Requirements of an instrument that enables mobile health assessments

Thomas van Eerden

Master thesis
1 May 2015
Master Business Informatics
Institute of Information and Computing Sciences
Utrecht University

Supervisors:
Dr. Ronald Batenburg
Dr. Marco Spruit

Utrecht University

Supervisors:
Jiska de Wit
Dr. Jaap Trappenburg

UMC Utrecht

Abstract

There is a major debate to what extent stakeholders adopt, use and benefit from mobile health. The time it takes to prove the efficacy and validity of mobile health applications is one of the reasons why mobile health has not been fully adopted. The technology is furthermore progressing at a much faster pace than the research on scientific validation of mobile health. This introduces the risks that applications that are possibly ineffective, harmful or iatrogenic will be used and deployed. The aim of this research is therefore to identify the requirements of a framework that enables a rapid, systematic and critical assessment of mobile health applications. A systematic and explorative literature study provides insights in the requirements for such an instrument. Existing methods are evaluated and compared to determine how these currently assess applications. Results and a first design of the framework were presented to experts during a focus group and various interviews for validation. An artifact has been created and tested by assessing four popular mobile health applications. Further research is necessary to improve and valorize the instrument.

Table of content

List of tables

List of figures

1 Introduction

This chapter provides an introduction to the problem statement. A general overview of the field is further illustrated by describing the current challenges of mobile health (mHealth) and the stakeholders of mHealth. These points are the input of the research question for this thesis.

1.1 Problem statement

A major debate in both the medical and information system literature is the question to what extent the intended stakeholders fully adopt, use and benefit from mHealth, given the large and growing amount of available applications. Nictiz and NIVEL did a research about the usage of electronic health (eHealth) in the Netherlands in 2014. Results from this report are that 11% of the internet users use the internet to do a medical related test, 9% used the internet or an mHealth app to keep track of personal health data, and only 3% used their phone to set automatic reminders to take their medicines (Krijgsman et al., 2014). Several promises of mHealth, such as better monitoring, better management, better outcomes and reduced disease burden still need to be adequately tested (Nilsen et al., 2012). An important barrier mentioned by Nilsen et al. (2012) is the time it takes to prove the efficacy and validity of mHealth applications. The technology of mHealth is progressing at a much faster pace than the research on scientific validation of mHealth. This introduces the risks that possibly ineffective, harmful or iatrogenic applications will be implemented. Users also find security, privacy and safety important aspects, all these issues contribute to the acceptance of mHealth (Tate et al., 2013).

1.1.1 State of practice

There are a number of Dutch organizations that try to create an overview of mHealth applications by reviewing them. The difficulty of these overviews is that they do not keep track of version numbers of the application and they do not test the validity and efficacy of the applications. Keeping track of version numbers is important since a newer version can have problems that did not occur in a previous version and vice versa.

Artsennet

An example of such an organization is the Dutch organization for physicians Artsennet. Artsennet has a knowledge section on their website[1] where the community of healthcare professionals can rate and comment on mHealth applications. The professionals can rate an application by giving it one to five stars. They can furthermore add a comment about the application. Most professionals choose to rate the application on the subject of user friendliness, innovation, and added value on a scale of 1 to 10. There are currently 300 (date: 11-9-14) applications on their website, not all of these applications are reviewed.

Digitalezorggids

Another knowledge platform that also present expert reviews is Digitalezorggids[2]. There are currently 95 (date: 11-9-14) expert reviews on their website. The ratings on this website are also based on a system with a maximum of five stars. The description of the application includes a small review of the following points:

1. User friendliness;
2. Look and feel;
3. Use and effectiveness;
4. Reliability;
5. Quality price ratio.

The difference with Artsennet is that the review is from one professional, instead from different members of the community.

NICTIZ

The national IT institute for healthcare in the Netherlands (NICTIZ) presents a roadmap for companies to check if their mobile health application is required to have a CE mark (Ekker & van Rest, 2013). All mHealth applications that categorized as a medical device are required to have the CE mark. However, the CE mark only indicates that the product meets certain safety and health requirements (EU, 2013). The mark does not guarantee the quality and the clinical relevance of the application.

Existing frameworks

There are various frameworks for mHealth available. These frameworks are focused on other issues than the assessment, such as the technical side of the application or the privacy of an application (Laakko, Leppänen, Lähteenmäki, & Nummiaho, 2008); (Kotz, Avancha, & Baxi, 2009). Current frameworks for the assessment and evaluation of mHealth applications are not adequate enough (Ouma, Herselman, & Vangrauen, 2011). Healthcare institutes are required to monitor their own quality according to the law on quality of healthcare (Dutch: Kwaliteitswet zorginstellingen)

[1] http://www.artsennet.nl/Kennisbank/Medische-apps/Alle-medische-apps.htm

[2] http://www.digitalezorggids.nl

(Gezondheidszorg, 2014). There is a lack of standardized methods or instruments for testing and validating mHealth. Validation is necessary since premature adoption of untested mHealth technologies may detract from, rather than contribute to, what is needed for true overall health improvement (Kumar et al., 2013). mHealth applications can even have harmful effects (Wolf et al., 2013). Examples are melanoma detection applications. Wolf et al. (2013) investigated four applications. Three of the four applications incorrectly classified 30% or more of the melanomas. Applications with the purpose of detecting melanomas are not subject to regulatory oversight. They can therefore delay the diagnosis of a melanoma and harm the users.

1.2 Challenges

This section will provide a closer look at the various challenges of mHealth. As stated above in section 1.1, a challenge is that research is needed to prove the validity and efficacy of mHealth applications (Nilsen et al., 2012). As described, the CE mark indicates that the product meets certain safety and health requirements (EU, 2013). However, the mark does not guarantee the quality and the clinical relevance of the application. A CE mark is required on an application when it is classified as a medical device. An application is a medical device if (European Commission, 2012):

1. A diagnosis is made by the application;
2. Energy is added to the human body (e.g. hearing aid), or;
3. If vital body functions are measured by the application.

The manufacturer has to determine if the application meets the requirements of a medical device. The purpose as stated by the manufacturer is leading in determining this (IGZ, personal communication, December 18, 2014). It is very likely that there are applications available that are required to have a CE mark but do not have one (Ekker & van Rest, 2013). This could jeopardize the safety of the user since the application does not have the effect the manufacturer claims it has.

A second challenge is the concern about privacy and security of personal healthcare information (Galpottage, 2005) (Tate et al., 2013). Stakeholders of mHealth find it important that the transmitted data is encrypted. A third challenge that has to be solved is the acceptance of the technology by patients and healthcare professionals (Norris, Stockdale, & Sharma, 2009). Patients are likely to adopt the new technologies faster than clinicians. The reason for this according to Norris et al. (2009) is that clinicians are more traditional when it comes to the acceptance of new technologies. mHealth, including network, hardware and software application should be integrated into the workflow of the clinician and add value to gain acceptance by clinicians (Yu, Wu, Yu, & Xiao, 2006).

Another challenge is the classification of mHealth applications. It is important that there is a taxonomy and risk assessment for each application. The U.S. Food and Drug Administration (FDA) and the Dutch Healthcare Inspectorate (IGZ) have classifications for medical applications. These classifications deal with the same critique as the CE mark; they do not guarantee the effect the developer claims it has.

The last challenge is that the applications are developed without the input or appraisal from the end users (Lindley & Fernando, 2013). These challenges should be solved to increase the likelihood of adoption of mHealth.

1.3 Who are the stakeholders of mHealth

Different stakeholders in the Dutch healthcare landscape are also stakeholders of mHealth. The main stakeholders in the Dutch healthcare landscape are depicted in Figure 1. The first stakeholder is the Dutch government, that supervises and regulates the healthcare market (Schäfer et al., 2010). The other stakeholders are the insured/patients, the insurers and the healthcare providers. Insurers offer insurance packages through the insurance market to insured/patients, who are all obliged to insure themselves. Insurers make agreements with providers through the purchasing market on price, quality and volume of healthcare. Insured/patients are connected with providers through the healthcare provision market, where providers offer healthcare to insured/patients. A stakeholder that is not in Figure 1 is the developer of mHealth applications. The figure with stakeholders from Schäfer et al. (2010) is based on the Dutch healthcare landscape in general, developers are therefore not included, they are however an important stakeholder regarding this research.

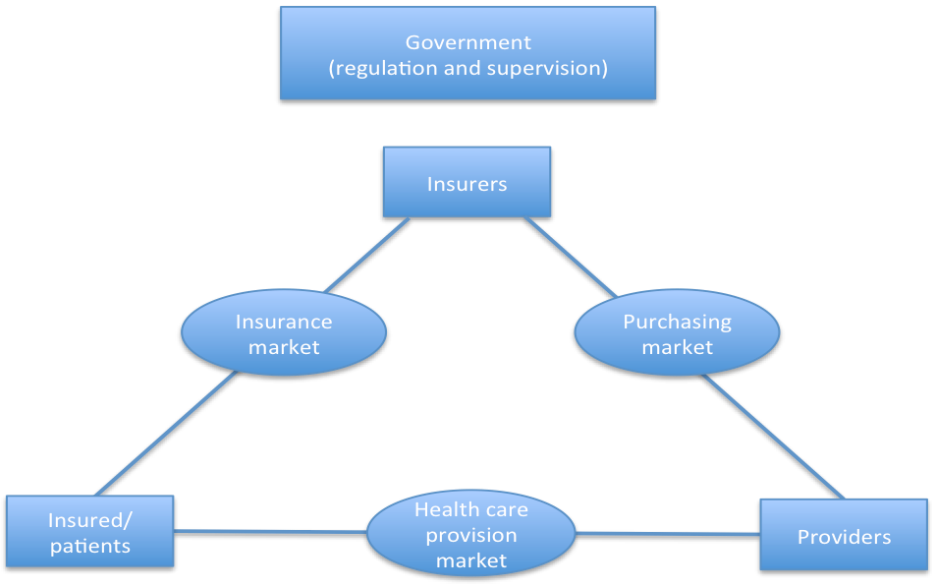


Figure 1 Stakeholders Dutch healthcare landscape (Schäfer et al., 2010).

How does mHealth add value to the stakeholders in the Dutch healthcare landscape and how are they positioned with regard to the problem of mHealth validity and efficacy? Value in healthcare can be defined as the patient health outcome achieved per dollar spent (Porter, 2010). Embedded in this definition of value are other goals such as quality, safety and cost reduction. For patients mHealth validity and efficacy means convenience, active engagement and personalization of healthcare (Steinhubl, Muse, & Topol, 2013). An example is enabling patients to self-diagnose their symptoms and by allowing them to monitor, track and communicate information to physicians. In the case of

insurers, this increases efficiency and will result in less visits to clinicians and therefore in a decrease in costs (Eysenbach, 2001). Eysenbach describes other advantages of eHealth that are also relevant for mHealth. Eysenbach describes the following three advantages that are related to the interest and position of the main stakeholders in healthcare:

Improving the quality of healthcare
The first advantage is that the quality of healthcare will improve when patients can compare different healthcare providers and when patient streams are redirected to the best providers.

Increasing the knowledge about healthcare
The second advantage is that eHealth can increase the knowledge of patients and healthcare professionals. The increase in knowledge can improve the relationship of the patient and the healthcare professional since the patient has a better understanding of the decisions of the healthcare professional.

Broadening the scope of healthcare
The third and last advantage is that eHealth also extends the scope of healthcare. People can obtain information and healthcare services from all over the globe.

However, Eysenbach states that the effectiveness and efficiency of eHealth should be proven scientifically. This is confirmed by a research from Kumar (2013). This study states that it is unknown if mHealth leads to better health outcomes and reduced disease burden. Research is needed to examine the potential and the challenges of using mobile technology to improve health outcomes (Kumar et al., 2013). These points, as well as the critical point of validity and efficacy leads to the research question, that is described next, and the related goal to create an instrument that enables a broad and general assessment of mobile health applications.

1.4 Research question

The research question is based on the problem that there is no general instrument, method or tool available for stakeholders to assess mHealth applications. The answer to this question should aid the stakeholders described in chapter 1.3. The main research question is therefore formulated as:

"What are the requirements of an instrument that enables a rapid, systematic and critical assessment of all types of mobile health applications, satisfying all relevant stakeholders, and how can this instrument be modeled and validated?"

The following sub-questions need to be answered in order to find an answer for the main research question:

SQ1: What types of mobile health applications can be distinguished and how has mobile health evolved?

The answer to this question will provide a general overview of the field including background information on mobile health.

SQ2: What do we know about the added requirements of mobile health?

This question need to be answered to determine what important is while evaluating mHealth applications.

SQ3: How are mobile health applications currently assessed?

It is necessary to study relevant literature since existing methods will be used as a basis for the instrument. Current assessment methods will provide information on what criteria are currently applied in the assessment of health applications. Scientific models and non-scientific models will be taken into account for this sub-question.

SQ4: What are assessment criteria and what level of evidence is needed for a rapid, systematic and critical appraisal instrument for mobile health applications?

The answer to the question will provide, together with the answer of sub-question three, a set of criteria and requirements that defines what 'rapid', 'systematic' and 'critical' implies for the appraisal instrument.

SQ5: How can the criteria be modeled into an instrument that can actually be tested?

The criteria from the previous questions need to be modeled into an instrument that can be used to test current mHealth applications. Testing of the model is necessary to find improvements. The model can be tested by:

- Assessments of mHealth applications;
- Conducting interviews;
- Organizing a focus group where experts evaluate the instrument.

SQ6: What are the results of applying the critical appraisal instrument on a sample of mobile health applications?

The results of testing will be analyzed and processed in the instrument.

SQ7: How can this instrument be applied for different stakeholders?

Different stakeholders will have interest in the instrument and they will probably want to use the instrument for different reasons. A developer may want to check what important is before developing, while a healthcare institute wants to check the application when it is developed.

1.5 Scope

Determining the scope of this research is important since mHealth as a subject is very elaborate. The following definition of mobile health or mHealth will be used for this research:

"The emerging mobile communications and network technologies for healthcare systems (Istepanian, Laxminarayan, & Pattichis, 2006)".

The following is included in this research:

- MHealth and related topics.
- Validation of the scientific instrument with interviews and by doing assessment analysis of medical mobile applications. The analysis will focus on patient to doctor applications.

1.6 Relevance

This thesis will contribute to the scientific community with the development of a critical appraisal instrument for mobile health applications. The instrument will make the validation of mobile health applications more efficient. Other researchers can develop the instrument further, by testing it with other applications or with serious games for healthcare. The social relevance is that patients and professionals will be able to safely use mobile health applications that have been tested. This will bring various benefits for the patients and professionals who are responsible if they advice or "subscribe" an application, as described in the problem statement.

1.7 Unovate

The research was executed in collaboration with Unovate. Unovate is an organization that is focused on innovation in healthcare and on the challenges of mHealth. It bridges the gap between the medical academic environment of University Medical Centre Utrecht and the business world (Unovate, n.d.). This resulted in the University Medical Center Utrecht (UMCU) having a leading position in the digital healthcare in the Netherlands.

The UMCU and Unovate want to strengthen this position by expanding their expertise on mobile healthcare. The UMCU and Unovate have established The Medical App Co-Creation Center (MAC^3) that stimulates scientific research in mHealth. The goals of MAC^3 are to improve the quality of healthcare, better utilize the time of healthcare professionals and to enhance the role of patients in their own healthcare through mHealth. MAC^3 aims to develop an instrument that connects with the current pace of innovation of mobile health applications. This is necessary since the development of mHealth applications is faster then the scientific validation of these applications (Nilsen et al., 2012). The reason for this is that mobile medical applications are not subject to the same high standards and regulations as e.g. medicines or medical devices (Barton, 2012). MAC^3 collaborates with the Utrecht University to solve the problem of the discrepancy between the pace of innovation of mHealth applications and the scientific validation of these applications.

2 Design of this research

This chapter provides a description of the different research steps and methods that are applied in this research. Followed by a more detailed explanation of each step. Additionally an explanation is given to link the steps to the sub-questions.

2.1 Design science

The answers to the sub-questions presented in chapter will provide an answer to the main research question. The methodology used in this research is design science. Design science intents to:

"Produce and apply knowledge of tasks or situations in order to create effective artifacts (March & Smith, 1995)."

Several researchers present steps for design science. The steps described below are from Vaishnavi and Kuchler (2004). The first step is the suggestion step; the goal of this step is to do a literature study to provide a foundation for the scientific validated instrument. The second step is to develop a first version of the instrument based on the information from step one. The third step is to evaluate the instrument with experts. The final step is the conclusion step where the whole project will be evaluated. Important to note is that the steps can be repeated to improve the final outcome of the research. An overview of the steps can be viewed in Figure 2.

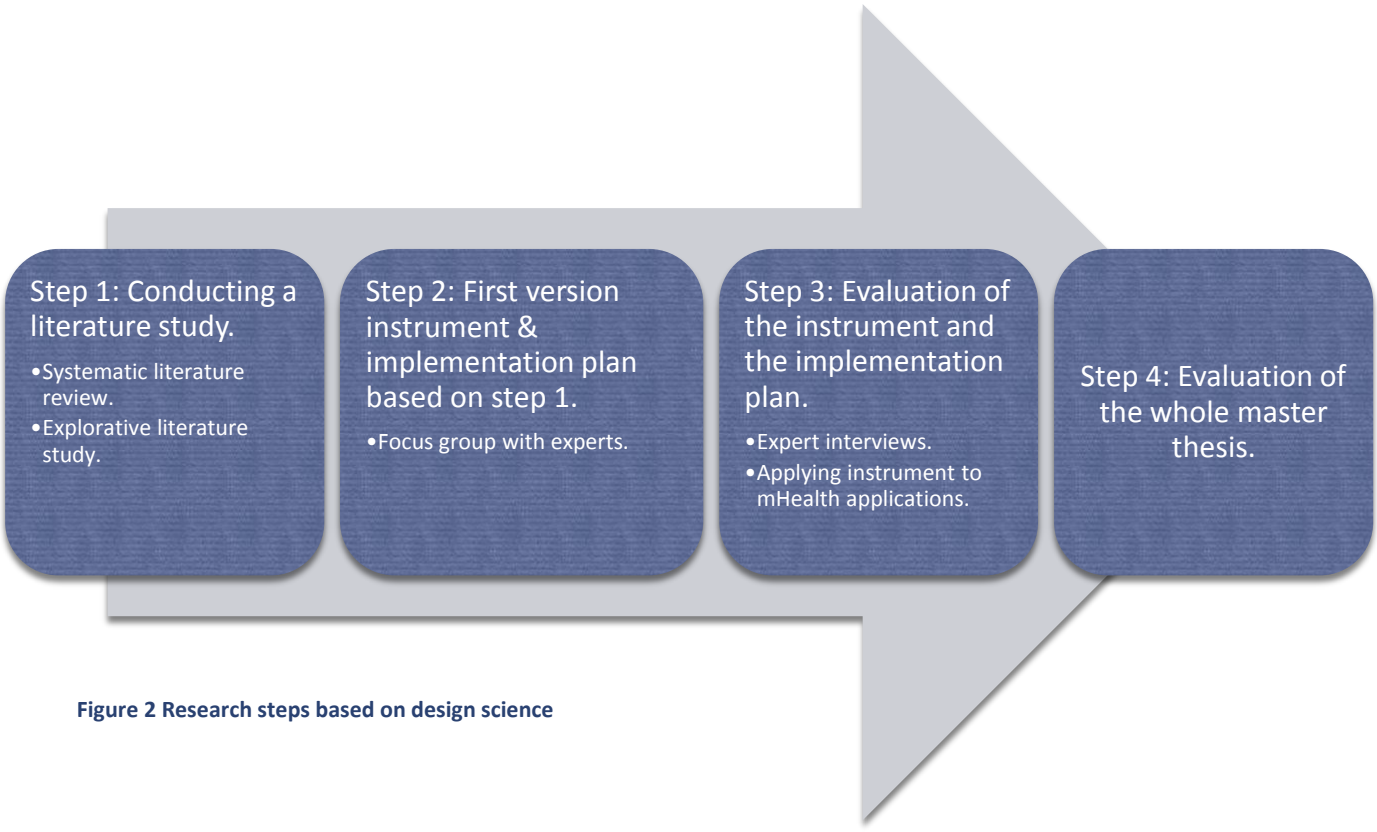


Figure 2 Research steps based on design science

In short, the research methods are:

- A systematic literature review;
- An explorative literature study;
- A focus group;
- Retrospective assessment of mHealth applications;
- Expert interviews.

The steps in Figure 2 match with the sections that are in the table of content.

2.2 Literature study

The literature study performed in step 1 is performed to gather information for the following sub-questions:

SQ1: What types of mobile health applications can be distinguished and how has mobile health evolved?

SQ2: What do we know about the added requirements of mobile health?

SQ3: How are mobile health applications currently assessed?

The answers to sub-question four and five are partly answered with literatur. Findings of the literature study will be checked with interviews, a focus group and retrospective analysis of a selection of mHealth applications.

Systematic literature review

The literature study is divided into a systematic literature review (SLR) and into an explorative literature study where the snowball method is applied. The SLR is used since it is an easy way to aggregate knowledge about a certain research topic (Kitchenham et al., 2010). Another strong point of a systematic literature review is that a SLR is efficient. An SLR makes it easy to recognize strong and weak points of past studies. It is furthermore a convenient way to search for consistency among results and it enables the researcher to select relevant information of critical studies (Mulrow, 1994).

The papers are in or excluded based on their title and abstract, the next selection is based on a full read of the articles. Duplicate and irrelevant papers will be excluded during this phase. All the results and advantages that are described are extracted from the papers if the paper is of sufficient quality. The literature study will use a framework created by Mosa et al. 2012 (Mosa, Yoo, & Sheets, 2012).

The following list of keywords will be used for the systematic literature review:

- Value of “MHealth” OR “Mobile health”
- Testing of “MHealth applications” OR “Mobile health applications”
- Assessment of “mobile health applications” OR ”MHealth applications”
- Assessment criteria “mobile health applications” OR “MHealth applications”
- Framework assessment requirements "mobile health" OR "MHealth"

Scholar and PubMed will be used for the systematic literature study. PubMed is focused on the medical domain and Scholar includes a wide variety of journals. Papers are excluded if there is no PDF file available, the article is not in English/Dutch or if the paper is not of decent quality.

Explorative literature study

The second part of the literature study is an explorative literature study, where the snowball method is applied (Biernacki & Waldorf, 1981). With the snowball method papers are found by searching for related papers and citations, papers that cite the paper you found or by searching in the reference section for relevant literature. Snowballing allows to find papers that were not found trough the SLR due to other keywords or because they are indexed in other databases. The difference between the two methods is that with a SLR keywords are determined beforehand. The papers are then selected based on their title, abstract and full text, which makes it easy to reproduce for other researchers.

2.3 First version instrument

Step 3 will answer the sub-questions four and five:

SQ4: What are assessment criteria and what level of evidence is needed for a rapid, systematic and critical appraisal instrument for mobile health applications?

SQ5: How can the criteria be modeled into a critical appraisal instrument that can be tested?

A first version of the instrument will be designed to answer these sub-questions. The first version is based on literature that is found in step 1. The design process is iterative, multiple versions will therefore be made before the research will proceed to step three (Simon, 1996). Later versions of the instrument will also have input from interviews.

Focus group

A focus group with experts that represent the stakeholders described in chapter 1.3 will be organized. The variation in the group of experts will hopefully contribute in the acceptance of the instrument among the different stakeholders. The goal of this meeting is to let experts discuss and vote on elements found during the literature study. The focus group should contribute to answering sub-questions four and five.

2.4 Evaluation

The fourth step is to evaluate and improve the instrument designed in step three. This will provide an answer to sub-questions six and seven.

SQ6: What are the results of applying the critical appraisal instrument on a sample of mobile health applications?

SQ7: How can stakeholders that are involved in mobile health use the critical appraisal instrument?

Retrospective assessment

The second version of the instrument will be used to assess mHealth applications. The Apple App Store and the Google Play Store will be used to find applications for assessment. The focus of the assessment is on patient to doctor applications. Sites such as www.artsennet.nl or www.digitalezorggids.nl present overviews that can be used to make a selection. Other applications will be searched for during the SLR, a quick search on PubMed resulted into a list of articles that reviewed mobile applications. Articles that describe relevant applications will be used for testing.

Expert interviews

The evaluation will happen by conducting interviews with experts. The expert interviews will exist of semi structured qualitative interviews and will be recorded and transcribed. The main objective of the expert interviews is to receive criteria for validation and evaluation of the instrument, resulting in an improved version of the instrument

3 The classification and evolution of Mhealth

This chapter will provide an answer to the first sub-question. As described in chapter 2, results from the literature study will be used to answer the first sub-question. The first sub-question is posed as:

"What types of mobile health applications can be distinguished and how has mobile health evolved?"

First, some background information is presented about mobile devices and mHealth. Followed by a short history of mHealth applications and different classifications that are found in scientific literature.

3.1 What is a mobile device

There are multiple definitions available in literature of a mobile device. A broad definition is from Derballa & Pousttchi (2004). They describe a mobile device as

"A device developed for mobile use (Derballa & Pousttchi, 2004)".

This is a broad term since it encompasses a wide spectrum of devices. The researchers do exclude laptops from the definition. Laptops can be easily moved, but they are usually not used during that process. Mobile phones are included in this definition.

A recent development in the mobile phone industry is the smartphone. A smartphone has advanced computing and communication capability, including a high resolution screen, internet access and geo-positioning systems next to the standard facilities such as voice and text communication (Boulos, Wheeler, Tavares, & Jones, 2011). The newer generation smartphones also have personal management tools, high quality cameras and recording devices. Mobile phones have some limitations in comparison with laptops (Boulos et al., 2011). Phones have a small internal storage capacity, less processing power and often a small screen. Free et al. (2010) highlights several key features of mobile phones over other communication devices. These advantages are continuous communication capabilities, portability and sufficient computing power to support multimedia software applications (Free et al., 2010).

In short, a modern smartphone is a high-end mobile phone that combines the functionality of a pocket-sized communication device with capabilities from a PC (Carroll & Heiser, 2010).

3.2 What is mHealth

Different papers, found with the explorative literature study, created an overview of the different synonyms that are used for mHealth. Krouse (2012) states that mHealth uses mobile applications to deliver health to patients and helps them make actual diagnosis (Krouse, 2012). Kumar (2013) describes that mHealth is the use of mobile application either by consumers or providers, for monitoring the health status or improving health outcomes, including wireless diagnostic and clinical decision support (Kumar et al., 2013).

Both the definition of Krouse (2012) and from Kumar (2013) do not encompass all of the mHealth applications that are available. There are many other uses of mHealth, as can be seen in section 3.4. The definition of mHealth used in this research is therefore very broad and is defined as:

"The emerging mobile communications and network technologies for healthcare systems (Istepanian et al., 2006)".

Other terms that are used instead of mHealth are health 2.0, connected health, eHealth, telehealth and medicine 2.0 are amongst other terms used in literature. All these terms have in common that they connect information technology with healthcare. A study from 2005 found 51 different definitions for eHealth, the oldest one dating back to 1999 (Oh, Rizo, Enkin, & Jadad, 2005). Another study from 2008 found 46 unique definitions of health 2.0 and four definitions for medicine 2.0 (Hughes, Joshi, & Wareham, 2008). The difference between health 2.0 and medicine 2.0 is that medicine 2.0 is focused on the relation between a professional and the patient. Health 2.0 on the other hand is focused on healthcare in general. There are also terms that are not used anymore because of more popular synonyms. Telehealth is for example a synonym for the older word telemedicine, which is rarely used nowadays (Maheu, Whitten, & Allen, 2001).

3.3 History of Mhealth

Mobile applications became immensely popular when smartphones were introduced. The first mobile applications however was already developed in 1992 (Krouse, 2012). Examples of some early applications are calendars and lists. Allowing the user to keep track of information with their mobile phone. These applications are simple compared with the applications that are developed nowadays, this is due to the constantly expanding field (Fiordelli, Diviani, & Schulz, 2013). Mobile phone applications as we know them today are rapidly developed with the help of technological innovations. One example of such an innovation are the online application stores. These stores are a place for third party developers to sell their applications. The two biggest stores are Google's Play store (Rowinski, 2013), with over one million applications, and Apple's App store, with more than 900,000 applications (Apple, 2013). Official statistics on the number of mHealth applications are not published by either of the companies. Other sources have collected data on the number of mHealth apps in the application stores. The website www.statista.com shows that the number of applications in the category "medical" grew with 156% in 2010 (Statista, 2010). They do not provide statistics about the number of applications. Other websites, such as www.pocketgamer.biz do provide statistics about the number of applications in the Apple app store (Pocketgamer, 2014). They state that there are around 35,000 Health & Fitness applications and around 27,000 Medical applications in the US app store in September 2014. In the Google play store were around 8,000 medical applications are available according to an article on www.imedicalapps.com (Aungst, 2013).

3.4 Classifications of mHealth

Different papers from the SLR mention the classification of mHealth applications. Some of these papers include a classification. Other papers only state the need for such a classification. This section describes the most popular classifications found in the SLR. A classification with different categories is needed since mHealth applications are used for different purposes (Riley et al., 2011). Patients would for example use applications for smoking cessation, weight loss, treatment adherence or disease management. While medical professionals use applications for monitoring of patients or for education.

Classification by Martínez-Pérez

Martínez-Pérez et al. (2013) included a classification that has three main categories where the applications are divided in the categories "open source", "commercial" and "research" applications (Martínez-Pérez, de la Torre-Díez, Candelas-Plasencia, & López-Coronado, 2013). These three categories all have the same sub-categories, namely "patients" and "healthcare staff". This classification is not used since the main categories are focused on how the application is developed, which is not relevant for this research. The classification can be viewed in Appendix B: classification Martinez-Perez et al..

Classification by Corral

Corral (2012) did not include a classification as elaborate as Martínez-Pérez et al. (2013) but made a distinction between three types of applications (Corral, 2012). The distinction made is based on the function of the application within the healthcare workflow. Corral defined **facilitation**, **assistance** and **extension** of care as functions in the healthcare workflow.

1. **Telemonitoring (facilitation of care)**
   Care is delivered through communication technology instead of a physical visit. It extends healthcare by expanding and continuing the collection of health related information at home. Telemonitoring has the possibility to automate the transmission of vital signs through email, portals or with a phone.

2. **Point-of-Care (assistance of care)**
   Point-of-care are applications that use portable devices to assist healthcare at the bedside or clinic. Providing on demand information, management, educational material or decision support. Hospitals or clinics can connect to these devices to collect this information.

3. **Telehealth (extension of care)**
   Telehealth collects information from the patient at their home or residence. Information such as blood pressure and blood glucose is collected with computers, webcams or other communication technology. Telehealth enables the patient to send high-resolution video and images.

The classification by Corral (2012) is insufficient for this research since it is only focused on applications that have a function within the healthcare workflow that Corral (2012) defined with (only) three categories.

Classification by Free

Another classification that is found during the explorative literature study is from Free et al. (2010). Free et al. (2010) present a classification of mobile applications, see Figure 3, with three main categories (Free et al., 2010). These categories are, “tools for health research”, “improving health services” and “improving health outcomes”. The classification of Free et al. (2010) is selected, since this classification takes different sort of applications into account by providing different categories and sub-categories. These (sub-)categories are based on the function of the mHealth application.

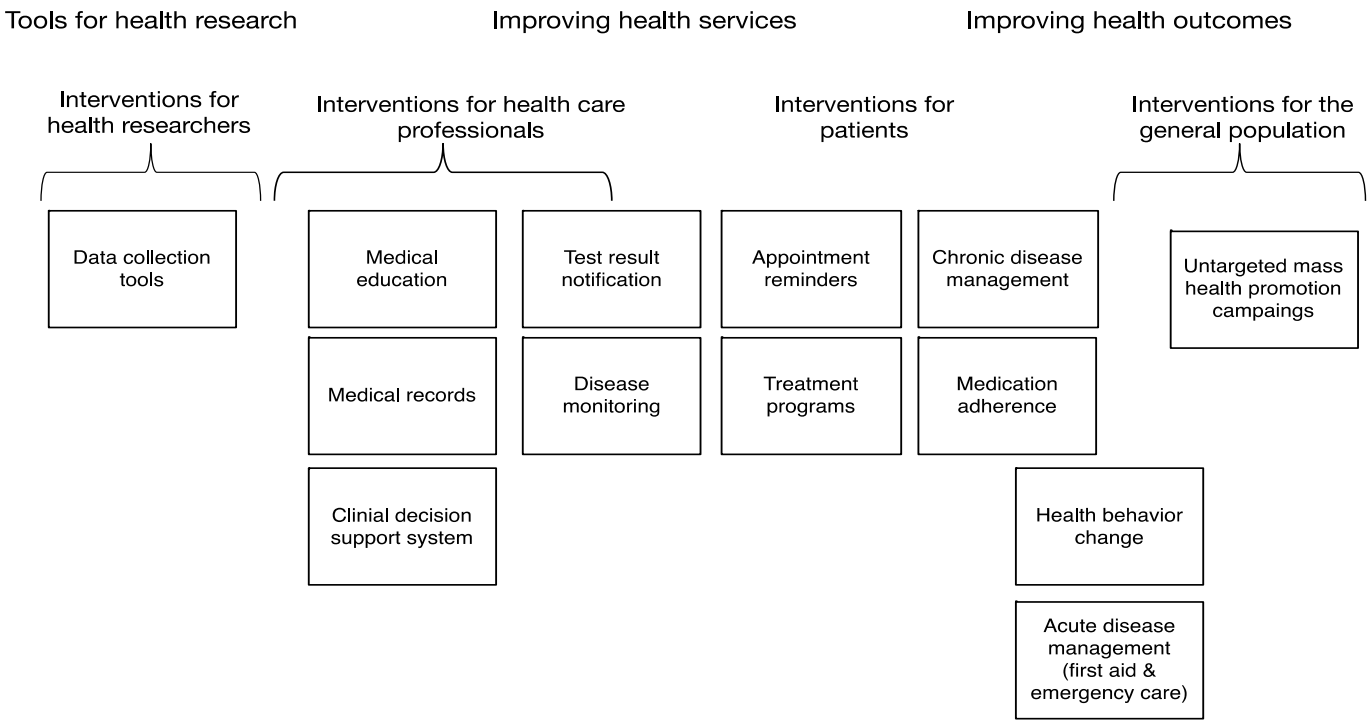


Figure 3 Mobile electronic device intervention classification (Free et al., 2010).

3.5 Conclusion

This chapter provided an overview of the different definitions of mHealth and synonyms that are used in scientific literature. Followed with a description of the history of mHealth and how applications can be classified. The classification and the definition presented in this chapter will be used in this research.

4 A literature study on mHealth assessment criteria

As described in chapter 2.2, the methods to find relevant articles are a SLR and an explorative literature study. The SLR will follow the method of Mosa et al. (2010). First keywords and databases are selected to find articles. These articles are scanned on title and abstract, the relevant articles that are left are fully read. Duplicate articles, articles without a PDF, and irrelevant articles are filtered out.

4.1 Existing systematic literature reviews

Various researchers already performed SLR's about mHealth. The following seven SLR's are some of the more popular and relevant SLR's.

1) Braun et al. (2013) performed a SLR that was focused on community health workers and mobile health. 25 papers were identified and the results of the SLR are that mobile technology has promising opportunities to improve the range and quality of services provided by community health workers (Braun, Catalani, Wimbush, & Israelski, 2013).

2) A SLR by Fiordelli et al. (2013) consisted of 117 articles between 2002 and 2013. The objective of the SLR was to provide a comprehensive view of the mHealth research field to date and to understand how new smartphones trigger research. Their conclusion is that the interest of researchers towards mHealth is growing. New opportunities of new mobile technologies still have to be explored. (Fiordelli et al., 2013)

3) The SLR from Free et al. (2013) aimed to systematically review papers that described controlled trials of mobile technology interventions to improve healthcare delivery processes. 42 trials were identified with the conclusion that some mHealth interventions have modest benefits but other interventions were of a lower quality when they were compared with the golden standard (Free et al., 2013).

4) Another SLR from Free et al. (2010) included articles published since 1990. The focus was also broader since it also included PDA's and MP3 players next to mobile phones (Free et al., 2010).

5) Mosa et al. (2012) SLR focused on existing healthcare applications for smartphones. They identified 83 applications and concluded that professionals and patients already use many healthcare applications. They furthermore concluded that smartphones can play an important role in patient education, disease self-management, and remote monitoring of patients (Mosa et al., 2012).

6) Gurman et al. (2012) performed an SLR that included 44 articles that are focused on behavior change communication in mHealth. The studies did not consistently demonstrate significant effects to behavior change communication in mHealth. They see mHealth as a promising tool

with the ability to promote behavior change. However, current interventions need to be tested to establish better evidence (Gurman, Rubin, & Roess, 2012).

7) A systematic review by Krishna et al. (2009) was focused on the use of short message services (SMS). This resulted in evidence that SMS is a valuable tool in improving healthcare outcomes. 25 studies with 38,060 participants were reviewed. 20 of theses studies were randomized controlled trails and 5 controlled studies. The studies reviewed had a different frequency of message delivery; some studies send 5 messages each day, other studies just one message each week. The results of the review are that standard care can be enhanced with reminders, disease monitoring and management, and education with cell phone voice and SMS. This will have consequences for patients and providers (Krishna, Boren, & Balas, 2009).

It can be concluded from the aforementioned SLR's that there is a general consensus that mHealth has promising opportunities. mHealth can have an important role in:

- Improve the range and quality of services provided by community health workers.
- Enhance the quality of care with reminders, disease monitoring and management.
- Improve education, self-management, behavior change and remote monitoring of patients.

However, current interventions need to be tested to establish better evidence on the efficacy and efficiency of these applications. Since the majority of mHealth interventions are of a lower quality when compared with current methods. The only tool that has been thoroughly tested is SMS. SMS is a valuable tool to improve healthcare outcomes. The last conclusion is that new opportunities of mobile technologies and how to use them for mHealth still need exploration.

Despite the fact that there are several SLR's that review the subject mHealth. It is necessary to conduct an additional SLR, since none of the existing SLR's researched:

- If there are assessment frameworks available for mHealth applications;
- What assessment criteria for mHealth applications are;

The points above and the fact that this SLR will be conducted a year later then the most relevant SLR are the reasons for another SLR.

4.2 Results of the SLR

Five queries were used for the SLR. Tables with more detailed information about the SLR can be found in Appendix A: Queries SLR. A short overview of the results of the SLR can be viewed in Figure 4. The results were first filtered on the title and the abstract of the article. The articles left were reviewed in full text. This resulted in a list with 42 relevant articles. The box with "other sources" includes the articles that were not found with the SLR. Searching for additional literature was needed when the information found during the SLR was insufficient.

