in fifth from 13th place [34]. These discrepancies may be explained by the fact that our study excluded Chinese databases or non-English papers. The Netherlands' numbers suggest that there is a greater attention to eHealth evaluations in general. According to the Healthcare Information and Management Systems Society (HIMSS), the Netherlands is one of the frontrunners in the digitalisation of healthcare [35].

Psychiatry/mental health was the most frequently encountered medical specialty involved in DHT research, followed by cardiology and neurology. A 2018 US-based weighted survey on the adoption of DHT by physicians in different specialties, showed a varied spectrum of adoption rates. Consistent with the data we collected in this scoping review, 27.8% of psychiatrists and 24.1% of cardiologists used DHT for patient interactions [36]. DHTs for mental health have seen a proliferation considering inherent benefits such as anonymity, accessibility and acceptability, possibly explaining why this is the leading speciality in DHT research studies [37].

When relating the medical specialties to the evaluation study phases, the medical specialties of cardiology and neurology showed an interestingly opposing result. Of the included cardiology studies, half were in the effectiveness phase, and less than a fifth were in the pilot study phase. As for neurology studies one third of the studies were in the effectiveness phase and almost half were in the pilot study phase. As such, perhaps there is less focus on the pilot phase within cardiology and fewer neurology studies that progress to the effectiveness study phase. This could also be an indication of the stage of technological development and application to each speciality and the year of publication we evaluated (2019). For example, technology has been used cardiology since the first implantable pacemaker in 1958, with studies on remote monitoring, Virtual Reality etc. surfacing between 2000 and 2010 and have sustained considerable scholarly attention ever since [38,39]. Whereas in neurology, for example there are very few papers on fall detection using IOT before 2010 as this is a relatively newer field of study [40].

Limitations

Although our scoping review was thorough, there are some limitations to mention. We only sought in

the PubMed database and only English, peer-reviewed literature was included, which excludes studies based in foreign contexts (e.g. Chinese as well as grey literature) which is employed by a large number of commercial institutions involved in the development of DHTs. However, this would have had questionable value regarding methodology and scientific rigour. Only papers published in 2019 were considered in the data extraction step. The decision to include studies from 2019 was made in order to have the most up-to-date data available, without considering the impact of the COVID pandemic on DHT research. Because we only extracted data from studies published in 2019, we might have missed novel evaluation approaches. Further, a possible shift in used evaluation approaches over time cannot be evaluates. It would be interesting for a future study the compare e.g. a 2015 vs. 2019 vs. 2023 cohort.

Due to unacceptable discrepancy between the researchers included data in the data extraction phase of the variables *functionality* and *evaluation study phase*, a second unblinded cross-checking was performed to facilitate concordance, which may have introduced a level of bias. Further, on data extraction, the reviewers standardised the terminology for evaluation approaches, which is open to a level of bias in interpretation. However, this should be minimal considering research uses common terminology in general.

Unfortunately we were not able to include a variable describing the DHT, e.g. ‘patient portal’ or ‘mHealth’. In the majority of the included studies the description of the DHT varied too much. Therefore, more in-depth (sub)analyses concerning the type of DHT were not possible. To our opinion, a well though through and easy to use DHT taxonomy would be of great help to classify and evaluate future DHT research. Lastly, the NICE evidence standards framework for DHT was updated in August 2022. However, for the majority the subgroups of our study are still used in the NICE framework.

Future

It would be valuable to track which technologies are evaluated in a single or multiple study phases. However, to do this, each technology would require a unique identification that would allow traceability across studies. This would allow visibility on which technologies that begin from a conceptual evaluation phase are eventually implemented and evaluated in subsequent phases for effectiveness and uptake. It would also provide clarity on how many technologies remain in the research stage and fail to see practical clinical implementation.

The results of this study may be used to pinpoint areas of DHT research that require more focus and

support the completion of multiple steps of the *eHealth evaluation cycle*. In such, future research could also look at a temporal view of research in relation to medical speciality and evaluation phase as this could provide further insights as to the stage of digital transformation and where certain fields are lagging behind.

The Excel data file (see Data Availability statement) with the extracted data of the 824 included studies is published online [9,41]. We would encourage interested researchers to use the filter options to look at subsets of data of the field of interest. For example, when one is interested in Cardiology studies in the pilot phase, one can select 'pilot' study phase, and Cardiology studies within the pilot study phase from 2019 will be shown. Finally, all encountered evaluation approaches are published in an online freely accessible document [Multimedia Appendix 6]. The document will a updated on a regular basis and we encourage readers of this paper to e-mail additions to the document if any evaluation approach is missing.

Conclusion

Improving the successful development and implementation of DHTs is crucial to enhancing the transition to an efficient and future-proof healthcare system. We developed a better understanding of the practise of the sequential evaluation study phases of the eHealth evaluation cycle and the use of the relative DHT evaluation approaches. In the *eHealth evaluation cycle*, most attention is paid to the pilot (feasibility) and effectiveness (impact) study phases. Whether the evaluation of the earlier study phases and the uptake (implementation) study phase indeed improve successful outcomes is yet to be evaluated. Surveys, interviews, and mixed methods dominate the earlier study phases of DHT research. While the majority of the evaluation approaches still use a RCT design, the iterative nature of technology may be better suited to more novel assessment approaches. The most often explored DHT were those focusing on treatment and communication. Interestingly, the specialties of psychiatry/mental health, cardiology and neurology, are more interested in DHT evaluation than others. This offers potential opportunities to focus on unaddressed specialties in the search for DHTs, which can provide novel ways to transform and improve our healthcare system. Finally, future research might benefit from tracking and sharing which technologies successfully proceed through all stages of the *eHealth evaluation cycle*.

Author Contribution

AR, HG, JM and PS were involved in screening and data extraction. AR performed the data analyses which was checked by author HG and PS. Authors PS and AR completed the write up of the paper. Review and editing was done by authors HG, JM, SZ, MS, MK and NC. All authors have read and agreed to the published version of the manuscript.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Declaration of Conflicting Interests

The Authors declare that there is no conflict of interest

Declaration of generative AI

The Authors declare that no generative AI was used in any portion of the manuscript writing.

