Quality criteria for health apps: a systematic review

Wouter Billiet, MD, UGent

I.s.m. Vanden Bûssche Birgit, MD, UGent

Promotor: Dr. Van Nuland Marc, MD, PhD, KU Leuven

Master of Family Medicine

Masterproef Huisartsgeneeskunde

ABSTRACT

Achtergrond

De komst van mobiele apps biedt veel nieuwe mogelijkheden en kan een revolutie betekenen voor onze gezondheidszorg op vlak van kwaliteit, efficiëntie en kosten. Om mHealth te kunnen toepassen moeten we echter zeker zijn dat de apps die we gebruiken van goede kwaliteit zijn. Apps van slechte kwaliteit zouden namelijk schade kunnen berokkenen aan patiënten op verschillende manieren (bv. financieel, privacy, gezondheid…). De vraag rijst wat die 'goede kwaliteit' dan precies inhoudt en of er reeds algemeen aanvaardde evaluatiemethodes zijn om de kwaliteit van apps aan af te toetsen.

Doelstelling

Een systematische review van artikels met evaluatietools voor gezondheidsapps werd uitgevoerd om na te gaan welke tools of criteria belangrijk zijn bij het evalueren van gezondheidsapps.

Methode

Op basis van een verkennend literatuuronderzoek werden zoekcriteria samengesteld voor een systematisch literatuuronderzoek in vier medische databases. Alle resultaten tot 5 februari 2017 werden door twee verschillende auteurs overlopen en een uiteindelijke selectie van bruikbare artikels werd gemaakt op basis van verschillende consensusmeetings. Deze artikels werden beoordeeld met behulp van een (aangepast) sjabloon om de kwaliteit ervan te bepalen. Daarna werden ook de verschillende evaluatietools en –criteria uit de artikels met elkaar vergeleken.

Resultaten

De uiteindelijke selectie bevat zeven artikels die een algemene evaluatietool voor apps lijken te bezitten. Deze artikels worden besproken op basis van het sjabloon. De meest onderbouwde evaluatietool is de MARS. De studie die deze tool valideert is van goede kwaliteit en de tool werd al vaak opgepikt in de verdere literatuur. Daarnaast komt ook de recentere ACDC-tool in aanmerking als valabele evaluatiemethode, maar een validering ontbreekt vooralsnog. De tools uit de andere studies zijn ofwel erg onvolledig ofwel te sterk toegepast op een specifiek type apps. Interessante elementen of criteria uit deze overige tools worden ook besproken.

Conclusie

Er is een grote nood aan een algemeen aanvaarde evaluatiemethode maar meer onderzoek is hiervoor nodig. De MARS heeft voorlopig de meeste evidentie om gebruikt te worden bij het evalueren van de kwaliteit van gezondheidsapps.

Table of Contents

ABSTRACT 3
Achtergrond 3
Doelstelling 3
Methode 3
Resultaten 3
Conclusie 3
VOORWOORD 5
INTRODUCTION 6
METHODS 8
LITERATURE SEARCH AND SELECTION METHODS 8
TEMPLATE DEVELOPMENT 8
APPLICATION OF THE TEMPLATE FOR THE ARTICLES 10
RESULTS 11
LITERATURE SEARCH 11
EVALUATION OF THE INCLUDED ARTICLES 13
COMPARISON OF SCALES 19
CRITERIA USED IN THE TOOLS 19
DISCUSSION 22
LIMITATIONS 24
CONCLUSION 25
REFERENCES 26
APPENDICES 28
APPENDIX 1: Construction of the template 28
APPENDIX 2: Manual for the template 30
Manual for scoring of qualitative part of the articles 30
Manual for scoring of quantitative part of the articles 34
APPENDIX 3: Suggestion 42
APPENDIX 4: Beschrijving van de werkverdeling 43
APPENDIX 5: Advies Ethisch Comité (OBC) 44

VOORWOORD

Coauteur Birgit kwam met het idee om samen onze thesis te maken over ‘apps in de huisartsenpraktijk’. Ik ben normaal niet echt een early adopter wat betreft het aanwenden van technologische revoluties. Toch vond ik het een interessant thema omdat ik meteen ook geloofde in het grote potentieel van mHealth voor onze samenleving en gezondheidszorg. Vooral in het overdragen van informatie en het bijhouden van data kunnen apps volgens mij een grote rol gaan spelen en deze trein wil ik dus niet missen. Momenteel staat mHealth nog in zijn kinderschoenen en de belangrijkste voorwaarde om apps te gaan gebruiken is dat we voldoende vertrouwen kunnen hebben in de kwaliteit ervan. Dit is meteen ook de reden waarom ik vaak achterdochtig ben bij het aangrijpen van nieuwe technologieën. Het bestuderen van kwaliteit van gezondheidsapps was voor mij dan ook een erg verrijkend thema.