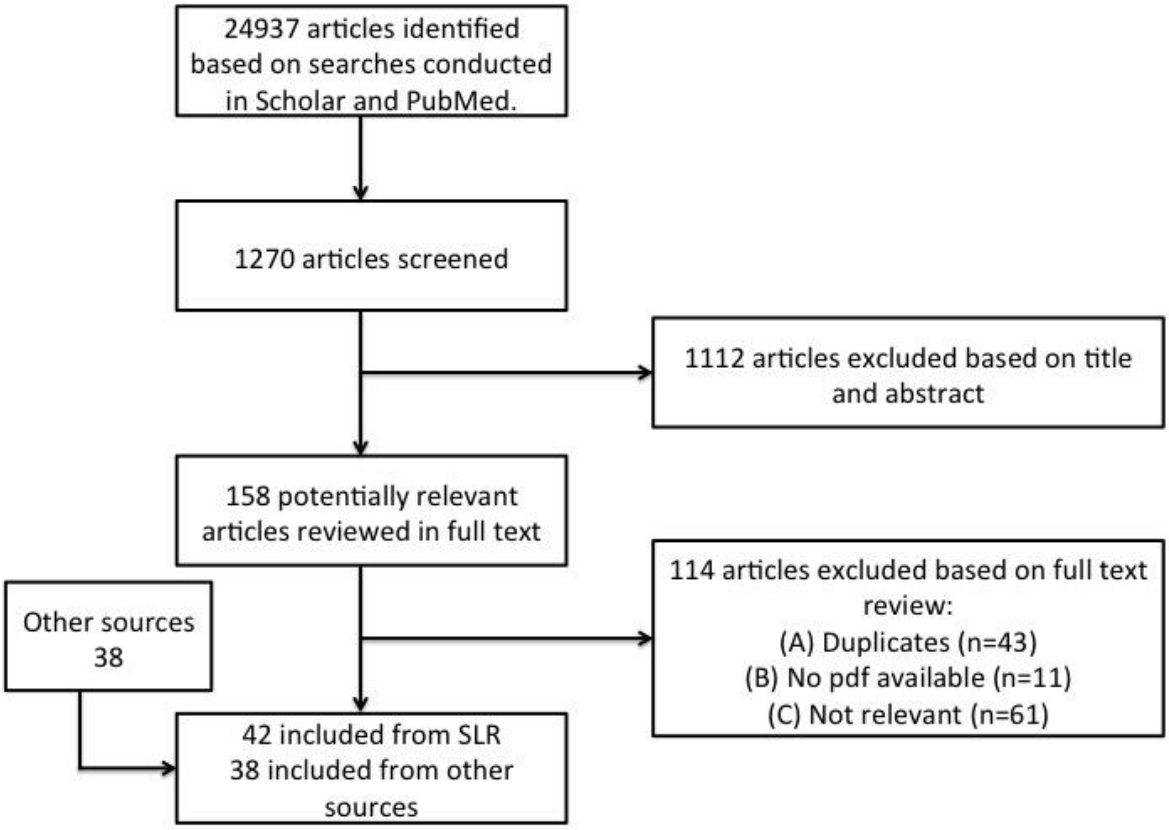


Figure 4 Results systematic literature review

4.3 Analysis of the selected papers

As can be seen in Figure 4, there are 42 papers included from the SLR. 33 of these papers are published in a journal. From the 9 papers that are left are 3 chapters of a book and the other 6 are either conference proceedings or generic papers that did not disclose where they were published or submitted. The majority of all the journal papers are published in a medical journal or a journal about information systems in the medical field. The graph in Figure 5 on the next page depicts the number of papers that are included from the SLR per year. The reasons that almost none of the papers are from before 2010 is that smartphones as we know them today were introduced in 2005. Hence, research only started after this year. However, the first paper included is from 2009, four years after the introduction of the smartphone. The reason for this could be that mHealth was not popular at the introduction of the smartphone, but the popularity of mHealth has grown ever since. The reason that only four articles of the year 2014 are included is that the SLR was conducted mid-2014. The number of papers included from 2014 would therefore be higher if the same SLR is conducted in a later year.

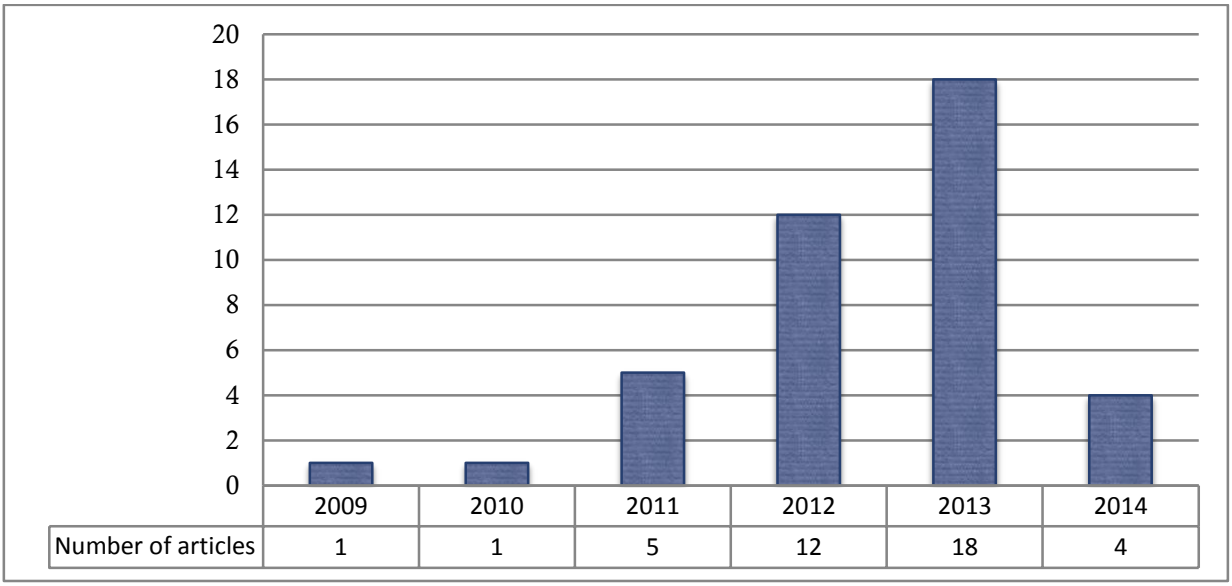


Figure 5 Number of articles per year

In Figure 6 are the number of articles per type that is presented. The group "generic" consists of articles where it is unclear what for type it is. Since additional information could not be found on the website or in the article self. The largest group with 34 articles is "Journal". Concluded from this can be that the literature in this research is based on scientific articles.

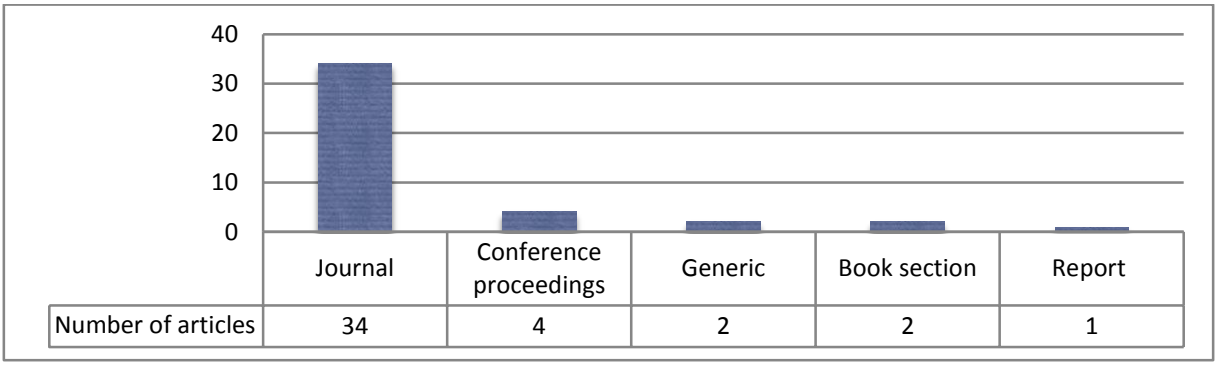


Figure 6 Number of articles per type

Table 2 and Table 1 on the next page include a list of topics that are relevant for this study and the number of times that the topic is found in one of the papers in a relevant context. Table 1 includes high level or general topics. Table 2 includes topics that are more specific; these are topics that can easily be translated to criteria for mHealth applications. An explanation of each table is given on the next page. The reason for this list of topics is that these topics were found during the 'review in full text' of the articles.

Table 1 Number of times high level topics are mentioned

Topic	Number of papers that mention the topic
Challenges	29
Assessment/assessed	28/13
Framework	22
Requirements	20
Barriers	17
Criteria	16
Regulation	13
Validation/Validated	9/9
Instrument	3

It can be seen that the topics in Table 1 are very general. These topics cannot be translated to direct criteria. Some of the papers only made a statement that there are 'challenges' or that an 'assessment' of mHealth applications is necessary. Other papers also gave an overview of 'challenges' or how an 'assessment' should be performed. This explanation applies for all of the topics in Table 1.

Table 2 Number of times low level topics are mentioned

Topic	Number of papers that mention the topic
Security	23
Privacy	21
Behavior change maintenance	17
Usability	16
Accessibility	12
Data transfer	11
Data security	6
Personalization	5
Timeliness	4
Reliability & validity and efficacy & effectiveness	2
App overload	2
Transparency	2
Speed of research vs. industry	1
Black boxes	1

Table 2 includes the low level topics that are relevant for an assessment according to scientific literature. It is clear from this table that 'security' and 'privacy' are both topics that are important. Topics that were mentioned less are timeliness of an application, personalization, app overload, transparency and personalization of applications. The next chapter includes a description of the topics mentioned above.

5 Potential requirements for mHealth assessments from literature

One of the goals of the extraction was to describe the potential criteria for requirements analysis of the mHealth instrument. Chapter 1.6 provides a short overview of the current challenges in mHealth. This chapter will give a more elaborate overview of these challenges and translate these challenges to requirements for an assessment instrument, providing an answer to sub-question two:

"What do we know about the added requirements of mobile health?"

Different papers describe challenges (or barriers) in the mHealth field. 29 papers of the SLR describe or mention challenges. The challenge most often mentioned is 'security' (23 times). Closely followed by 'privacy', 21 times, and 'usability' with 16 times. Table 3 presents an overview of challenges and the sub-chapter in which they are described.

Table 3 Number of times challenges are mentioned in articles

Challenge	Number of papers that mention the challenge	Described in sub-chapter
(Data) Security	23	5.1
Privacy	21	5.1
Behavior change maintenance	17	5.2
Usability	16	5.3
Data transfer	11	5.4
Trans disciplinary science	3	5.5
App overload	2	5.6
Transparency	2	5.7
Reliability & validity and efficacy & effectiveness	2	5.8
Speed of research vs. industry	1	5.9
Black boxes	1	5.10

5.1 Data security and participant privacy

An important issue according to many papers is privacy. 21 papers from the SLR acknowledge privacy as a challenge in mHealth. Some papers combine data security with privacy. Whittaker (2012) state that the privacy of users and the data security are important issues (Whittaker, 2012). It is important that GPS data, medical records and other mHealth related information is secured. Encryption of multimedia should be enforced since wireless technology can increase the risk of unauthorized access by third parties (Luxton, Kayl, & Mishkind, 2012). Another attention point are the application developers, since they tend to collect names, email addresses and other personal information when applications are downloaded. A solution by Luxton et al. (2012) is a data security policy that is standardized with specifications for encryption and secures communication. An advantage of a

standardized data security policy is interoperability between systems with a minimal risk for the users. The problem however is that most groups do not aim to create interoperable systems. Luxton et al. (2012) identified non-commercial and commercial developers; both groups are free to design interoperable systems. Commercial developers may attempt to design interoperable systems. Another option is that a government institution forces regulation on medical mobile applications.

One paper found with the explorative literature study also states that mHealth applications raise concerns about health data security and confidentiality (Patrick, Griswold, Raab, & Intille, 2009). Patrick et al. (2009) includes capturing personal health-related data from a mobile phone, up-linking it to a server, transmitting it to a web-based or other form of electronic personal health or medical record, using the data for interpretation and professional judgments in the care of that individual, and responses back to the person via, for example, a SMS message. This chain of events can be very complex and outside of the electronic medical record environment. Another problem is that other persons than the owner of the phone can use the mobile phone. Patrick et al. (2009) do not provide an answer in their article for this rising problem.

5.2 Maintenance of behavior change

17 papers from the SLR mention behavior change, this is important since a challenge with behavior change interventions is to sustain for example the weight loss after an intervention has ended. Burke et al. (2012) describes benefits of mHealth for weight loss interventions (Burke et al., 2012). These benefits could be applied to other behavior change interventions as well. A benefit of mHealth is that users can use their own device to run applications that could help after a program has ended. This extends healthcare beyond what is possible with personal approaches. Another benefit is that mobile apps can collect data frequently, interactive and unobtrusively, which can help boost self-monitoring (Burke et al., 2012). Studies without the use of mHealth were unable to record long-term progress. mHealth interventions are able to collect data over a long time, the only drawback is that the interventions are dependent on the devices used (Tate et al., 2013).

5.3 Usability

Usability is an important aspect of mobile applications and is mentioned 16 times in the papers from the SLR. However, it is difficult to develop a seamless and autonomous application that provides a service to a target group that is not familiar with technology (Boulos et al., 2011). Usability is a critical issue for a target group that is not familiar with technology and who also has physical (e.g. bad eyes) and/or cognitive disabilities (e.g. dementia). A paper by Brown et al. (2013) state that there are two challenges that have delayed the rigorous usability testing for mHealth technology (Brown, Yen, Rojas, & Schnall, 2013). The first challenge is that mobile devices present unique challenges: small, low-resolution screens, no mouse or keyboard, slow operating system and variable connectivity. The second challenge is that the “technology is rapidly advancing, but end-user testing equipment and software is making a much slower progress (Brown et al., 2013).”

5.4 Data transfer issues

According to eleven papers from the SLR is data transfer a notable challenge. Wrong information or feedback can be shown to participants if data or server transfer issues occur. Reasons of transfer issues are low battery power on the device, or other technical difficulties (Jordan-Marsh, 2011). Issues like this can demotivate and frustrate users or weaken the program efficacy. There is a risk that other persons than the persons who enrolled for the study fill in survey data or provide biometric input, since participants of mHealth studies are not supervised during data collection. If this happens, even when data are transferred successfully, the data will be incorrect for the research participant. Another issue is missing data as a result of technological errors; this could cause difficulties for data analysis.

5.5 Challenges of trans disciplinary science

A collaboration of experts from different fields is needed when mHealth applications are developed (Jordan-Marsh, 2011). These different fields have different scientific processes and have their own jargon. The challenge for interdisciplinary teams is to create a scientific process that incorporates the processes and languages of the different fields to solve problems. The industry may help to guide the development of standards for researchers from different disciplines. However, introducing another stakeholder could bring more challenges. Since stakeholders from different fields may have other guidelines for success in testing and dissemination.

5.6 App overload

Only two papers from the SLR mention app overload as a challenge in the mHealth field. The challenge with app overload is to find the right mHealth app in the app stores, since the number of applications grows every day. This is a problem for both medical professionals and patients (van Velsen, Beaujean, & van Gemert-Pijnen, 2013). Patients stated that they do not want too many health apps on their smartphone. Another study by Franko et al. (2012) state that physicians have difficulty in keeping an overview of all the available medical applications (Franko & Tirrell, 2012). Professionals find that the applications that they find have limited value for them (Payne, Wharrad, & Watts, 2012). Professionals also declared that separate applications do not contain enough material to keep them interested.

In short, the applications overload described above has the following consequences.

- Users have difficulty finding the right app.
- Too much fragmentation of information and/or features over too many applications, resulting in an added value that is too low.

The application overload is only getting larger since the number of applications grows.

5.7 Transparency

Another research states that transparency of its information towards the users is important. The following four points should be in every health application (Boulos, Brewer, Karimkhani, Buller, & Dellavalle, 2014).

1. Provide authorship information, including detailed information about the affiliations and credentials and about any medical professional involvement in content preparation.
2. List all references or sources of content (attribution).
3. Disclose any app sponsorship or other commercial funding arrangements, and any potential conflicts of interests.
4. Ensure a balanced, non-biased coverage of facts and information currency (up-to-datedness).

The (high-quality) content displayed in the application is of limited value if it is presented in such a way that the user does not understand it. Developers should address usability, accessibility, readability (reading with understanding) and health literacy needs of the user. Important is to focus on different stakeholders, demographic and if the user is physical or cognitive impaired.

5.8 Reliability & validity and efficacy & effectiveness

mHealth methods can only be recommended when they are tested. Rigorous evidence of effectiveness need to be objectively measured where possible (Kumar et al., 2013; Whittaker, Merry, Dorey, & Maddison, 2012). Kumar et al. (2013) group the testing in reliability and validity, and in efficacy and effectiveness of mHealth. They state that the golden standard is a RCT to determine the efficacy of a health intervention but that it takes too long for mHealth applications. The technology will probably be obsolete when the RCT is finished. They therefore included other designs to evaluate the efficacy and effectiveness of an application. Another solution is to use the method “Continuous Evaluation of Evolving Interventions” (CEEI). With CEEI substantively new versions are deployed along with the previous version, users are randomized to available versions. The version that is most efficacious is kept. This method can also be applied to ongoing evaluation of interventions as they are scaled up. The application can be continuously improved over time and adept to the changing technology.

5.9 Speed of research versus industry

The speed of technology is much faster than the scientific research needed to validate the applications. By the time the research process is finished, the technology researched may have become obsolete. The mobile technology industry aims to create devices that are attractive with software that persuade the customers. This results in a product that looks and feels professional. These products are not routinely tested for accuracy, reliability, validity, or use in clinical preventive settings. Researchers on the other hand develop empirically validated programs, but the technology that they use in outdated and cannot compete with the newer industry products. A problem is that these outdated products made for health research are poor at capturing and holding the interest of

patients such as children who are used to the new and sleek looking commercial devices, software and technology.

Other papers stated that there is an increasing amount of literature documenting mHealth studies. A number of these reviews noted the lack of quality evidence and research in the mHealth domain (A.T. Kearney, 2012; Mechael & Searle, 2010; Needs, Health, & Philbrick, 2013; Tamrat & Kachnowski, 2012). Lee identifies two reasons for the varying levels of rigor research in mHealth (Lee, 2013). The first reason is the multi-disciplinary nature of mHealth, since it combines the health and technology sector. These sectors use different methodologies to assess intervention, different speed and ways of distributing their findings. The second reason is the different pace of development in these sectors. The technology sector has a fast pace and a lot of iterations. Results are published in grey literature, whitepapers, conference papers, presentations and blogs. The health sector moves at a much slower pace. The focus is on methodological rigor research, which takes longer than the research in the technology sector. The majority of the results in the health sector are published in peer-reviewed journals and conference papers.

This discrepancy between the sectors is an issue, since the time for a rigorous methodological research to be completed and pear reviewed can take a couple of years. A peer-reviewed journal has furthermore much higher standards than a white paper. This result in different quality of evidence between the sectors (Lee, 2013).

5.10 Black boxes

Black boxes are only mentioned one time in the papers from the SLR. Tomlinson et al. (2013) call the current mHealth apps the equivalent of black boxes (Tomlinson, Rotheram-Borus, Swartz, & Tsai, 2013). Each entrepreneur or researcher includes whatever they like in an attempt to demonstrate the efficacy (Tomlinson et al., 2013). This results in a lot of small pilot studies finding out if e.g. text messaging works. A solution to this problem would be gateway applications. Gateway applications are applications that serve as a gateway to medical information and features. A gateway application can function as a portal that can lead the user to the demanded health content. Such an application should prevent a situation where users have a lot of single applications that offer help with single issues.

It is furthermore important to prevent information overload within these applications. It is probably difficult to find the right information when there are just a couple of applications with all the medical information. A solution to this problem is to personalize the information and the features. Only the information should be shown that align with the characteristics of the user (van Velsen et al., 2013). It is necessary for third-party developers to gain access to medical content if they need to create high quality health applications.

5.11 Miscellaneous criteria

A paper, found during the explorative literature study, from van Gemert-Pijnen et al. (2011) presents principles that should be integrated in a framework for eHealth (van Gemert-Pijnen et al., 2011). These guidelines are based on 16 existing frameworks and on a literature study. The focus of the paper is on the improvement of the uptake and impact of eHealth technologies. Gemert-Pijnen et al. drafted the following guidelines:

1. It is important that the stakeholders are involved during the development of e-health applications (Van De Belt, Engelen, Berben, & Schoonhoven, 2010).
2. The development should be iterative and flexible. Stakeholders must be able to give feedback.
3. The development is intertwined with implementation. Implementation issues must be identified and these issues must be addressed during development.
4. eHealth can change the organization of healthcare. There could be a shift from hospital based care to home based care.
5. eHealth applications should involve persuasive design principles. The technology should be able to motivate the users.
6. Methods are needed to assess the impact of e-health technology.

It is also important that the system meets the changing need of the users (Pingree et al., 2011). Users that are diagnosed with diseases such as cancer are likely to use the eHealth system for a long time. It is important that the eHealth system conforms the whole time span of the intervention.

5.12 Conclusion potential mHealth assessment

This chapter described potential criteria for mHealth applications and the challenges that coexist with these points. The points encompass a wide variety of topics; relevant topics and points need to be translated to criteria that can be included in the instrument. The content in this chapter was extracted from the systematic literature review. The results will be taken into account when developing the instrument.

6 Overview of existing guidelines and frameworks

One of the goals of the literature study was to determine how mHealth applications are currently assessed. This chapter will therefore provide an answer to the third sub-question:

"How are mobile health applications currently assessed?"

25 papers out of the SLR papers mention frameworks or an instrument to test applications. 28 papers write about assessment of mHealth applications and 13 papers wrote about the need of regulation. Not all of these papers give explicit guidelines on how to do an assessment or present a framework/instrument. Some of the papers only mention that an assessment should be executed before an application can be used or that a framework should be developed to make an assessment easier. Most of the papers that write about guidelines do this referring to government institutes such as the FDA. However, there are also papers that made a comparison between existing frameworks.

There are next to the scientific literature also organizations that present guidelines for assessing mHealth applications. These guidelines are presented after the scientific frameworks in chapter 6.4.

6.1 Existing frameworks

The guidelines and frameworks found during the SLR will be compared in this section. Table 4 below presents the theme of the framework and the number of frameworks that are focused on that theme. A total of 21 existing frameworks were found during the SLR.

Table 4 Numbers of frameworks per theme

Theme	Number of frameworks
Privacy	1
Development	2
Implementation	9
Assessment	4
Usability	1
Quality	1
Other	3
Total	**21**

None of the frameworks includes a full assessment of the application and also the frameworks that are focused on assessments are not sufficient. However, parts of these existing frameworks can be used as input. A long list of papers that have a framework/model found during the SLR is depicted in Table 5 below. A more elaborate table including the author(s), title, and year of the paper can be found in Appendix C: Long list models and frameworks.

Table 5 Long list scientific frameworks/models

#	Name framework/model
	Privacy
1	A threat taxonomy for mHealth privacy
	Development
2	Development and intervention process mHealth
3	Simplified evaluation process
	Implementation
4	Barriers to the diffusion of telemedicine
5	eHealth readiness assessment tools
6	Unified Theory of Acceptance and Use of Technology (UTAUT) applied to telehealth
7	Seven Core Principles for the successful development of telemedicine systems
8	Lessons in telemedicine service innovation
9	Framework for Assessing the Health System Challenges to Scaling up mHealth
10	Comprehensive Model for the Evaluation of Telemedicine
11	The Layered Telemedicine Implementation Model
12	The Khoja-Durrani-Scott (KDS) Evaluation Framework
	Assessment
13	Model for Assessment of Telemedicine Applications: MAST
14	Research stages and standards for the dissemination of MHealth
15	HOT-fit
16	Evaluation Framework
	Usability
17	Health IT Usability Evaluation Model
	Quality
18	Perceived service quality of mHealth
	Other
19	MHealth Grading Tool
20	Framework for Evaluating Mobile Applications for Cardiac Rehabilitation
21	Dimensions of eHealth Research

The long list will be subjected against in- and excluding criteria. The models that remain will be compared among each other.

6.2 In- and exclusion criteria

Different in- and exclusion criteria were created to make the long list manageable and to filter the list of models that are less relevant for this study.

The following inclusion criteria were created:

In-1
Frameworks that include assessments on efficiency and effectiveness are included since healthcare institutes will only use applications that are tested on these points.

In-2
Frameworks that assess applications on multiple levels are interesting since they provide insight in what important is for different levels of testing. Which criteria are important to test on what level and what is the reason for these levels? Examples are frameworks that assess during different phases of the application (development, integration, sustained operation) or have different levels of analysis (individual and society).

The **inclusion** criteria in short:

- In-1 Includes efficiency/effectiveness testing of applications;
- In-2 Framework has multiple levels of assessment.

The following exclusion criteria are established to filter the long list of frameworks that are not relevant for the comparison.

Ex-1
It is unlikely that frameworks from before 2005 are relevant for mHealth as we know it today. Modern smartphones were introduced in 2005; frameworks from before this year are not relevant for this research and are excluded.

Ex-2
Frameworks are excluded when there is such a lack of information that a comparison with other frameworks is impossible to make.

Ex-3
Frameworks are excluded when their focus is too narrow. Frameworks that focus on a single item/subject are not relevant for comparison. Such frameworks could only be compared with other frameworks that have the same narrow focus of a certain subject. The goal of this research is to have an instrument that can do a full assessment of mHealth applications. The purpose of the comparison is to get an overview of existing methods to assess mHealth applications. Focusing on frameworks with a single subject do not give a full overview of existing methods and are therefore not relevant. An example would be a framework that is only focused on the encryption of data in the application.

Ex-4
Some of the frameworks in the long list have a target group that is not relevant. Examples are frameworks that are focused purely on developers.

Ex-5
Some frameworks are found through other articles, these articles provide a description or summary of that framework. The frameworks that are described are excluded when the original article is not available for free.

The **exclusion** criteria in short:

- Ex-1 Paper is published before 2005.
- Ex-2 Framework is too brief and the article does not provide enough information.
- Ex-3 The focus of the framework is too narrow.
- Ex-4 Focus of the framework is not relevant.
- Ex-5 Paper is not free and not available in English or Dutch.

Table 6 below gives an overview of all the frameworks and the reason of their exclusion. Framework number 10 is included despite conforming to exclusion criteria number 5. The reason for this is that a review of the framework was available. This review was found to be elaborate enough.

Table 6 Reason of exclusion frameworks/tools/models

#	Frameworks/tools/methods	Reason of exclusion
1	A threat taxonomy for mHealth privacy	Ex-3
2	Development and intervention process	Ex-4
3	Simplified evaluation process	Ex-4
4	Barriers to the diffusion of telemedicine	Ex-1
5	eHealth readiness assessment tools	Ex-4
6	Unified Theory of Acceptance and Use of Technology applied to telehealth	Ex-1
7	Seven Core Principles for the successful development of telemedicine systems	Ex-1
8	Lessons in telemedicine service innovation	Included
9	Framework for assessing the health system challenges to Scaling up mHealth	Included
10	Comprehensive Model for the Evaluation of Telemedicine	Included (Ex-5)
11	The Layered Telemedicine Implementation Model	Ex-3 & 4
12	The Khoja-Durrani-Scott (KDS) Evaluation Framework	Included
13	Model for assessment of Telemedicine Applications: MAST	Included
14	Research stages and standards for the dissemination of mHealth	Ex-3
15	HOT-fit	Ex-2
16	Evaluation framework	Ex-4
17	Health IT Usability Evaluation Model	Included
18	Perceived service quality of mHealth	Ex-2
19	mHealth Grading tool	Included
20	Framework for Evaluating Mobile Applications for Cardiac Rehabilitation	Ex-2, 3 & 4
21	Dimensions of eHealth Research	Ex-2

In Table 7 is a list of the included scientific frameworks/methods and the sub-chapter where they are described.

Table 7 Included frameworks

	Frameworks/methods	Sub-chapter
1	Lessons in telemedicine service innovation	6.2.1
2	Framework for assessing the health system challenges to Scaling up mHealth	6.2.2
3	Comprehensive Model for the Evaluation of Telemedicine	6.2.3
4	The Khoja-Durrani-Scott (KDS) Evaluation Framework	6.2.4
5	Model for assessment of Telemedicine Applications: MAST	6.2.5
6	Health IT Usability Evaluation Model	6.2.6
7	MHealth Grading tool	6.2.7

6.2.1 Lessons in telemedicine service innovation

A study of 8 years was conducted (1997-2005) to identify factors that promote successful use of teledermatology (Finch, Mair, & May, 2007). This resulted in a list of five themes, which are focused on the implementation of telemedicine.

Theme 1: Policy context

Policies should be introduced to ensure that innovation through telehealth is encouraged instead of discouraged. An important step is to translate the policies to resources.

Theme 2: Perceived benefit and related commitment

There is a link between the perceived benefits of technology and the commitment of role players. The article states that the benefits clearly outweighed the effort and commitment that was required to make the system work in their research.

Theme 3: Evidence gathering, proving safety and managing risk

The most successful systems were those where the risks were identified and safeguards were integrated in the systems.

Theme 4: Reconfiguring services

It is important that the technology is accepted. This can be achieved by modifying the service so that is matches the users need and by providing support. It is not about the technology but about the way the service is delivered to the user.

Theme 5: Professional roles and boundary crossing

It is important that the role of the professional is altered so that it can include the use of a telemedicine service. Current role descriptions of professional often do not mention the use of such a technology.

Unfortunately mHealth was and could not be included in this study. However, these themes can be taken into consideration since the focus is on implementation of a service and not solely on telemedicine.

6.2.2 Framework for assessing the health system challenges to Scaling up mHealth

This framework exists of four dimensions with requirements for each dimension (Figure 7) (Leon, Schneider, & Daviaud, 2012). An explanation of the dimensions and the requirements are in Table 8. The framework is designed for areas with a weak ICT environment and limited implementation capacity. The requirements are points that should be evaluated before mHealth is implemented so that possible risks and pitfalls are known.

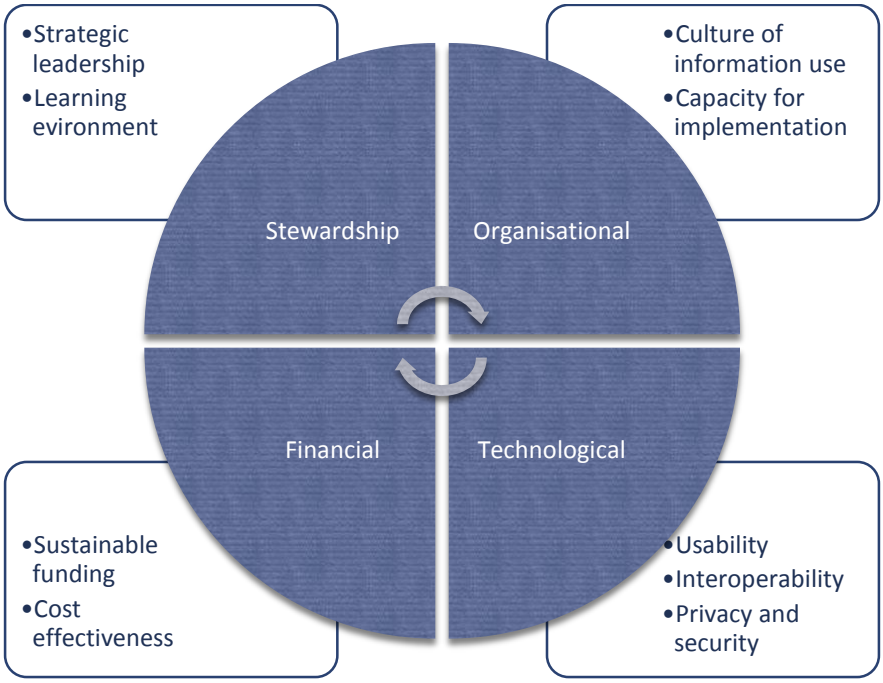


Figure 7 Dimensions framework for assessing health system challenges (Leon et al., 2012)

Table 8 Explanation of the dimensions in the Scaling up mHealth framework

Health system dimensions	Health system capacity requirements
Government stewardship: **Is there a policy environment supportive of mHealth?**	**Strategic leadership:** Are there policy guidelines that promote health goals, funding sources, ICT standards, and collaboration partnerships. **Learning environment:** Systematic evaluation of projects, a learning environment and sharing of knowledge.
Organizational: **Is there a culture of and capacity for using information technology for management?**	**Capacity for implementation:** Capacity to implement mHealth interventions, a functional ICT environment, assessment of eReadiness and effective mechanism for implementation, support and evaluation **Culture of information use:**

	Organizational culture of using health information for management.
Technological: How useable, integrated and sustainable is the chosen technology?	**Usability:** Technology has ease of use, flexibility, and durability and is beneficial to end-users. **Interoperability:** Can communicate with other systems across technological and information platforms, integrates with existing work practices, use common standards and is financial sustainable. **Privacy and security:** Privacy and security of data is ensured.
Financial: is adequate financial provision being made for the medium to long term use of mHealth	**Sustainable funding:** Sustainable funding for large-scale implementation, clear business and funding plans. **Cost-effectiveness:** Cost-effectiveness of mHealth is evaluated. Intervention is weighed up against other evidence-based interventions.

6.2.3 Comprehensive model for evaluating telemedicine

Van Dyk (2014) made a short summary of the comprehensive model for evaluating telemedicine (van Dyk, 2014). This is used since the original article is not available for free. The model has the following three dimensions:

Level of analysis
Existing of three categories (individual, community and society) each category has multiple elements. The acceptability may vary across the three levels, since benefits and costs may vary.

Focus of analysis
In this dimension the consideration between cost, quality and access to healthcare is made.

Activities of analysis
This dimension gives an overview of the uses of telehealth. It can be used for education, research, consultation, discussion or other clinical purposes.

All the categories of each dimension can be combined with another category. Some examples provided by van Dyk (2014):

- The cost of education/research concerning the service at the level of an individual;
- The level of community access to clinical services;
- The level of quality of the clinical services to society;
- An individuals level of access to administrative services;
- The quality of education/research concerning the service at the community level;
- The cost of administrative services at the society level.

6.2.4 The Khoja-Durrani-Scott (KDS) Evaluation Framework

The KDS evaluation framework has seven themes that can be matched against the stages of an eHealth lifecycle (Khoja, Durrani, Scott, Sajwani, & Piryani, 2013). There are three sets of questionnaires, depending on the role of the respondent. There is a set for “managers”, “healthcare providers” and for “clients”. Separate tools were developed for every role. However, these tools are not available.

The following themes are identified:

1. **Health services outcomes:** this theme is based on the principle that the intervention has to produce a change in the health status of a patient or community and that this change can be measured on the basis of:
    a. Change in disease or health status
    b. Impact on quality of life
    c. Change in health indicators
2. **Technology Outcomes:** this theme refers to software, hardware and connectivity infrastructure used for the eHealth solution. This can be measured in terms of appropriateness, relevance, use, safety and effectiveness of the technology.
3. **Affordability and cost-effectiveness:** this considers the extent to which a service is affordable for the user. This is measured by the willingness to pay for the service by the user.
4. **Social and behavioral impact:** focused on the impact of the intervention, this includes the processes of analyzing, monitoring and managing the consequences both negative and positive.
5. **Ethics:** what is good for the user and for the society. Points that should be paid attention to are security, research, resource allocation, use and access to the solution.
6. **Readiness and change:** this refers to the readiness of healthcare institutions and individuals to implement or use the eHealth program. It may be necessary to change business processes and ensure appropriate training and support for adopting this new process.
7. **eHealth policy:** a set of statements, directives, regulations and laws that direct and mange the life cycle of eHealth interventions. This is required to have a structured and consistent eHealth practice.

The seven themes above can be matched against the stages of the eHealth lifecycle. The four stages that are identified:

1. Development
2. Implementation
3. Integration
4. Sustained Operation

6.2.5 Model for assessment of telemedicine applications: MAST

Mast is a model with the following three elements (Kidholm et al., 2012):

1. Preceding considerations
2. Multidisciplinary assessment
3. Transferability assessment

The elements with their sub elements can be viewed in Figure 8.

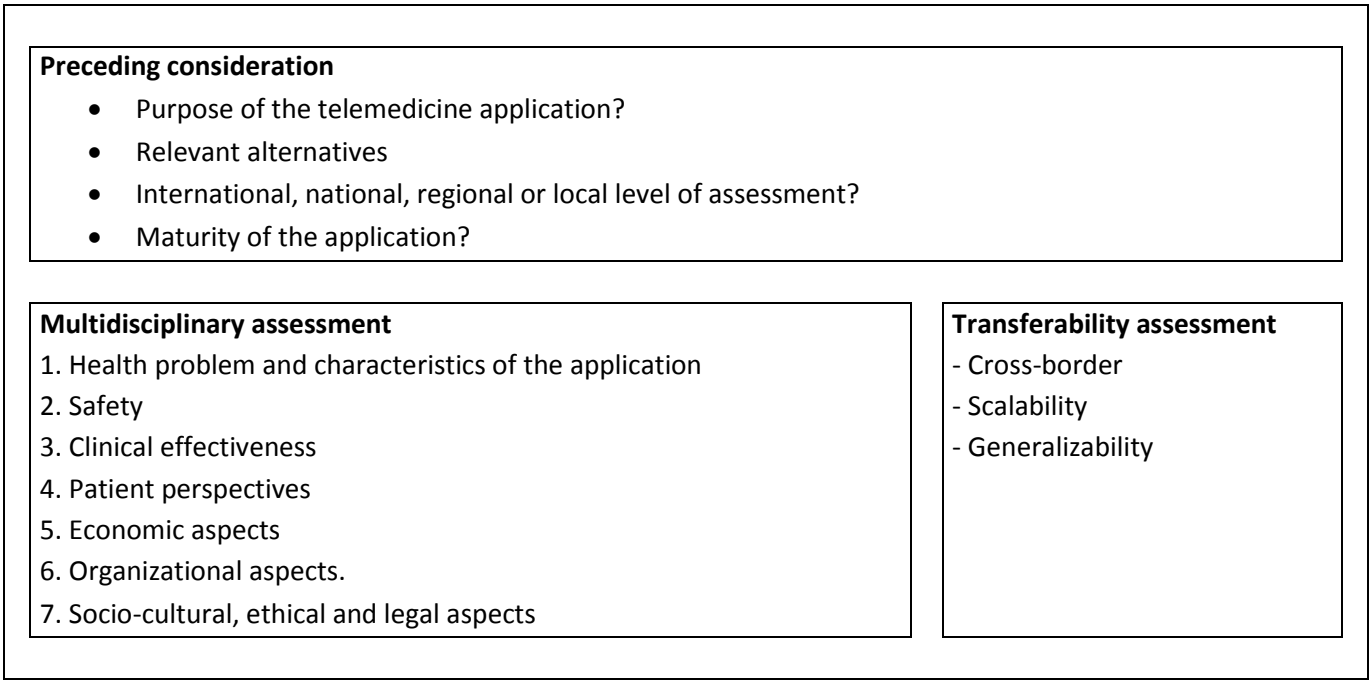


Figure 8 MAST model

Step 1: Preceding Considerations

The first step is to determine the aim of the application and if there are any relevant alternatives. The following conditions should be considered to identify any barriers or issues that must be addressed before an assessment of the application.