Data Availability

The data sets generated during and/or analysed during this study are available in the online ‘Evaluation Approaches of Digital Health Technologies’ URL: https://docs.google.com/spreadsheets/d/18TRbIUKE5lqChFsrKli3yo4hhlzQ0oxG6qMRH6cUQIA/edit?usp=sharing.

Multimedia Appendix 1.

Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist.

Multimedia Appendix 2.

WHO classification scheme.

Multimedia Appendix 3.

PubMed Search.

Multimedia Appendix 4.

Data extraction sheet.

Multimedia Appendix 5.

Results variables country, medical specialty and journals.

Multimedia Appendix 6.

Complete list of evaluation approaches.

Multimedia Appendix 7.

All performed cross tabulations.

References

1. WHO. WHO - Ageing and Health.
2. Lau F, Kuziemsky C. Introduction - Handbook of eHealth Evaluation: An Evidence-based Approach - NCBI Bookshelf [Internet]. [cited 2022 Dec 21]. Available from: https://www.ncbi.nlm.nih.gov/books/NBK481610/
3. WHO. Classification of Digital Health Interventions v 1.0 [Internet]. [cited 2022 Dec 21]. Available from: http://who.int/reproductivehealth/topics/mhealth/en/.
4. NICE. Evidence Standards Framework for Digital Health Technologies Contents. 2019;
5. Unsworth H, Dillon B, Collinson L, Powell H, Salmon M, Oladapo T, et al. The NICE Evidence Standards Framework for digital health and care technologies – Developing and maintaining an innovative evidence framework with global impact. Digit Health. 2021 Jun 24;7.
6. Campbell M, Fitzpatrick R, Haines A, Kinmonth AL, Sandercock P, Spiegelhalter D, et al. Framework for design and evaluation of complex interventions to improve health. BMJ : British Medical Journal. 2000 Sep 9;321(7262):694.
7. Craig P, Dieppe P, Macintyre S, Mitchie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. The BMJ. 2008 Oct 25;337(7676):979–83.
8. Bonten TN, Rauwerdink A, Wyatt JC, Kasteleyn MJ, Witkamp L, Riper H, et al. Online Guide for Electronic Health Evaluation Approaches: Systematic Scoping Review and Concept Mapping Study. J Med Internet Res 2020;22(8):e17774 https://www.jmir.org/2020/8/e17774 [Internet]. 2020 Aug 12 [cited 2023 Jan 27];22(8):e17774. Available from: https://www.jmir.org/2020/8/e17774
9. NFU. eHealth methodology guide [Internet]. [cited 2022 Dec 21]. Available from: https://www.citrienfonds-ehealth.nl/e-health-toolkit/onderzoek/e-health-evaluation-methodology/overview-of-methods/
10. Trochim W, Kane M. Concept mapping: an introduction to structured conceptualization in health care. International Journal for Quality in Health Care. 2005;17(3):187–91.
11. Safavi KC, Cohen AB, Ting DY, Chaguturu S, Rowe JS. Health systems as venture capital investors in digital health: 2011–2019. npj Digital Medicine 2020 3:1 [Internet]. 2020 Aug 4 [cited 2023 Feb 4];3(1):1–5. Available from: https://www.nature.com/articles/s41746-020-00311-5
12. Maramba I, Chatterjee A, Newman C. Methods of usability testing in the development of eHealth applications: A scoping review. Int J Med Inform [Internet]. 2019 Jun 1 [cited 2023 Feb 4];126:95–104. Available from: https://pubmed.ncbi.nlm.nih.gov/31029270/
13. Ross J, Stevenson F, Lau R, Murray E. Factors that influence the implementation of e-health: A systematic review of systematic reviews (an update). Implementation Science. 2016 Oct 26;11(1):1–12.
14. Tomlinson M, Rotheram-Borus MJ, Swartz L, Tsai AC. Scaling up mHealth: where is the evidence? PLoS Med [Internet]. 2013 [cited 2023 Dec 17];10(2). Available from: https://pubmed.ncbi.nlm.nih.gov/23424286/
15. Aminoff H, Meijer S. Context and Complexity in Telemedicine Evaluation: Work Domain Analysis in a Surgical Setting. JMIR Perioper Med [Internet]. 2021 Sep 16 [cited 2023 Dec 17];4(2):e26580. Available from: /pmc/articles/PMC8485199/
16. Rayyan. Rayyan – Intelligent Systematic Review - Rayyan [Internet]. [cited 2022 Dec 21]. Available from: https://www.rayyan.ai/
17. NICE. Evidence Standards Framework for Digital Health Technologies Contents. 2019;
18. Cho H, Flynn G, Saylor M, Gradilla M, Schnall R. Use of the FITT framework to understand patients' experiences using a real-time medication monitoring pill bottle linked to a mobile-based HIV self-management app: A qualitative study. Int J Med Inform. 2019 Nov 1;131.