Het was eerst de bedoeling om een literatuurstudie en een trial met twee verschillende apps uit te voeren. Deze apps werden als kwaliteitsvol voorgesteld door het artikel van Bijma (1). Tijdens het uitvoeren van de literatuurstudie bemerkten we echter dat er reeds veel kleinschalige onderzoeken met apps waren gerealiseerd en dat deze weinig toegevoegde waarde hadden voor de wetenschap. Apps updaten namelijk regelmatig en de ontwikkeling van nieuwe apps is niet bij te houden. Dit betekende dat ons onderzoek mogelijks al voorbijgestreefd zou zijn nog voor het afgewerkt was. We bemerkten dat dat vooruitzicht ons frustreerde en na overleg met de promotor besloten we om het anders aan te pakken. We zouden wat extra inspanning leveren en een systematisch literatuuronderzoek over kwaliteitscriteria voor apps uitvoeren en onze thesis in het Engels schrijven met oog op een eventuele publicatie.

Ik ben dan ook trots dat we uiteindelijk wel iets nuttigs aan de wetenschap bijgedragen hebben en ik denk dat er ook echt wel nood was aan onderzoek zoals het onze. Daarom wens ik de promotor Dr. Marc Van Nuland en coauteur Birgit Vanden Bûssche te bedanken voor hun hulp bij het realiseren van deze masterproef. Verder bedank ik ook graag mijn ouders om mijn studies mogelijk te maken en mijn vriendin die mij hierin zo goed gesteund heeft.

INTRODUCTION

eHealth is the use of information and communication technologies (ICT) for health. Mobile Health (mHealth) regards the use of mobile and wireless devices to improve health outcomes, health research and healthcare services. (2-4)

Nowadays cell- and smartphones are widely used, even in low and middle income countries, increasing the possibilities to implement mHealth. For example, in Belgium over 1.1 mobile phone contracts per citizen are registered. (2)

Certain mHealth apps have proven to be cost efficient and were shown to improve health outcomes. Mobile Health has the potential to improve health and healthcare and can play an important role in the transformation to qualitative health and Universal Health Coverage. (2, 3, 5-7)

If an app serves as a medical device (i.e. apps that aid diagnostic or treatment choices, interpret results or make calculations without external control of a health worker [e.g. insulin calculator]), clear recommendations and a control authority are already established. However, for apps informing patients or health workers, saving patient health information or using tele-medicine (radiology/dermatology) the quality control is less clear. (8, 9)

The WHO tried to elaborate a framework for mHealth apps, with a subdivision of twelve different kinds of apps. Nevertheless, a standard evaluation method is not yet implemented. (4)

Uncontrolled use of mHealth apps certainly has its limitations. One of the biggest obstacles concerns the lack of trust in these new technologies by patients and clinicians. This lack of trust is influenced by different risk factors related to these health apps. (10) One of the major risks concerns the poor quality of the information in many of the available apps in the app stores. (9, 11, 12) Equally important is the security and privacy of the data that are shared in health apps. This is of particular interest when dealing with information as vulnerable as medical information. Furthermore, an app has to be appealing to be used and needs to have a sustainable impact on health care and health behaviour. (2, 13-18)

Medical apps are high-end technology on the intersection of very different types of disciplines in an ever more demanding society. (18, 19) This observation implicates that the development and evaluation of health apps is a difficult task. To generate a high-quality app the developer needs to be fully skilled in IT development, privacy and security issues, behaviour change strategies and medicine. Difficulties in combining the knowledge might be a feasible reason why many apps are perceived to be of low-quality.

If mobile apps are to be used to their full potential - to improve our health care - we should only use high-quality apps. This requires a selection procedure. (20) Users are rarely capable to define high-quality, bearing in mind all the different aspects to be considered especially when related to health care apps. The star-criteria of app stores, based on their popularity among users, are an unreliable measure compared to the evidence-based quality assessments that are used for other healthcare interventions. (21, 22) Some (for-profit) organizations have tried to gain authority in this subject by creating certification programs evaluating apps based on their own criteria. However, these attempts generally failed due to the fact that certificates have been granted to apps which appeared to show serious problems. (14, 23) These organizations rarely release the criteria they use, or don't provide a transparent evaluation method. Another ethical consideration is the non-compatibility of the possible economic incentives and the vulnerability of medical information being collected by the apps. Big data lead to big interest by different companies, and not all companies have the patients' best intentions in mind. (11) Governmental initiatives tried to select some trustworthy apps, but compared to the fast progress of mHealth, the multi-layer evaluation methods for every new (update of an) app are too time-consuming. (23) Governmental regulation of app development is feasible, but may be hard to enforce and may delay the improvement of mHealth. (5) The vast majority of health apps are 'health and lifestyle' apps for which no rules exist. (18, 23)

In Belgium, a pilot project to invest in a five health apps has recently been launched, but results aren't yet available. The purpose is to implement a juridical framework and a fee model for health workers. (5)

No waterproof system has been globally accepted to assess app quality so far. For the moment automation of quality control is too complex. Therefore, criteria for a reproducible evaluation method need to be created. (21) This should not be too extensive and time-consuming, nor should it be too minimized and narrow-scoped. It is a huge challenge considering the multitude of functionalities apps can have. (18, 22)

mHealth apps are starting to be an important part of health and primary health care. They have the potential to improve patient education, monitoring and treatment of patients. (19, 24) On the other hand, providing wrong or unclear information and leakage of health information can be harmful. (10, 25)

The purpose of this systematic literature review is to provide an overview of published quality criteria for health and lifestyle apps and to assess its quality. Results may guide clinicians and patients to distinguish trustworthy from untrustworthy mHealth apps.