Legislation: before the application is introduced it should be determined if the application is in accordance with legislation.

Reimbursement: who is going to pay for the application. The business case and if the application is economically feasible can change when the application is expensive and insurance is not going to cover.

Maturity: Development takes time and it is important to know in what face the current application is. Evaluation should establish safety of the application. Kidholm et al. (2012) refers with safety to the question if the application is an immediate risk or harm to patients. After safety, evaluations of feasibility should be carried out. The last evaluation should be to test the effectiveness of the application. It is important that the application is in a state of development where there are not substantial changes anymore. Testing with e.g. a RCT could become problematic when there are to

many (large) changes. Other tests should be carried out if the application is not yet mature enough for a RCT. These tests should have the intention to improve the service.

Number of patients: Implementation can involve large investments in equipment and integration with existing information systems. Education of staff and change in the existing workflow can bring substantial costs. It is necessary to make an assessment of the number of patients that will use the application to test if the implementation is economically feasible.

Step 2: Multidisciplinary Assessment

As can be seen in Figure 8, the assessment involves multiple disciplines, with outcomes in seven domains. Kidholm et al. (2012) made definitions for each domain, including topics that are important in for each domain (Kidholm et al., 2012). These can be viewed in Table 9; the topics are issues for consideration within the domain according Kidholm et al. (2012). The paper further state that the methods and study designs used for data collection in each domain, are in general methods and designs that follow the state of the art research methods. Another requirement is that these methods and design produce valid and reliable estimates. The choice of outcome measures can differ per application, patient group or the organization that is using the application.

Step 3: Assessment of Transferability

Important is that the results from a study of healthcare technology can be transferred from one setting to another setting. These assessments should include information that could be relevant for other situations. This could be information regarding costs per patients, or problems with the validity and reliability of the studies that were included during the research. Information regarding these subjects should be mentioned for other healthcare institutes.

Table 9 The domains in MAST, copied from (Kidholm et al., 2012)

Domain	Definition	Topics
1. Health problem and description of the application	Description of the health problem of the patients expected to use the telemedicine application and the application being assessed incl. description of the current use.	• Health problem • Description of the application • Technical characteristics • Current use of the application
2. Safety	Identification and assessment of harms	• Clinical safety (patients and staff) • Technical safety (technical reliability)
3. Clinical effectiveness	Effects on the patients health	• Effects on mortality • Effects on morbidity • Effects on health related quality of life (HRQL) • Behavioral outcomes

		• Usage of health service
4. Patients perspectives	Issues related to the perception of the patient or the relatives of the telemedicine application including the patients and relatives acceptance of the technology	• Satisfaction and acceptance • Understanding of information • Confidence in the treatment • Ability to use the application • Access and accessibility • Empowerment, self-efficacy
5. Economic aspects	A societal economic evaluation comparing a telemedicine application with relevant alternatives in terms of costs and consequences and a business case describing the expenditures and revenues for the healthcare institutions using the telemedicine application.	Economic evaluation: • Amount of resources used when delivering the application and comparators • Prices for each resource • Related changes in use of healthcare • Clinical effectiveness Business case: • Expenditures per year • Revenue per year
6. Organizational aspects	Assessment of what kind of resources has to be mobilized and organized when implementing a new technology, and what kind of changes or consequences the use can further produce in the organization.	• Process • Structure • Culture
7. Socio-cultural, ethical, and legal aspects	The socio-cultural aspects include the social-cultural arenas where the patient lives and acts during use of the application. The ethical analysis appraises the ethical questions raised by the application itself and by the consequences of implementing it or not. Legal aspects focus on the legal obligations that must be met and any specific legal barriers that may exist to the implementation of the application.	• Ethical issues • Legal issues • Social issues

6.2.6 Health IT Usability Evaluation Model

A paper by Brown et al. (2013) demonstrates the applicability of the Health IT Usability Evaluation Model (Health-ITUEM) (Brown et al., 2013). The reason for this study is that there are many studies conducted to explore usability requirements, problems and design solutions. However, almost none of these studies evaluated the usability of mobile technologies. The Health-ITUEM was developed using existing models. Definitions from TAM and ISO 9241-11 were used together with Nielsen's ten heuristics, Shneiderman's eight rules for user interface design and Norman's seven principles for design. Interviews with users and developers provided additional concepts for the framework. The following nine concepts are included: Error prevention, Completeness, Memorability, Information needs, Flexibility/Customizability, Learnability, Performance speed and Competency. The four overarching constructs are: Quality of (Work) Life, Perceived Usefulness, Perceived Ease of Use, and User control. Table 10 gives a description of all the concepts. Each of these nine concepts should be measured as negative, neutral or positive by interviewing users of the application. The authors claim that this model produces a robust usability evaluation framework. The framework can be viewed in Figure 9.

Table 10 Description concepts Health-ITUEM

Title	Description
Error prevention	System offers error management, such as error messages as feedback, error correction through undo function, or error prevention, such as instructions or reminders, to assist users performing tasks.
Completeness	System is able to assist users to successfully complete tasks. This is usually measured objectively by log files for complete rate.
Memorability	Users can remember easily how to perform tasks through the system.
Information needs	The information content offered by the system for basic task performance, or to improve task performance.
Flexibility/ Customizability	System provides more than one way to accomplish tasks, which allows users to operate system as preferred.
Learnability	Users are able to easily learn how to operate the system.
Performance speed	Users are able to use the system efficiently.
Competency	Users are confident in their ability to perform tasks using the system, based on Social Cognitive Theory.
Other outcomes	Other system-specific expected outcomes representing higher level of expectations. (Uses of non-phone app technology, non-mobile resources, other health related entities outside of study protocol)

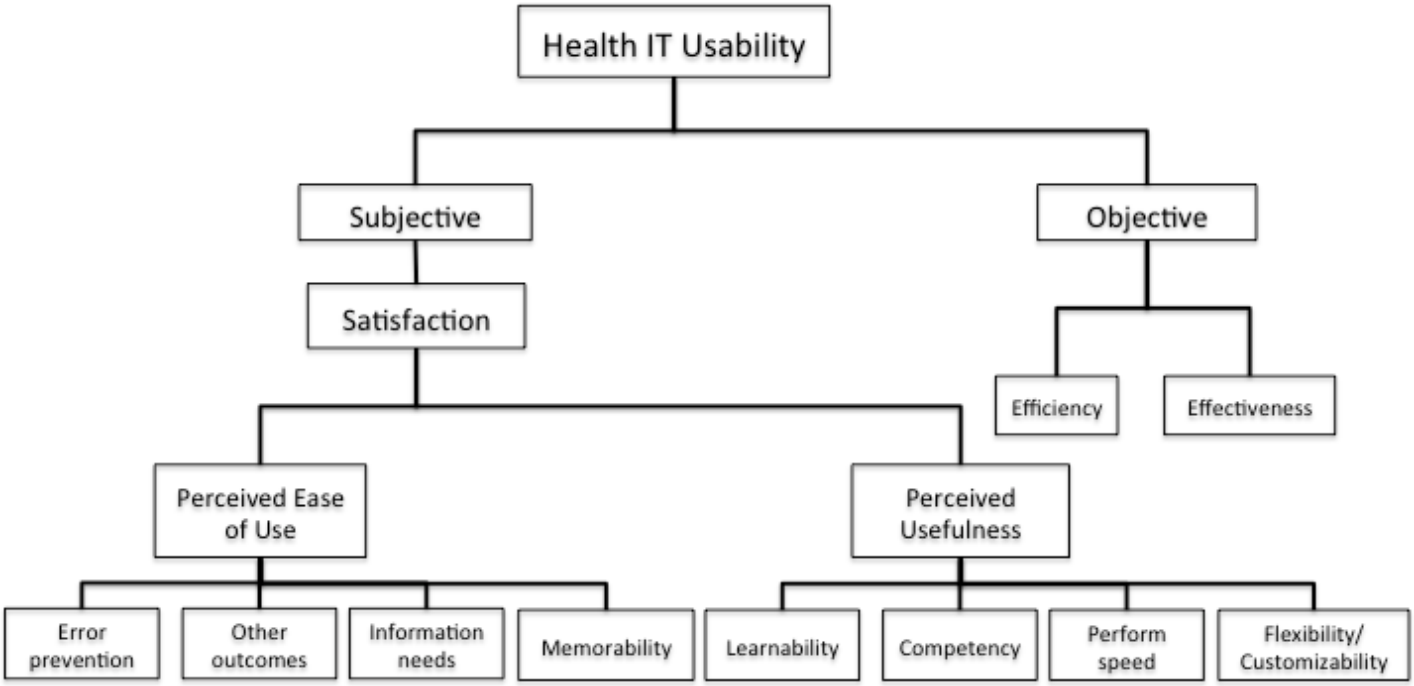


Figure 9 Health-ITUEM framework

6.2.7 MHealth Grading tool

A paper by Lee (2013) presents a tool to examine the quality of information in mHealth studies (Lee, 2013). This is necessary according to Lee (2013) since the majority of the studies lack clarity, transparency and rigor. The tool itself is not relevant for this research. However, a part of the tool is focused on mHealth applications. These criteria should be measured in an mHealth application according to Lee (2013). The criteria including a description can be viewed in Table 11. Important is that the criteria as well as the description are designed as a check if an mHealth study did cover all the important aspects.

Table 11 Criteria mHealth grading tool from Lee (2013)

Criteria	Description
Infrastructure	Clearly presents the availability or kind of infrastructure to support technology operations (e.g. electricity, access to power, connectivity).
Technology architecture	Describes the technology architecture including the software and hardware.
Intervention	mHealth intervention is clearly described with frequency and mode of delivery of intervention (i.e. SMS, face-to- face, interactive voice response) for replication. Details of the content of the intervention are clearly described or link is presented and content is publically available.
Usability	Clearly describes the ability of different user groups to successfully use the technology in a given context e.g. literacy, computer/Internet literacy, ability to use device.
User feedback	Describes user feedback about the intervention.
Identifies constraints	mHealth solution states one or more constraints in the delivery of current service, intervention, process or product.
Access and affordability	Presents data on the access and affordability of the mHealth solution from varying user perspectives.
Cost assessment	Presents basic costs assessment of the mHealth intervention from varying perspectives.
Training inputs	Clearly describes the training inputs for the adoption of the mHealth solution.
Strengths and limitations	Clearly presents mHealth solution considerations, both strength and limitations, for delivery at scale.
Language adaptability	Describes the adaptation, or not, of the solution to the local language.
Replication	Clearly presents the source code/screenshots/flowcharts of the algorithms/ examples of messages to ensure replication.
Data security	Describes the data security procedures/ confidentiality protocols.

6.3 Comparison criteria from selected frameworks

Table 12 on the next page presents a comparison from the selected frameworks. The categories or a synonym of these categories were subtracted from the selected frameworks. The categories are high level and the majority of the frameworks have sub-categories that are not presented in this table. The last column "total" presents how many times each category is found in the compared frameworks. The bottom row "total" is the total of categories that were found in each individual framework. It can be concluded from Table 12 on the next page that "Financial" and "Evidence" are the two most important items, each found five times in the existing frameworks. Followed by "Organizational/Readiness", "Usability" and "Security", which are found four times. The frameworks that are the most elaborate are:

1. MAST
2. KDS framework

MAST has nine out of ten criteria included. Followed by the KDS framework that has seven criteria included. However, this does not mean that these two frameworks are almost complete. This comparison only includes criteria that are found in existing frameworks. It could be that there are criteria missing or that the framework needs more detail on certain criteria. All of the categories found in the selected frameworks will be used as input while developing the assessment instrument.

Table 12 Comparison scientific frameworks

	1	2	3	4	5	6	7	Total:
Categories	Telemedicine innovation	Scaling up mHealth	Evaluation of telemedicine	KDS	MAST	Health IT Usability	mHealth Grading tool	
Organizational /Readiness	X	X		X	X			**4**
Technological		X		X			X	**3**
Financial		X	X	X	X		X	**5**
Policy/Legal		X		X	X			**3**
Evidence	X		X	X	X		X	**5**
Ethics				X	X			**2**
Usability		X			X	X	X	**4**
Security		X		X	X		X	**4**
Preliminary research					X			**1**
Transferability					X		X	**2**
Total:	**2**	**6**	**2**	**7**	**9**	**1**	**6**	

6.4 Guidelines by organizations and associations

The first part of this chapter included frameworks and models found in scientific literature. The second part of this chapter is based on the guidelines from organizations. These guidelines are non-scientific.

There are various organizations and associations that published evaluation criteria for mHealth applications. The criteria of the following organizations is listed below:

1. The Agency for Healthcare Research and Quality (www.ahrq.gov)
2. The Arizona Health Sciences Library (http://libguides.library.arizona.edu/ahsl)
3. The Medical Library Association (https://www.mlanet.org/)
4. The Healthcare Information and Management System Society (HIMMS) (http://www.himss.org/)

The organizations and associations where found by searching for "mHealth evaluation tools", "how to evaluate medical apps" and other similar queries. An important criterion was that it could not be a commercial organization, since this could create conflicts of interest. AHRQ is part of the U.S. Department of Health and Human Services. The Arizona Health Sciences Library is part of the University of Arizona. The Medical Library Association and HIMMS are non-profit organizations with members worldwide. Both organizations have their headquarters in Chicago. This list is biased for the following reasons:

- The search was focused on organizations and associations;
- Only English guides are included.

6.4.1 The Agency for Healthcare Research and Quality

The Agency for Healthcare Research and Quality[3] published the following list with important evaluating criteria for health information on the internet (Agency for Healthcare Research and Quality, 1999). They published this list since the amount of information regarding health on the web is growing. It is important according to the AHRQ that internet users can evaluate this information in an objective, reproducible manner. This list or parts of it can be adopted for mobile applications.

- **Credibility:** includes the source, currency, relevance/utility, and editorial review process for the information.
- **Content:** must be accurate and complete, and an appropriate disclaimer provided.
- **Disclosure:** includes informing the user of the purpose of the site, as well as any profiling or collection of information associated with using the site.
- **Links:** evaluated according to selection, architecture, content, and back linkages.
- **Design:** encompasses accessibility, logical organization (navigability), and internal search capability.

[3] www.ahrq.gov

- **Interactivity:** includes feedback mechanisms and means for exchange of information among users.
- **Caveats:** clarification of whether site function is to market products and services or is a primary information content provider.

6.4.2 The Arizona Health Sciences Library

The Arizona Health Sciences Library[4] presents a guide with the following focus points for students to evaluate applications. The authors of this guide do not provide any sources that helped them creating this guide; the guide is nonetheless still useful and provides information on how to assess mobile applications. The focus points of this guide are:

- Accuracy
    - Is the information correct, are there references, does the app what it intends to do?
- Bias/Objectivity
    - Objective information, what kind of organization sponsored the app, are there advertisements, is the app unbiased?
- Authority
    - Who developed the app and did experts review the app/content?
- Currency/Timeliness
    - Is the app regularly updated, medical information still relevant?
- Usability
    - Does the app work stable, smooth navigation and is the app easy to understand?
- Scope/Completeness
    - Is the medical information complete and are there sources?

The guide also mentions that it is important to test an application before it should be used. Multiple clinical scenarios should be created and tested. The application should pass the test on the points mentioned above in multiple scenarios. It does not provide information on how to create medical clinical scenarios and the requirements of such a test. The complete guide can be viewed in Appendix D: AZHIN Guide.

[4] http://libguides.library.arizona.edu/ahsl

6.4.3 The Medical Library Association

The Medical Library Association[5] also presents tips that should help users in selecting a website with health information. The tips start with some general remarks on the search terms that users should use and that users should focus on sites from the government or from acknowledged health institutes. The tips have a lot in common with the guide from the Arizona Health Sciences Library, but are not as elaborate and complete. The main points of the Medical Library Association are:

- Who is the owner of the site?
- Is the information updated?
- Is the information unbiased?
- Who is the intended audience?

6.4.4 The Healthcare Information and Management System Society

The Healthcare Information and Management System Society[6] (HIMMS) published an elaborate guide on evaluating the usability of medical applications. The guide is focused on healthcare providers and on the IT staff members that are engaged with the selection of medical applications for their healthcare organization. HIMMS notes that the selection of an application depends on various factors, including price, user reviews, usability and security.

The HIMMS guide focuses on the usability of mobile applications, a list with mobile design tenets is included in this guide. The following four areas are the most important design tenets according HIMMS.

- **Data**, the focus should be on the data. A simple and intuitive interface with clear design elements that are not interfering with the data on a (possibly) small screen.
- **Layout**, a consistent layout that is meaningful for the user.
- **Feedback**, users can provide feedback to the developer in an easy way.
- **Interactions** with the interface are with the content that is currently presented, keeping the cognitive burden of the user to a minimum.

The four design tenets that are focused on by HIMMS will result in a higher score on usability if they are designed correctly. However, a full usability test should be conducted in order to test the usability of mHealth applications instead of focusing just on these four points.

The points from HIMMS are all related with usability. Usability is defined by ISO as:

“The extent to which a product can be used by specified users to achieve specified goals with **effectiveness**, **efficiency** and **satisfaction** in a specified context of use (Jokela, Iivari, Matero, & Karukka, 2003)”.

The point’s effectiveness, efficiency and user satisfaction can all be measured. These points are explained below.

---

[5] https://www.mlanet.org/

[6] http://www.himss.org/

The difference between efficiency and effectiveness is that efficiency is the amount of time that it takes for users to complete their tasks. Effectiveness is how accurate and complete users can complete their tasks. Important is that effectiveness also includes how quick an error is made by users. User errors can lead to inaccurate or incomplete patient records. Satisfaction is a subjective issue, questionnaires should be provided among users to gain insight in the problems users have with the system. HIMMS also includes ease of learning and platform optimization in their guide to application usability. Platform optimization is how effective and efficient the application is using the capabilities of the device it is running on. HIMMS furthermore suggests to redesign smartphone applications that also available for tablets, since tablets have bigger screens and a different resolution.

It is important that effectiveness, efficiency and satisfaction are handled as one component instead of separate components. All three components influence each other and should be balanced according to the goals and priorities of the application.

The usability of an application also depends on the visual design of the interface. Applications that have a messy design, use a lot of colors, or visual elements have a lower usability then applications that have a design that is well thought out. A simple design in appearance does not increase cognitive burden for the user, e.g. they will not get tired that easily. A good design will increase the usability of the application instead of lowering the usability. The guide furthermore points out that functionality is different from usability. Users of the guide should understand the difference.

The mobile design tenets from HIMMS are adopted from design tenets for software in general. The design tenets from HIMMS can be viewed in Appendix E: HIMMS Mobile design tenets.

The guide furthermore describes the following three steps to evaluate a medical application.

1. **What are the goals of your healthcare organization and how do these goals relate to usability?**
    a. Focus points can be to improve the efficiency of prescription refill requests, using patient data for education or fast training of clinicians.
    b. Baseline measures should be set that are related to the goals of the healthcare organization.
    c. Efficiency can be measured by how long a certain tasks take to complete.
    d. User satisfaction can be measured with a questionnaire about the application.
    e. Ease of learning can be measured by having users attempt the same task and objectively measure learning based on the number of attempts.
2. **Are there other resources that can be checked for reviews of the applications?**
    a. Other sources on the internet should be checked if the application is publicly available. There are websites that rate applications and there is the user rating in the app store. Other professional organizations can also assist in checking the usability of a medical application.
3. **Assess the usability of the application with clinical scenarios.**
    a. It is necessary to perform tests to assess the usability of an application thoroughly. HIMMS suggest creating a set of clinical scenarios that are representative for the healthcare organization. These scenarios should include the essential and frequent tasks that the app will support and if possible also complex tasks.

6.5 Comparison guidelines by organizations

Table 13 below gives a comparison between the four different guidelines from chapter 4.1. An empty cell means that the guide does not mention the item. A minus (-) sign means that the guide does mention it but does not provide any explanation on how to do it. A plus (+) sign means that the guide does mention it and provide some information with what is important and on how to evaluate these items. The double plus (++) sign means that the guide provides an elaborate explanation of the item, what important is to focus on and also include examples. The AZHIN and the HIMMS guide are the most elaborate ones. HIMMS is more elaborate then AZHIN since a major part of the guide is focused on testing the usability of the application. HIMMS furthermore provides suggestion on how to do clinical testing and how this can differ for each health organization.

Table 13 Comparison guidelines by organizations

	1. AHRQ	2. AZHIN	3. MLA	4. HIMMS	Total
Credibility	-	+		+	**3**
Accuracy	-	+		+	**3**
Objectivity	-	+		+	**3**
Authority	-	+	+	+	**4**
Currency	-	+	+	+	**4**
Scope	-	+		+	**3**
Usability	-	+		++	**4**
Assess with clinical scenarios		+		++	**3**
Feedback to vendor				+	**1**
Total	**7**	**8**	**2**	**11**	

6.6 Conclusion

This chapter presented an overview of existing frameworks and guidelines that are used for assessing mHealth applications. The comparison made from the selected frameworks shows that there is not a single framework that is complete. However, it can be concluded that a wide variety of subjects are relevant while assessing mHealth applications. The compared frameworks all had multiple themes or categories to make a distinction between different topics. The structure of categories that is used in existing frameworks can be used as an inspiration. Examples of such frameworks are the KDS and the MAST model. KDS identified seven important themes, while MAST identified three main themes.

The second part of this chapter focused on guidelines from organizations. The advantage of these guidelines is that, unlike the scientific frameworks, they are focused on assessing mHealth applications. The comparison shows that there is quite some difference between the guidelines. It is furthermore clear that the topics covered by scientific frameworks differ from the topics covered by guidelines. It could be that this is because the focus of the scientific frameworks is not on assessment of mHealth applications. Another possible reason is the difference in target audience. The scientific frameworks are developed for organizations, while the guidelines are focused on the users of mHealth applications. HIMMS is the only guideline that can be excluded from this, since their target audience is healthcare providers and IT staff.

The criteria from the selected frameworks and guidelines will be used as input for this research. Findings from the literature study and this chapter are combined in the next chapter.

7 Towards a first version of the assessment instrument on mHealth

This chapter will present a conceptual framework by combining the relevant findings of the previous chapters. This is visualized in Figure 10 below. A short list is made from the combined results. This list is used as input for a focus group where experts representing the various stakeholders joined. This should provide an answer to the following sub-questions:

"What are assessment criteria and what level of evidence is needed for a swift and critical appraisal instrument for mobile health applications?"

"How can the criteria be modeled into a critical appraisal instrument that can be tested?"

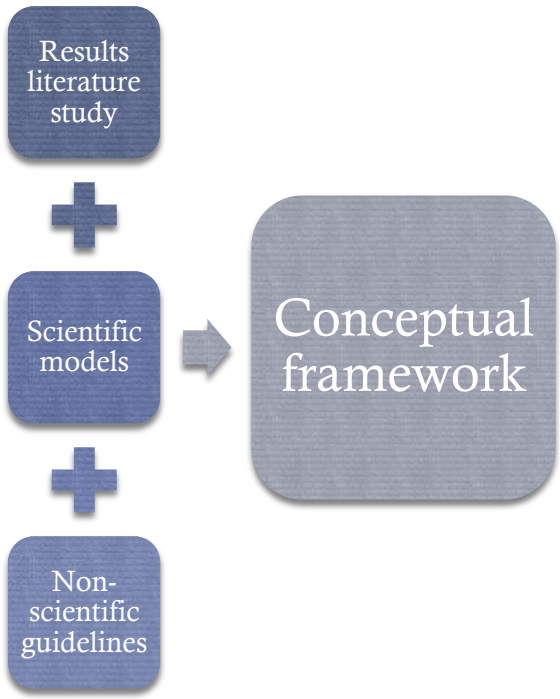


Figure 10 Separate elements conceptual framework

7.1 Long list potential criteria

The long list of potential criteria exists of all the criteria found in the previous chapters. There are 40 items identified, these items are found in literature, selected frameworks and guidelines. Not all of the criteria found are relevant for the final instrument. Exclusion criteria were established to keep

the list relevant. The criteria and if they are in or excluded can be viewed in Table 14. The list was filtered with the following exclusion criteria:

Ex-1
Criteria that focus on **implementing** applications are excluded since implementation is out of scope of this research.

Ex-2
Criteria that focus on **developing** applications are excluded since development is out of scope of this research.

The exclusion criteria in short:

- Ex-1 focused on implementation of applications.
- Ex-2 focused on development of applications.

The exclusion criteria in chapter 6.2 were created to exclude frameworks and models. The criteria in this chapter are created to exclude criteria. The reason for this is that elements of frameworks focused on implementation or development could be relevant. However, the specific criteria focused on these subjects are irrelevant.

Duplicates of criteria are included in the long list but ignored in the short list. Criteria that are synonyms or that have the same description are merged together. This can be viewed in Table 16 on the next page. In Table 15 on the next page is an overview of excluded criteria.

Table 14 Long list potential criteria

Criteria from guidelines	Criteria selected models	Criteria literature
Credibility **(Included)**	Organizational/Readiness **(Ex-1)**	Transparency **(Included)**
Accuracy **(Included)**	Technological safe, relevant, effective & appropriate **(Included)**	Reliability & validity **(Included)**
Objectivity **(Included)**	Financial **(Ex-1)**	Efficacy & effectiveness **(Included)**
Authority **(Included)**	Policy/Legal **(Included)**	Development iterative and flexible **(Ex-2)**
Currency **(Included)**	Evidence **(Included)**	Development intertwined with implementation **(Ex-2)**
Scope **(Included)**	Ethics **(Included)**	Persuasive design **(Included)**
Usability **(Included)**	Usability **(Included)**	Impact of eHealth technology **(Ex-1)**
Assess with clinical scenarios **(Included)**	Security **(Included)**	Data security **(Included)**
Feedback to vendor **(Ex-1)**	Preliminary research **(Included)**	Participant privacy **(Included)**
	Transferability **(Included)**	Behavior change maintenance **(Included)**
	Alternatives **(Ex-1)**	Usability **(Included)**
	Maturity of the application **(Included)**	App overload **(Ex-2)**
	Purpose of the applications **(Included)**	Black boxes **(Ex-2)**
		Safety **(Included)**
		Stakeholder involvement **(Ex-2)**
		System must meet changing need user **(Included)**
		User Interface **(Included)**
		Readability **(Included)**

In Table 15 below is an overview of the criteria and the reason of their exclusion.

Table 15 excluded criteria

#	Excluded criteria	Reason exclusion
1	Feedback to vendor	(Ex-1)
2	Organizational/Readiness	(Ex-1)
3	Financial	(Ex-1)
4	Alternatives	(Ex-1)
5	Development iterative and flexible	(Ex-2)
6	Development intertwined with implementation	(Ex-2)
7	Impact of eHealth technology	(Ex-1)
8	App overload	(Ex-2)
9	Black boxes	(Ex-2)
10	Stakeholder involvement	(Ex-2)

In Table 16 are the criteria that are included and the criteria that are combined with each other according to their description. A total of 18 criteria are included, 10 criteria are excluded and 12 criteria are combined with one of the 18 criterions that are included.

Table 16 Included criteria

#	Included criteria	Combined with
1	Credibility	-
2	Accuracy	Scope
3	Objectivity	Authority
4	Authority	-
5	Currency	Maturity of the application
6	System must meet changing need user	-
7	Usability	Usability, usability, user interface, readability
8	Participant privacy	-
9	Technological safe, relevant, effective & appropriate	-
10	Policy/Legal	-
11	Evidence	Efficacy & effectiveness, Reliability & validity, assess with clinical scenarios
12	Ethics	-
13	Behavior change maintenance	-
14	Data security	Security
15	Persuasive design	-
16	Transferability	-
17	Transparency	-
18	Purpose of the applications	Preliminary research

7.2 Short list potential criteria

The short list exists of 18 items. These items are grouped into categories to create an overview. An explanation of each category is given below. The different categories are based on the scientific models that include main categories/themes. This short list of criteria can be viewed in Table 17 on page 60.

1) Preliminary research

This category is based on the "Preceding considerations" theme from the MAST model (Kidholm et al., 2012). The purpose of this category is to quickly rate an application before doing an extensive, in depth, assessment.

2) Evidence

Most of the compared frameworks included evidence in their assessment. This is hardly surprising since systems with identified risks and integrated safeguards were the most successful (Finch et al., 2007). Criteria focused on evidence are therefore grouped together.

3) Usability

Usability is an important subject for mHealth applications. Even though not all of the frameworks and guidelines give the same amount of attention to the subject. Examples are the guidelines from HIMMS or the Health-ITUEM model. The different subjects that are part of usability are combined to ensure that the list is manageable. This also implies that there is no usability test included. There are various tests in existence that can test the usability of mHealth applications. Such a test could be a step in the assessment instrument.

4) Security

Results from the literature review shows that security is one of the most important subjects. Scientific literature also states that it is a complex problem (Patrick et al., 2009). Security is furthermore included in the majority of the frameworks. Security is included as a separate category for these reasons.

5) Transparency

This category is mainly based on the guidelines HIMMS and AZHIN. Both empathized subjects focused on the credibility and the objectivity of the sources. They argued that next to the application itself, it is also important to know the source of the information and if the application is unbiased.

7.3 Conclusion short list

The short list in this chapter is the first version of the assessment instrument. The exclusion of various criteria was necessary to ensure that the focus of the instrument is on assessing mHealth applications. This version is generic and the inclusion of different categories is the first step to a multi level instrument. It is generic to ensure that it can be used on all of the existing available mHealth applications and by all of the different stakeholders of mHealth.

Table 17 Categorized short list

Category	#	Criteria	Description
Preliminary research	1	Purpose	What is the purpose of the application, what should it do?
	2	Currency	Are the application and the information presented up-to-date? Is the information relevant?
	3	Technology	What for technology is used and is this technology: 1. Effective; 2. Appropriate; 3. Relevant.
	4	Policy/Legal	Are there legal obligations to be met and are there any legal barriers?
	5	Ethics	Is it ethical to let patients use the application?
	6	Safety	Is it safe for the patients and/or the staff to use the application?
Evidence	7	Accuracy	Is the information in the application correct?
	8	Evidence	How is the application tested? Where clinical scenarios involved in the testing? Are the reliability and the validity tested? Is the test data available? How is the efficacy and the effectiveness of the application tested? What are the results of these tests? Are there negative effects? How are the efficacy and effectiveness compared with current methods?
	9	Transferability	Can results of testing be transferred to other scenarios?
Usability	10	Usability	Is the usability of the application tested and what are the results? Was the target group involved in testing the application?
	11	Changing need user	Can the application be used for a longer period of time? Does it adapt to the changing need of the user?

	12	Readability	Is the application understandable for people without a medical background?
	13	Behavior change maintenance	Is one of the goals of the application to change behavior of the user? If so, how is this change maintained?
	14	Persuasive design	Are there persuasive design principles included?
Security	15	Data security	Is the application secure? Is the data encrypted before it is send to the server?
	16	Participant privacy	What happens with the data gathered? Who is gathering data? Does the developer have access to the data? Is it being sold to 3th parties?
Transparency	17	Objectivity	Does the application provide authorship information? • Including detailed information about the affiliations, credentials and about any medical professional involvement in content preparation? E.g. is the information balanced and non-biased? Is there a list of references and sources included? Does the application provide information about any app sponsorship (including advertisements) or other commercial funding arrangements, and any potential conflicts of interests?
	18	Credibility	What is the credibility of the sources involved in developing the application?

8 Validation of the first version of the assessment framework

Various experts validated the first version of the instrument. A focus group with experts was organized, a questionnaire was sent and various interviews were conducted. The findings of the validation are described in this chapter.

8.1 Focus group

A focus group with experts from multiple fields was organized. The goal of the focus group was to receive input from experts from multiple fields. The reason for experts from a variety of fields is that this would contribute to the acceptance among different stakeholders. The software Thinktank was used during the focus group as decision support software. With Thinktank participants can simultaneously and anonymously provide input. This chapter will elaborate on the information gathered during the focus group. The supervisors of this research helped with preparing the focus group and shared contact information of potential participants. The advantage of group sessions is that they are useful for dealing with complex, unstructured problems in which the actors have various interests, different backgrounds and knowledge areas (Herik & Vreede, 2000). This makes it more productive than individual interviews.

A maximum of fourteen experts could participate in the focus group. 23 invitations were sent; eleven experts attended the focus group.

The participants and their professional background can be found in Table 18 below. As stated above, the goal was to have participants that represent different stakeholders. The following stakeholders were unfortunately not present:

- Insurance companies.
- Inspection of healthcare (IGZ, inspectie gezondheidszorg).
- Medical professionals.

Table 18 Background participants' focus group

Background of the participant	Number of expert(s)
Researcher (University) Medical Center	2
Researcher University	3
Developer mHealth applications	1
Representative Dutch patient organization	1
Society of serious games & simulation in Healthcare	1
Lawyer mHealth and ehealth	1
Innovation manager or consultant	2
Total number of participants	**11**

The agenda of the focus group is shown in Table 19. Results of activity number 3, 5, 7 and 8 can be found in Appendix F: Results focus group. The transcription of the discussions is left out for the privacy of the participants.

Table 19 Agenda focus group.

#	Activity	Time
1	Introduction participants	10 minutes
2	Presentation literature study	10 minutes
3	Input and feedback on criteria from participants	15 minutes
4	Discussion feedback	25 minutes
5	Arranging criteria with most important criteria on top.	5 minutes
6	Discussion arranging criteria	25 minutes
7	Scoring criteria per type of application	15 minutes
8	Feedback & conclusion focus group	5 minutes
	Total time	**2 hours**

8.1.1 Feedback and discussion criteria

Participants could provide input and feedback on the criteria that is presented in Table 17. The participants were given 15 minutes to complete this task. This was the only activity were the feedback gathered was not anonymously. Most of the comments were given on the category "Preliminary research". This could be explained by the fact that it is the first category and that the majority of the time given for feedback was spend at this category. A discussion with the feedback from the participants as input was the next activity. Results of these activities are processed below. First, the general remarks about the framework are described, thereafter the input about the criteria is shown.

8.1.2 General remarks about the assessment

Purpose of the assessment

Multiple participants found it important to determine the context and the purpose of the assessment before the assessment starts. The following questions where suggested during the discussion and in the feedback:

- What is the purpose of the technical assessment?
- In what context will the technical assessment be used?
- Who are involved?

Feasibility

Another comment made during the discussion is that applications that did a proper feasibility study or a context analysis should get more attention. The developer should perform the feasibility study. This check should be carried out before the category "preliminary research". This could be the first step if the assessment of applications exists of different levels. The purpose of the application could also be determined during the feasibility test. This is important since not all applications require the highest level of evidence.

A feasibility test that was suggested is the COrETeSt (Meulendijk et al., 2013). This test is designed to test the feasibility of medical informatics projects. The researchers claim it is not an all-encompassing test, but the following points are worth exploring:

1. Conceptual feasibility;
2. Organizational feasibility;
3. Economic feasibility;
4. Technological feasibility;
5. Societal feasibility.

A short description of each of these points is given below. However, future research should determine if this test is sufficient as a feasibility test.

1. Conceptual feasibility

The first step in the feasibility study is to develop a conceptual model of the intended venture. It is essential that consensus is reached when there are multiple partners included.

2. Organizational feasibility

An important step is to explore the market and its key players. It is important to know potential partners next to the competitors.

3. Economic feasibility

The economic feasibility can be determined based on the previous steps. This includes optional partnerships with key market players, but the development strategies for the intended venture should also be drawn.

4. Technological feasibility

The researchers mention that this step may be difficult since detailed technological approaches cannot yet be modeled in preliminary stages. However, the usability approach and the information exchange between the new product and existing ones can be visualized.

5. Societal feasibility

The last step is to do predictive calculations to determine the societal gains resulting from the implementation. In the medical field, societal gains would mean reduced mortality and morbidity rates and therefore improved health.

Levels of evidence

Participants mentioned that the final instrument should have different levels. An example is the feasibility study mentioned above. However, the different levels and what they should include were not discussed during the focus group.

What should be tested

The last general remark was about what should be tested. Since certain parts of the application such as determining the usability could be ask towards the users of the application. However, other items such as the theory, calculations or other processes that are not visible or clear for users should be tested with an assessment. This is where this assessment could make a difference according to one of the participants. A next step could therefore be to categorize the criteria into groups of how they could be tested.

8.1.3 Input criteria from participants

The input and comments from the participants are combined with the existing short list. The table that resulted can be viewed in Appendix G: List of criteria focus group. This table is used as input for interviews. The shortlist that was handed to the participants included 18 criteria; the new list exists of 24 criteria. Next to adding criteria the participants also elaborated on the description of the existing criteria. There were no new categories added.

8.1.4 Arranging criteria

The participants had to arrange the criteria with the most important criteria on the first place. One of the participants commented before we started that it is impossible to establish a "perfect" ranking of criteria for an mHealth assessment. It is important while assessing mHealth applications that the different criteria interact simultaneous with each other. The participant implies with this that the different criteria are equally important and mHealth applications should get a sufficient score on all of the criteria. During the exercise multiple participants mentioned that some of the criteria were equally important to them. However, the system does not allow voters to put two or more criteria on the same rank. Five minutes were scheduled for the participants to complete this task. This was not enough and the time given was extended. Some of the participants lacked time even after the extension, but due to time constraints the exercise had to end.