19. Dijkstra A, Heida A, Van Rheenen PF. Exploring the Challenges of Implementing a Web-Based Telemonitoring Strategy for Teenagers With Inflammatory Bowel Disease: Empirical Case Study. J Med Internet Res. 2019 Mar 1;21(3).
20. Van Gemert-Pijnen JEWC, Nijland N, Van Limburg M, Ossebaard HC, Kelders SM, Eysenbach G, et al. A Holistic Framework to Improve the Uptake and Impact of eHealth Technologies. J Med Internet Res 2011;13(4):e111 https://www.jmir.org/2011/4/e111 [Internet]. 2011 Dec 13 [cited 2023 Feb 7];13(4):e1672. Available from: https://www.jmir.org/2011/4/e111
21. Kho SES, Lim SG, Hoi WH, Ng PL, Tan L, Kowitlawakul Y. The Development of a Diabetes Application for Patients With Poorly Controlled Type 2 Diabetes Mellitus. Comput Inform Nurs. 2019 Feb 1;37(2):99–106.
22. Dial HR, Hinshelwood HA, Grasso SM, Hubbard HI, Gorno-Tempini ML, Henry ML. Investigating the utility of teletherapy in individuals with primary progressive aphasia. Clin Interv Aging. 2019;14:453.
23. Hoth AB, Shafer C, Dillon DB, Mayer R, Walton G, Ohl ME. Iowa TelePrEP: A Public-Health-Partnered Telehealth Model for Human Immunodeficiency Virus Preexposure Prophylaxis Delivery in a Rural State. Sex Transm Dis. 2019 Aug 1;46(8):507–12.
24. WHO. Medical devices: Managing the Mismatch An outcome of the Priority Medical Devices project 2010. 2010;
25. Royle JK, Hughes A, Stephenson L, Landers D. Technology clinical trials: Turning innovation into patient benefit. Digit Health [Internet]. 2021 Apr 30 [cited 2022 Dec 21];7. Available from: https://journals.sagepub.com/doi/full/10.1177/20552076211012131
26. Guo C, Ashrafian H, Ghafur S, Fontana G, Gardner C, Prime M. Challenges for the evaluation of digital health solutions—A call for innovative evidence generation approaches. npj Digital Medicine 2020 3:1 [Internet]. 2020 Aug 27 [cited 2023 Jan 3];3(1):1–14. Available from: https://www.nature.com/articles/s41746-020-00314-2
27. Hrynyschyn R, Prediger C, Stock C, Helmer SM. Evaluation Methods Applied to Digital Health Interventions: What Is Being Used beyond Randomised Controlled Trials?-A Scoping Review. Int J Environ Res Public Health [Internet]. 2022 May 1 [cited 2023 Jan 10];19(9). Available from: https://pubmed.ncbi.nlm.nih.gov/35564616/
28. Unsworth H, Dillon B, Collinson L, Powell H, Salmon M, Oladapo T, et al. The NICE Evidence Standards Framework for digital health and care technologies – Developing and maintaining an innovative evidence framework with global impact. Digit Health [Internet]. 2021 Jun 24 [cited 2022 Dec 21];7. Available from: https://journals.sagepub.com/doi/full/10.1177/20552076211018617
29. Van Velsen L, Wentzel J, Van Gemert-Pijnen JEWC. Designing eHealth that Matters via a Multidisciplinary Requirements Development Approach. JMIR Res Protoc [Internet]. 2013 Jan 1 [cited 2023 Jan 27];2(1). Available from: https://pubmed.ncbi.nlm.nih.gov/23796508/
30. England I, Stewart D, Walker S. Information technology adoption in health care: when organisations and technology collide. Aust Health Rev [Internet]. 2000 [cited 2023 Jan 27];23(3):176–85. Available from: https://pubmed.ncbi.nlm.nih.gov/11186051/
31. Golinelli D, Boetto E, Carullo G, Nuzzolese AG, Landini MP, Fantini MP. Adoption of Digital Technologies in Health Care During the COVID-19 Pandemic: Systematic Review of Early Scientific Literature. J Med Internet Res 2020;22(11):e22280 https://www.jmir.org/2020/11/e22280 [Internet]. 2020 Nov 6 [cited 2023 Jan 27];22(11):e22280. Available from: https://www.jmir.org/2020/11/e22280
32. Tossaint-Schoenmakers R, Kasteleyn MJ, Rauwerdink A, Chavannes N, Willems S, Talboom-Kamp EPWA. Development of a Quality Management Model and Self-assessment Questionnaire for Hybrid Health Care: Concept Mapping Study. JMIR Form Res [Internet]. 2022 Jul 1 [cited 2023 Feb 9];6(7). Available from: https://pubmed.ncbi.nlm.nih.gov/35797097/
33. Esmaeilzadeh P, Sambasivan M, Kumar N, Nezakhati H. Adoption of Technology

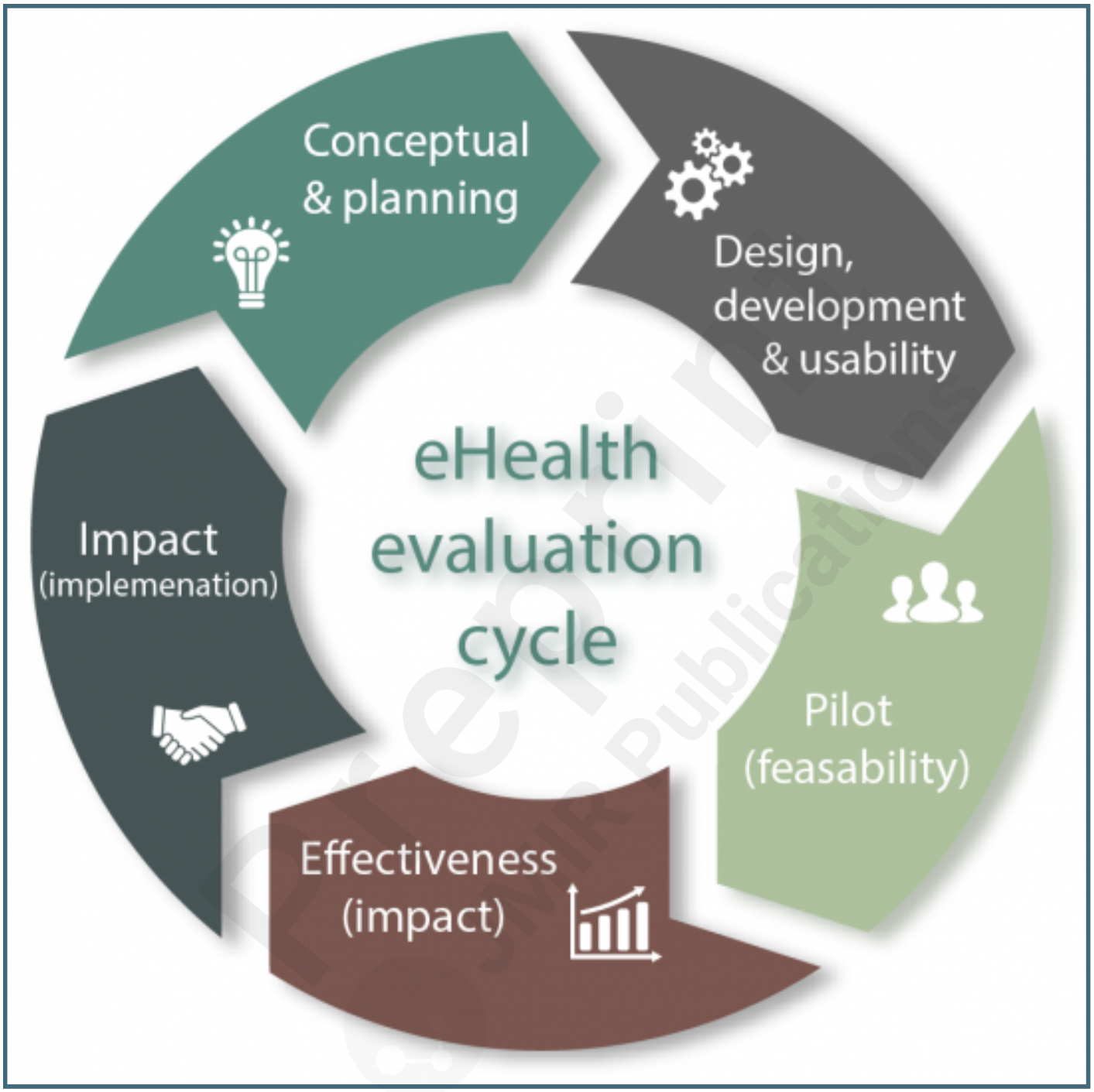
Applications in Healthcare: The Influence of Attitude toward Knowledge Sharing on Technology Acceptance in a Hospital BT - U- and E-Service, Science and Technology. In: Kim T hoon, Adeli H, Ma J, Fang W chi, Kang BH, Park B, et al., editors. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011. p. 17–30.