The following criteria were voted as the top 5:

1. Purpose
2. Evidence
3. Safety
4. Usability
5. Participant privacy

The software automatically calculated if there was consensus in the group for each criterion. There was consensus on the criteria evidence (number: 2), usability (number: 4), technology (number: 25) and risk assessment (number: 26). However, the software pointed out that the level of consensus is lower when the list of criteria is large. The criteria with a high level of consensus all had a standard deviation lower than 4.5. Another reason for the diversity in voting is that the participants all had different backgrounds (diversity of stakeholders). This could indicate that different criteria are important for each group of stakeholders. The final instrument could therefore have different sets of criteria for each stakeholder.

The full list and the vote distribution can be viewed in Appendix F: Results focus group.

During the discussion the comment was made that it is difficult to discuss criteria such as objectivity and transparency. An application can have multiple levels and the same criteria can score different on each level. For example, the application can be objective in general, but to what extent is this true for the individual user? It is therefore important to define different levels in a mHealth applications.

8.1.5 Scoring the criteria on different types of applications

The participants had to decide for every criterion how important it is for a certain type of application. The options were “low”, “medium” or “high”. There were three types of applications. The participants first voted on the main categories, and used the time left to vote on the criteria. The types of application used and the reason can be found in Table 20 below. The applications were selected from the classification in Figure 3.

Table 20 Types of applications used in focus group

Type of application	Reason
Clinical decision support systems	The application has to assist a professional in making decisions. E.g. what medicine to use, the dose of the medicine.
Health behavior change	Should help the user in changing their behavior. E.g. applications for losing weight, or to quit smoking.
Chronic disease management	Application has to be used for a longer period of time. E.g. applications for diabetes or cancer patients.

Some of the participants mentioned that the types of applications could be more distinct when the exercise ended. They found that all of the above applications should have a high level of evidence. However, other participants did not agree with this and gave an explanation why the different applications are distinct enough. This discussion could be prevented if the exercise was better explained to the participants. This discussion also made it clear that participants had different sorts of applications in mind when they started the exercise, this could have affected the outcome.

The results of the voting can be viewed in the Table 21, Table 22 and Table 23 below. The number of total votes shows that not every participant voted on every category. The tables only show the votes on the main categories. The complete lists can be found in Appendix F: Results focus group.

1) Chronic disease management

Table 21 shows that the categories “preliminary results” and “transparency” are the most important, and that “evidence” is less important. During the discussion it became clear that patients with a chronic disease have a low level of intrinsic motivation, and that they often do not know the effects of their actions on their disease. It is therefore important that the application is easy for the users to use (high usability), that the purpose is clear (what should the application do), and the effects of the application should be known for the users. This corresponds with the votes in Appendix F: Results focus group. The vote distribution shows that the individual criteria “purpose”, “accuracy”, “usability” and “long term engagement” are the top voted criteria. It is understandable that “long term engagement” is important for chronic disease management. Since the application is meant to be used for a longer period of time. “Accuracy” is important since it entails if the information shown is correct.

Table 21 Vote distribution chronic disease management

Chronic disease management				
	Low	Medium	High	**Total votes**
Preliminary research	-	2	7	**9**
Evidence	-	4	4	**8**
Usability	-	2	6	**8**
Security	1	1	6	**8**
Transparency	1	-	7	**8**

2) Health behavior change

The votes for health behavior change are more distributed in comparison with chronic disease management and with clinical decision support system. Table 22 shows that the most important category is "usability" and that the categories "evidence" and "security" are the least important. This corresponds with the vote distribution in the appendix. It shows that the criteria with the highest score are all in the category usability.

There was a discussion about the distribution of "evidence". As can be seen in Table 22, a slight majority of the participants voted "low" for this category. Other participants were worried that applications without evidence will effect applications that do have evidence. It could be that users will not use health behavior change applications anymore because the application that they used did not have any effect. This does not only harm other applications but also harms the users. Since the user did not get the expected effect, i.e. a change in their behavior.

Table 22 Vote distribution health behavior change

Health behavior change				
	Low	Medium	High	**Total votes**
Preliminary research	1	5	3	**9**
Evidence	4	3	1	**8**
Usability	-	1	7	**8**
Security	4	3	1	**8**
Transparency	1	3	4	**8**

3) Clinical decision support system

In Table 23 are the votes for clinical decision support system applications. "Preliminary research" is the most important category, followed by "evidence" and "transparency". The vote distribution in the appendix shows that the criteria "safety", "accuracy" and "evidence" are the most important. In the discussion participants agreed that "transparency" is important. An example that was given was that calculations for dosing an opiate are very precise. The application should be transparent in how this is calculated, since errors can be fatal for the patients. This is in line with "safety", "accuracy" and "evidence". Applications that are used as decision support should have evidence and be safe for patients.

Table 23 Vote distribution clinical decision support system

Clinical decision support system				
	Low	Medium	High	**Total votes**
Preliminary research	-	-	9	**9**
Evidence	-	-	8	**8**
Usability	3	1	4	**8**
Security	-	5	3	**8**
Transparency	2	-	6	**8**

8.1.6 Questionnaire

A questionnaire was sent to the participants of the focus group and the interviewees, the list was sent to a total of 15 participants. The goal of this questionnaire was to distill a top 10 of criteria per stakeholder for the following stakeholders:

1) Healthcare professional
2) Patient
3) Insurance companies
4) Developers

The list of criteria that was included in the questionnaire was based on the interviews and the focus group. Criteria were combined where possible to ensure that the list was manageable. The list that resulted from this can be viewed in Table 27 on page 80. The participants of the questionnaire also had the possibility to give comments. Comments that were given about making a top 10 are:

"One criteria could lead to the next criteria"

"Very difficult. All of the criteria are important."

"The question is if a top 10 of criteria is all encompassing. Maybe it is better to keep it more complex to make sure it is complete."

These quotes show that the questionnaire was difficult and that it is hard to make a list of criteria per stakeholder.

8.1.7 Results questionnaire

The questionnaire was filled in by a total of 4 respondents. This is not enough for significant results. However, the results are in Table 24 on the next page.

Table 24 Vote distribution questionnaire

		Healthcare professional	Patient	Insurance	Developer
	Preliminary research				
1	Purpose	3	3	4	4
2	Currency	1	1	1	
3	Technology	1	1	1	3
4	Legislation	2	1	2	3
5	Risk assessment	1		3	
6	Feasibility	1		2	1
7	Financial information	1		3	
	Evidence				
8	Accuracy	3	1	1	1
9	Evidence	3	2	3	
10	Transferability	1	1	1	2
	Usability				
11	Usability	2	4		2
12	Behavior change maintenance	2	3	1	2
13	Interoperability	3	2	1	2
14	Satisfaction	1	4	1	1
	Security				
15	Data security	3	3	2	2
16	Participant privacy	1	4	1	1
	Transparency				
17	Objectivity	3	3	1	
18	Credibility	3	1	3	1
19	Certification	3	1	4	1
20	Ethics	4	2	2	2

A top list of criteria per stakeholder is in Table 25 on the next page. Criteria with the same number of votes are randomly ordered. Only the criteria with three or with four votes are included, none of the stakeholders have therefore a top 10 of criteria. The stakeholders could start with these criteria during the assessment, ensuring that the most important criteria are covered. However, there should be a questionnaire with more respondents to get significant results.

Table 25 Criteria with the most number of votes per stakeholder

#	Healthcare professional	Patient	Insurance	Developer
1	Ethics	Participant privacy	Certification	Purpose
2	Certification	Satisfaction	Purpose	Technology
3	Credibility	Usability	Credibility	Legislation
4	Objectivity	Objectivity	Evidence	-
5	Data Security	Data security	Financial information	-
6	Interoperability	Behavior change maintenance	Risk assessment	-
7	Evidence	Purpose	-	-
8	Accuracy	-	-	-
9	Purpose	-	-	-
10	-	-	-	-

8.1.8 Conclusion focus group

Results from the focus group show that the original short list was incomplete according to the experts. The results gathered from the prioritization assignment shows that there is no consensus among the group. As pointed out by the meeting the reason could be that the list was too long. The vote distribution on the different types of applications and the discussions held show that different criteria are important for different applications, and that stakeholders can have different interests. An instrument for the assessment of mHealth applications should take this in consideration. The list with criteria that resulted from the focus group (Appendix G: List of criteria focus group) needs to be checked and reduced to a list that is manageable. Redundant criteria will be deleted and criteria that are synonymous to each other will be combined. It was unfortunately not discussed how the criteria should be modeled into an instrument. In short, the points that will be processed are:

- The criteria and descriptions suggested by the experts;
- The suggestions of multiple levels of assessment;
- A pre-instrument level where the purpose of the assessment is determined.

The instrument will not have different lists of criteria for different stakeholders and for different types of applications. As commented on the questionnaire, a selection of the criteria does not insure an all-encompassing instrument for every stakeholder. The reason that there is no distinction in different types of applications is that the focus group proved that some criteria are more important than other criteria. However, there was never a selection round where criteria were voted out for being less significant.

8.2 Expert interviews

A number of interviews were conducted to validate the instrument. Interviews were conducted with stakeholders that could not attend the focus group. The interviews gave the opportunity to ask more in depth questions and to focus on specific subjects that were not discussed during the focus group. The feedback of the focus group on the first version of the instrument was processed in the version that was sent to the interviewees. The organizations and the functions of the interviewee are in Table 26 below.

Table 26 Function interviewees and their organization

Organization	Function
Synappz	CEO Synappz
The Dutch Healthcare inspectorate (Inspectie Gezondheidszorg, IGZ)	Senior Inspector Healthcare IT
Achmea	Program manager Achmea Care division
Deloitte	Head of Deloitte Assuring Medical Apps

The interviews conducted were semi-structured. The questions and the list of criteria were sent to the interviewees beforehand, the questions can be viewed in Appendix H: Interview questions. This helped the interviewee with preparation and prevented too much side tracking of the actual subject during the interview. The interviews were semi-structured, making it possible to ask in depth questions based on the answers of the interviewee. The transcript of the interviews can be provided upon request.

8.2.1 Synappz

A comparison has been made with the list from Appendix G: List of criteria focus group and with the development process of Synappz[7], a developer of (award winning) mHealth applications in the Netherlands. The comparison was made to test if their development process includes the different criteria that are important according to this research and if there are criteria missing in the instrument. Criteria that are missing in the instrument but are focused on the development of mHealth applications were excluded during this comparison. Two interviews, where the second interview was by phone, were conducted to make the comparison. Both of the interviews were unstructured. The following points have been covered during the interview:

1. General information on the development process of Synappz;
2. A check whether the criteria are incorporated in the development process of Synappz.

Point two was partly handled during the second interview by phone. First, the different phases of development will be explained, with subsequently a description of the comparison. Additional tables with information about the comparison with Synappz can be found in Appendix I: Comparison Synappz.

[7] http://www.synappz.nl/en/

The development phases of Synappz

Synappz applications are developed according to a blueprint. This blueprint is based on "The elements of user experience" from J.J. Garrett. The phases in this method are as follows:

Phase 1: Strategy

The strategy is the most abstract phase during development. Issues such as the purpose of the application, the problem that the application needs to solve, if the application changes the healthcare process, the technology that will be used and the feasibility of the project are discussed in this phase.

Phase 2: Scope

Answers questions such as: how are the requirements of the users translated to functionalities? What are the technical requirements? What is in- and excluded of the application?

Phase 3: Structure

In the structure phase questions about the structure of the database, the information architecture, the number of screens and the validation of the application are answered.

Phase 4: Skeleton

The skeleton phase answers questions about the elements of the user interface, if the navigation in the application is properly organized and if all of the interactions are necessary.

Phase 5: Surface

The surface phase is the least abstract phase. Issues such as the user experience is tested, and the user interface of the different screens is finalized.

Different criteria are incorporated in different phases; the next section provides a comparison between the phases and the main categories.

Main categories matched with development phases

It was discussed during the interview where the categories and the criteria of the instrument would fit in the development process of Synappz. A description of how the main categories of the instrument could fit in the development process is shown below.

Preliminary research

Preliminary research is largely answered in the **strategy phase**. Questions such as the purpose, currency and what the stakeholders of the application are, are all answered in this phase. Safety is answered in a risk assessment, which is carried out for every application. The difference with a risk assessment for applications with a CE mark is that these assessments require more documentation.

Evidence

It depends on the application if evidence is necessary. The main issue with evidence is that it takes a lot of money and time. Synappz also develops for third parties, limiting them to the budget of their clients. Most companies do not bother with evidence if the application is a side product or an extra service for the customer. This is different when an application is the core product of a company. In

this case it is much more important for that company to have a sound application and therefore have evidence. If evidence is required it is incorporated in the **strategy and structure phase**.

Usability

Usability is important for every application. It is therefore taken into account during the whole development process. This starts in the **strategy phase**, where the target group is determined and if the application should cause a behavior change. Functionalities of an application are determined during the **scope**. A flowchart and test protocols of the application are made during the **structure phase**. The wireframing and the final graphical elements are developed during the last two phases, **skeleton and surface**. Synappz further involves patient profiling, which means that different users of an application can have different information due to demographic differences.

Security

Security is handled during the **scope and structure phase.** Part of security is to decide if anonymization and pseudonymization is necessary, encryption of data, and other safety aspects of the databases. Synappz furthermore makes sure that they abide to the data protection act. A comment during the interview was that developers can say that the application is secure and that they find the privacy of the user important. However, this is impossible to check without a certificate from a third party. An example of a third party that was given is "Truste"[8]. Companies that have the ISO-9001 certificate are also more trusted. The certificate does not guarantee everything, but it takes time and money to obtain it. Showing other parties that the company is willing to invest in quality management and documentation of problematic issues. The ISO-9001 certificate can be simplified as:

1. Tell what you are going to do;
2. Do what you were saying;
3. Prove it (with documentation).

Transparency

Transparency and all the criteria included in it are important for Synappz. However, it is a difficult subject since competitors are also able to see the information. Certain information such as a calculation could be important to know for a professional who uses the application, while it is at the same time unique for the application. Synappz does value transparency and tries to develop objective applications. However, issues about transparency always need to be discussed with the client, this happens during the **strategy phase**.

8.2.2 Interview Dutch Healthcare Inspectorate

The Dutch Healthcare Inspectorate could not attend the focus group since they cannot engage in the development of a framework for assessing mHealth applications. Other organizations could interpret their attendance as an approval for the instrument. The Dutch healthcare inspectorate has a controlling function. Meaning that they will only inspect existing methods, devices, frameworks etc. and that they will not contribute in the development since that could lead to a conflict of interest. The persons interviewed were a senior inspector ICT and a senior inspector.

[8] http://www.truste.com/

During the interview it became clear that there are different laws relevant for this research. The three most important laws are:

1. The medical devices Act (Wet medische hulpmiddelen)
2. The Dutch care institutions (quality) Act (Kwaliteitswet zorginstellingen)
3. The data protection Act (Wet bescherming persoonsgegevens)

The medical devices act

Some mHealth applications can be qualified as medical devices. mHealth applications are qualified as a medical device based on the intended purpose as described by the manufacturer of the application. The name of the device does not have any effect on this process nor does the risk related to a malfunction of the software. An example given during the interview was that a heart rate monitor used in a fitness center is not qualified as a medical device. Since the monitor is used by individuals who want to measure their heart rate during a workout. There are also no consequences when the monitor is not precisely calibrated. However, the same heart rate monitor is qualified as a medical device if it is placed in a hospital to measure the heart rate of patients. The reason for this is that it now can have an influence on the health of a patient. An mHealth application is placed under the medical devices act:

- If the application makes a diagnosis;
- Energy is added to the human body by the application (e.g. hearing aid), or;
- If the application monitors vital body functions whereby variations are a direct threat for the live of the patient.

A more elaborate description of the medical devices act can be found in Appendix J: Elaboration of MED DEV 2.1/6.

The Dutch care institutions Act

This act obliges the healthcare institute to give quality care to patients. This act can therefore be viewed as a sort of safeguard. Medical devices are subject to a lot of requirements. However, not all mHealth applications are classified as medical devices. This act ensures that healthcare institutes uses application of a proper quality, even when the application is not a medical device.

mHealth applications that are a medical device and that are used by a healthcare institute are subject to both laws.

The data protection act

The data protection act ensures the privacy of the users of the application. It states the conditions for gathering personal data. Personal data is defined as data that is about a person or can be used to identify a person (Article 1, Data protection act). Common personal information is name, address, place of residence or cellphone numbers. Medical information about a person is called “special personal information”. Rules for this information are extra strict. mHealth applications that are partly excluded are applications for research or educational purposes. It is always important that the manufacturer of the application has taken steps that prevent theft of the gathered data.

Other comments

This section will describe other comments from the interview.

The framework could be split into two sections, based on the law of medical devices and on the law of quality care in institutions. There is the technical side of the framework, and the application side as can be seen in Figure 11. The technical side is the law of medical devices and the application side is the Dutch care institutions quality act. Most errors occur on the application side. People on the application side have to use the application. Most of the times they are not trained in how to use the application, and it is unclear what the goal of the application is. Another problem is that the inspectorate cannot monitor applications that are not a medical device and that are used by users without the intervention of a healthcare provider. Since applications are in this situation not subject to both laws.

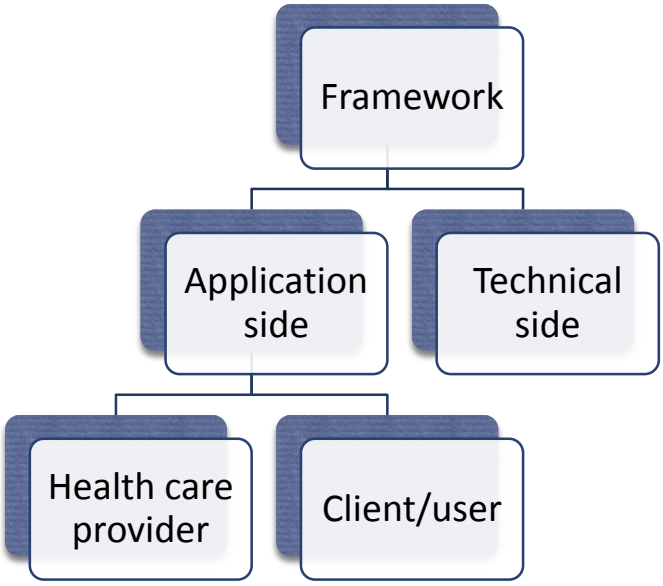


Figure 11 Framework divided based on relevant laws

The inspectorate furthermore suggested that the different levels of assessment could be based on the medical devices act. Different categories are specified in this law; each one of these categories could be used as a level of assessment. However, not all mobile applications are medical devices according to the medical devices act. This categorization would therefore exclude a group of mHealth applications.

8.3 Interview Achmea

An interview with the insurance firm Achmea[9] was scheduled since they were unable to join the focus group. The position of the interviewee is program manager. During the interview the list of criteria that resulted from the focus group was discussed on a high level. The feedback on this list was positive. The different categories were perceived as positive. However, it became clear that more attention should be given to "Finance". Finance is always an important point before Achmea decides to reimburse an application. Applications should have a solid business model, since applications that do not earn any money will most likely cease to exist within a few months. Finance should also include information for the users, such as:

- How much do I need to pay?
- Do I need to pay a one-time fee, or do I have to take a subscription to the application?

It is furthermore important that there is adequate evidence provided by the developer or responsible organization of the application before reimbursement is considered. The interviewee explained that Achmea is not responsible for gathering evidence.

8.4 Interview Deloitte

It was suggested during the focus group to seek contact with the assuring medical apps project from Deloitte[10]. The focus of the assuring medical apps project is security and privacy of mHealth applications. The interviewee, a manager at Deloitte, reviewed the list with criteria that resulted from the focus group. The review focused on the security and privacy. The difficulty with these criteria is that it is difficult to test these criteria. A third party is needed to test them adequately. There are however a few steps that almost every user can perform to review the privacy and security of an application. The following tips were given:

1. Does the application have a clear security and privacy policy that is easily available for users?
2. Is it easy for users to report security and privacy related problems to the developer?
3. Is it possible to set a pin code for the application?
4. Does the application use services from third parties and is the privacy guaranteed? This should be explained in a clear way in the user agreement or the privacy policy of the application.
5. Does the application explicitly ask to use data that is entered by the user?
6. Is the application transparent about the usage of user data?
7. Is the application designed in such a way that it only asks data of the user that is necessary for the functionality of the application?
8. Is it possible to delete your account and the related data that has been gathered by the application?
9. Is there an ethical hack test performed[11]?

[9] https://www.achmea.nl/

[10] www.assuringapps.com

[11] An ethical hack test is a hack ordered by the developer to test the security of the system.

10. Is there a responsible disclosure policy[12]?

The feedback received will be taken into account for the final version. An important note that was given is that the questions should not be answered with “yes” or “no”. It is important that there is documentation available.

8.5 Conclusion interviews

It can be concluded from the interviews that some topics, such as the different laws or finance, need more attention in the framework. Other criteria, such as the security of an application are likely too difficult for individual users to check. The reason for this is that most users do not have the knowledge necessary to determine if an application is secure enough. The interviewees from the inspectorate and Achmea also mentioned that current legislation is up-to-date and that different levels could be derived from the current medical devices directive. Additional information on finances, legislation and tips to test the security of an application are processed into the final list of criteria. The different levels in the final instrument will not be derived from current legislation. Since the legislation is not focused on assessing mHealth applications, but is focused on determining if an application is a medical device.

8.6 Final list criteria

The final list with criteria is based on the first version and the input received during the focus group and the various interviews. There are a total of 20 criteria, which are divided into categories. Each category is in a separate table with a short explanation of the changes compared to the first version. On the next page, in Figure 12, is a Venn diagram with the 20 criteria sorted into People, Process, or Technology. The Venn diagram provides an overview of how the criteria are related to each other.

[12] A responsible disclosure policy is a disclosure where stakeholders agree to allow a period of time for the vulnerability to be patched before publishing details.

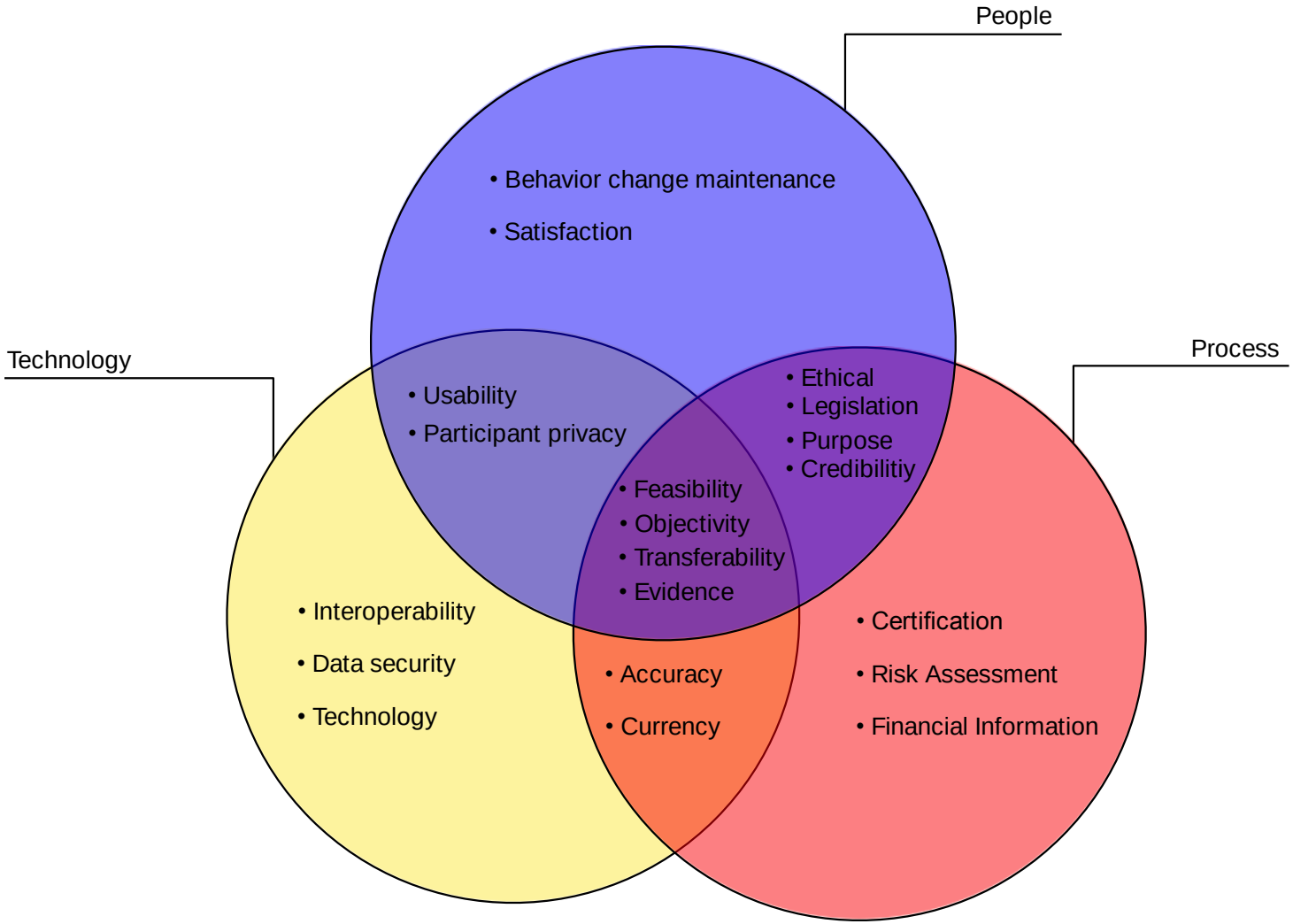


Figure 12 Venn diagram people, process, and technology with the criteria placed inside

8.6.1 Preliminary research

On the next page is Table 27 with the final list of criteria for the category ‘preliminary research’. The biggest change compared to the first version, which was based on literature, is that the following criteria are added:

- Feasibility;
- Financial information;
- Risk assessment.

Feasibility and risk assessment are added based on the discussion of the focus group. Participants added these criteria and agreed on the importance of these criteria during the discussion that followed. Financial information is added based on the interview with Achmea. As described in the previous chapter, financial information is valuable information for Achmea. It should be clear for every stakeholder what the involved costs of an application are and if it is likely that the application will exist in a few months. The description of other criteria is furthermore elaborated. The elaboration is based on the suggestions and discussions from the focus group.

Table 27 Final list of criteria included in Preliminary research

Preliminary research Criteria	Description
Purpose	What is the target group of the application? What is the intended purpose of the application for: • Patients; • Doctors; • Insurance companies; • Other stakeholders. What problem is solved by the application? Was the target group involved in the development of the application?
Currency	What is the life cycle of the application? • When is the application made? • How frequently is the application updated? • What was the last time the application was updated? • Are the developers actively updating and improving the application? Is the information used in the application up-to-date/relevant?
Technology	What kind of technology is used? Is this technology: • Effective; • Appropriate; • Relevant; Is the technology being updated? Is the application compatible with the newest phones?
Legislation	Does the application need a CE mark? This is the case if the application: • Makes a diagnosis • Add energy to the human body (e.g. hearing aid), or; • Monitor vital body functions whereby variations are a direct threat for the live of the patient. Is there any other legislation that is relevant for the application?
Risk assessment	Which possible harms are identified during the risk assessment? Is it safe for the patients and/or the staff to use the application? Non-inferiority should be the minimum threshold for large-scale implementation, especially if the mHealth solution replaces the usual care. Was security of the application part of the risk assessment?
Feasibility	Is there a feasibility study performed? COrETeSt (Meulendijk et al., 2013) is a feasibility test for medical informatics projects. This test includes: 1. Conceptual feasibility 2. Organizational feasibility 3. Economic feasibility; is there funding that will keep the application updated? What is the health economic value 4. Technological feasibility 5. Societal feasibility Is there a business case?
Financial information	Does the application specify financial information for the users? Is the application covered by insurance companies What are the costs of using the application? A one-time payment or a monthly fee? What is the business model of the application?

8.6.2 Evidence

Table 28 includes the criteria from the category 'evidence'. There are no criteria added compared to the first version. However, the description of each criterion is elaborated. The reason for this is that it became clear during the focus group that the description of the first version lacked. More information about evidence is necessary to determine how the evidence of an application was established.

Table 28 Final list of criteria included in Evidence

Evidence Criteria	Description
Accuracy	Is the information in the application correct? Are the following subjects specified: • What theory is used; • Which frameworks are used; • Which models are used? What part of the application does not rely on theory? What or who is the input for this information? • Users; • Professionals; • Sensors; • Other input.
Evidence	How is the application tested? Specify: • Design of the test(s); e.g. Clinical scenarios • Sample(s) used; • Total population; • Outcome of the test(s). Are the reliability and the validity tested? How is the efficacy and the effectiveness of the application tested? What, if any, are there negative effects of the application? How are the efficacy and effectiveness compared with current methods? Is the test data available? Can adjustments/updates be introduced without any loss of evidence? What is the added value of the application?
Transferability	Can results of testing be transferred to other scenarios?

8.6.3 Usability

There were seven categories after the results from the focus group were processed in the category 'usability'. These criteria and the descriptions are combined in the four categories that are in Table 29 below. Usability encompasses a wide variety of subjects. However, the instrument is not solely focused on usability and it is not a usability test. Criteria are therefore combined to keep the list manageable. The most important points are described in the table below since usability is important according to literature and experts.

Table 29 Final list of criteria included in Usability

Usability Criteria	Description
Usability	How is the usability of the application tested? Was the target group involved in testing the application? Both the paper prototype and the final version What is the adherence of the application to existing standards? How well does the application fit in the regular life of the user? Did the developer use the human interface guidelines from the platform the application is published on? E.g. Google, Apple, etc. Are there persuasive design principles included in the application? Is the application understandable for the target group? E.g. is an application for patients understandable and without jargon? How is the adherence and compliance of the application? Can the application be used for a longer period of time? Does it adapt to the changing need of the users?
Behavior change maintenance	Is one of the goals of the application to change behavior of the user? • If so, how is this change maintained?
Interoperability	What is the interoperability of the application with other systems and how is this measured? Are there standards used?
Satisfaction	What is the satisfaction of the user and how is this measured?

8.6.4 Security

The criteria in the category 'security' remained the same as with the first version, the description however is elaborated. It requires expert knowledge to test security and privacy thoroughly. The descriptions with a '⍰' in front are questions that can be easily answered in most scenarios by an average user. These are added based on the interview with Deloitte to ensure that every stakeholder can use the instrument.

Table 30 Final list of criteria included in Security

Security Criteria	Description
Data security	❖ Is there a clear security policy that is easily available for users? ❖ Is it easy to report security related problems? ❖ Is there an ethical hack test performed? ❖ Is there a responsible disclosure policy? Is the data encrypted before it is send to the server? Is the developer legally allowed to work with the patient data? How does the developer handle anonymization/pseudonymization? What are the security requirements of: • The application; • The phone; • The servers; • Other devices that use data from the application? Where are the servers located? Are there other applications running simultaneously on the device and what type of interaction might you expect between the application that is under assessment and other applications?
Participant privacy	❖ Is there a clear privacy policy that is easily available for users? ❖ Is it easy to report privacy related problems? ❖ Is it possible to set a pin code for the application? ❖ Is it possible to delete your account and the related data that has been gathered by the application? ❖ Can users download their data? ❖ Does the application use services from third parties and is the privacy guaranteed? This should be explained in a clear way in the user agreement or the privacy policy of the application. What happens with the collected data? How long is the collected information stored? Is the application designed in such a way that it only asks data of the user that is necessary for the functionality of the application? Does the application explicitly ask to use data that is entered by the user? Is the application transparent about the usage of user data? E.g. Who is collecting data (developer, sold to 3th parties)? • Who is able to see the data? • Is the data being sold to 3th parties? Is it possible to link the collected data with medical records?

8.6.5 Transparency

Additional criteria and descriptions are added in the category 'transparency'. 'Ethical' was moved to this category from preliminary research and the criterion 'certification' is added. These can be viewed in Table 31 below. Certification was suggested multiple times during the focus group and is therefore added. Ethical is moved to this category after the interview with Synappz. Synappz suggested this change since it suits better with transparency than with preliminary research.

Table 31 Final list of criteria included in Transparency

Transparency Criteria	Description
Objectivity	Does the application provide authorship information? Including detailed information about: • Affiliations • Credentials • Any medical professional involvement in content preparation? Is the information balanced and non-biased? Is there a list of references and sources included? Does the application provide information about any app sponsorship (including advertisements) or other commercial funding arrangements, and any potential conflicts of interests? Is the application/developer transparent in: • Development process • Algorithms used • Degree to which end-user and experts are used
Credibility	What is the credibility of the sources involved in developing the application?
Certification	Which certificates does the application have? What is the credibility of these certificates?
Ethical	Is it ethical to use the application for: • Patients; • Medical professionals; • Society at large; • Other stakeholders?

8.7 Proof of concept

The list with criteria is translated into an artifact in Excel were the user can select answers from a drop down list. This review does not provide an evaluation of the content of the application but presents an indication about the amount of information that is present or that is made publicly available. The answers are filled in on the first sheet; a report is automatically generated on the second sheet.

The proof of concept is composed as follows.

- Users fill in the name of the application
- The main and sub-category can be picked from a drop down list, these categories are based on the classification in Figure 3 on page 24. The drop down list with sub-categories changes depending on the answer of the main category.
- Users fill in the date of the review.

The next step is to fill in the questions for the categories:

1. Preliminary
2. Evidence
3. Usability
4. Security
5. Transparency

All the questions can be answered with “yes” or “No/Unknown”, which can be selected from a drop down menu. A report is generated on the second Excel worksheet. A spider graph gives a visual representation of the level of available information. A warning will appear that the amount of available information is low when more than 50% of the answers are answered with “No”. Screenshots from the instrument are in Figure 13 on the next page. The reported generated is in Figure 14.

However, this version should be developed further. Not all of the questions are relevant for all types of applications. For example, applications that do not gather any data or information will have less information available about their security. However, the proof of concept does not take this into account. It is therefore important that the report from the instrument is supported with a review where the answers are explained. The review could also explain if something (for example security) is important or not.

Version 0.1 instrument

The results are not an evaluation of the content of the application but are an indication about the amount of information that is publicly available.

Select an answer from the drop down menu or type in the date when required. An overview will be generated on the second tab.	Anwers should be inserted in this column

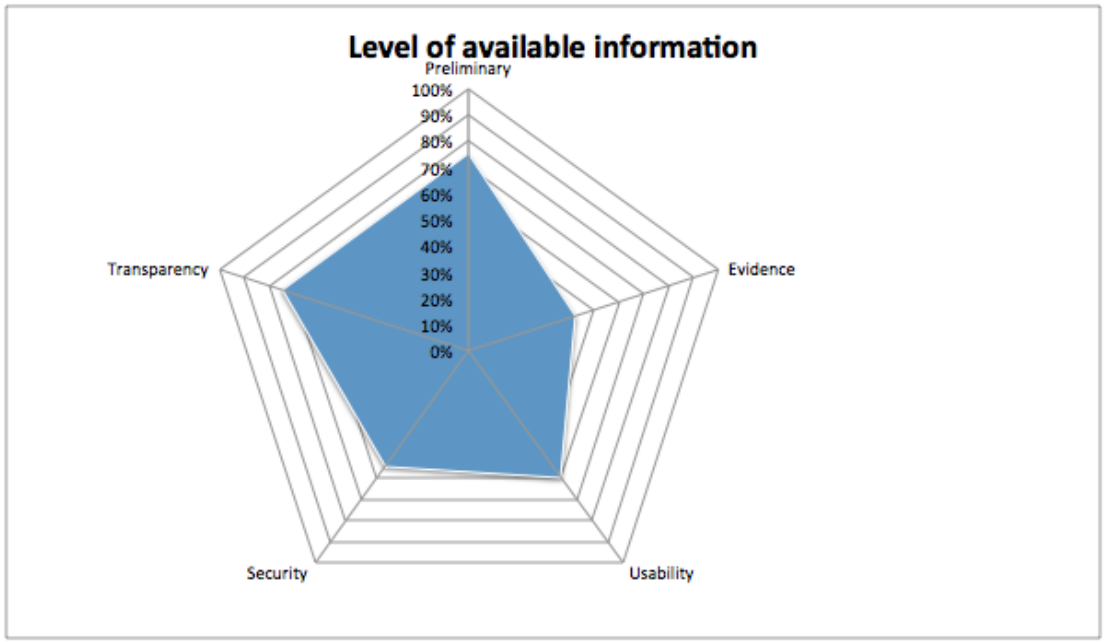
Start of the review

	Fill in the name of the application	**Example**
	Main category of the application:	Health_care_professional
	Sub-category of the application: (First choose a main category)	Clinical decision support system
	Date of review:	**20-02-15**
	Preliminary research	
1	What is the target group of the application? ☐ Patients ☐ Healthcare professional / researcher ☑ General population	
2	In what year was the application published? (yyyy)	2010
3	What was the last time the application was updated? (dd-mm-yy)	02-02-14
4	Is the information in the application up-to-date and relevant?	No/unknown
5	For which operating systems is the application available? ☐ Android ☑ Apple iOS ☐ Windows Phone	
6	Is the application compatible with the newest versions of the operating systems?	Yes
7	Is the application required to have a CE mark?	Yes
8	What are the costs of using the application?	Advertisements are shown
9	Did the developer perform a feasibility study?	Yes
10	Did the developer perform a risk assessment	Yes
	Evidence	
11	Is there a description of the theory that is used?	No/unknown
12	Is there a description of the frameworks/models that are used?	No/unknown
13	Is there information available regarding clinical tests?	Yes
14	Is there information regarding the testing of the reliability and the validity?	Yes
15	Are the negative effects of the application stated?	No/unknown
16	Is any of the test data available?	No/unknown
17	Is it clear what the added value of the application is?	Yes
	Usability	
18	Is the usability of the application tested?	Yes
19	Does the application adhere to existing standards?	No/unknown
20	Is the application interoperable with other systems?	Yes
21	Is the satisfaction of the users measured?	Yes
22	Is the application easy to understand for the target group?	No/unknown
	Security	
23	Is there a clear security policy available?	Yes
24	Is it easy to report security related problems?	No/unknown
25	Is there an ethical hack test performed?	No/unknown
26	Is there a responsible disclosure policy?	Yes
27	Is it clear where the servers are located?	No/unknown
28	Is there a clear privacy policy available?	No/unknown
29	Is it easy to report privacy related problems?	Yes
30	Is it possible to set a pin code for the application?	Yes
31	Is it possible to delete your account and the related data that been gathered by the application?	No/unknown
32	Can users download their data?	Yes
33	Is it clear if the application uses services from third parties?	Yes
	Transparency	
34	Does the application provide authorship information?	Yes
35	Is there a list of references?	Yes
36	Does the application provide information about any sponsorship?	No/unknown
37	Does the application have any certificates?	Yes

Figure 13 Example answers instrument

Application reviewed is: **Example**
The main category of the application is: Health_care_professionals
The sub category is: Clinical decision support system
Current date: 20-02-15

Category	Level of available information:
Preliminary	75,0%
Evidence	42,9%
Usability	60,0%
Security	54,5%
Transparency	75,0%



The target group of the application is/are:
Healthcare professional or researcher
General population
Patients

The application was published in: 2010
The last time the application is updated 02-02-14

The application is available for:
Windows Phone
Apple iOS
Android
The application is available for the newest operating systems

Financial information:
The application displays advertisements to earn money

CE mark
The application is required to have a CE mark, the developer is required to identify the risks of the application by performing a risk assessment.

Figure 14 Example report generated instrument

9 Results of applying the instrument

This chapter describes how the current instrument is applied to various existing mHealth applications. The criteria that are the result of the previous chapter were used to review applications. As described before, it can be difficult to test certain criteria without expert knowledge and without the cooperation of the developer. The applications that were selected for review were therefore applications that should have a lot of information publicly available. Applications that were chosen all won the Health app award from Artsennet. These applications got attention in the media and there should be more information available about these applications. The data that was generated and possibly collected by the applications was also captured to test if the application did not upload any irrelevant data to the servers. This chapter provides an answer to sub-question six:

"What are the results of applying the critical appraisal instrument on a sample of mobile health applications?"

9.1 Selected applications

The following applications that won the health app award have been selected for testing.

1. "Huidmonitor" (Won in: 2012);
2. "IP Plaslijst" (Won in: 2012);
3. "Moet ik naar de dokter?" (Won in: 2013);
4. "EHBO – Rode Kruis" (Won in: 2014).

All of the applications were reviewed with information that could be found easily on the internet. The purpose of the assessment was to test the instrument. All of the applications have been tested on an iPad and when available, on a phone with Android. Some reviews are more elaborate; this is due to the information that was publicly available. The data transmitted by the applications was also captured to test if the applications did not sent or receive any irrelevant data.

The results of the data capture are presented first. Followed by the results of applying the instrument on various applications.

9.2 Results data capturing

The data transmitted and received by the applications was captured while testing the applications. The reason for the data capture was to test if the applications did not send irrelevant data and if they did send the data encrypted when the applications stated that they would. Applications were installed on an iPad and a smartphone with Android. These devices were not connected with the internet but trough Bluetooth with a laptop. The laptop was connected with the internet and ran a program called "WireShark". WireShark is a program that can capture data packets. This set up is visualized in Figure 15 on the next page.

Figure 15 Set-up data capture

It became clear during the data capture that the devices generated a lot of data traffic on the background. This could be from other applications, or from Apple (on iOS devices), or Google (on Android devices). Packets that were not relevant or interesting were therefore filtered out. The following filters were set on the data capture:

- IP source, only packets were captured of the device that was currently used for testing.
- IP destination, only packets were captured send to the servers of the application being tested.
- HTTP(s) packets, these packets included information send to the servers.

IP source could easily be found on each of the devices that were used for testing. IP destination was found with trial and error. An example of the data capture, where the selected package requested an image, is in Figure 16 on the next page. One of the columns in Figure 16 is called “Length”. The number in this column represents the length of the captured data package. Packets that are longer include more information. The selected package in the example has a length of 1418 and included a request for an image, which was used in the application. An example of information in a shorter package is a link that was clicked on during testing. All of the functionalities were tested during the data capture and there were no packages found with unnecessary information. Nor was there unnecessary information send to the devices.

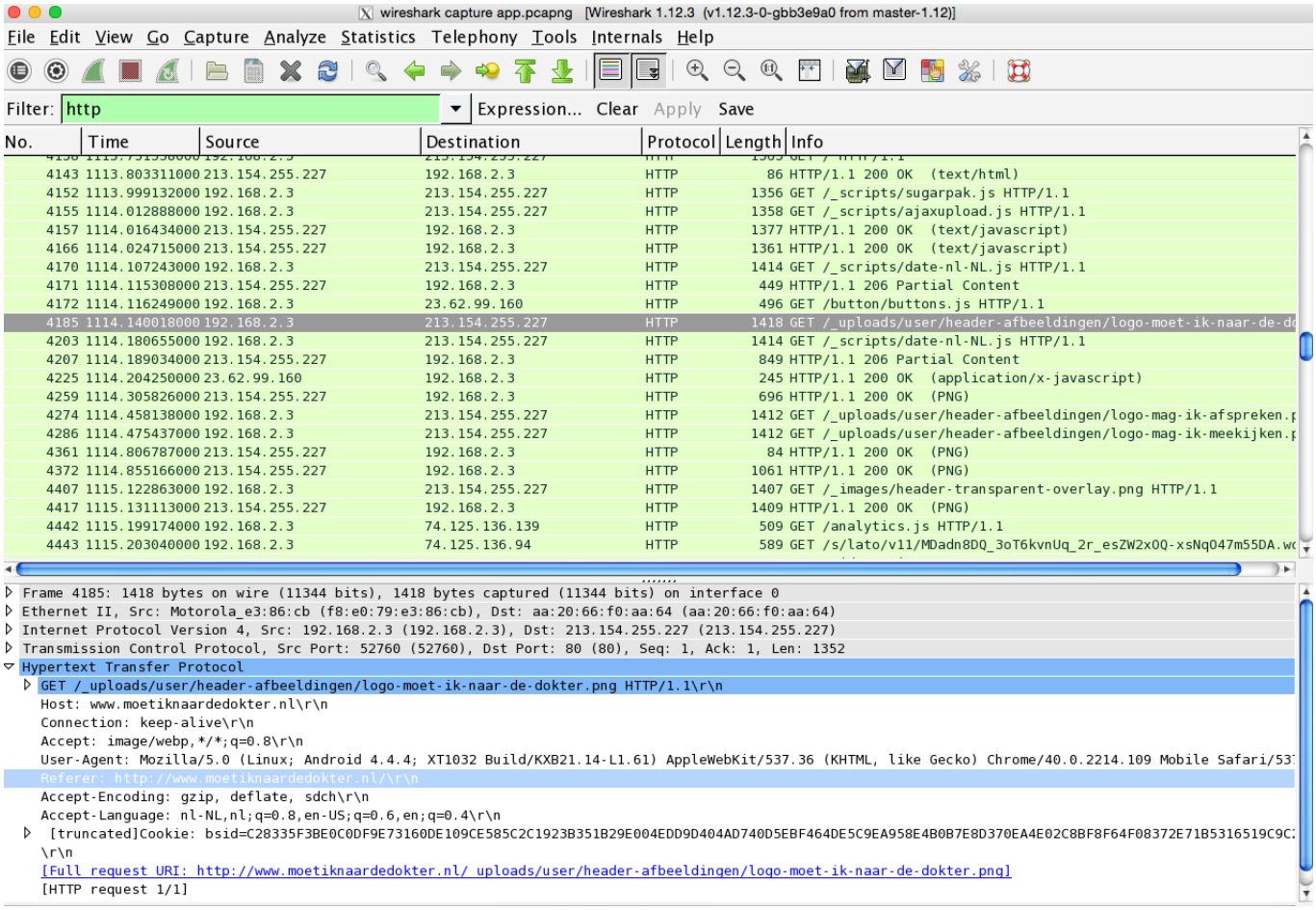


Figure 16 Example data capture

9.3 Huidmonitor

Preliminary research:

Huidmonitor is a free application and was tested on an iPad and on a smartphone with Android. The application clearly states when it is opened for the first time that the purpose is educating users and that it does not make a diagnosis. The main reason for this is probably that an application with the purpose of educating do not need a CE mark. While applications that do make a diagnosis are required to have a CE mark. The website of the application states for which platforms the application is available. The currency and the date of the latest update are standard information in both app stores, this was therefore found easily. We were unable to find information about a risk assessment or about a feasibility study.

Evidence:

It is important that the information is correct for an application that aims to educate their users. The application does state persons and companies that contributed to the development of the application. These organizations include hospitals, foundations and various healthcare companies.

They furthermore specify names of certain healthcare professionals that possibly provided information. There was no information found about used literature.

Usability:
There is no information available about usability testing. There are reviews on Artsennet from users and professionals that have improvement points for the application. The application is furthermore easy to understand and the terminology is explained. An advantage of the application is that text can be made smaller or bigger to increase the readability.

Security:
The data capture shows that the application does not send or receive unnecessary data. Photos taken in the application are only saved on the phone and are not send to a database. The application does not have a clear security or privacy policy and it is impossible to set a pin code. Users can download the photos to their computer and they can delete all of the photos.

Transparency:
The application is transparent about sponsorship and about the medical professionals that assisted during development. The credibility of the sources is high since the main sponsors are non-profit organizations focused on skin cancer and hospitals. Other sponsors are healthcare/pharmaceutical companies, which could be biased. However, it is not clear whether the sponsors provided input for the presented information, or if they only contributed financially.

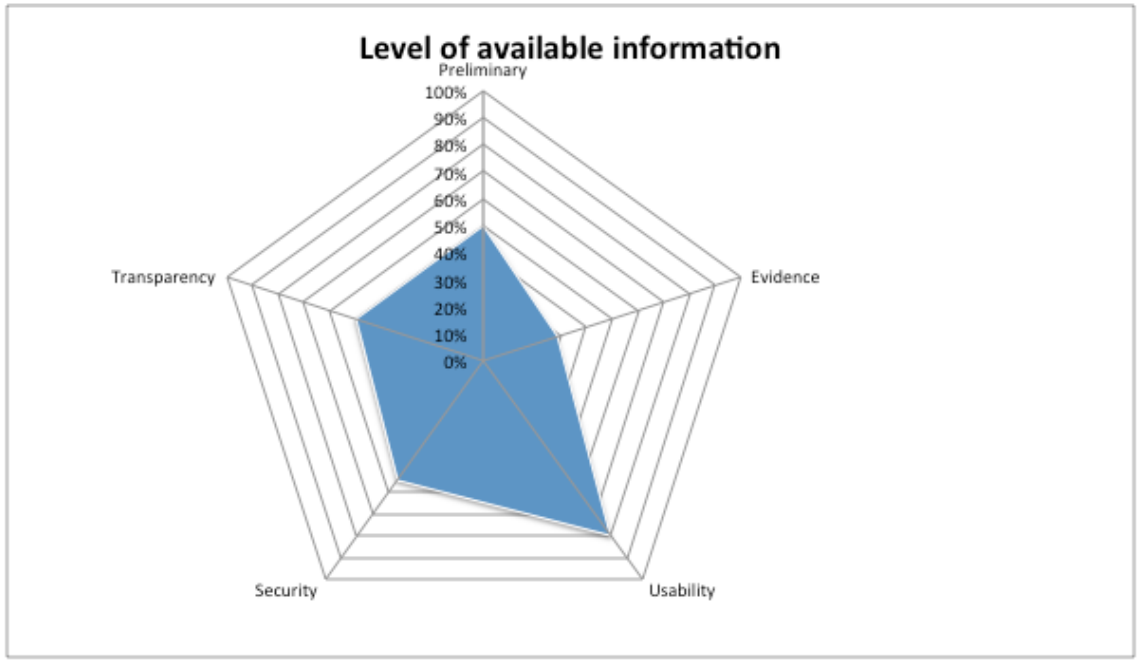
Report instrument
In Figure 17 are the filled in answers for the application Huidmonitor. In Figure 18 is the generated report. It can be easily read from the report that there is not much information available about Evidence. However, the overall amount of available information is high. It is furthermore clear that the application has not been updated since May 2012. This could be an indication that the application is not maintained anymore.

	Version 0.1 instrument	
	The results are not an evaluation of the content of the application but are an indication about the amount of information that is publicly available.	
	Select an answer from the drop down menu or type in the date when required. An overview will be generated on the second tab.	Anwers should be inserted in this column
	Start of the review	
	Fill in the name of the application	**Huidmonitor**
	Main category of the application:	Patients
	Sub-category of the application: (First choose a main category)	Disease monitoring
	Date of review:	**28-03-15**
	Preliminary research	
1	What is the target group of the application? ☑ Patients ☐ Healthcare professional / researcher ☑ General population	
2	In what year was the application published? (yyyy)	2012
3	What was the last time the application was updated? (dd-mm-yy)	17-05-12
4	Is the information in the application up-to-date and relevant?	Yes
5	For which operating systems is the application available? ☑ Android ☑ Apple iOS ☐ Windows Phone	
6	Is the application compatible with the newest versions of the operating systems?	Yes
7	Is the application required to have a CE mark?	No/unknown
8	What are the costs of using the application?	Free
9	Did the developer perform a feasibility study?	No/unknown
10	Did the developer perform a risk assessment	No/unknown
	Evidence	
11	Is there a description of the theory that is used?	No/unknown
12	Is there a description of the frameworks/models that are used?	No/unknown
13	Is there information available regarding clinical tests?	No/unknown
14	Is there information regarding the testing of the reliability and the validity?	No/unknown
15	Are the negative effects of the application stated?	Yes
16	Is any of the test data available?	No/unknown
17	Is it clear what the added value of the application is?	Yes
	Usability	
18	Is the usability of the application tested?	Yes
19	Does the application adhere to existing standards?	Yes
20	Is the application interoperable with other systems?	Yes
21	Is the satisfaction of the users measured?	No/unknown
22	Is the application easy to understand for the target group?	Yes
	Security	
23	Is there a clear security policy available?	No/unknown
24	Is it easy to report security related problems?	Yes
25	Is there an ethical hack test performed?	No/unknown
26	Is there a responsible disclosure policy?	No/unknown
27	Is it clear where the servers are located?	Yes
28	Is there a clear privacy policy available?	Yes
29	Is it easy to report privacy related problems?	Yes
30	Is it possible to set a pin code for the application?	No/unknown
31	Is it possible to delete your account and the related data that been gathered by the application?	Yes
32	Can users download their data?	Yes
33	Is it clear if the application uses services from third parties?	No/unknown
	Transparency	
34	Does the application provide authorship information?	Yes
35	Is there a list of references?	No/unknown
36	Does the application provide information about any sponsorship?	Yes
37	Does the application have any certificates?	No/unknown

Figure 17 Answers instrument Huidmonitor

Application reviewed is: **Huidmonitor**
The main category of the application is: Patients
The sub category is: Disease monitoring
Current date: 28-03-15

Category	Level of available information:
Preliminary	50,0%
Evidence	28,6%
Usability	80,0%
Security	54,5%
Transparency	50,0%



The target group of the application is/are:
Healthcare professional or researcher
General population
Patients

The application was published in: 2012
The last time the application is updated 17-05-12

The application is available for:
Windows Phone
Apple iOS
Android
The application is available for the newest operating systems

Financial information:
The application is free to use

CE mark
The application does not have a CE mark, the developer is therefore not required to perfrom a risk assessment.

Figure 18 Report instrument Huidmonitor

9.4 IP Plaslijst

Preliminary research:

IP plaslijst is a free application from Synappz for the iPad. The application is for persons who have urinary incontinency. Users can keep a digital diary with the application about their drinking and bladder behavior. The application has a CE mark since the application gives a diagnosis. It is mandatory for applications with a CE mark to perform a risk assessment. However, this assessment is not publicly available. The application is available for free for iOS devices; users need to pay a one-time fee if they want to save their diary.

Evidence:

The application is a digital version of a micturition diary. The application does not provide sources about this method nor of the presented information.

Usability:

The website or the application does not specify if the usability of the application is tested. However, the website of the developer claims to involve end users in the development and during testing of the application. The generated reports can easily be mailed to GP's or other medical professionals.

Security:

It is mentioned in the instructions of the application that the generated reports are encrypted and uploaded to the servers of the developer. This could not be tested during the data capture since the free version was used. The data capture did show that the application does not gather irrelevant data from the device nor does it send unnecessary data. The application furthermore clearly states what it does with the collected data and that the data is uploaded to Amazon servers. The only negative point is that users cannot provide feedback trough the application. There is a phone number and email mentioned that users could use to give feedback.

Transparency:

The development method is clearly described on the website. An improvement point would be to include a list of references and sources in the application or on the website. The application furthermore has the CE mark and won the Health app award, which gives the application credibility.

Report instrument

On the next page in Figure 19 are the answers that are the input for the report in Figure 20. The report shows clearly that the level of information available about security is the lowest. It is furthermore easy to see that the application is only available for iOS. This application is also required to perform a risk assessment since it has a CE mark.

	Version 0.1 instrument	
The results are not an evaluation of the content of the application but are an indication about the amount of information that is publicly available.		
Select an answer from the drop down menu or type in the date when required. An overview will be generated on the second tab.		Anwers should be inserted in this column
Start of the review		
	Fill in the name of the application	**IP Plaslijst**
	Main category of the application:	Patients
	Sub-category of the application: (First choose a main category)	Disease monitoring
	Date of review:	**28-03-15**
Preliminary research		
1	What is the target group of the application? ☑ Patients ☑ Healthcare professional / researcher ☐ General population	
2	In what year was the application published? (yyyy)	2013
3	What was the last time the application was updated? (dd-mm-yy)	28-01-14
4	Is the information in the application up-to-date and relevant?	Yes
5	For which operating systems is the application available? ☐ Android ☑ Apple iOS ☐ Windows Phone	
6	Is the application compatible with the newest versions of the operating systems?	Yes
7	Is the application required to have a CE mark?	Yes
8	What are the costs of using the application?	One time payment
9	Did the developer perform a feasibility study?	Yes
10	Did the developer perform a risk assessment	Yes
Evidence		
11	Is there a description of the theory that is used?	Yes
12	Is there a description of the frameworks/models that are used?	Yes
13	Is there information available regarding clinical tests?	No/unknown
14	Is there information regarding the testing of the reliability and the validity?	No/unknown
15	Are the negative effects of the application stated?	No/unknown
16	Is any of the test data available?	No/unknown
17	Is it clear what the added value of the application is?	Yes
Usability		
18	Is the usability of the application tested?	Yes
19	Does the application adhere to existing standards?	Yes
20	Is the application interoperable with other systems?	Yes
21	Is the satisfaction of the users measured?	No/unknown
22	Is the application easy to understand for the target group?	Yes
Security		
23	Is there a clear security policy available?	Yes
24	Is it easy to report security related problems?	No/unknown
25	Is there an ethical hack test performed?	No/unknown
26	Is there a responsible disclosure policy?	No/unknown
27	Is it clear where the servers are located?	No/unknown
28	Is there a clear privacy policy available?	Yes
29	Is it easy to report privacy related problems?	No/unknown
30	Is it possible to set a pin code for the application?	No/unknown
31	Is it possible to delete your account and the related data that been gathered by the application?	No/unknown
32	Can users download their data?	Yes
33	Is it clear if the application uses services from third parties?	Yes
Transparency		
34	Does the application provide authorship information?	Yes
35	Is there a list of references?	No/unknown
36	Does the application provide information about any sponsorship?	Yes
37	Does the application have any certificates?	Yes

Figure 19 Answers instrument IP Plaslijst

Application reviewed is: **IP Plaslijst**
The main category of the application is: Patients
The sub category is: Disease monitoring
Current date: 28-03-15

Category	Level of available information:
Preliminary	100,0%
Evidence	42,9%
Usability	80,0%
Security	36,4%
Transparency	75,0%

The target group of the application is/are:
Healthcare professional or researcher
General population
Patients

The application was published in: 2013
The last time the application is updated 28-01-14

The application is available for:
Windows Phone
Apple iOS
Android
The application is available for the newest operating systems

Financial information:
A one time payment is required to use the application

CE mark
The application is required to have a CE mark, the developer is required to identify the risks of the application by performing a risk assessment.

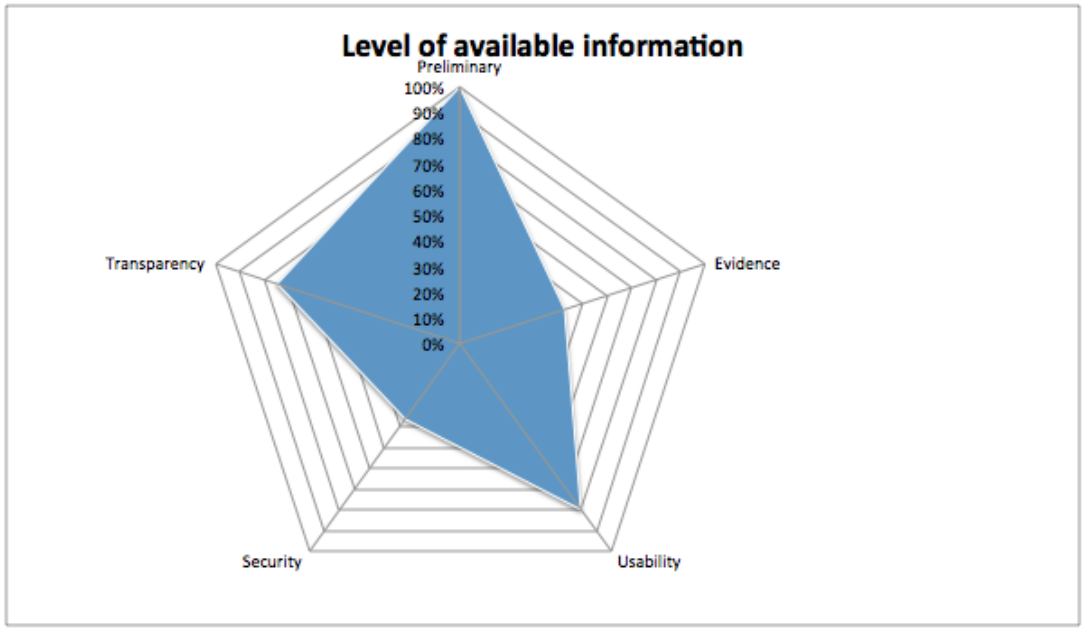


Figure 20 Report instrument IP Plaslijst

9.5 Moet ik naar de dokter

Preliminary research:

Moet ik naar de dokter is a free application tested on the iPad and on a phone with Android. The application, which is available for iOS and Android devices, gives users advice about starting health complaints. The application also suggests if you should go to a doctor. A CE mark is necessary since it generates a diagnosis based on the input of the user.

Evidence:

Information about tests and sources are missing in the application and on the website. Partners are mentioned, but it is unclear what their input was for the application.

Usability:

There is no information available about the usability. There is also no information on the site of the developer. The jury review on the www.artsennet.nl is positive about the application and about the user friendliness. The majority of the user reviews are also positive.

Security:

There is a privacy policy section in the application which redirects to the website of the application. The problem is that this page does not exist on the website. Feedback can only be provided trough a form on the website. The data capture showed that the application does not collect unnecessary data.

Transparency:

The website include a page about their partners, other sources are not mentioned. The application has a CE mark and has won the Health app award 2013. The information in the application seems unbiased since it is made with the help of the Dutch GP's association.

Report instrument

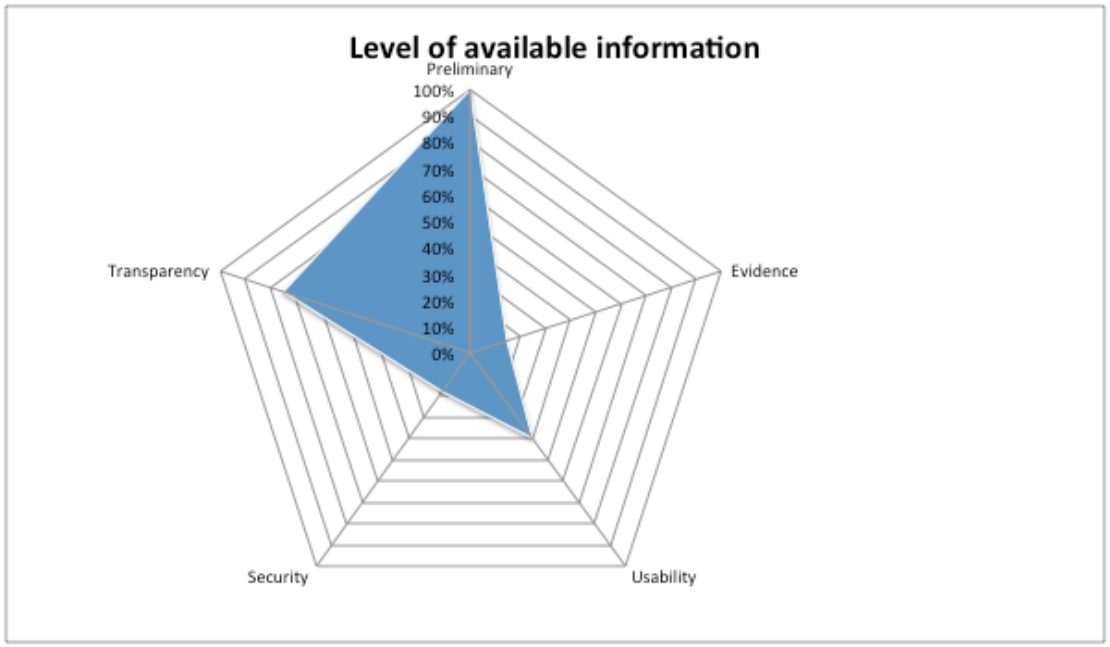
In Figure 21 are the answers that are filled in the instrument for the application "Moet ik naar de dokter". It is already clear from the answers that the level of available information is low. This can also be seen in the report in Figure 22. A warning appeared since the average level of available information is low. It is unclear what the theory is that has been used and there is also no information available about the security of the application. The latter one can be explained by the fact that the application does not collect any information. However, the application could improve by allowing users to easily report feedback and by having clear policies.

	Version 0.1 instrument	
	The results are not an evaluation of the content of the application but are an indication about the amount of information that is publicly available.	
	Select an answer from the drop down menu or type in the date when required. An overview will be generated on the second tab.	Anwers should be inserted in this column
	Start of the review	
	Fill in the name of the application	**Moet ik naar de dokter**
	Main category of the application:	General_population
	Sub-category of the application: (First choose a main category)	Acute disease management (first aid & emergency care)
	Date of review:	**28-03-15**
	Preliminary research	
1	What is the target group of the application? ☐ Patients ☐ Healthcare professional / researcher ☑ General population	
2	In what year was the application published? (yyyy)	2013
3	What was the last time the application was updated? (dd-mm-yy)	05-01-15
4	Is the information in the application up-to-date and relevant?	Yes
5	For which operating systems is the application available? ☑ Android ☑ Apple iOS ☐ Windows Phone	
6	Is the application compatible with the newest versions of the operating systems?	Yes
7	Is the application required to have a CE mark?	Yes
8	What are the costs of using the application?	Free
9	Did the developer perform a feasibility study?	Yes
10	Did the developer perform a risk assessment	Yes
	Evidence	
11	Is there a description of the theory that is used?	No/unknown
12	Is there a description of the frameworks/models that are used?	No/unknown
13	Is there information available regarding clinical tests?	No/unknown
14	Is there information regarding the testing of the reliability and the validity?	No/unknown
15	Are the negative effects of the application stated?	No/unknown
16	Is any of the test data available?	No/unknown
17	Is it clear what the added value of the application is?	Yes
	Usability	
18	Is the usability of the application tested?	No/unknown
19	Does the application adhere to existing standards?	Yes
20	Is the application interoperable with other systems?	No/unknown
21	Is the satisfaction of the users measured?	No/unknown
22	Is the application easy to understand for the target group?	Yes
	Security	
23	Is there a clear security policy available?	No/unknown
24	Is it easy to report security related problems?	No/unknown
25	Is there an ethical hack test performed?	No/unknown
26	Is there a responsible disclosure policy?	No/unknown
27	Is it clear where the servers are located?	No/unknown
28	Is there a clear privacy policy available?	Yes
29	Is it easy to report privacy related problems?	No/unknown
30	Is it possible to set a pin code for the application?	No/unknown
31	Is it possible to delete your account and the related data that been gathered by the application?	No/unknown
32	Can users download their data?	No/unknown
33	Is it clear if the application uses services from third parties?	Yes
	Transparency	
34	Does the application provide authorship information?	Yes
35	Is there a list of references?	No/unknown
36	Does the application provide information about any sponsorship?	Yes
37	Does the application have any certificates?	Yes

Figure 21 Answers instrument Moet ik naar de dokter

Application reviewed is: **Moet ik naar de dokter**
The main category of the application is: General_population
The sub category is: Acute disease management (first aid & emergency care)
Current date: 28-03-15

Category	Level of available information:
Preliminary	100,0%
Evidence	14,3%
Usability	40,0%
Security	18,2%
Transparency	75,0%



The target group of the application is/are:
Healthcare professional or researcher
General population
Patients

The application was published in: 2013
The last time the application is updated 05-01-15

The application is available for:
Windows Phone
Apple iOS
Android
The application is available for the newest operating systems

Financial information:
The application is free to use

CE mark
The application is required to have a CE mark, the developer is required to identify the risks of the application by performing a risk assessment.

Warning: The average amount of available information is low. Try to find more information before using the application or consider using another application

Warning appears when the average level of transparency is lower than 50%

Figure 22 Report instrument Moet ik naar de dokter

9.6 EHBO

Preliminary research:

EHBO is a free application from the Red Cross (Dutch: Rode Kruis). The application was tested on an iPad and a smartphone with Android. The application is developed for smartphone users in general. The Red Cross aims to educate people in first aid. Users can furthermore search for AED locations and locations of first aid posts. The application is available for all major platforms; only the Windows platform lacks updates. A CE mark is unnecessary since it only gives instructions on how to perform first aid.

Evidence:

The information in the application is based on evidence-based guidelines. All of the sources are included in the application.

Usability:

There is no information available on how the usability is tested. The application is easy to use and all of the terminology is explained. The jury from the Health app award and the user reviews in the app stores are positive. The application also has a function to give the instructions by voice. This is very convenient for issues such as reanimation. The sound of the phone is automatically turned on to full volume. It then gives instructions on how to reanimate and also plays the rhythm for the chess compressions.

Security:

It is stated in the application that name, postcode and email address are transferred to the servers of EHBO in a secure way. There is furthermore a privacy policy included and users can provide feedback to the developers. The data capture showed that the application did not collect unnecessary data.

Transparency:

The application provides detailed information about their sources and sponsor. The sponsor has a commercial interest in selling first aid kits.

Report instrument

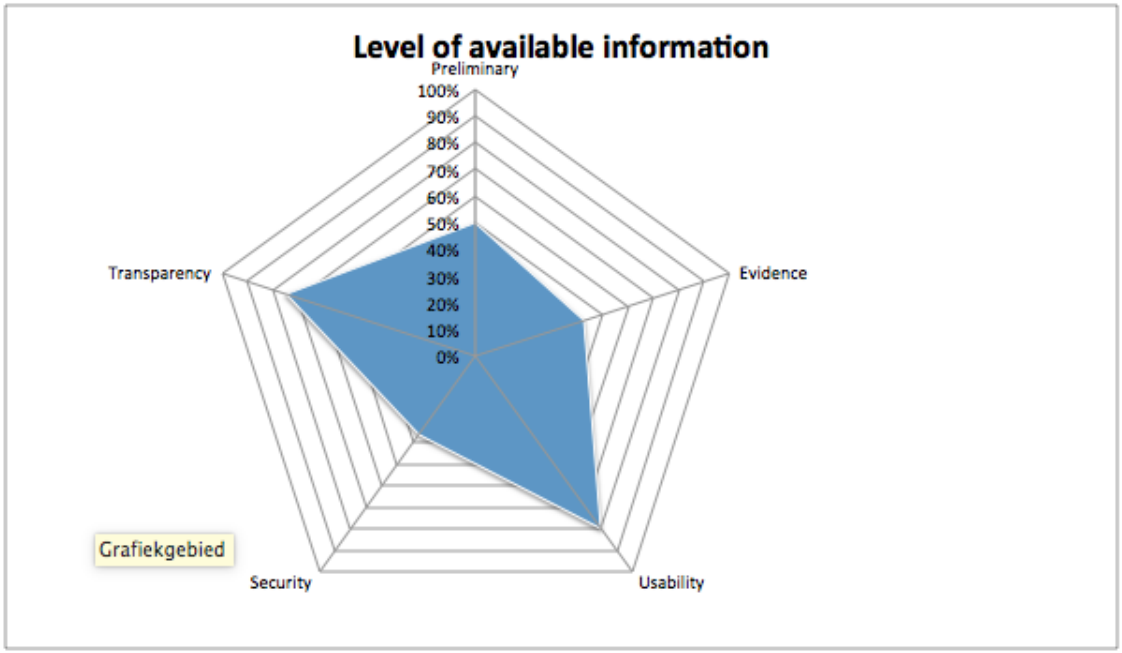
In Figure 23 are the answers from the EHBO application. In Figure 24 is the generated report. This application also scores low on the category security. However, this is also an application that does not gather information. The application does score high on usability and transparency.

☐ Version 0.1 instrument		
The results are not an evaluation of the content of the application but are an indication about the amount of information that is publicly available.		
Select an answer from the drop down menu or type in the date when required. An overview will be generated on the second tab.		Anwers should be inserted in this column
Start of the review		
	Fill in the name of the application	**EHBO**
	Main category of the application:	Health_researchers
	Sub-category of the application: (First choose a main category)	Acute disease management (first aid & emergency care)
	Date of review:	**28-03-15**
	Preliminary research	
1	What is the target group of the application? ☐ Patients ☑ Healthcare professional / researcher ☑ General population	
2	In what year was the application published? (yyyy)	2012
3	What was the last time the application was updated? (dd-mm-yy)	19-05-14
4	Is the information in the application up-to-date and relevant?	Yes
5	For which operating systems is the application available? ☑ Android ☑ Apple iOS ☑ Windows Phone	
6	Is the application compatible with the newest versions of the operating systems?	Yes
7	Is the application required to have a CE mark?	No/unknown
8	What are the costs of using the application?	Free
9	Did the developer perform a feasibility study?	No/unknown
10	Did the developer perform a risk assessment	No/unknown
	Evidence	
11	Is there a description of the theory that is used?	Yes
12	Is there a description of the frameworks/models that are used?	Yes
13	Is there information available regarding clinical tests?	No/unknown
14	Is there information regarding the testing of the reliability and the validity?	No/unknown
15	Are the negative effects of the application stated?	No/unknown
16	Is any of the test data available?	No/unknown
17	Is it clear what the added value of the application is?	Yes
	Usability	
18	Is the usability of the application tested?	Yes
19	Does the application adhere to existing standards?	Yes
20	Is the application interoperable with other systems?	Yes
21	Is the satisfaction of the users measured?	No/unknown
22	Is the application easy to understand for the target group?	Yes
	Security	
23	Is there a clear security policy available?	No/unknown
24	Is it easy to report security related problems?	Yes
25	Is there an ethical hack test performed?	No/unknown
26	Is there a responsible disclosure policy?	No/unknown
27	Is it clear where the servers are located?	No/unknown
28	Is there a clear privacy policy available?	Yes
29	Is it easy to report privacy related problems?	Yes
30	Is it possible to set a pin code for the application?	No/unknown
31	Is it possible to delete your account and the related data that been gathered by the application?	No/unknown
32	Can users download their data?	No/unknown
33	Is it clear if the application uses services from third parties?	Yes
	Transparency	
34	Does the application provide authorship information?	Yes
35	Is there a list of references?	Yes
36	Does the application provide information about any sponsorship?	Yes
37	Does the application have any certificates?	No/unknown

Figure 23 Answers instrument EHBO

Application reviewed is: **EHBO**
The main category of the application is: Health_researchers
The sub category is: Acute disease management (first aid & emergency care)
Current date: 28-03-15

Category	Level of available information:
Preliminary	50,0%
Evidence	42,9%
Usability	80,0%
Security	36,4%
Transparency	75,0%



The target group of the application is/are:
Healthcare professional or researcher
General population
Patients

The application was published in: 2012
The last time the application is updated 19-05-14

The application is available for:
Windows Phone
Apple iOS
Android
The application is available for the newest operating systems

Financial information:
The application is free to use

CE mark
The application does not have a CE mark, the developer is therefore not required to perfrom a risk assessment.

Figure 24 Report instrument EHBO

9.7 Conclusion per category

The assessment on the different application was carried out to validate the instrument. The results are described per category in this section.

Preliminary research

The information for the category preliminary research was the easiest to find. Most of the Information about the purpose, currency, technology and financial information is standard included in the application stores. Information about a risk assessment and a feasibility study were not found for any of the applications. The organizations responsible for the application should be contacted to find out if they performed a risk assessment and feasibility study. It is furthermore important that this information is shared. It can also be difficult to determine if there is relevant legislation for an application. To determine if an application needs a CE mark is clear for most applications. Since most applications do not measure vital body functions, do not add energy to the human body and they do not make a diagnosis. However some applications are in a grey area, it is in this case unclear for te average user if an application needs a CE mark. It is furthermore impossible for an average user to determine if the application abides the data protection act (Wet bescherming persoonsgegevens). The points that could not be found in the applications are:

- Was the target group involved in the development of the application? (**Purpose**)
- Are the developer actively updating and improving the application? (**Currency**)
- Is the information used in the application up-to-date/relevant? (**Currency**)
- Is there any other legislation that is relevant for the application? (**Legislation**)
- **Risk assessment;**
- **Feasibility**;
- Do insurance companies cover the application?

It can be concluded for preliminary research that most information can be found easily. However, information about a risk assessment and a feasibility study could not be found at all. Most users would probably not contact the developer for this information. However, healthcare organizations should ensure that they have a good understanding of the application before they ‘subscribe’ applications to patients.

Evidence

Information about evidence was most of the times not included in the application or on the website. Two of the applications included information about accuracy. The EHBO application specified the used theory and Huidmonitor specified the healthcare professionals included in the development of the application. Information about testing or transferability was not mentioned in any application. There were also no papers found about the applications on Google scholar or PubMed. The points that could not be found are:

- **Evidence;**
- **Transferability.**

It can be concluded that the information included about evidence lacks in all of the tested applications. The main reason for this is probably that the information is not there, or that they do not want to share this information. If there were papers that tested the application they would probably be easy to find. The lack of these papers and the lack of information in the applications suggest that they are not tested. Although testing for evidence is not always necessary.

Usability

The least amount of information was available about usability. The tested application had the same look and feel as any other iOS/Android application, i.e. they probably used the same human interface guidelines published by those platforms. The category was tested by using the application and by reading comments of users and of the jury of the Health app award. However, more interesting would be to have information about usability testing. None of the points that are included in the category usability were explicitly described. However, it can be concluded from the user and the jury reviews that the developers considered how to ensure that the usability is as high as possible. The websites of the developers furthermore claim that usability is an important point that has attention in their development process.