34. Scimag. SJR - International Science Ranking [Internet]. [cited 2022 Dec 21]. Available from: https://www.scimagojr.com/countryrank.php?area=3600&order=itp&ord=desc&year=2019
35. HIMSS. HIMSS Annual European Digital Health Survey | HIMSS [Internet]. [cited 2022 Dec 21]. Available from: https://www.himss.org/resources/himss-annual-european-digital-health-survey
36. Kane CK, Gillis K. The Use Of Telemedicine By Physicians: Still The Exception Rather Than The Rule. Health Aff (Millwood) [Internet]. 2018 Dec 1 [cited 2023 Jan 27];37(12):1923–30. Available from: https://pubmed.ncbi.nlm.nih.gov/30633670/
37. Basnet S, Tamminen M, Lahti T. The Feasibility of eHealth in Mental Health Care. Journal of Addiction Research & Therapy 2014 5:4. 2014 Dec 20;5(4):1–4.
38. van Hemel NM, van der Wall EE. 8 October 1958, D Day for the implantable pacemaker. Netherlands Heart Journal [Internet]. 2008 Jan [cited 2023 Dec 17];16(Suppl 1):S3. Available from: /pmc/articles/PMC2572009/
39. Zwack CC, Haghani M, Hollings M, Zhang L, Gauci S, Gallagher R, et al. The evolution of digital health technologies in cardiovascular disease research. npj Digital Medicine 2023 6:1 [Internet]. 2023 Jan 3 [cited 2023 Dec 17];6(1):1–11. Available from: https://www.nature.com/articles/s41746-022-00734-2
40. Newaz NT, Hanada E. The Methods of Fall Detection: A Literature Review. Sensors 2023, Vol 23, Page 5212 [Internet]. 2023 May 30 [cited 2023 Dec 17];23(11):5212. Available from: https://www.mdpi.com/1424-8220/23/11/5212/htm
41. NeLL Projects [Internet]. Available from: https://nell.eu/projecten

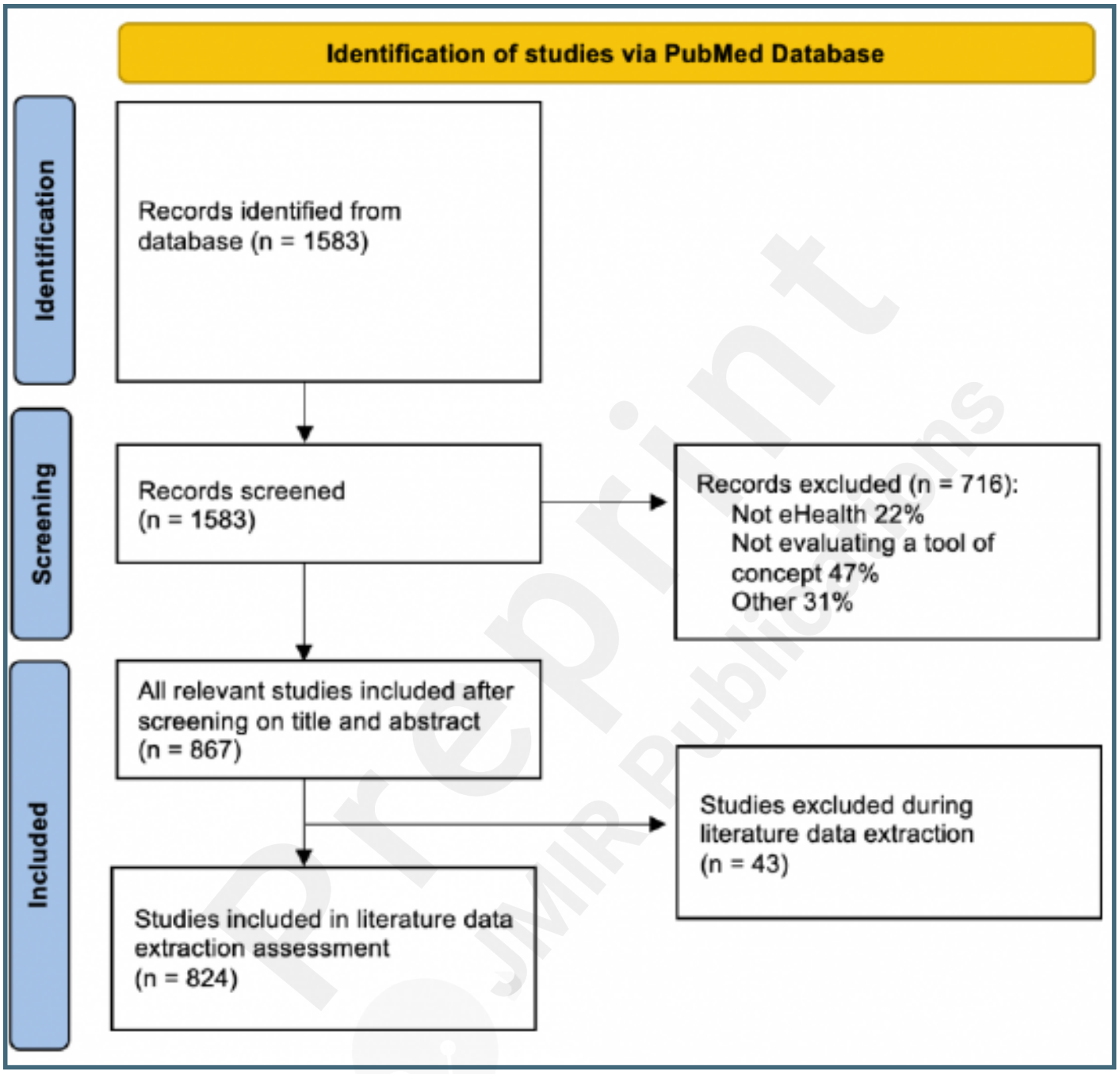
Supplementary Files

Figures

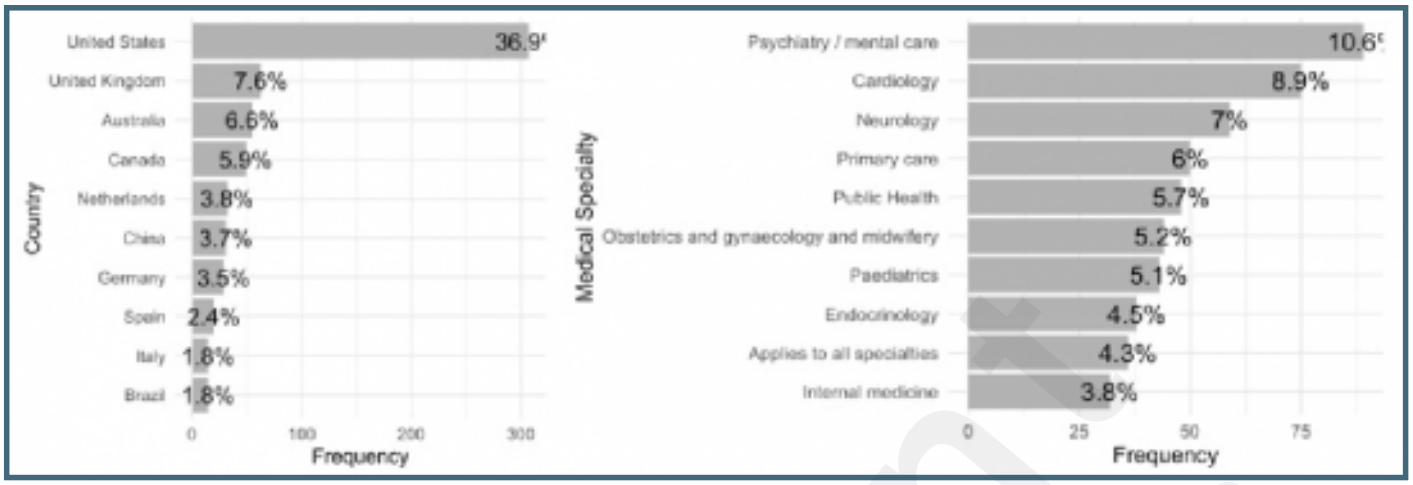
eHealth evaluation cycle.



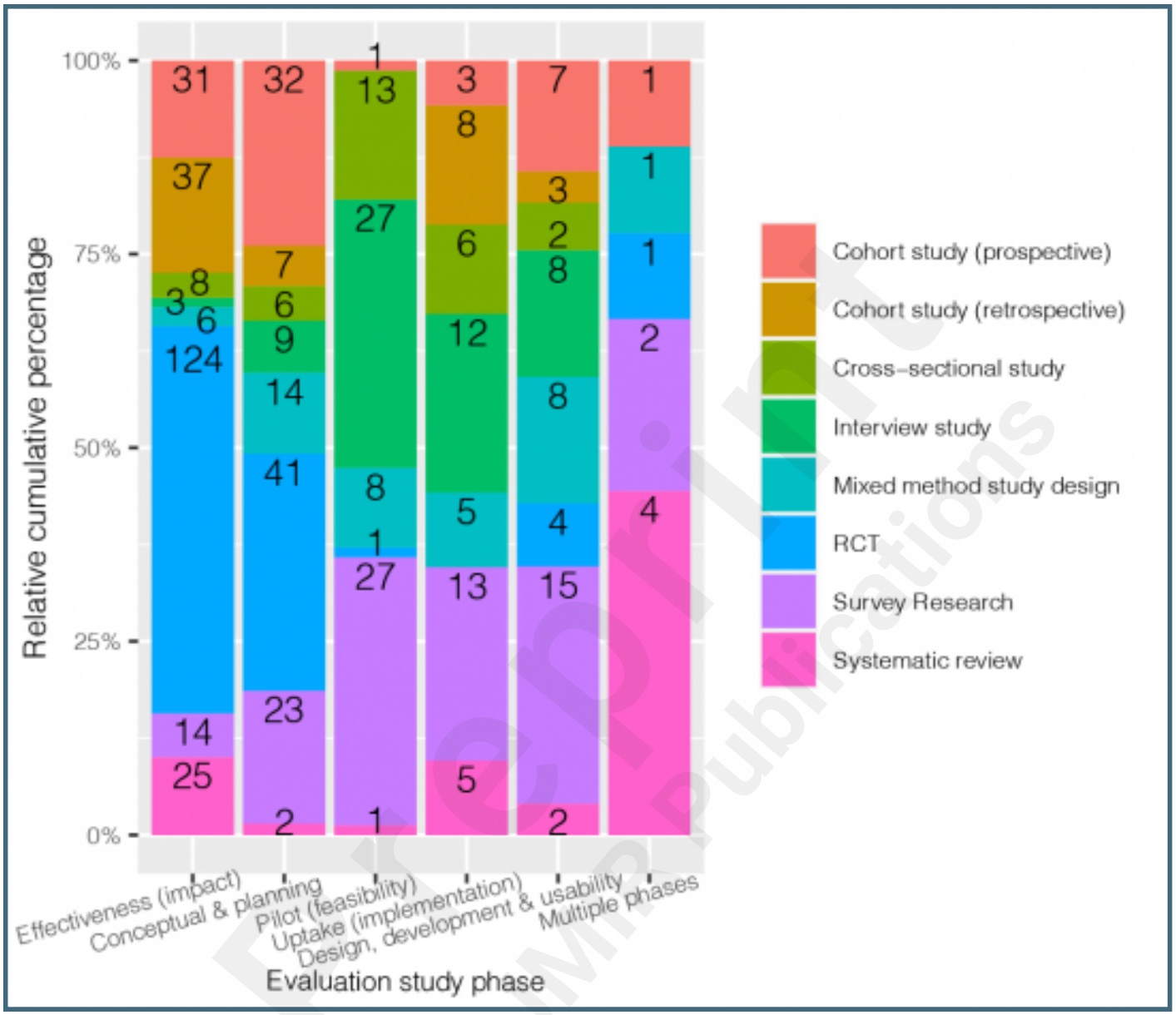
Prisma Flowchart.