Security

Security and privacy policies were easily available in most of the applications. They clearly stated what information was saved on the device and what information was uploaded to the servers. The data capture furthermore showed that the applications did not send any irrelevant data while they were tested. However, it remains difficult to truly test the security and privacy of an application. Most of the times points are not mentioned at all. It was assumed that points such as “is there an ethical hack test performed?” where not executed when they were not mentioned. Other points such as, “does the application explicitly ask to use the data that is entered?” where mentioned, but not explicitly asked for. The following information could not be found:

- Is the developer legally allowed to work with the patient data? (**Data security**)
- How does the developer handle anonymization/pseudonymization? (**Data security**)
- What are the security requirements of: (**Data security**)
    - The phone;
    - The application;
    - The servers
    - Other devices that use data from the application?
- Where are the servers located? (**Data security**)

- Are there other applications running simultaneously on the device and what type of interaction might you expect between the application that is under assessment and other applications? (**Data security**)
- How long is the information stored? (**Participant privacy**)

It can be concluded that not all of the information in the category security is specified. Organizations probably do not want to share all of this information. However, it remains an important subject that should be paid attention to when applications are assessed. Cooperating with the developer of the application could help identify information about these points.

Transparency

Information such as authorship, sponsors and affiliations could be easily found in the reviewed applications. It is hard to determine for the average user if the information is non-biased and balanced. Experts could more easily give an advice about the bias in applications. Information about certification is generally also easy to find. It can be concluded that an application is not transparent when there is no information available. However, there was no information found about the following points:

- Algorithms used (**Transparency**);
- Degree to which end-user (and sometimes experts) are used (**Transparency**);

9.8 Final conclusion of applying the instrument

A final conclusion is that there is quite some information available for the average user to determine if they want to assess a certain application. However, detailed information about topics such as risk assessments is not publicly available. The assessments of the applications in this chapter were performed quite fast. Information for each application was found within an hour. This can be found even faster if only the application store and the website of the application are searched for available information. However, only searching these two aforementioned sources could mean an incomplete overview. More difficult for most users is to perform a data capture of each application. The set-up for this can a lot of time when users are not familiar with the technology.

A quick assessment of an application, as in this chapter, could be a first step in assessing applications by healthcare institutes. Determine what information is available, i.e. how transparent is the application? Then, subsequently do a more in depth assessment of an application together with the developer. The in depth assessment is necessary since evidence remains an important issue for mHealth applications, and as previously described. All of the assessed applications lacked information about evidence.

The instrument could be improved on the following points:

- Not all questions are relevant for all applications; some categories score therefore sometimes low without any reason.
- The instrument only gives an indication about the amount of available information. It does not test the quality of information. The next version should also test the quality of the available information.

10 The final instrument and suggestions for utilization

As described in the first chapter, different stakeholders have different interest. This chapter includes the final instrument and a short overview of suggestions on how this instrument can be applied for different stakeholders. The information in this chapter is based on the previous chapters. These suggestions should ensure that each stakeholder effectively uses the instrument. This chapter will therefore answer the last sub-question:

How can this instrument be applied for different stakeholders?

10.1 The final instrument

The final instrument consists of two parts. First, there is the artifact/the proof of concept that is developed in Excel. Screenshots of this artifact are included in chapter 9. Users can select answers from a drop down menu and the generated report gives an indication about the level of available information. The second part is the additional explanation that should be added to the report when an application is assessed. The additional explanation is a piece of text explaining the individual criteria. The assessment of applications will only be complete when the generated report is combined with an explanation.

The assessment is furthermore divided in different steps, the steps can be seen in Figure 25 below. The different steps should ensure that every stakeholder could use the instrument to assess available mHealth applications. The different steps are explained on the next page.

The instrument has been realized by conducting a literature study, organizing a focus group and by conducting several interviews. The preparatory step and the first two steps are also clearly visible in the (Excel) artifact. The third step requires cooperation with the developer and could not be included in the proof of concept.

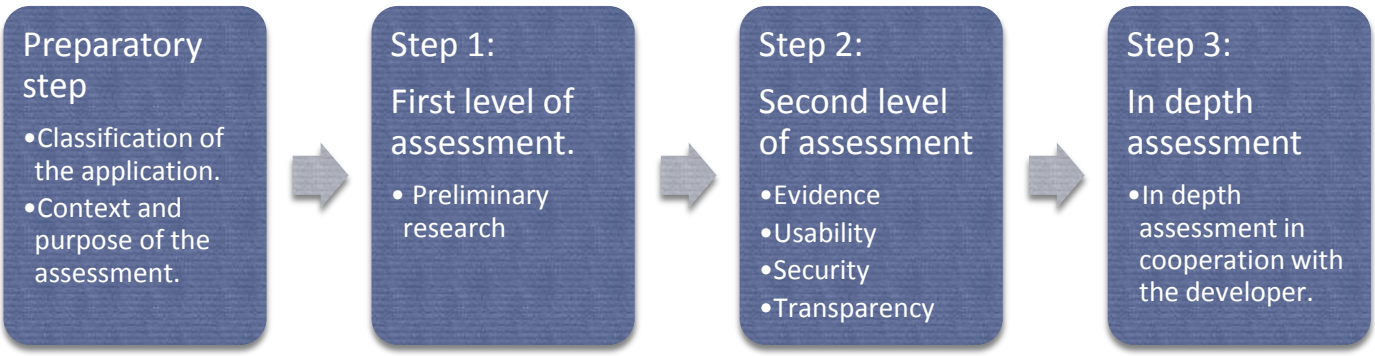


Figure 25 Steps in the final instrument

10.1.1 Preparatory step

This step is a preparatory step that should be performed before the assessment starts. This step is based on the input from the focus group. Participants agreed that the purpose of the assessment should be described before the assessment starts. Stakeholders included and the goal of the assessment should furthermore be clear before the assessment starts. It is furthermore important to classify the application. This classification used in this research can be viewed in Figure 3 on page 13. This is important since the results from the focus group suggested that some criteria are more important for different types of application. Extra attention could be given to these criteria during an assessment. However, further research should provide a clear answer to the criteria that are important for different types of applications. The classification that is used in this research is also included in the proof of concept. Users can indicate the main and the sub-category of an application.

10.1.2 Step 1

The first step is created to get a general overview of the application. As described earlier in chapter 7.2 on page 59, the purpose of the category preliminary research is to quickly rate an application before doing an extensive, in depth, assessment. Criteria from the category “preliminary research” should therefore be answered to create a quick and general overview. A short overview of the criteria that are in this category are in Table 32 below. Information that is publicly available on the internet or in the application should be used to answer the questions in this category. What is the purpose of the application and what is the currency? It is for example important that the developer updates the application to ensure that it is compatible with the newest devices and with up-to-date information. Other criteria in this category determine if the application is in accordance with the right legislation and if there is adequate financial information available about the application. The last two criteria from this category are if there is a risk assessment and a feasibility study performed. It could be that this information is not publicly available. However, websites of developers could give insight in these issues and applications with a CE mark are required to perform and to document a risk assessment. A full overview of the criteria and the questions that should be answered is in Table 27 on page 80. Preliminary research is also processed in the artifact, users can indicate if the information is present and if the application requires a CE mark.

#	Criteria Preliminary research
1	Purpose
2	Currency
3	Technology
4	Legislation
5	Risk assessment
6	Feasibility
7	Financial information

Table 32 Overview criteria Preliminary research

Results step 1

This step should result in a document where the different criteria from the category preliminary research are described. This document can be used to determine if further assessment is necessary. It is important that an explanation of each point is added.

10.1.3 Step 2

The next step in the instrument is to describe the categories that are in Table 28, Table 29, Table 30 and Table 31 on pages 81 to 84. These categories are called “evidence”, “usability”, “security” and “transparency”. A short overview of these criteria is in Table 33 below. These categories and their criteria are based on the literature study, focus group and interviews. The initial list was constructed after the literature review. The focus group and the interviewees reviewed these findings. Information that is publicly available should be used to answer the criteria in these categories. This could be more difficult since detailed information about these subjects is not always available. However, a lot of information can be found by using the application and by searching on the website of the developer.

#	Evidence
1	Accuracy
2	Evidence
3	Transferability
	Usability
4	Usability
5	Behavior change maintenance
6	Interoperability
7	Satisfaction
	Security
8	Data security
9	Participant privacy
	Transparency
10	Objectivity
11	Credibility
12	Certification
13	Ethical

Table 33 Overview criteria step 2

The category evidence starts with the criterion accuracy. This criterion is about the information in the application, who or what the input is for this information and if this information is correct. The next two criteria are evidence and transferability. Applications that tested for evidence are likely to describe this in the application or on the website and should therefore be easy to find. Medical professionals could also search for scientific articles that tested the applications.

The next category is usability. The criteria in this category are usability, user satisfaction, interoperability with other systems, and if the application ensures that a behavior change is maintained.

Security is the next category. It can be difficult for an average user to determine if the privacy and the security of an application are reliable. However, the description in Table 30 includes questions that can be answered by the average user. Privacy and security policies are a good source for finding information about these criteria.

The last category is transparency. Criteria about the objectivity and the credibility are included in this category. Are the sources, sponsorship and other affiliations included in the application or on the website and what is the credibility of these sources? The last two criteria are if the application has certain certifications and if it is ethical to use the application.

Results step 2

Step 2 should result in a document with information about the evidence, usability, security and the transparency of the application. The categories and the criteria should be described instead of short yes/no answers. This document should determine if it is necessary to continue with step 3, or to quit the assessment of the application. The assessment should continue if the application has potential based on the available information, but when there is not enough information available to create a solid overview. The assessment could stop when the information found indicate a low level of quality.

Step 2 furthermore results in a report generated from the answers when they are filled in proof of concept. This gives an overview of the level of transparency of the application.

Not all of the criteria can be answered in this step since the information is most of the time not publicly available. This lack of information could indicate a low level of transparency.

10.1.4 Step 3

The third step is to perform an in depth assessment. This step should be performed together with the developer/responsible organization of the application. It is likely that not all of the information was found in step 1 and step 2. The goal of the third step is therefore to collect information that was not found during the previous two steps.

Results step 3

The documentation that resulted from the previous steps can be used as input for this step. The result of this step is an elaboration on the documents from the previous steps.

10.2 Suggestion per stakeholder

Healthcare providers and insurance companies could at least perform the preparatory step and the first three steps of the instrument, performing an in depth assessment of an application. The incentive for healthcare providers is to decide if the application is suitable to subscribe to patients. While insurance companies would use the assessment to determine which applications are eligible

for reimbursement. It is important in both cases to involve the developer or responsible organization in the assessment. Reason for this is that not all of the information is publicly available. If necessary, they could do step 4, assisting with further research.

Unlike organizations, users (patients) could use the instrument for a quick assessment by performing the preparatory step and the first two steps. Determining if they should use a certain application. Their assessment could be based on information available in the application, application stores and on the website of the application. They can furthermore rate the application on usability after using the application. The advantage of using the instrument is that they get a clear overview of reasons of advantages and disadvantages of an application.

Developers could use the instrument during the development phase of an mHealth application. Turning the instrument in a checklist of points that should be kept in mind during development.

11 Conclusion

This chapter includes the conclusion of this research. First, the sub-questions are answered followed with a conclusion of the main research question.

11.1 Conclusion per sub-question

The basis of this research was the literature study performed at the start. The sources found with the literature study were sufficient to answer the first three sub-questions. The results of the literature study were combined and presented to experts during a focus group. The focus group and the interviews provided input to answer sub-question four and five. Sub-question six was answered by reviewing mHealth applications. The last sub-question was answered by combining information from the previous chapters. The sub-chapters below provide a conclusion for every sub-question.

11.1.1 Sub-question one

The first question focused on the different types of applications that exist and how mobile health has evolved. The question was posed as:

"What types of mobile health applications can be distinguished and how has mobile health evolved?"

The different types of mHealth applications that can be distinguished can be seen in Figure 3 on page 24. This classification divided the applications in three main categories, which are called:

- Tools for health research;
- Improving health services;
- Improving health outcomes.

Each category has one or two sub-categories. Lastly, the sub-categories include different types of mHealth applications. The second part of the question focused on the evolution of mHealth applications. The first "mHealth applications" were simple reminders or lists that users could keep track of on their mobile phone. Applications as we know them today only appeared after the modern smartphones and the application stores were introduced. Figure 6, together with the explanation of the history of mHealth give an answer to sub-question one.

11.1.2 Sub-question two

The second sub-question was also answered with literature. The second sub-question was posed as:

"What do we know about the added requirements of mobile health for different stakeholders?"

Various criteria were found during the literature study. The criteria that most frequently appeared in the literature are 'security' and 'privacy'. Users find it important that the application and their data are safe. Another frequently reoccurring requirement is 'behavior change maintenance'. Applications that aim to change behavior should ensure that this change is lasting, even if the application is not

used anymore. Other requirements found are ‘usability’, ‘evidence’ and ‘transparency’. These are all points that are found important by stakeholders according to scientific literature.

11.1.3 Sub-question three

The last sub-question that was answered with literature is sub-question three. This question was drafted as:

“How are mobile health applications currently assessed?”

A long list of 21 existing frameworks and models was extracted from the literature. The short list, existing of seven models, are described and compared with each other. The majority of the existing models and frameworks are focused on the implementation of mHealth applications in organizations. The result of this focus is that these frameworks also include topics such as the financial health of the organization and if the employees are willing to use a mHealth application. However, a comparison resulted in a list of criteria that could be used for the assessment of mHealth applications. Important criteria that are also relevant for this research are “evidence”, “usability” and “security”.

The guidelines from organizations were compared with each other next to the models found in literature. This comparison shows that their main focus is on “usability”, “authority” (who made the application) and on “currency” (is the application relevant).

The results from scientific literature and the guidelines from organizations show that there is not a solid method or instrument available to adequately test a mHealth application. However, they do provide criteria that could be important to include. Some of these criteria match with the criteria found with sub-question two.

11.1.4 Sub-question four and five

The results from the previous sub-questions were combined and used as input for sub-question four and five. These questions were posed as:

“What are assessment criteria and what level of evidence is needed for a swift and critical appraisal instrument for mobile health applications?“

“How can the criteria be modeled into a critical appraisal instrument that can be tested?”

The assessment criteria found with the previous questions was used as input for this question. Interviews were conducted and a focus group with experts was organized to establish a final list with criteria. The reason for this was that there could be criteria missing or that some of the included criteria were not relevant. The list with assessment criteria resulting from the focus group can be found in Appendix G: List of criteria focus group. However, this is not the final list of criteria, this list can be found in Table 27 on page 80. The list exists of five categories, based on the categories found in scientific frameworks. The categories are called:

1. Preliminary research;

2. Evidence;
3. Usability;
4. Security;
5. Transparency.

Each of the categories exists of multiple criteria.

The participants of the focus group agreed that the instrument should be modeled in such a way that there are different levels of evidence. A first assessment could exist of a check if the developer has performed a feasibility study. A next step could be if the application complies with the list of criteria resulted from this research. The last step could be a randomized controlled trial. These levels are suggestions and should be researched for confirmation.

11.1.5 Sub-question six

The sixth sub-question was to validate the instrument on a sample of mHealth applications. The sixth sub-question was drafted as:

"What are the results of applying the critical appraisal instrument on a sample of mobile health applications?"

The results show us that not all of the information for an assessment is publicly available. However, an assessment with publicly available information is relatively fast. This could therefore be a first step for the assessment of mHealth applications. The next step would than be to cooperate with the developer to fill in the remaining points. However, case studies are needed to confirm these findings.

11.1.6 Sub-question seven

The last sub-question was posed as:

"How can stakeholders that are involved in mobile health use the critical appraisal instrument?"

It became clear during this research that each stakeholder has his own interest. The different steps in the instrument should make it easier for each stakeholder to use the instrument. Healthcare professionals and insurance companies could do an assessment that is in depth, preferably in cooperation with the developer of the application. They can decide based on the results if the application is suitable for patients. Users could use the instrument to do a quick assessment, based on the information that is publicly available. Resulting in an overview of reasons with advantages and disadvantages of an application. Developers could use the instrument during the development phase of an mHealth application.

11.2 Conclusion main research question

The main research question of this research is:

"What are the requirements of an instrument that enables a rapid, systematic and critical assessment of mobile health applications, and how can this instrument be modeled and validated?"

The main research question is answered by the aforementioned sub-questions. The different requirements for an instrument are discussed and described during this research. Different levels of assessment should exist in the instrument to ensure that the assessment is rapid. Since elaborate testing for evidence can be time consuming. A preliminary study towards the application or a feasibility test provided by the developer could therefore be used as a first level. A step that should be taken before the assessment starts is to decide what the purpose of the assessment is. The list of criteria that resulted from this research should ensure that every application is assessed systematic and critical. These criteria were validated with a focus group with experts and various interviews. Important is that the criteria are not answered with yes/no answers; an explanation should be given for each point. The instrument was validated by reviewing four mHealth applications. The reviews showed that an initial assessment could be performed quite fast. The steps in the final instrument can be viewed in Figure 25, screenshots from the created artifact are in chapter 9.

12 Discussion

The limitations are described per sub-question, followed by further work that can result from this thesis.

12.1 Limitations

It became clear during the research process that there were several limitations. The research relied heavily on the literature study. This was not a problem for sub-question one, but interviews could have had an added value for answering sub-question two and three. Interviews with stakeholders could have resulted in requirements for the final instruments. Furthermore, interviews with Dutch organizations could have given an overview of how these organizations currently assess medical applications. None of the organizations included in sub-question three are solely focused on the Dutch market. It was attempted to conduct an interview with Zelfzorg Ondersteund[13], but they did not respond to the inquiry. Furthermore, the selection of literature during the SLR could have been stricter. The number of articles that were eventually included was low compared to the number of articles that have been read in full text.

The feedback from the participants of the focus group suggested that the focus group could be improved by having fewer activities. Since some participants needed more time for certain tasks and discussions had to be aborted due to the lack of time. The results of the focus group were furthermore not validated with other focus groups. There was also no time for an in depth discussion on how the instrument should be modeled and what the different levels of evidence should include.

The instrument was validated by a reviewing four mHealth applications. Only the first two steps of the instrument were tested due to a lack of time. Further research should test the instrument with multiple case studies in cooperation with developers.

The last limitation of the research is that the response on the questionnaire was too low (N=4). Results of this questionnaire can therefore not be used.

12.2 Future work

There is a lot of research possible in the field of mHealth. New research questions also arose during this research. Firstly, the results of this research should be tested in multiple case studies. The applications that are used in this case studies should encompass the whole spectrum of mHealth applications that are defined in the classification of Free et al. (2013). The assessments should be performed in cooperation with the developers to receive as much information as possible.

The second point that could be researched is different modeling techniques. The emphasis of this research was on the requirements that are part of such a model. Scientific literature and experts should give an answer regarding the modeling of these requirements in a solid, easy to use instrument. Clinimetric literature could be studied as a first step to determine how to model a proper

[13] http://zelfzorgondersteund.nl/

instrument for the medical field. Important is that the instrument not only tests if the information is available, but also the quality of the information.

The feasibility of the instrument should also be tested. It is unclear at this moment if people will use the artifact to assess their applications. It could be that people think that it takes too much time and that they are only interested in the results of the assessment.

The last point is to research the valorization potential of the final instrument. Important for this point is that the instrument has been fully developed. Different scenarios for valorization can be researched.

1. Independent organizations could test applications with the instrument. Stakeholders could pay the organization for access to the database of tested applications.
2. The instrument could also be available for free on the internet. Everybody could assess applications and share the results however they would like.
3. Another scenario for valorization is when healthcare institutes test the applications. Subscribing applications to patients could save money. The institutes could furthermore sell the data to other healthcare institutes.
4. It is also possible that insurance companies use the instrument and determine with the instrument if they reimburse applications. Profit is made when applications are subscribed that are cheaper or better than the current method.

Bibliography

A.T. Kearney. (2012). *Improving the evidence for mobile health* (p. 27).

Agency for Healthcare Research and Quality. (1999). Assessing the Quality of Internet Health Information: Summary. Retrieved from http://www.ahrq.gov/research/data/infoqual.html

Apple. (2013). WWDC Apple keynote. Retrieved from http://www.apple.com/nl/apple-events/june-2013/

Aungst, T. (2013). Apple app store still leads Android in total number of medical apps. Retrieved from http://www.imedicalapps.com/2013/07/apple-android-medical-app/

Barton, A. J. (2012). The regulation of mobile health applications. *BMC Medicine*, *10*(1), 46. doi:10.1186/1741-7015-10-46

Biernacki, P., & Waldorf, D. (1981). Snowball sampling: Problems and techniques of chain referral sampling.

Boulos, M. N. K., Brewer, A. C., Karimkhani, C., Buller, D. B., & Dellavalle, R. P. (2014). Mobile medical and health apps: state of the art, concerns, regulatory control and certification. *Online Journal of Public Health Informatics*, *5*(3), 229. doi:10.5210/ojphi.v5i3.4814

Boulos, M. N. K., Wheeler, S., Tavares, C., & Jones, R. (2011). How smartphones are changing the face of mobile and participatory healthcare: an overview, with example from eCAALYX. *Biomedical Engineering Online*, *10*(1), 24. doi:10.1186/1475-925X-10-24

Braun, R., Catalani, C., Wimbush, J., & Israelski, D. (2013). Community health workers and mobile technology: a systematic review of the literature. *PloS One*, *8*(6), e65772. doi:10.1371/journal.pone.0065772

Brown, W., Yen, P.-Y., Rojas, M., & Schnall, R. (2013). Assessment of the Health IT Usability Evaluation Model (Health-ITUEM) for evaluating mobile health (mHealth) technology. *Journal of Biomedical Informatics*, *46*(6), 1080–7. doi:10.1016/j.jbi.2013.08.001

Burke, L. E., Styn, M. a, Sereika, S. M., Conroy, M. B., Ye, L., Glanz, K., … Ewing, L. J. (2012). Using mHealth technology to enhance self-monitoring for weight loss: a randomized trial. *American Journal of Preventive Medicine*, *43*(1), 20–6. doi:10.1016/j.amepre.2012.03.016

Carroll, A., & Heiser, G. (2010). An Analysis of Power Consumption in a Smartphone. *USENIX Annual Technical Conference*, 14. Retrieved from https://www.usenix.org/event/usenix10/tech/full_papers/Carroll.pdf

Corral, E. (2012). Comprehensive & Seamless mHealth Best Practices.

Derballa, V., & Pousttchi, K. (2004). Extending knowledge management to mobile workplaces. *Proceedings of the 6th International Conference …*, (3616). Retrieved from http://dl.acm.org/citation.cfm?id=1052294

Ekker, A., & van Rest, B. (2013). *Medische apps, is certificeren nodig?* (pp. 1–10). Den Haag.

EU. (2013). Enterprise & Industry Magazine Fairer, safer, simpler: New rules for pressure equipment. Retrieved from http://ec.europa.eu/enterprise/magazine/articles/single-market-goods/article_11093_en.htm

European Commission. (2012). Guidelines on the qualification and classification of stand alone software used in healthcare within the regulatory framework of medical devices - MEDDEV 2.1/6.

Eysenbach, G. (2001). What is e-health? *Journal of Medical Internet Research*, *3*(2), 20. doi:10.2196/jmir.3.2.e20

Finch, T. L., Mair, F. S., & May, C. R. (2007). Teledermatology in the UK: lessons in service innovation. *The British Journal of Dermatology*, *156*(3), 521–7. doi:10.1111/j.1365-2133.2006.07608.x

Fiordelli, M., Diviani, N., & Schulz, P. J. (2013). Mapping mHealth research: a decade of evolution. *Journal of Medical Internet Research*, *15*(5), e95. doi:10.2196/jmir.2430

Franko, O. I., & Tirrell, T. F. (2012). Smartphone app use among medical providers in ACGME training programs. *Journal of Medical Systems*, *36*(5), 3135–9. doi:10.1007/s10916-011-9798-7

Free, C., Phillips, G., Felix, L., Galli, L., Patel, V., & Edwards, P. (2010). The effectiveness of M-health technologies for improving health and health services: a systematic review protocol. *BMC Research Notes*, *3*(1), 250. doi:10.1186/1756-0500-3-250

Free, C., Phillips, G., Watson, L., Galli, L., Felix, L., Edwards, P., … Haines, A. (2013). The effectiveness of mobile-health technologies to improve health care service delivery processes: a systematic review and meta-analysis. *PLoS Medicine*, *10*(1), 26. doi:10.1371/journal.pmed.1001363

Galpottage, P. a. B. (2005). Patient consent principles and guidelines for e-consent: a New Zealand perspective. *Health Informatics Journal*, *11*(1), 5–18. doi:10.1177/1460458205050681

Gezondheidszorg, I. voor de. (2014). Kwaliteitswet zorginstellingen (KWZ). Retrieved from http://www.igz.nl/onderwerpen/handhavingsinstrumenten/wetten/kwaliteitswet_zorginstellingen/

Gurman, T. a, Rubin, S. E., & Roess, A. a. (2012). Effectiveness of mHealth behavior change communication interventions in developing countries: a systematic review of the literature. *Journal of Health Communication*, *17 Suppl 1*(September 2014), 82–104. doi:10.1080/10810730.2011.649160

Herik, K. W. van den, & Vreede, G. J. de. (2000). Experiences with facilitating policy meetings with group support systems. *International Journal of Technology Management*, *19*, 246. doi:10.1504/IJTM.2000.002819

Hughes, B., Joshi, I., & Wareham, J. (2008). Health 2.0 and Medicine 2.0: tensions and controversies in the field. *Journal of Medical Internet Research*, *10*(3), e23. doi:10.2196/jmir.1056

Istepanian, R., Laxminarayan, S., & Pattichis, C. (2006). *M-health* (p. 619). Springer. Retrieved from http://link.springer.com/content/pdf/10.1007/b137697.pdf

Jokela, T., Iivari, N., Matero, J., & Karukka, M. (2003). *The Standard of User-Centered Design and the Standard Definition of Usability : Analyzing ISO 13407 against ISO* (pp. 53–60).

Jordan-Marsh, M. (2011). *Health Technology Literacy: A Transdisciplinary Framework for Consumer Oriented Practice* (p. 448).

Khoja, S., Durrani, H., Scott, R. E., Sajwani, A., & Piryani, U. (2013). Conceptual framework for development of comprehensive e-health evaluation tool. *Telemedicine Journal and E-Health : The Official Journal of the American Telemedicine Association*, *19*(1), 48–53. doi:10.1089/tmj.2012.0073

Kidholm, K., Ekeland, A. G., Jensen, L. K., Rasmussen, J., Pedersen, C. D., Bowes, A., … Bech, M. (2012). A model for assessment of telemedicine applications: mast. *International Journal of Technology Assessment in Health Care*, *28*(1), 44–51. doi:10.1017/S0266462311000638

Kitchenham, B., Pretorius, R., Budgen, D., Pearl Brereton, O., Turner, M., Niazi, M., & Linkman, S. (2010). Systematic literature reviews in software engineering – A tertiary study. *Information and Software Technology*, *52*(8), 792–805. doi:10.1016/j.infsof.2010.03.006

Kotz, D., Avancha, S., & Baxi, A. (2009). A Privacy Framework for Mobile Health and Home-Care Systems. *SPIMACS*, 1–12.

Krijgsman, J., Peeters, J., Burghouts, A., Brabers, A., de Jong, J., Beenkens, F., … van Gennip, L. (2014). *eHealth, Op naar meerwaarde!* (p. 160). Den Haag.

Krishna, S., Boren, S. A., & Balas, E. A. (2009). Healthcare via cell phones: a systematic review. *Telemedicine Journal and E-Health : The Official Journal of the American Telemedicine Association*, *15*(3), 231–40. doi:10.1089/tmj.2008.0099

Krouse, A. (2012). iPads, iPhones, Androids, and Smartphones: FDA Regulation of Mobile Phone Applications as Medical Devices. *Ind. Health L. Rev.*, *9*(2), 731–764. Retrieved from http://heinonlinebackup.com/hol-cgi-bin/get_pdf.cgi?handle=hein.journals/inhealr9&section=29

Kumar, S., Nilsen, W. J., Abernethy, A., Atienza, A., Patrick, K., Pavel, M., … Swendeman, D. (2013). Mobile health technology evaluation: the mHealth evidence workshop. *American Journal of Preventive Medicine*, *45*(2), 228–36. doi:10.1016/j.amepre.2013.03.017

Laakko, T., Leppänen, J., Lähteenmäki, J., & Nummiaho, A. (2008). Mobile Health and Wellness Application Framework. *Methods of Information in Medicine*, 217–222. doi:10.3414/ME9113

Lee, J. (2013). Grading the Quality of Information and Synthesis of mHealth Evidence.

Leon, N., Schneider, H., & Daviaud, E. (2012). Applying a framework for assessing the health system challenges to scaling up mHealth in South Africa. *BMC Medical Informatics and Decision Making*, *12*, 123. doi:10.1186/1472-6947-12-123

Lindley, J., & Fernando, J. (2013). Being Smart : Challenges in the use of mobile applications in clinical settings, (November), 5–14.

Luxton, D. D., Kayl, R. a, & Mishkind, M. C. (2012). mHealth data security: the need for HIPAA-compliant standardization. *Telemedicine Journal and E-Health : The Official Journal of the American Telemedicine Association*, *18*(4), 284–8. doi:10.1089/tmj.2011.0180

Maheu, M., Whitten, P., & Allen, A. (2001). *E-health, telehealth, and telemedicine: a guide to startup and success* (p. 380). John Wiley & Sons.

March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems*, *15*(4), 251–266. doi:10.1016/0167-9236(94)00041-2

Martínez-Pérez, B., de la Torre-Díez, I., Candelas-Plasencia, S., & López-Coronado, M. (2013). Development and evaluation of tools for measuring the quality of experience (QoE) in mHealth applications. *Journal of Medical Systems*, *37*(5), 9976. doi:10.1007/s10916-013-9976-x

Mechael, P., & Searle, S. (2010). Barriers and Gaps Affecting mHealth in Low and Middle Income Countries : Policy White Paper, 1–79.

Meulendijk, M. C., Drenth-van Maanen, C., Jansen, P., Brinkkemper, S., Numans, M., & Spruit, M. (2013). Introducing the COrETeSt feasibility analysis in medical informatics: a case study of a decision-supportive knowledge system in the Dutch primary care sector. In M. M. Cruz-Cunha, I. M. Miranda, & P. Gonçalves (Eds.), *Handbook of Research on ICTs and Management Systems for Improving Efficiency in Healthcare and Social Care* (pp. 805–829). IGI Global. doi:10.4018/978-1-4666-3990-4

Mosa, A. S. M., Yoo, I., & Sheets, L. (2012). A systematic review of healthcare applications for smartphones. *BMC Medical Informatics and Decision Making*, *12*(1), 67. doi:10.1186/1472-6947-12-67

Mulrow, C. D. (1994). Rationale for systematic reviews, *309*(September), 597–599.

Needs, S., Health, C., & Philbrick, W. C. (2013). mHealth and MNCH : State of the Evidence, (January).

Nilsen, W., Kumar, S., Shar, A., Varoquiers, C., Wiley, T., Riley, W. T., … Atienza, A. a. (2012). Advancing the science of mHealth. *Journal of Health Communication*, *17 Suppl 1*(April 2014), 5–10. doi:10.1080/10810730.2012.677394

Norris, a C., Stockdale, R. S., & Sharma, S. (2009). A strategic approach to m-health. *Health Informatics Journal*, *15*(3), 244–53. doi:10.1177/1460458209337445

Oh, H., Rizo, C., Enkin, M., & Jadad, A. (2005). What is eHealth (3): a systematic review of published definitions. *Journal of Medical Internet Research*, *7*(1), e1. doi:10.2196/jmir.7.1.e1

Ouma, S., Herselman, M., & Vangrauen, D. (2011). Factors that influence m-health implementations in resource constrained areas in the developing world (pp. 1–9). Retrieved from http://hdl.handle.net/10204/5528

Patrick, K., Griswold, W. G., Raab, F., & Intille, S. S. (2009). Health and the Mobile Phone, *35*(2), 177–181. doi:10.1016/j.amepre.2008.05.001.Health

Payne, K. F. B., Wharrad, H., & Watts, K. (2012). Smartphone and medical related App use among medical students and junior doctors in the United Kingdom (UK): a regional survey. *BMC Medical Informatics and Decision Making*, *12*(1), 121. doi:10.1186/1472-6947-12-121

Pingree, S., Hawkins, R., Baker, T., Dubenske, L., Roberts, L. J., & Gustafson, D. H. (2011). The Value of Theory for Enhancing and Understanding e-Health, *38*(1), 103–109. doi:10.1016/j.amepre.2009.09.035.The

Pocketgamer. (2014). Application Category Distribution. Retrieved from http://www.pocketgamer.biz/metrics/app-store/categories/

Porter, M. (2010). What is value in health care. *The New England Journal of Medicine*, 2477–2481.

Riley, W. T., Rivera, D. E., Atienza, A. a, Nilsen, W., Allison, S. M., & Mermelstein, R. (2011). Health behavior models in the age of mobile interventions: are our theories up to the task? *Translational Behavioral Medicine*, *1*(1), 53–71. doi:10.1007/s13142-011-0021-7

Rowinski, D. (2013). Google Play Hits One Million Android Apps. Retrieved from http://readwrite.com/2013/07/24/google-play-hits-one-million-android-apps

Schäfer, W., Kroneman, M., Boerma, W., Berg, M. Van Den, Westert, G., Devillé, W., & Ginneken, E. Van. (2010). The Netherlands: Health system review. *Health Systems in Transition*, *12*(1).

Simon, H. A. (1996). *The Sciences of the Artificial* (Third edit., p. 240). MIT Press.

Statista. (2010). Growth of the top Apple App Store categories for iPhone apps in the U.S. in 2010.

Steinhubl, S., Muse, E., & Topol, E. (2013). Can Mobile Health Technologies Transform Health Care ? *The Journal of the American Medical Association*, *310*(22), 2395–2396. doi:10.1001/jama.2013.281078.Conflict

Tamrat, T., & Kachnowski, S. (2012). Special delivery: an analysis of mHealth in maternal and newborn health programs and their outcomes around the world. *Maternal and Child Health Journal*, *16*(5), 1092–101. doi:10.1007/s10995-011-0836-3

Tate, E. B., Spruijt-Metz, D., O'Reilly, G., Jordan-Marsh, M., Gotsis, M., Pentz, M. A., & Dunton, G. F. (2013). mHealth approaches to child obesity prevention: successes, unique challenges, and next directions. *Translational Behavioral Medicine*, *3*(4), 406–15. doi:10.1007/s13142-013-0222-3

Tomlinson, M., Rotheram-Borus, M. J., Swartz, L., & Tsai, A. C. (2013). Scaling up mHealth: where is the evidence? *PLoS Medicine*, *10*(2), e1001382. doi:10.1371/journal.pmed.1001382

Unovate. (n.d.). Unovate. Retrieved from http://www.umcutrecht.nl/subsite/unovate/english/

Vaishnavi, V. K., & Kuchler, W. (2004). *Design science research methods and patterns: innovating information and communication technology* (p. 227). New York: Auerbach publications.

Van De Belt, T. H., Engelen, L. J. L. P. G., Berben, S. A. A., & Schoonhoven, L. (2010). Definition of Health 2.0 and Medicine 2.0: a systematic review. *Journal of Medical Internet Research*, *12*(2), e18. doi:10.2196/jmir.1350

Van Dyk, L. (2014). A review of telehealth service implementation frameworks. *International Journal of Environmental Research and Public Health*, *11*(2), 1279–98. doi:10.3390/ijerph110201279

Van Gemert-Pijnen, J. E. W. C., Nijland, N., van Limburg, M., Ossebaard, H. C., Kelders, S. M., Eysenbach, G., & Seydel, E. R. (2011). A holistic framework to improve the uptake and impact of eHealth technologies. *Journal of Medical Internet Research*, *13*(4), e111. doi:10.2196/jmir.1672

Van Velsen, L., Beaujean, D. J. M. a, & van Gemert-Pijnen, J. E. W. C. (2013). Why mobile health app overload drives us crazy, and how to restore the sanity. *BMC Medical Informatics and Decision Making*, *13*, 23. doi:10.1186/1472-6947-13-23

Whittaker, R. (2012). Issues in mHealth: findings from key informant interviews. *Journal of Medical Internet Research*, *14*(5), e129. doi:10.2196/jmir.1989

Whittaker, R., Merry, S., Dorey, E., & Maddison, R. (2012). A development and evaluation process for mHealth interventions: examples from New Zealand. *Journal of Health Communication*, *17 Suppl 1*(April 2014), 11–21. doi:10.1080/10810730.2011.649103

Wolf, J. a, Moreau, J. F., Akilov, O., Patton, T., English, J. C., Ho, J., & Ferris, L. K. (2013). Diagnostic inaccuracy of smartphone applications for melanoma detection. *JAMA Dermatology*, *149*(4), 422–6. doi:10.1001/jamadermatol.2013.2382

Yu, P., Wu, M., Yu, H., & Xiao, G. (2006). The Challenges for the Adoption of M-Health. In *International Conference on Service Operations and Logistics and Informatics* (pp. 181–186).