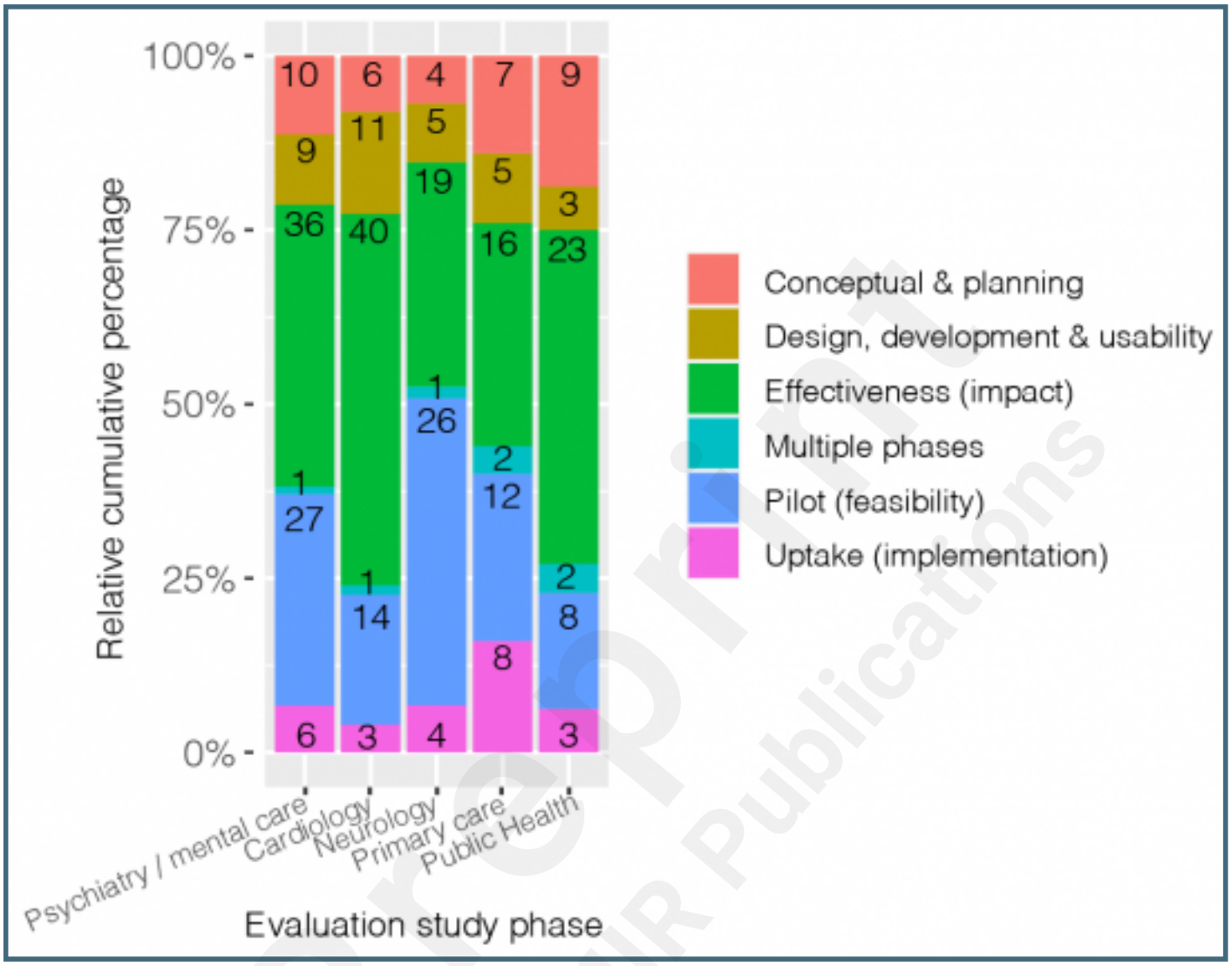
Bar charts country top 10 (total n = 73) and medical specialty top 10 (total n = 44).