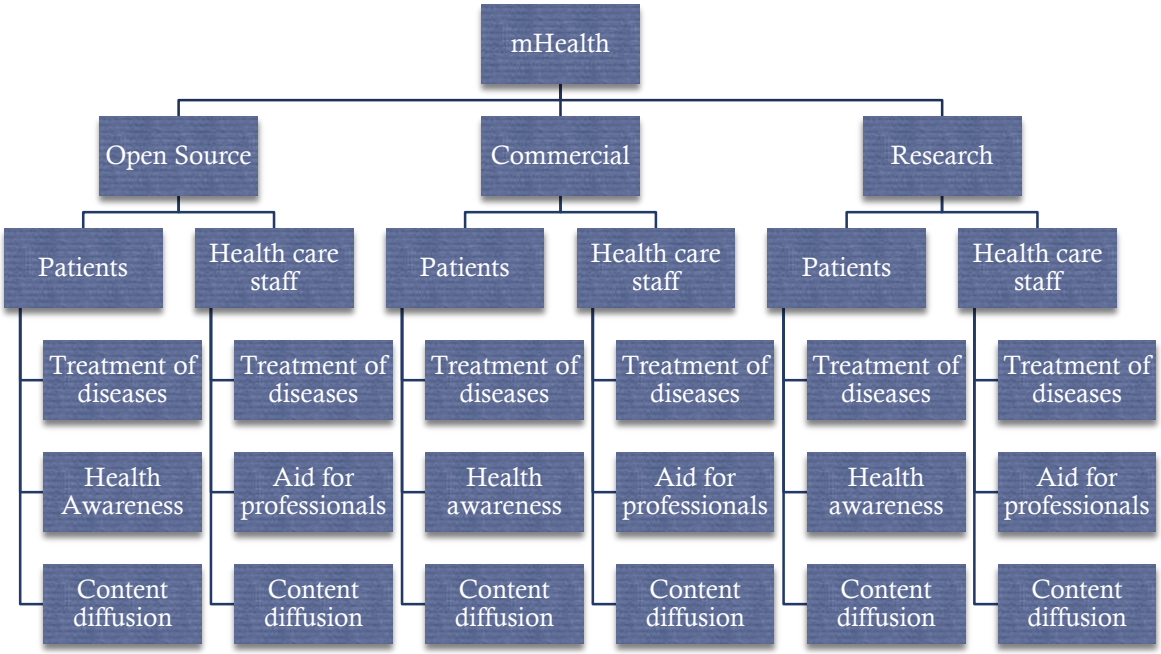
Appendices

Appendix A: Queries SLR

	Scholar.google.com		
#	Query	Number of results	Number of articles reviewed
1	Assessment of "mobile health applications" OR "mHealth applications"	922	200
2	Assessment Criteria “Mobile health applications” OR “MHealth applications”	399	200
3	Framework assessment requirements "mobile health" OR "MHealth"	5,590	200
4	Value of “MHealth” OR “Mobile health”	13,300	200
5	Testing of “MHealth applications” OR “mobile health applications”	1,030	200

PubMed			
#	Query	Number of results	Number of articles reviewed
1	Assessment of "mobile health applications" OR "mHealth applications"	22	22
2	Assessment Criteria “Mobile health applications” OR “MHealth applications”	20	20
3	Framework assessment requirements "mobile health" OR "MHealth"	256	100
4	Value of “MHealth” OR “Mobile health”	3370	100
5	Testing of “MHealth applications” OR “mobile health applications”	28	28

Appendix B: classification Martinez-Perez et al.



Appendix C: Long list models and frameworks

#	Framework/Model	Author	Title	Year
	Privacy			
1	A threat taxonomy for mHealth privacy	D. Kotz	A threat taxonomy for mHealth privacy	2011
	Development			
2	Development and intervention process mHealth	Whittaker et al.	A development and evaluation process for mHealth interventions: examples from New Zealand.	2012
3	Simplified evaluation process	Catwell et al.	Evaluating eHealth Interventions: The Need for Continuous Systemic Evaluation	2009
	Implementation			
4	Barriers to the diffusion of telemedicine	Tanriverdi et al.	Knowledge Barriers to Diffusion of Telemedicine	1998
5	eHealth readiness assessment tools	Khoja et al.	e-Health Readiness Assessment Tools for Healthcare Institutions in Developing Countreis	2007
6	Unified Theory of Acceptance and Use of Technology (UTAUT) applied to telehealth	Venkatesh et al.	User acceptance of information technology: Toward a unified view	2003
7	Seven Core Principles for the successful development of telemedicine systems	Yellowlees	Successfully developing a telemedicine system.	2005
8	Lessons in telemedicine service innovation	Finch et al.	Teledermatology in the UK: Lessons in service innovation	2006
9	Framework for Assessing the Health System Challenges to Scaling up mHealth	Leon et al.	Applying a framework for assessing the health system challenges to scaling up mHealth in South Africa	2012
10	Comprehensive Model for the Evaluation of Telemedicine	Hicks et al.	A comprehensive model for evaluating telemedicine	2004
11	The Layered Telemedicine Implementation Model	Broens et al.	Determinants of successful telemedicine implementations: A literature study	2007
12	The Khoja-Durrani-Scott (KDS) Evaluation Framework	Khoja et al.	Conceptual Framework for Development of Comprehensive e-Health Evaluation Tool	2013
	Assessment			
13	Model for Assessment of Telemedicine Applications: MAST	Kidholm et al.	A model for Assessment of Telemedicine Applications: MAST	2012
14	Research stages and standards for the dissemination of MHealth	Tomlinson et al.	Scaling up mHealth: Where is the evidence	2013

15	HOT-fit	Yusof et al.	An evaluation framework for Health Information Systems: human, organization and technology-fit factors (HOT-fit)	2008
16	Evaluation Framework	Kaufman et al.	Applying an Evaluation Framework for Health Information System Design, Development, and Implementation	2006
	Usability			
17	Health IT Usability Evaluation Model	Brown et al.	Assessment of the Health IT Usability Evaluation Model (Health-ITUEM) for evaluating mobile health (mHealth) technology.	2013
	Quality			
18	Perceived service quality of mHealth	Akter et al.	Development and validation of an instrument to measure user perceived service quality of mHealth	2013
	Other			
19	MHealth Grading Tool	J Lee	Grading the Quality of Information and Synthesis of mHealth Evidence	2013
20	Framework for Evaluating Mobile Applications for Cardiac Rehabilitation	Beatty et al.	Using mobile technology for cardiac rehabilitation: a review and framework for development and evaluation.	2013
21	Dimensions of eHealth Research	Dansky et al.	A framework for evaluating eHealth research	2006

Appendix D: AZHIN Guide

Accuracy

- Is the medical information contained in the app based on sound medical research and evidence? Can the information in the app be verified by another source?
- Are there references/sources included so that you can verify the information? Are these references reliable? (For example, a citation to a drug company website does not have the same weight as an article from *JAMA*.)
- Are there grammatical and spelling errors? (This may be a "tell" - if the information isn't even spelled correctly, maybe the information itself isn't correct.)
- Does the app do what it intends to do? Is there any potential for patient harm?

Bias/Objectivity

- Is the information showing just one point of view or is it sponsored by a company that is trying to sell something?
- What kind of organization sponsored the app? A pharmaceutical company? A non-profit organization? A reputable journal?
- Is advertising clearly marked and distinguishable from the informational/medical content? Can you tell if the information you are reading is advertisement?
- Does the app use data improperly to promote a position or a product, or is it unbiased/neutral?

Authority

- Who developed the app? What are the person's or sponsoring organization's credentials? Are they an expert in the content presented in the app? What do you know about them?
- Is the person backed by a known organization? (Be careful here... some "organizations" may simply be unreliable groups operating out of someone's basement; try to go with authoritative sources, like the National Library of Medicine.)
- Do experts review the content provided in the app and are these "experts" real authorities on the content?
- Can you easily find contact information in the app or on its download/information page? Check the about us link/seller information, usually found on the app's download page. What is the purpose of the organization? Is it trying to sell something or is it an unbiased, peer-reviewed information source?

Currency/Timeliness

- When was the app created and/or last updated?
- Does the app provide regular updates when new content or technological upgrades are required?
- Has there been more recent research on the content in the app? Many medical treatments change with the publication of new studies. What was published a year ago may be outdated now.

Usability

- Does the app work reliably and stably on the device you are using?
- Is navigation smooth and intuitive?
- Is the app efficient and effective? For example, is the type of content usable on a small screen (e.g., radiological images)? Is data entry easy?
- Is the app appropriate for the target audience (e.g., patient info apps are in plain language)?
- Does the app author provide technical support for the app?
- Is the app stand-alone (meaning you can use it without a wi-fi or Internet connection)? This is just a good thing to note so you are aware about whether the app can be used without an Internet connection.

Scope/Completeness

- Is the medical information presented in the app complete?
- Are there sources given for additional information?
- Who is the target audience - is the app targeted for use by medical professionals, patients, others?

Be sure to ask yourself:

- Why did the person/organization create the app?
- What's in it for them or are they trying to sell me something?
- Is the creator of the app an expert in the content presented in the app?
- Can I verify the information being presented to me in the app and is it accurate?
- Is there a way I can contact the app developer to provide feedback or ask a question?
- Are there any login requirements or privacy issues that I need to know about if I choose to use this app? Will my use of this app be tracked in any way?
- Is there a disclaimer that states any impact on clinical decision making, patient safety?

***TEST** before you use:*

- TEST the app before you use it in clinical care - create clinical scenarios and test.
- As you test, observe and evaluate the app according to the above ABACUS framework. Does it pass the Accurate, un-Biased, Authoritative, Current, Usable, Scope/Completeness benchmarks in multiple case scenarios?

Appendix E: HIMMS Mobile design tenets

Usability Principle	Mobile Design Tenet	Example attributes of a "usable" app
Simplicity	Let data scream	Only information, visual elements and functionality necessary to core tasks and decisions are included. Important information stands out, and function options are easy to understand. The focus is the data. The app has a clear, clean, uncluttered screen design.
Naturalness	Speak my sign. What interface?	Screen metaphors are familiar to everyday life, or commonly expected computer experiences for the clinician. Workflows match the clinical practice needs. The app is intuitive and easy to learn; minimal, if any, training is required. Iconography and symbols speak "naturally."
Consistency	Grid it	Graphic design and layout have the same look and feel, consistent placement of screen elements (e.g., gutters, columns, margins and captions). Terminology and data entry fields are used consistently. Understanding how one screen works helps you understand how other screens work.
Forgiveness and Feedback	Date your users	It's hard to lose data or destroy time-consuming effort with a wrong tap or wrong choice of buttons. If you make a mistake, the app helps you avoid it or provides a method to recover from errors gracefully (the system is "forgiving"). The app provides informative feedback about actions you are about to take or have taken. The app displays explanatory messages when processing information and describes how long it might take.
Effective Use of Language	Speak my sign Date your users	The app uses the same words that you use (while providing mapping to standardized codes and terms used for data retrieval). List or entry-form choices are clear and unambiguous. Sentences read like your native language (e.g. English).
Efficient Interactions	What interface? Get physical	The app minimizes the number of steps/gestures it takes to complete tasks; appropriate defaults are always provided. The app provides navigation options such as shortcuts for frequent and/or experienced users. Navigation methods minimize movements, such as scrolling and switching between typing and tapping. Although input methods vary from OS to OS and even within devices, the design ensures that a usable method

		is present whether the desired input is through typing or the use of gestures. Gestures include single tap, multi-tap, swipe (where the method of input never leaves the screen tracing the letters to input) and auto complete.
Effective Information Preservation	Type less + less type Color carefully	Information on screens includes sufficient white-space and large enough fonts to be read easily with high comprehension. No information is in all upper case. Colors are used to convey meaning (e.g., red to indicate medical urgency), not just for visual appeal.
Preservation of context	What interface?	The app keeps screen changes and visual interruptions to a minimum during completion of a particular task. Visual interruptions include anything (e.g., dialog boxes) that force you to shift visual focus away from the area on the screen where you are currently reading and/or working to address something else, and then re-establish focus afterward. For example, dialog boxes should be kept to a minimum and should appear in context (adjacent to or just below the control that triggered it).
Minimize cognitive overload	What interface?	Information needed for a particular task or decision is grouped together on a single screen, rather than requiring you to mentally synthesize information from multiple screens. Alerts presented are clear, concise and informative. The app performs calculations automatically.

Appendix F: Results focus group

PRELIMINARY RESEARCH

- Purpose
    - *I would further specify this to 'intended purpose' as this is relevant for legal/regulatory qualification of medical software.*
    - *What is the target group of the application?*
    - *Purpose for... patients? lay persons? healthcare professionals?*
    - *Specify for whom*
    - *Would add 'Which problem does it solve?'*
    - *Is it an new, an addition to existing therapy or subsitute?*
    - *is the targetgroup involved in the process of making the application.*
    - *wie is de eindgebryiker?*
- Currency
    - *Specify 'information' ; what type of information? Include for example What-Who-When-Why-for Whom?*
    - *wat is up-to-date? How to determine what is relevant?*
    - *life cycle technisch*
- Technology
    - *Measurement, behavior change and/or human included?*
    - *regular updates*
- Policy/Legal
    - *In Meulendijk et al (2014) werd Certification als criterium gevonden. Dat valt hieronder? Certification is volgens mij veel specifieker en slechts deels overlappend, maar wel vaak verplicht.*
    - *Beter: Certifiability*
    - *Medical Devices Directive and MEDDEV 2.1/6 are important as far as EU is concerned (CE marking). Other legal aspects: data protection, e-privacy, consumer rights. Note: divergence in EU Member States. For example France has specific legislation in force regarding the hosting of health data.*
- Ethics
    - *and/ or professionals. Maak generiek naar end users*
    - *Why only patients? Add other stakeholders, like medical professionals but also society at large*
    - *Ethics is not only about usage, it involves issues like (conditional solidarity due to the use of mHealth, autonomy, and privacy issues*

- Safety
    - *include and specify possible harms*
    - *ben benieuwd naar de operationalisatie van dit item*
    - *Non-inferiority as the minimum threshold for large scale implementation --> especially in case of replacement of usual care.*
- Feasibility
    - *Dit heeft meerdere aspecten gebruik/verkoop data, structurele financiering waardoor de app geupdate blijft en te gebruiken blijft, of de patient wellicht een vergoeding krijgt uit zorgbudgetten voor gebruik van de app, etc.*
    - *Ook algemeen business model*
    - *Gezondheidseconomische waarde*
    - *incl business case*
    - *Which problem does the app/mHealth concept solve?*
    - *De zog. CORETEST Feasibility analysis methode bijv.*
    - *Target user: wie is de beoogde eindgebruiker*
- Theory identification / theoretical framework / underlying models
- **What's the purpose of the TA? In what context will the TA be used? Whom are involved. These are questions one should ask oneself prior to the application of the TA.

EVIDENCE

- Accuracy
    - *Reliability, to what extent does the app rely on information given by others (i.e. user, professional, sensors, etc) and to what extent is this information reliable?*
- Evidence
    - *Order by level of evidence*
    - *there is more to testing than just clinical scenarios, depending on the user group and the content of the application*
    - *Only clinical evidence, or do you include other types of evidence too?*
    - *Where does the evidence come from?*
    - *ow is the application tested? specify desgin, sample (n), population- outcome.*
    - *Evidence of therapy or evidence of system*
    - *Denk aan ontwikkel loop. Als je iets aanpast in de app moet je wel kijken dat de klinische validatie nog actueel is.*
- Transferability
    - *Het moet heel duidelijk zijn in welke context en voor welke groep gebruikers deze applicatie getest is*

    - *Don't know what Transferability in this context means but may be include interoperability here? Eg working with standards such as HL7, SNOMED etc.*
    - *Transferability = transfer of data*
- Beste practices of literatuur waarop mechanism of action is gebaseerd
    - *Bij 3: ook andere aandoeningen of gebruikersgroepen*
    - *Causal mechanism - sequence of effects known*

USABILITY

- Usability
    - *Both paper prototyping / and field/lab usability with end-users*
    - *Adherence to standards for dyslexia etc. special needs groups*
    - *How well does it fit in life 'rituals'*
    - *Uses Human Interface Guidelines (Apple/Google)?*
    - *targetgroup involved in making the application?*
    - *adherence and compliance*
- Changing need user
    - *Wanneer je de app aanpast moet je dan weer terug naar evidence?*
- Readability
    - *Is dat nodig? hangt van de eindgebruiker af*
- Behavior change maintenance
- Persuasive design
- Interoperability
- Long term engagement
    - *Is de applicatie in staat patienten langere tijd te engagen (voorbeeld fitness apps 3 mnd gebruik)*
- Satisfaction

SECURITY

- Data security
    - *voldoen servers aan de eisen*
    - *What other applications might run simultaneously on the device and what type of interaction might you expect between the app under assessment and other apps?*
    - *Where are servers located (Patriot act)*
    - *Is the developer allowed (legal) to work with patient data (NEN7510/7512 etc)*
- Participant privacy
    - *Is it possible for users to remove their own data?*

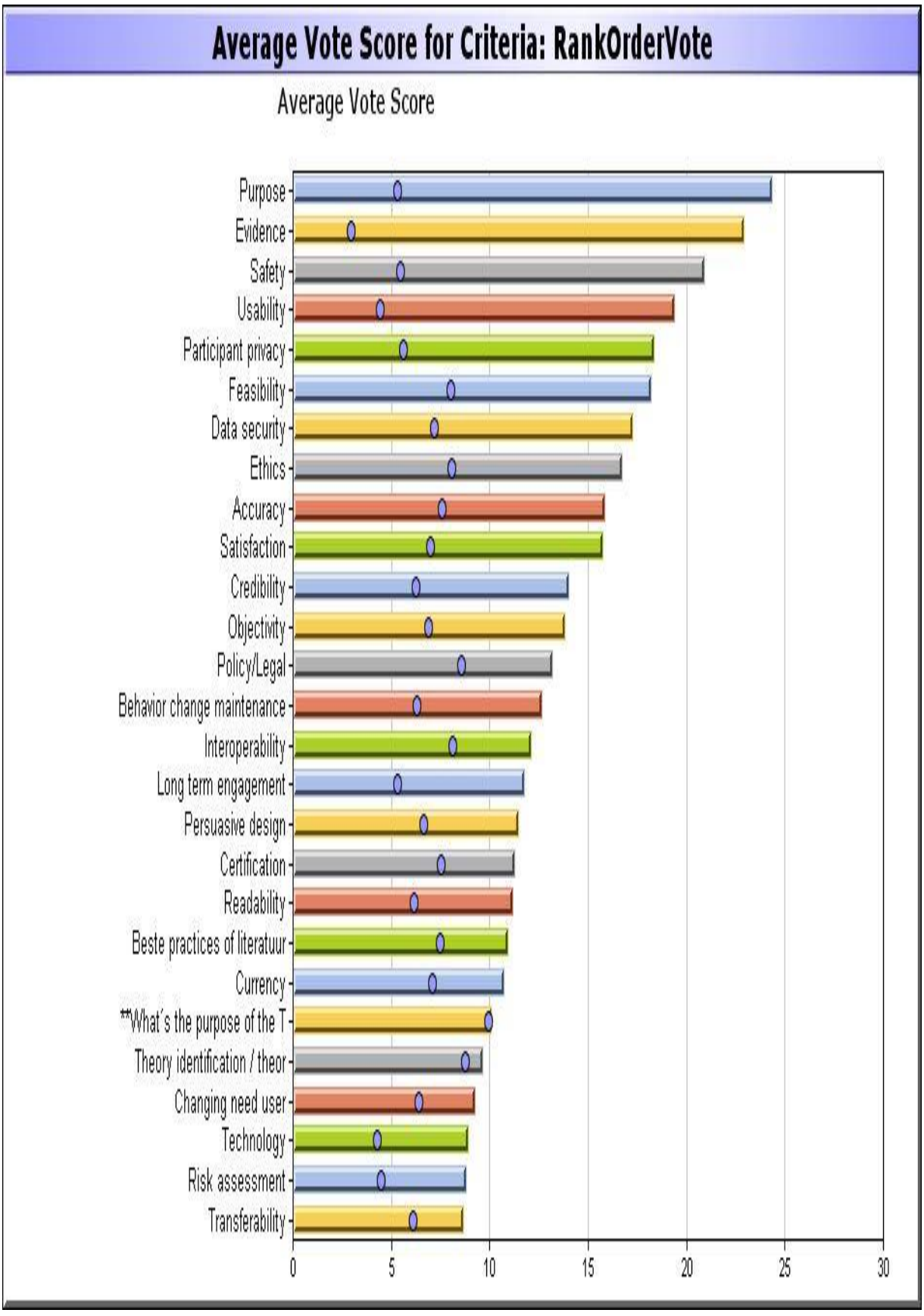
- *Or to download their data (Blue Button)?*
- *transparancy who can enter/ see the data*
- *How is anonymization or pseudonimization taken care of?*
- *proportionality; for what period do you retain the information and what information do you really need to store/use?*
- *is it possible to link it with the medical records*

- Risk assessment

TRANSPARENCY

- Objectivity
  - *Transparent in: development process, algorithms used, degree to which end-user and experts are used.*
  - *is the result of the scoring on the checklist published somewhere?*
  - *informed consent*
- Credibility
- Certification

De lijst met criteria - rangordenen



De lijst met criteria - rangordenen Criteria: RankOrderVote

					Vote Distribution																											
#	Ballot Items	Score	Avg. Rank	STD	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
1.	Purpose	24.36	3.64	5.32	6	3	-	-	-	-	-	-	-	-	1	-	-	-	-	-	1	-	-	-	-	-	-	-	-	-	-
11.	Evidence	22.91	5.09	2.95	-	3	1	-	4	-	1	-	1	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6.	Safety	20.91	7.09	5.47	-	2	1	2	1	1	1	-	-	-	-	1	-	-	1	-	-	1	-	-	-	-	-	-	-	-	-
14.	Usability	19.36	8.64	4.41	-	1	-	1	1	-	1	3	1	-	-	-	-	2	-	1	-	-	-	-	-	-	-	-	-	-	-
23.	Participant privacy	18.36	9.64	5.59	-	-	-	2	1	-	3	-	-	-	-	3	1	-	-	-	-	-	-	-	-	-	1	-	-	-	-
7.	Feasibility	18.18	9.82	8.02	2	1	-	1	-	1	-	1	-	-	1	-	-	1	-	-	1	-	-	1	-	-	-	1	-	-	-
22.	Data security	17.27	10.73	7.18	-	-	2	-	-	3	1	-	-	-	-	-	2	-	-	1	-	-	-	-	-	1	1	-	-	-	-
5.	Ethics	16.73	11.27	8.08	-	-	3	-	-	-	-	3	-	1	-	-	1	-	-	-	-	-	1	-	-	-	-	1	1	-	-
10.	Accuracy	15.82	12.18	7.57	-	-	1	2	-	1	-	-	-	1	-	-	1	-	1	1	1	-	1	-	-	-	-	-	-	-	1
21.	Satisfaction	15.73	12.27	6.99	-	-	-	1	1	1	-	-	3	-	-	-	1	-	-	1	-	1	-	1	-	-	-	-	-	1	-
26.	Credibility	14.00	14.00	6.24	-	-	-	-	-	1	2	-	-	-	1	1	-	1	1	1	-	-	1	-	-	1	-	-	1	-	-
25.	Objectivity	13.82	14.18	6.88	1	-	-	-	-	1	-	-	-	-	1	1	1	-	2	-	1	1	-	-	-	-	-	2	-	-	-
4.	Policy/Legal	13.18	14.82	8.57	-	-	1	1	-	1	-	1	-	-	1	-	-	-	-	-	1	-	-	1	-	1	1	1	1	-	-
17.	Behavior change maintenance	12.64	15.36	6.30	-	-	-	-	-	-	1	1	-	-	-	3	1	-	-	-	1	1	1	-	-	-	-	1	-	-	1
19.	Interoperability	12.09	15.91	8.13	-	-	1	-	1	-	-	1	-	-	1	-	-	-	-	-	2	-	1	-	2	-	-	-	-	1	1
20.	Long term engagement	11.73	16.27	5.29	-	-	-	-	-	-	-	-	1	1	-	1	-	2	1	-	1	-	1	-	1	1	-	-	-	1	-
18.	Persuasive design	11.45	16.55	6.62	-	-	-	-	-	-	-	-	1	1	3	-	-	-	-	1	-	1	-	1	-	-	-	-	2	1	-
27.	Certification	11.27	16.73	7.51	-	-	-	-	1	-	1	-	-	1	1	-	-	-	1	-	-	-	1	1	-	1	-	1	1	1	-

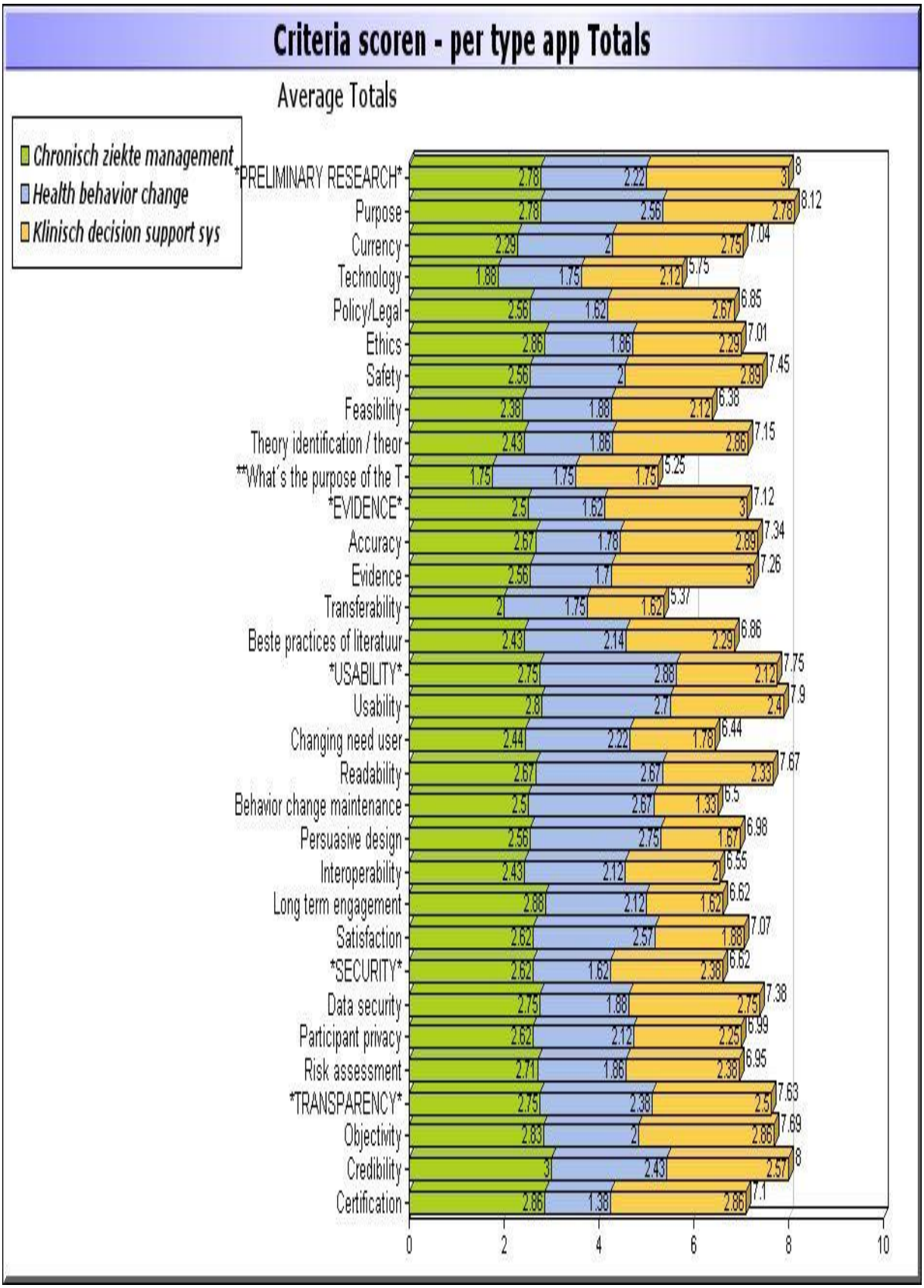
16.	Readability	11.18	16.82	6.15	-	-	-	-	-	-	-	-	1	2	-	-	1	1	1	-	-	-	-	-	2	1	-	1	-	1	-
13.	Beste practices of literatuur waarop mechanism of action is gebaseerd	10.91	17.09	7.48	-	-	1	-	-	-	-	-	1	1	-	-	-	1	1	-	-	-	1	-	1	-	2	-	1	1	-
2.	Currency	10.73	17.27	7.10	-	-	-	-	-	1	-	1	-	1	-	-	-	-	1	1	-	1	-	2	-	-	-	1	-	1	1
9.	**What's the purpose of the TA? In what context will the TA be used? Whom are involved. These are questions one should ask oneself prior to the application of the TA.	10.09	17.91	9.92	2	-	-	-	-	-	-	-	-	-	-	-	1	-	-	2	1	-	-	-	-	-	-	-	1	-	4
8.	Theory identification / theoretical framework / underlying models	9.64	18.36	8.77	-	1	-	-	-	-	-	-	1	1	-	-	1	-	-	1	-	-	-	-	-	1	-	-	2	1	2
15.	Changing need user	9.27	18.73	6.37	-	-	-	-	1	-	-	-	-	1	-	-	-	-	-	1	-	1	1	-	2	-	3	-	-	-	1
3.	Technology	8.91	19.09	4.28	-	-	-	-	-	-	-	-	-	-	-	1	-	2	-	-	-	1	1	2	1	-	1	1	1	-	-
24.	Risk assessment	8.82	19.18	4.49	-	-	-	-	-	-	-	-	1	-	-	-	-	-	1	-	1	1	1	1	1	3	-	-	-	1	-
12.	Transferability	8.64	19.36	6.09	-	-	-	1	-	-	-	-	-	-	-	-	-	-	1	-	-	-	2	-	1	1	1	2	1	-	1	-

Voting Details

Criteria Statistic: Mean. Votes Cast: 11, Abstained:

Criteria scoren - per type app

1. Criteria scoren - per type app Totals



Criteria scoren - per type app Totals

		Criteria:					
		Chronisch ziekte management	Health behavior change	Klinisch decision support system			
	Voting Method:	HighMedLow	HighMedLow	HighMedLow			
	Weight:	1.00	1.00	1.00			
#	Ballot Items				Weighted Total	Total	Avg. Score
1.	*PRELIMINARY RESEARCH*	2.78	2.22	3.00	8.00	8.00	2.67
1.1.	Purpose	2.78	2.56	2.78	8.12	8.12	2.70
1.2.	Currency	2.29	2.00	2.75	7.04	7.04	2.36
1.3.	Technology	1.88	1.75	2.12	5.75	5.75	1.92
1.4.	Policy/Legal	2.56	1.62	2.67	6.85	6.85	2.31
1.5.	Ethics	2.86	1.86	2.29	7.01	7.01	2.33
1.6.	Safety	2.56	2.00	2.89	7.45	7.45	2.48
1.7.	Feasibility	2.38	1.88	2.12	6.38	6.38	2.12
1.8.	Theory identification / theoretical framework / underlying models	2.43	1.86	2.86	7.15	7.15	2.38
1.9.	**What's the purpose of the TA? In what context will the TA be used? Whom are involved. These are questions one should ask oneself prior to the application of the TA.	1.75	1.75	1.75	5.25	5.25	1.75
2.	*EVIDENCE*	2.50	1.62	3.00	7.12	7.12	2.38
2.1.	Accuracy	2.67	1.78	2.89	7.34	7.34	2.44
2.2.	Evidence	2.56	1.70	3.00	7.26	7.26	2.41

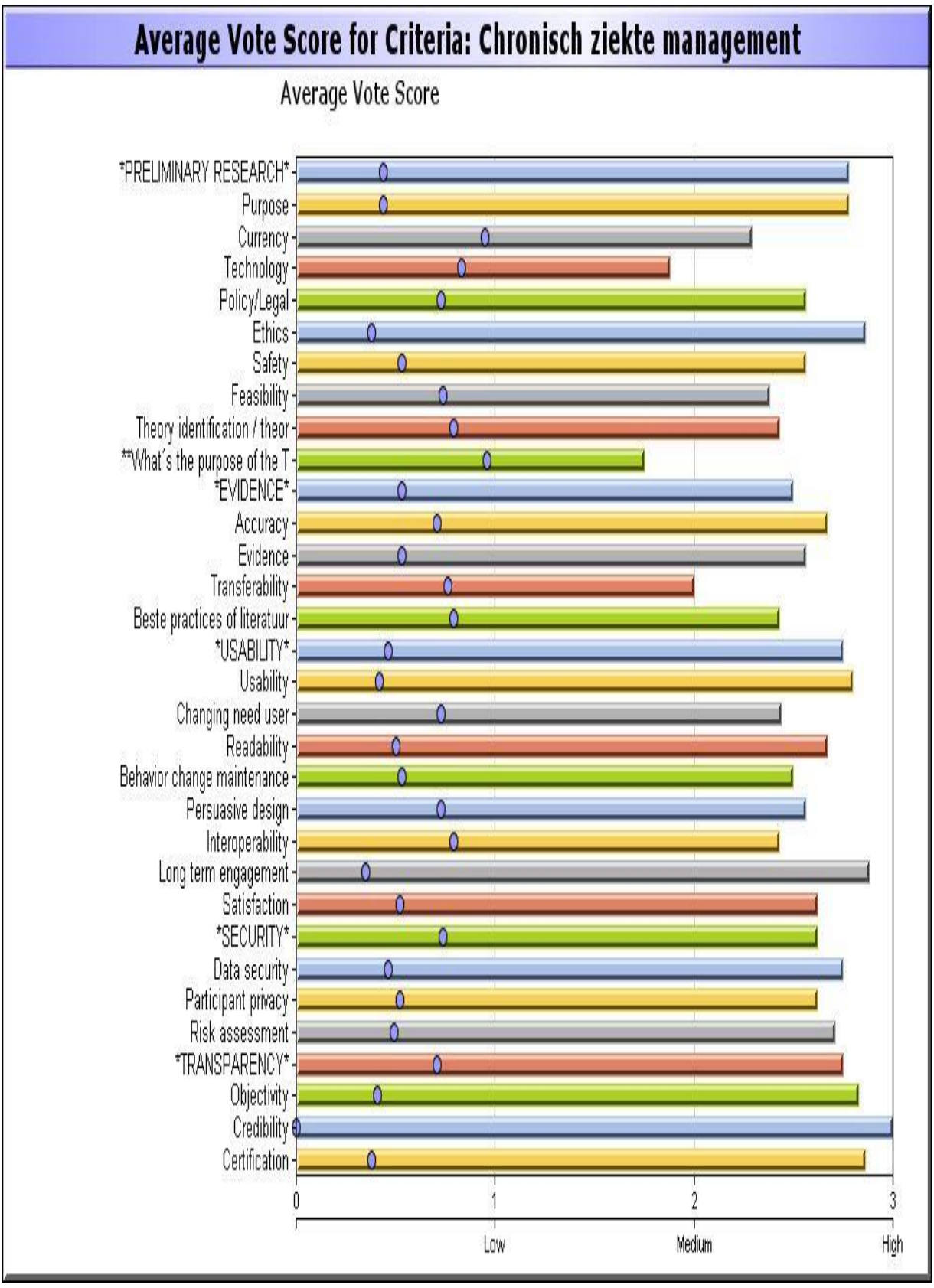
2.3.	Transferability	2.00	1.75	1.62	5.37	5.37	1.79
2.4.	Beste practices of literatuur waarop mechanism of action is gebaseerd	2.43	2.14	2.29	6.86	6.86	2.29
3.	*USABILITY*	2.75	2.88	2.12	7.75	7.75	2.58
3.1.	Usability	2.80	2.70	2.40	7.90	7.90	2.63
3.2.	Changing need user	2.44	2.22	1.78	6.44	6.44	2.15
3.3.	Readability	2.67	2.67	2.33	7.67	7.67	2.56
3.4.	Behavior change maintenance	2.50	2.67	1.33	6.50	6.50	2.15
3.5.	Persuasive design	2.56	2.75	1.67	6.98	6.98	2.31
3.6.	Interoperability	2.43	2.12	2.00	6.55	6.55	2.18
3.7.	Long term engagement	2.88	2.12	1.62	6.62	6.62	2.21
3.8.	Satisfaction	2.62	2.57	1.88	7.07	7.07	2.35
4.	*SECURITY*	2.62	1.62	2.38	6.62	6.62	2.21
4.1.	Data security	2.75	1.88	2.75	7.38	7.38	2.46
4.2.	Participant privacy	2.62	2.12	2.25	6.99	6.99	2.33
4.3.	Risk assessment	2.71	1.86	2.38	6.95	6.95	2.32
5.	*TRANSPARENCY*	2.75	2.38	2.50	7.63	7.63	2.54
5.1.	Objectivity	2.83	2.00	2.86	7.69	7.69	2.55
5.2.	Credibility	3.00	2.43	2.57	8.00	8.00	2.65
5.3.	Certification	2.86	1.38	2.86	7.10	7.10	2.32

Voting Details

Criteria Statistic: Mean. Votes Cast: 10, Abstained: 0

2. Criteria scoren - per type app Criteria: Chronisch ziekte management

Vote Method: HighMedLow



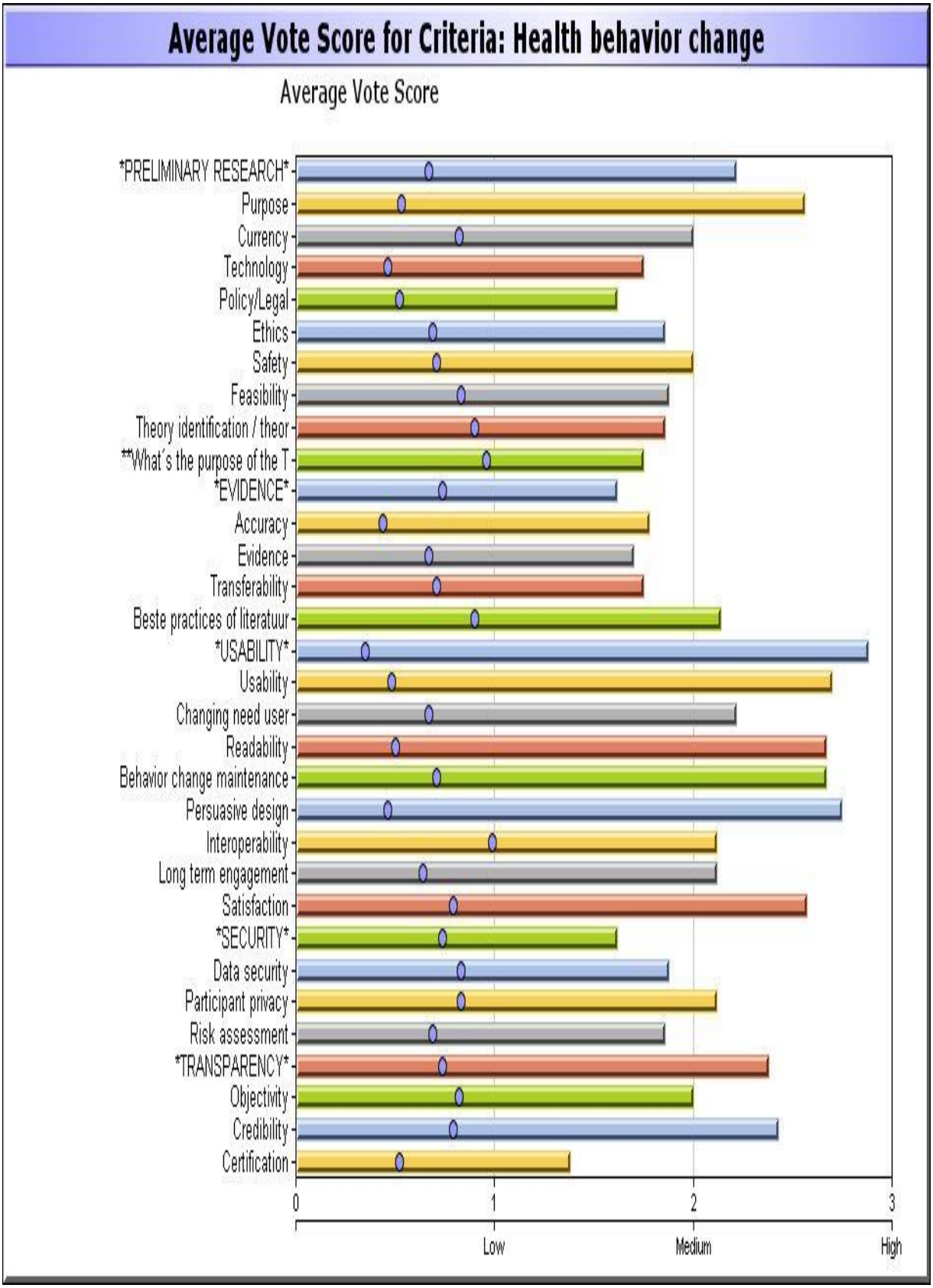
Criteria scoren - per type app Criteria: Chronisch ziekte management