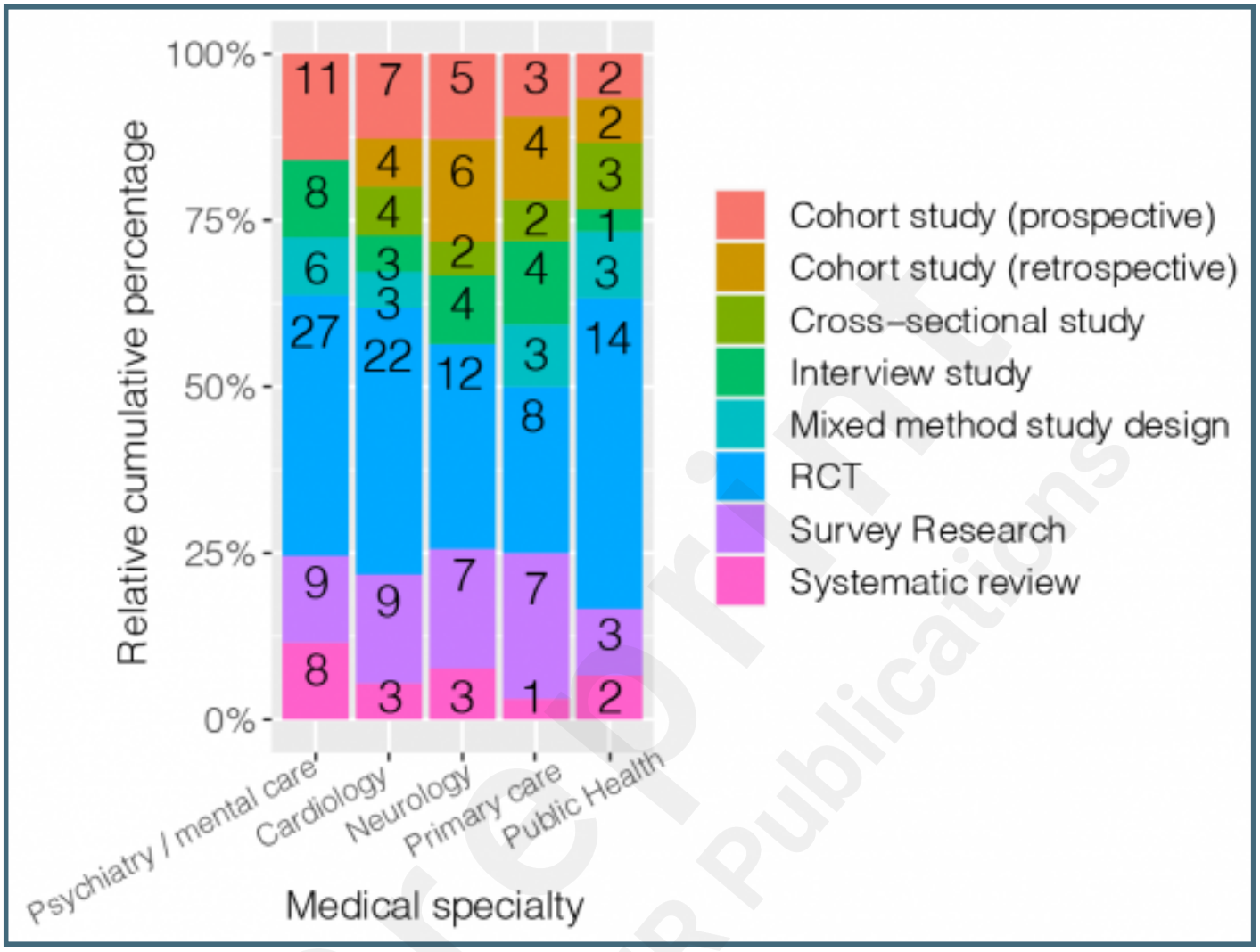
Bar chart evaluation study phase vs. top 8 evaluation approach.



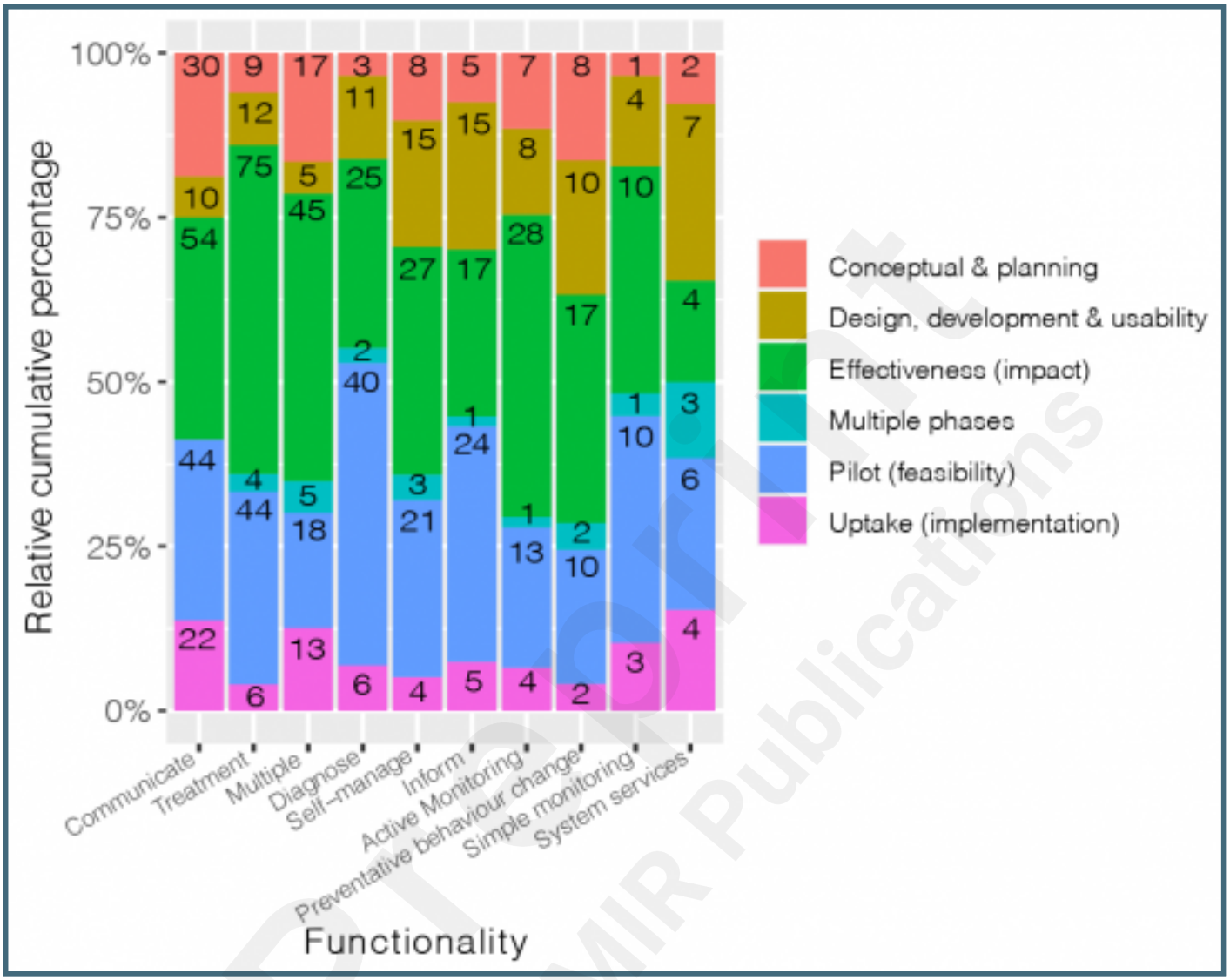
Bar chart evaluation study phase vs. medical specialty.