		Vote Distribution						
#	Ballot Items	L	M	H	Avg. Score	Total	STD	Votes
1.	*PRELIMINARY RESEARCH*	-	2	7	2.78	25.00	0.44	9
1.1.	Purpose	-	2	7	2.78	25.00	0.44	9
1.2.	Currency	2	1	4	2.29	16.00	0.95	7
1.3.	Technology	3	3	2	1.88	15.00	0.83	8
1.4.	Policy/Legal	1	2	6	2.56	23.00	0.73	9
1.5.	Ethics	-	1	6	2.86	20.00	0.38	7
1.6.	Safety	-	4	5	2.56	23.00	0.53	9
1.7.	Feasibility	1	3	4	2.38	19.00	0.74	8
1.8.	Theory identification / theoretical framework / underlying models	1	2	4	2.43	17.00	0.79	7
1.9.	**What's the purpose of the TA? In what context will the TA be used? Whom are involved. These are questions one should ask oneself prior to the application of the TA.	2	1	1	1.75	7.00	0.96	4
2.	*EVIDENCE*	-	4	4	2.50	20.00	0.53	8
2.1.	Accuracy	1	1	7	2.67	24.00	0.71	9
2.2.	Evidence	-	4	5	2.56	23.00	0.53	9
2.3.	Transferability	2	4	2	2.00	16.00	0.76	8
2.4.	Beste practices of literatuur waarop mechanism of action is gebaseerd	1	2	4	2.43	17.00	0.79	7
3.	*USABILITY*	-	2	6	2.75	22.00	0.46	8
3.1.	Usability	-	2	8	2.80	28.00	0.42	10

3.2.	Changing need user	1	3	5	2.44	22.00	0.73	9
3.3.	Readability	-	3	6	2.67	24.00	0.50	9
3.4.	Behavior change maintenance	-	4	4	2.50	20.00	0.53	8
3.5.	Persuasive design	1	2	6	2.56	23.00	0.73	9
3.6.	Interoperability	1	2	4	2.43	17.00	0.79	7
3.7.	Long term engagement	-	1	7	2.88	23.00	0.35	8
3.8.	Satisfaction	-	3	5	2.62	21.00	0.52	8
4.	*SECURITY*	1	1	6	2.62	21.00	0.74	8
4.1.	Data security	-	2	6	2.75	22.00	0.46	8
4.2.	Participant privacy	-	3	5	2.62	21.00	0.52	8
4.3.	Risk assessment	-	2	5	2.71	19.00	0.49	7
5.	*TRANSPARENCY*	1	-	7	2.75	22.00	0.71	8
5.1.	Objectivity	-	1	5	2.83	17.00	0.41	6
5.2.	Credibility	-	-	6	3.00	18.00	0.00	6
5.3.	Certification	-	1	6	2.86	20.00	0.38	7

3. Criteria scoren - per type app Criteria: Health behavior change

Vote Method: HighMedLow



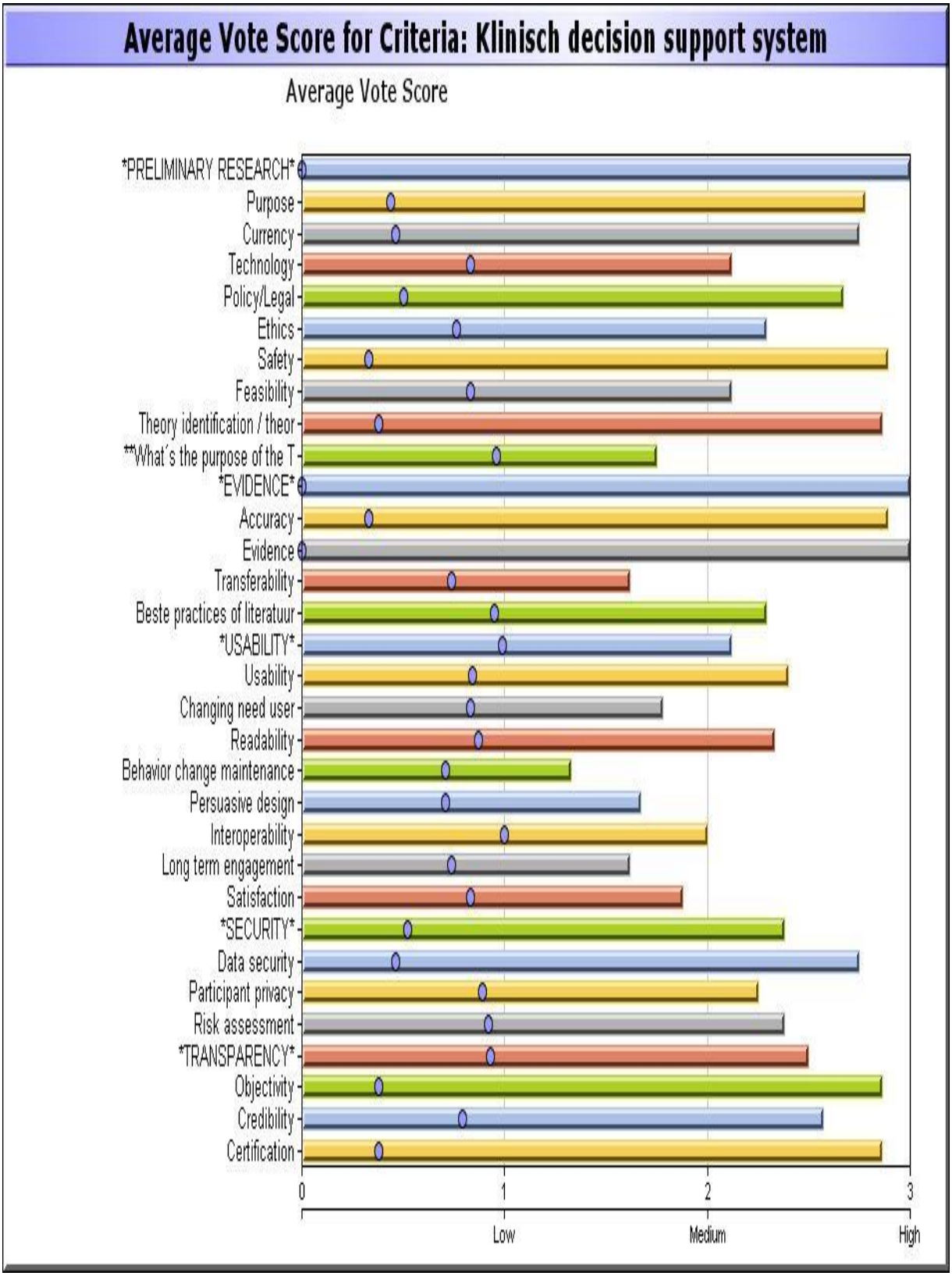
Criteria scoren - per type app Criteria: Health behavior change

		Vote Distribution						
#	Ballot Items	L	M	H	Avg. Score	Total	STD	Votes
1.	*PRELIMINARY RESEARCH*	1	5	3	2.22	20.00	0.67	9
1.1.	Purpose	-	4	5	2.56	23.00	0.53	9
1.2.	Currency	2	3	2	2.00	14.00	0.82	7
1.3.	Technology	2	6	-	1.75	14.00	0.46	8
1.4.	Policy/Legal	3	5	-	1.62	13.00	0.52	8
1.5.	Ethics	2	4	1	1.86	13.00	0.69	7
1.6.	Safety	2	5	2	2.00	18.00	0.71	9
1.7.	Feasibility	3	3	2	1.88	15.00	0.83	8
1.8.	Theory identification / theoretical framework / underlying models	3	2	2	1.86	13.00	0.90	7
1.9.	**What's the purpose of the TA? In what context will the TA be used? Whom are involved. These are questions one should ask oneself prior to the application of the TA.	2	1	1	1.75	7.00	0.96	4
2.	*EVIDENCE*	4	3	1	1.62	13.00	0.74	8
2.1.	Accuracy	2	7	-	1.78	16.00	0.44	9
2.2.	Evidence	4	5	1	1.70	17.00	0.67	10
2.3.	Transferability	3	4	1	1.75	14.00	0.71	8
2.4.	Beste practices of literatuur waarop mechanism of action is gebaseerd	2	2	3	2.14	15.00	0.90	7
3.	*USABILITY*	-	1	7	2.88	23.00	0.35	8
3.1.	Usability	-	3	7	2.70	27.00	0.48	10

3.2.	Changing need user	1	5	3	2.22	20.00	0.67	9
3.3.	Readability	-	3	6	2.67	24.00	0.50	9
3.4.	Behavior change maintenance	1	1	7	2.67	24.00	0.71	9
3.5.	Persuasive design	-	2	6	2.75	22.00	0.46	8
3.6.	Interoperability	3	1	4	2.12	17.00	0.99	8
3.7.	Long term engagement	1	5	2	2.12	17.00	0.64	8
3.8.	Satisfaction	1	1	5	2.57	18.00	0.79	7
4.	*SECURITY*	4	3	1	1.62	13.00	0.74	8
4.1.	Data security	3	3	2	1.88	15.00	0.83	8
4.2.	Participant privacy	2	3	3	2.12	17.00	0.83	8
4.3.	Risk assessment	2	4	1	1.86	13.00	0.69	7
5.	*TRANSPARENCY*	1	3	4	2.38	19.00	0.74	8
5.1.	Objectivity	2	3	2	2.00	14.00	0.82	7
5.2.	Credibility	1	2	4	2.43	17.00	0.79	7
5.3.	Certification	5	3	-	1.38	11.00	0.52	8

4. Criteria scoren - per type app Criteria: Klinisch decision support system

Vote Method: HighMedLow



Criteria scoren - per type app Criteria: Klinisch decision support system

		Vote Distribution						
#	Ballot Items	L	M	H	Avg. Score	Total	STD	Votes
1.	*PRELIMINARY RESEARCH*	-	-	9	3.00	27.00	0.00	9
1.1.	Purpose	-	2	7	2.78	25.00	0.44	9
1.2.	Currency	-	2	6	2.75	22.00	0.46	8
1.3.	Technology	2	3	3	2.12	17.00	0.83	8
1.4.	Policy/Legal	-	3	6	2.67	24.00	0.50	9
1.5.	Ethics	1	3	3	2.29	16.00	0.76	7
1.6.	Safety	-	1	8	2.89	26.00	0.33	9
1.7.	Feasibility	2	3	3	2.12	17.00	0.83	8
1.8.	Theory identification / theoretical framework / underlying models	-	1	6	2.86	20.00	0.38	7
1.9.	**What's the purpose of the TA? In what context will the TA be used? Whom are involved. These are questions one should ask oneself prior to the application of the TA.	2	1	1	1.75	7.00	0.96	4
2.	*EVIDENCE*	-	-	8	3.00	24.00	0.00	8
2.1.	Accuracy	-	1	8	2.89	26.00	0.33	9
2.2.	Evidence	-	-	10	3.00	30.00	0.00	10
2.3.	Transferability	4	3	1	1.62	13.00	0.74	8
2.4.	Beste practices of literatuur waarop mechanism of action is gebaseerd	2	1	4	2.29	16.00	0.95	7
3.	*USABILITY*	3	1	4	2.12	17.00	0.99	8
3.1.	Usability	2	2	6	2.40	24.00	0.84	10

3.2.	Changing need user	4	3	2	1.78	16.00	0.83	9
3.3.	Readability	2	2	5	2.33	21.00	0.87	9
3.4.	Behavior change maintenance	7	1	1	1.33	12.00	0.71	9
3.5.	Persuasive design	4	4	1	1.67	15.00	0.71	9
3.6.	Interoperability	3	1	3	2.00	14.00	1.00	7
3.7.	Long term engagement	4	3	1	1.62	13.00	0.74	8
3.8.	Satisfaction	3	3	2	1.88	15.00	0.83	8
4.	*SECURITY*	-	5	3	2.38	19.00	0.52	8
4.1.	Data security	-	2	6	2.75	22.00	0.46	8
4.2.	Participant privacy	2	2	4	2.25	18.00	0.89	8
4.3.	Risk assessment	2	1	5	2.38	19.00	0.92	8
5.	*TRANSPARENCY*	2	-	6	2.50	20.00	0.93	8
5.1.	Objectivity	-	1	6	2.86	20.00	0.38	7
5.2.	Credibility	1	1	5	2.57	18.00	0.79	7
5.3.	Certification	-	1	6	2.86	20.00	0.38	7

Appendix G: List of criteria focus group

The following table with criteria is based on the feedback from the focus group.

#		---Preliminary research---
1	Purpose	What is the target group of the application? What is the intended purpose of the application for: • Patients; • Doctors; • Insurance companies; • Other stakeholders. What problem is solved by the application? Was the target group involved in the development of the application?
2	Currency	What is the life cycle of the application? • When is the application made? • How frequently is the application updated? • What was the last time the application was updated? • Are the developers actively updating and improving the application? Is the information used in the application up-to-date? Is the information relevant?
4	Technology	What for technology is used and is this technology: • Effective; • Appropriate; • Relevant; Is the technology being updated? E.g. is the application compatible with the newest phones?
5	Policy/Legal	Are there legal obligations to be met and are there any legal barriers? Is the application in line with the Medical Devices Directive and MEDDEV 2.1/6?
6	Ethics	Is it ethical to use the application for: • Patients; • Medical professionals; • Society at large; • Other stakeholders?
7	Safety	Is it safe for the patients and/or the staff to use the application? What are possible harms of the applications? Non-inferiority should be the minimum threshold for large scale implementation, especially if the mHealth solution replaces the usual care.
8	Feasibility	Is a feasibility study performed? • If so, are the following points included? ○ Usage/sales of the application; ○ Data;

		○ Structural funding that will keep the application updated. ○ Do insurance companies pay for the application? ○ What is the business model ○ What is the health economic value? ○ What problem is solved by the application? ○ CORETEST feasibility analysis method ▪ Conceptual feasibility ▪ Organizational feasibility ▪ Economic feasibility ▪ Technological feasibility ▪ Societal feasibility ○ For whom is the application developed?
8	Theory	Is the application based on theory? • What theory is used; • What frameworks are used; • Which models are used?
		---Evidence---
9	Accuracy	Is the information in the application correct? To what extend does the application rely on information other than theory? What or who is the input for this information? 1. Users; 2. Professionals; 3. Sensors; 4. Other input. Which information is this?
10	Evidence	How is the application tested? Specify: • Design of the test(s); • Sample(s) used; • Total population; • Outcome of the test(s). Where clinical scenarios involved in the testing? Are the reliability and the validity tested? How is the efficacy and the effectiveness of the application tested? What, if any, are there negative effects of the application? How are the efficacy and effectiveness compared with current methods? Is the test data available? Can adjustments/updates be introduced without any loss of evidence? What is the added value of the application?
11	Transferability	Can results of testing be transferred to other scenarios?
		---Usability---
12	Usability	Is the usability of the application tested and what are the results? Was the target group involved in testing the application? Is both the paper prototype and field/lab usability test done with end-

		users? What is the adherence of the application to existing standards? How well does the application fit in the regular life of the user? Did the developer use the human interface guidelines from the platform the application is published on? E.g. Google, Apple, etc. How is the adherence and compliance of the application?
13	Changing need user Long term engagement	Can the application be used for a longer period of time? Does it adapt to the changing need of the user?
14	Readability	Is the application understandable for the target group? E.g. is an application for patients understandable and without jargon?
15	Behavior change maintenance	Is one of the goals of the application to change behavior of the user? If so, how is this change maintained?
16	Persuasive design	Are there persuasive design principles included in the application?
17	Interoperability	What is the interoperability of the application with other systems and how is this measured? Are there standards used?
18	Satisfaction	What is the satisfaction of the user and how is this measured?
		---Security---
19	Data security	Is the application secure? Is the data encrypted before it is send to the server? What are the security requirements of: • The application; • The phone; • The servers; • Other devices? Where are the servers located? Are there other applications running simultaneously on the device and what type of interaction might you expect between the application that is under assessment and other applications? Is the developer allowed to work with the patient data?
20	Participant privacy	What happens with the collected data? Who is collecting data? Who is able to see the data? Does the developer have access to the data? Is the data being sold to 3th parties? Can users remove their data? Can users download their data? How does the developer handle anonymization/pseudonymization? How long is the collected information stored? What information is necessary to store and what information is not necessary to store? Is it possible to link the collected data with medical records?

21	Risk assessment	Did the developer carry out a risk assessment on security of the application?
		---Transparency---

22	Objectivity	Does the application provide authorship information?
		• Including detailed information about:
		  ○ Affiliations
		  ○ Credentials
		  ○ Any medical professional involvement in content preparation?
		Is the information balanced and non-biased?
		Is there a list of references and sources included?
		Does the application provide information about any app sponsorship (including advertisements) or other commercial funding arrangements, and any potential conflicts of interests?
		Is the application transparent in:
		• Development process
		• Algorithms used
		• Degree to which end-user and experts are used
		Who made the application?

23	Credibility	What is the credibility of the sources involved in developing the application?
24	Certification	Which certificates does the application have?

Appendix H: Interview questions

General information

1. Name interviewee
2. Organization
3. Function interviewee
4. Experience
5. Daily routines

1. Wat is u ervaring met medische applicaties (mhealth apps?)
- Testen jullie op het moment medische apps voor mobiele telefoons? Zo ja, hoe?
- Belangrijkste toetsing punten bij het testen van mhealth applicaties?
- Wie is er verantwoordelijk voor dat de applicatie van goede kwaliteit is?
    - Is dat de ontwikkelaar of de persoon/organisatie die de applicatie gebruikt?
- Bij wie moet de bewijslast worden neergelegd?

1. Hoe wordt er bepaald wat een goede mhealth applicatie is?

2. Zijn er al voldoende wetten?
    a. Of houden wetten de innovatie van mhealth tegen?
3. Wat moet er veranderen qua wetgeving in de toekomst?
    a. Wat mist er m.b.t. mhealth?
4. Mogen ziekenhuizen op het moment apps voorschrijven aan patiënten?
5. Wat vinden jullie ervan dat er veel medische apps voor consumenten beschikbaar zijn?
6. Moeten er gradaties zijn in het beoordelen van applicaties? Bepaalde type apps die strenger moeten worden beoordeeld?

1. Wat voor rol denken jullie dat mhealth of ehealth in de toekomst krijgt in de gezondheidszorg?
2. Denken jullie dat een framework voor het testen van mhealth applicaties nut heeft?
3. Waar moeten een framework voor mhealth aan voldoen?
4. Zijn jullie zelf bezig met een dergelijk framework?

Appendix I: Comparison Synappz

The criteria could not be mapped one to one with the different actions/processes in the development phases of Synappz. This is due to the development method of Synappz, which is more than a list of criteria. Instead, the main focus points of each phase are identified. Some of the criteria can be matched with these focus points; an overview of this comparison is in Table 34.

Table 34 Development stages Synappz

Phase	Main focus points in the development phase	Criteria that could be matched
Strategy	Context	Purpose
	Target audience	Purpose
	User story	Purpose/Usability
	Objectives	Feasibility/Purpose/Transparency
	Behavior change, if applicable	Usability
	Validation, if applicable	Theory/Evidence
Scope	Functionalities	Usability
	Technical requirements	Technology
	Privacy and security	Data security/participant privacy
	CE certification, if applicable	Legal
Structure	Flowchart	Usability
	Information architecture and data management	Security
	Validation and evidence	Evidence/Accuracy
	Implementation	Feasibility
	Test protocols	Usability
Skeleton and surface	Wireframing	Usability
	Graphical elements	Usability
	Approval from customer for development	-
	CE certification analysis	Legal

Table 35 below includes a checklist of the criteria and whether they are included in one of the development phases. All of the criteria except for 'transferability of evidence'; are included. The reason that this is excluded is probably because it is related to 'evidence'. Meaning that it will be included if the client has the time and budget for it.

Table 35 Checklist criteria with development Synappz

Criteria	Strategy	Scope	Structure	Skeleton	Surface
Purpose	X				
Currency	X				
Technology	X				
Policy/legal		X			
Ethics	X				
Safety		X			
Feasibility	X				
Theory			X		
Accuracy			X		
Evidence			X		
Transferability of evidence					
Usability			X	X	X
Changing need user			X	X	X
Long term engagement		X	X	X	X
Readability	X			X	X
Behavior change maintenance	X		X	X	
Persuasive design			X	X	X
Interoperability		X	X		
Satisfaction					X
Data security	X	X	X		
Participant privacy		X	X		
Risk assessment	X	X			
Objectivity	X				
Credibility	X				
Certification	X	X			

Appendix J: Elaboration of MED DEV 2.1/6

The medical devices act is based on the European Medical Devices Directive. There are three important definitions in the Medical Devices Directive 2.1/6 that are adopted in the Dutch medical devices act. The following definitions are important for this research:

Medical device

"'Medical device' means any instrument, apparatus, appliance, software, material or other article, whether used alone or in combination, including the software intended by its manufacturer to be used specifically for diagnostic and/or therapeutic purposes and necessary for its proper application, intended by the manufacturer to be used for human beings for the purpose of:

- Diagnosis, prevention, monitoring, treatment or alleviation of disease,
- Diagnosis, monitoring, treatment, alleviation of or compensation for an injury or handicap,
- Investigation, replacement or modification of the anatomy or of a physiological process,
- Control of conception, and which does not achieve its principal intended action in or on the human body by pharmacological, immunological or metabolic means, but which may be assisted in its function by such means."

Stand alone software

"'Stand alone software' means software which is not incorporated in a medical device at the time of its placing on the market or its making available." **MHealth applications fall under this definition.**

Active Medical device

"Any medical device operation of which depends on a source of electrical energy or any source of power other than that directly generated by the human body or gravity and which acts by converting this energy. Stand alone software is considered to be an active medical device."

However, not all mHealth applications are medical devices. Important is that only the intended purpose as described by the manufacturer of the product is relevant for the qualification and classification of any device. The way the device is called does not have any effect on this process nor does the risk related to a malfunction of the software. Figure 26 depicts a decision diagram from MED DEV 2.1/6 that can be used to determine if an mHealth application is qualified as a medical device.

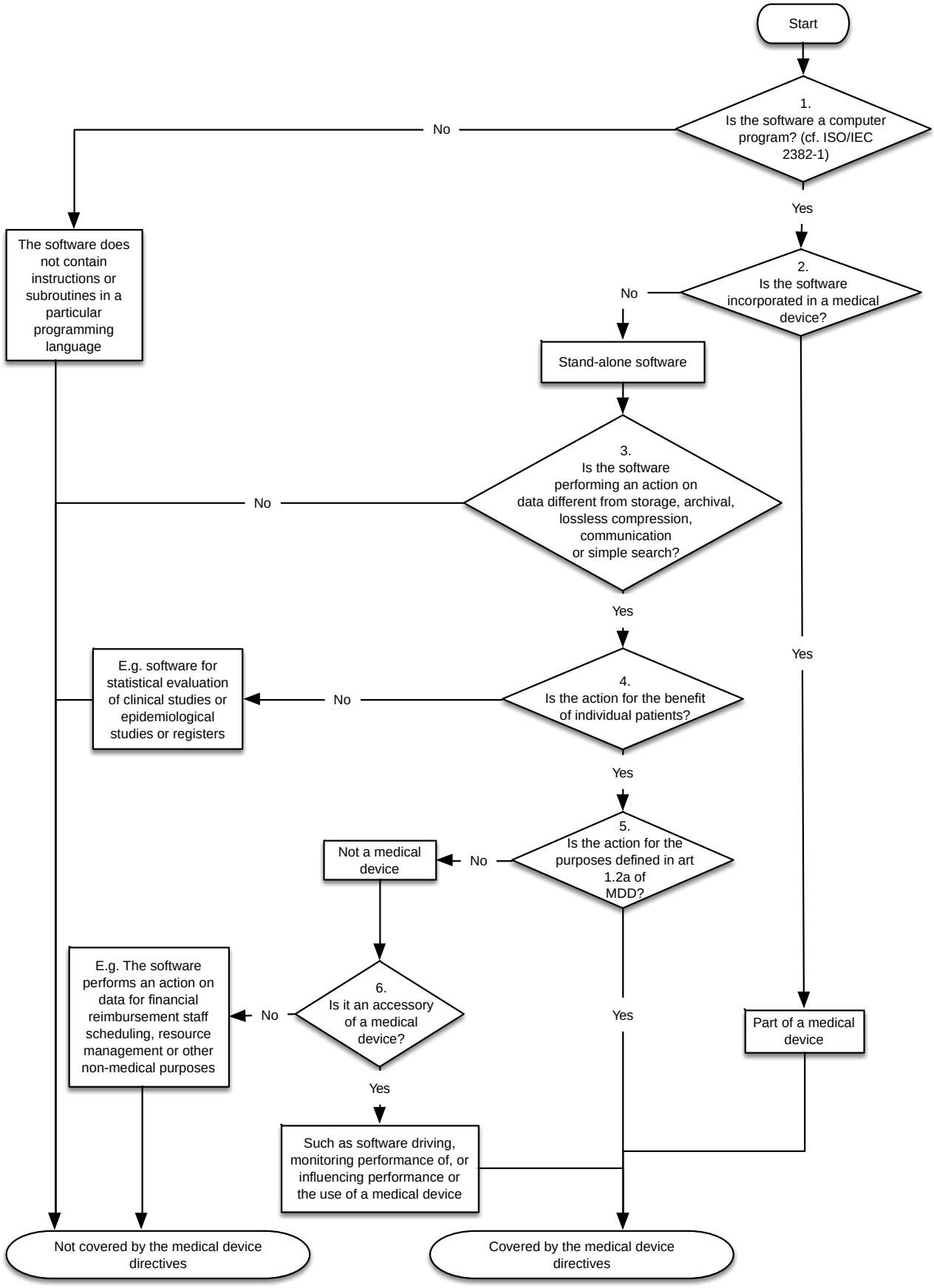


Figure 26 Decision diagram to assist qualification of software as medical device.

Medical device

The following steps must be taken to determine if an application is qualified as a medical device:

Step 1:
The first step is to determine if the application is a computer program. The definition from ISO/IEC 2382-1 is used, this definitions reads as:

"A computer program is defined as syntactic unit that conforms to the rules of a particular programming language and that is composed of declarations and statements or instructions needed to solve a certain function, task, or problem."

If the software is not a computer program it is a digital document according to MED DEV and is therefore not a medical device.

Step 2:
If the software is included into a medical device instead of stand alone software, it is seen as part of the medical device.

Step 3:
The software is not a medical device, if it **does not** perform an action on data, or performs an action limited to storage, archival, communication, simple search or lossless compression.

Simple search means the retrieval of records by matching record metadata against record search criteria. This does not include software that provides interpretative search results.

Altering the representation of data for enhancement purposes does not make the application a medical device. However, the application could be a medical device when the data is enhanced for a medical purpose.

Applications that intend to create or modify medical information could be a medical device. But only if these modifications are made to facilitate the perception and/or interpretation tasks of the healthcare professional when reviewing this information. A full list of modifications is included in MED DEV 2.1/6.

Step 4:
Is the application for the benefit of the individual patient? An example of this given by MED DEV 2.1/6 is software to be used for the evaluation of patient data to support or influence the medical care provided to that patient. Examples of applications that are not beneficial to the individual patients are applications that:

- Aggregate population data;
- Provide generic diagnostic;
- Treatment pathways;
- Scientific literature;
- Medical atlases;
- Models and templates;

- Applications for epidemiologic studies or registers.

Step 5:
The application is a medical device if the manufacturer specifically intends the software to be used for any of the purposes listed in Article 1(2)a of Directive 93/42/EEC. The application is not a medical device if it is intended for non-medical purposes such as invoicing, staff planning, emailing, web or voice messaging, data parsing, word processing or making back-ups.

Step 6:
The application is not a medical device when it is an accessory to a medical device. If the application is an accessory of a medical device, it is seen as part of that medical device and the application will fall under Directive 93/42/EEC. Applications that are available over the internet or via *in vitro* diagnostic commercial services are subject to MED DEV 2.1/6.

The application falls under the medical devices act if the application is a medical device according to the decision diagram in Figure 26. The medical devices act places requirements on a medical device.

- Prove is required
- Approval of a certified institute
- An appropriate description of the application
- Should obey the regulations and laws that are applicable.

It is also important what the **intended purpose** of the application is as described by the manufacturer. An example given during in interview was that a heart rate monitor used in a fitness center is not qualified as medical device. Since the monitor is used by individuals who want to measure their heart rate during their workout. There are also no consequences when the monitor is not precisely calibrated. However, the same heart rate monitor is qualified as a medical device if it is placed in a hospital to measure the heart rate of patients. The reason for this is that it now can have an influence on the health of a patient.

The intended purpose described by MED DEV93/42/EEC is:

“Intended purpose” means the use for which the device is intended according to the data supplied by the manufacturer on the labeling, in the instructions and/or in promotional materials.”

Classification of medical devices

The next step is to determine what class the application is part of. The following classes are identified in the medical devices act:

- **Low risk – class I**
    - A diagnosis is made by the mobile application.
- **Medium risk – class IIa and IIb**
    - Energy is added to the human body by the application, or applications that monitor vital body functions whereby variations are a direct threat for the live of the patient.

- **High risk – class III**
    - Not applicable for mobile health applications.

There are 18 rules for the classification of medical devices. However, only the rules that are applicable to active medical devices are relevant for mHealth applications. Those are rules 9, 10, 11 and 12. Figure 27 represents a visualization of these rules. The MED DEV 2.4/1 gives a general explanation with examples for each rule.

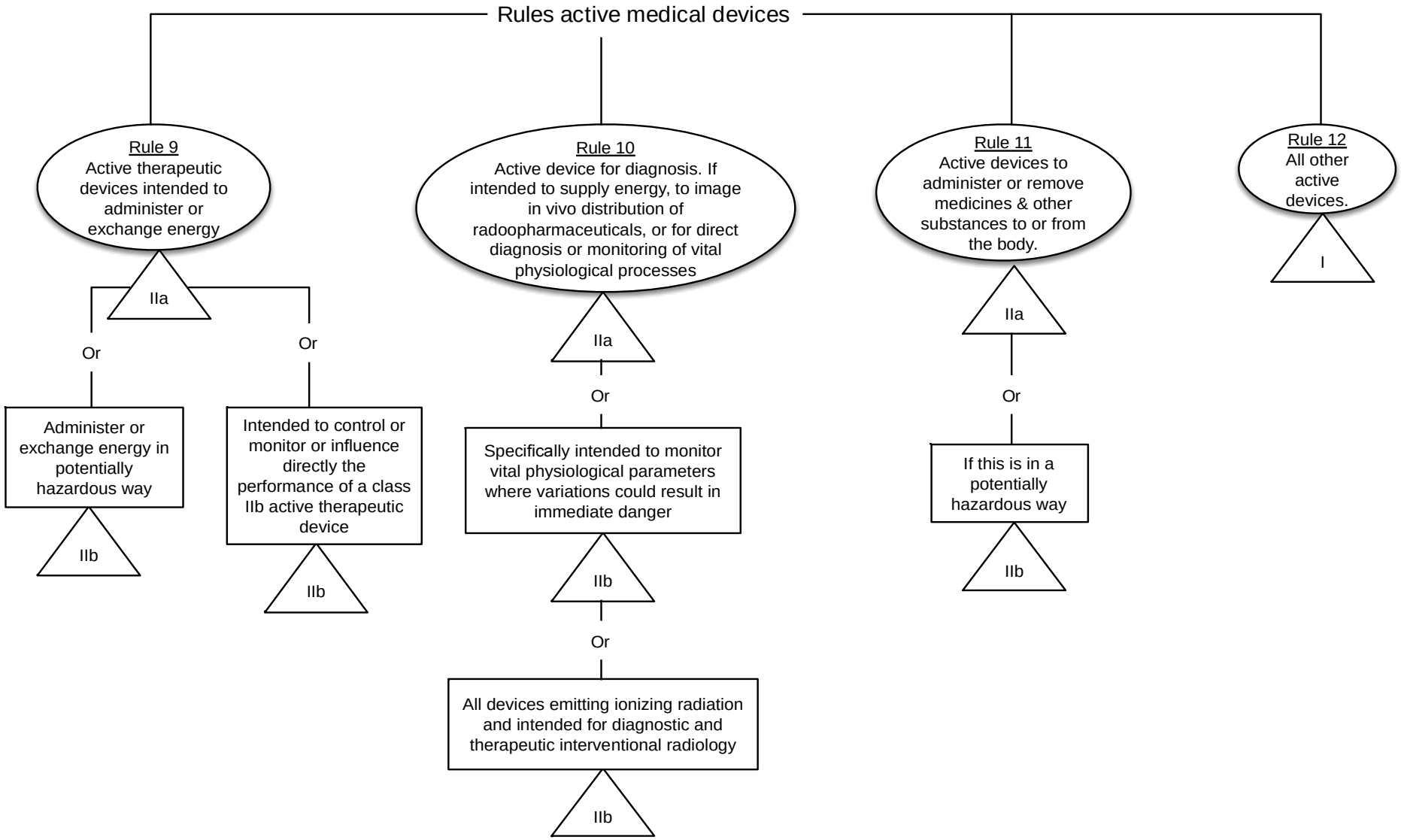


Figure 27 Rules active medical devices

Rule 9 - Therapeutic devices intended to administer or exchange energy

Most of the devices classified by this rule are electrical equipment used in surgery. Examples given are lasers, surgical generators or radiation treatment. Example that are relevant for mHealth applications are apps that provide hearing aids or applications that use light treatment. All of these devices are in **class IIa**.

There is also the option that the active medical device administers or exchange energy to and from the human body in a potentially hazardous way. In this case the application would be in class IIb. However, this is not relevant for mHealth applications. Since examples that are given are devices that use ultrasound, ionizing radiation or thermal energy.

The last exclusion of this rule is that the active medical device intends to control or monitor the performance of active therapeutic devices or influence the performance of devices in class IIb. In this situation the device is in **class IIb**. This is applicable for mHealth applications. There could be for example an application to control another medical device through a smartphone.

Other rules apply when an application controls a medical device that is an in vitro diagnostic medical device. Regulations for these devices can be found in Directive 98/79/EC. An explanation of these regulations is out of scope of this research.

Rule 10 – Active devices for diagnosis

This is a broad rule that encompasses a range of widely used equipment. Examples that are relevant will be given below. Applications for diagnosis are in **class IIa** if the application is intended to allow direct diagnosis or monitoring of vital physiological processes. An example is an application that measures the heart rate of the user. There are exceptions where devices for diagnosis are in class IIb. However, it is unlikely that there are mHealth applications that are subject to these exceptions.

Rule 11 – Devices intended to administer and/or remove medicines, or body liquids

This rule is intended to cover drug delivery systems and anesthesia equipment. This rule is not applicable for mHealth applications.

Rule 12 – All other active devices

This rule is to cover all active devices not covered by one of the previous rules. The majority of mHealth applications that are medical devices will be covered by this rule. These applications will be classified as **class I** devices.

Requirements medical device

Medical applications that are classified as a medical device are subject to a whole set of rules. The following steps are the most important steps that should be taken into account when the application is a medical device. An overview of all of the requirements can be found in the medical devices directive. The most important difference between a class I, class IIa and class IIb is that the

manufacturer of a class I application can do most of the steps described below himself. Applications in a higher class are required to involve an external party who will review documentation or the entire application.

Step 1: Meet the Essential Requirements
Medical devices are subject to the "essential requirements". These are part of the medical device directive. As an example, some of the requirements are listed here. The full list of requirements can be viewed in the medical device directive.

- Eliminate or reduce risks for the users;
- Technical knowledge, experience, medical condition etc. of the intended user should be taken into account while designing;
- Include alarms where necessary
- Measuring functions should provide accuracy;
- Validated according to the state of the art taking into account the principles of development lifecycle, risk management, validation and verification.
- Include information on how to use it safely and properly.
- Instructions for use must be included in the packaging for every medical device.
    - However, the instructions can be left out for class I or class IIa devices if the device can be safely used without any instructions.

Step 2: Prepare technical documentation
The technical documentation describes the application and relevant issues such as the risk analysis performed or the clinical data and how this is gathered. An external party called a "notified body" needs to review this documentation if the application is classified as class IIa or as class IIb (MED DEV 2.4/1). The list below includes subjects that should be described in the documentation. The full list can be viewed in the medical device directive.

1. General description of the product;
2. Packaging and labeling documentation;
3. Design verification;
4. Results of risk analysis;
5. Clinical data;
6. List of relevant essential requirements and harmonized standards.

Step 3: Request notified body intervention
Applications that are in class IIa or class IIb need to be certified by a notified body. The reason for this is that these applications have an increased risk compared with class I applications.

Step 4: Affix the CE marking
All medical devices that are on the market should have the CE marking.

Step 5: Record, evaluate and notify incidents
It is important that there is a procedure for informing authorities about incidents that occurred with the medical device. The manufacturer of the application is obliged to make an investigation of any incident. All incidents need to be recorded and evaluated according to this procedure. Actions as a result from an incident should also be recorded and reported to the relevant authority.

Step 6: Post-market surveillance

It is important that there is a post-market surveillance. This implies a procedure for reviews, experience and feedback gained from the application. Actions that result from this procedure need to be documented. Any clinical evaluation and its documentation need to be actively updated with data gathered during the post-market surveillance.