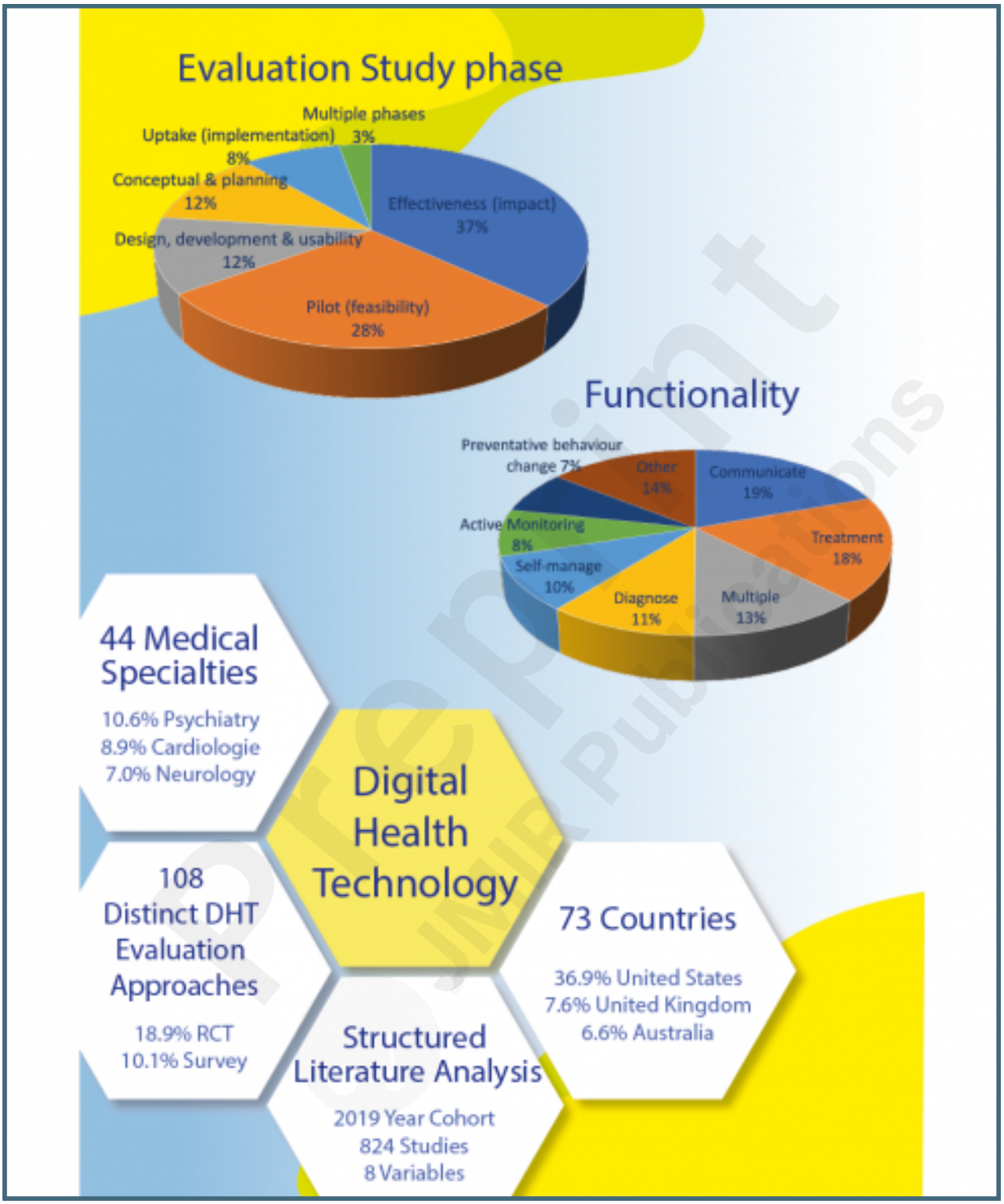
Bar chart medical specialty vs. top 8 evaluation approach.



Bar chart functionality vs. evaluation study phase NB: functionality categories ‘unclear’, ‘calculate’ and ‘other’ were left out due to limited results.



Visual summary of main findings.



Multimedia Appendixes

Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) Checklist.
URL: http://asset.jmir.pub/assets/abaff7552a45bda57e58b106cedbd72e.docx

WHO classification scheme.
URL: http://asset.jmir.pub/assets/e961972b09d6921307653f4f47b0336d.docx

PubMed Search.
URL: http://asset.jmir.pub/assets/5f964b4533c7fd9d89bb7c9a50c93ced.docx

Data extraction sheet.
URL: http://asset.jmir.pub/assets/cd80801f5e006eb83e72e737cf0cb6d8.xlsx

Results variables country, medical specialty and journals.
URL: http://asset.jmir.pub/assets/360e0f1363c582c400ee6d6ec5640f5f.docx

Complete list of evaluation approaches.
URL: http://asset.jmir.pub/assets/10b159126cc27e0d40239c25b0fd45bf.docx

All performed cross tabulations.
URL: http://asset.jmir.pub/assets/d0c133c4caa4ddeecb76c15639e9020a.docx