METHODS

LITERATURE SEARCH AND SELECTION METHODS

After an initial exploratory literature search, we formulated the following search string in a first consensus meeting: "(m-health OR (mobile AND health)) AND ((health AND app) OR (lifestyle AND app) OR (wellness AND app) OR (wellbeing AND app)) AND (regulation OR evaluat* OR rating OR scale OR asses* OR quality OR criteria OR certification)".

We applied this search query in four databases: Embase, Cochrane Library, Medline and Psychinfo. Articles were then independently screened and selected by two researchers (BV & WB). A first selection took into account the title and abstract, while in a second step the entire content of the retrieved articles was assessed. Inclusion criteria were English-language articles which dealt with evaluation or development of quality criteria for mobile mHealth apps. Quantitative, qualitative and descriptive research were included. We excluded articles who only discussed evaluation tools for one specific disease or problem and articles for which no full text was available.

TEMPLATE DEVELOPMENT

In a next step a suitable template for quality assessment and risk of bias for research for health improvement tools was searched for on Pubmed, Google, Cochrane and The Equator Network. We also sought advice by an expert in the field (TB). A prototype of the template was made out of the results and was tested with one of the articles (20). A consensus meeting (BV, WB & MVN) was held to decide on the final content of the template. More information on the construction of the final template can be found in appendix 1.

The final template consists of a qualitative and a quantitative part in which we merged a selection of items from three different templates. The qualitative part of the template, fifteen items, analyses the set-up of the evaluation tool, while the quantitative part, eighteen items, digs deeper into the actual testing and validation of the tool (see Table 1).

TEMPLATE		
Qualitative		
	Kmet Q 1 + EQUIPH	Question/objective sufficiently described (A/ aim of the health promotion tool
1	IIb&c + BH7	mentioned?; B/ target population mentioned?) incl. target group
2	Kmet Q 2	Study design evident and appropriate
3	Kmet Q 3	Context for the study clear
4	Kmet Q 4	Connection to a theoretical framework/wider body of knowledge
	Kmet Q 5 +	
5	BH5	Sampling strategy described, relevant and justified incl. in- and exclusion criteria
	Kmet Q 6 +	
6	BH1	Data collection methods clearly described and systematic, incl. timeframe
7	Kmet Q 7	Data analysis clearly described and systematic
8	Kmet Q 8	Use of verification procedure(s) to establish credibility
9	Kmet Q 9	Conclusions supported by the results
10	SRQR 19	limitations (trustworthiness and limitations of findings)
11	Kmet Q 10	Reflexivity of the account (Researchers' characteristics that may influence the research (personal attributes, qualifications/experience, relationship with participants, assumptions/presuppositions)
12	SRQR 20	conflict of interest (potential sources of influence or perceived influence on study conduct and conclusions; how was this managed)
13	SRQR 21	funding (sources, role, interpretation, reporting)
14	BH 6	Does the study identify the country in which the search was conducted?
15	BH 8	Does the study provide a list of the apps included in the review?
Quantitative		
	Kmet K 1 + EQUIHP	Question/objective sufficiently described (A/ aim of the health promotion tool
1	IIb&c + BH 7	mentioned?; B/ target population mentioned?) incl. target group
2	Kmet K 2	Study design evident and appropriate
	Kmet K 3 +	Method of subject /comparison group selection or source of information/input
3	BH1 + BH5	variables described and appropriate incl. time frame) incl. in- and exclusion criteria
4	Kmet K 4	subject (and comparison group if applicable) characteristics sufficiently described
5	Kmet K 6	reporting of blinding of investigators
6	Kmet K 8	Outcome and (if applicable) exposure measure(s) well defined and robust to measurement/misclassification bias? means of assessment reported
7	Kmet K9	sample size appropriate
8	Kmet K 10	analytic methods described/justified and appropriate
9	Kmet K11	some estimate of variance is reported for the main results
10	Kmet K 12	controlled for confounding
11	SRQR 20	conflict of interest (potential sources of influence or perceived influence on study conduct and conclusions; how was this managed)
12	SRQR 21	funding (sources, role, interpretation, reporting)
13	Kmet K13	results reported in sufficient detail
14	Kmet K 14	conclusions supported by the results
15	SRQR 19	limitations (trustworthiness and limitations of findings)
16	BH 2	Does the study specifically mention that the reviewed app was downloaded and analysed based on its content?
17	BH 6	Does the study identify the country in which the search was conducted?
18	BH 8	Does the study provide a list of the apps included in the review?

Table 1: Final template for quality assessment and risk of bias for research for health improvement apps

To avoid any misunderstandings or misinterpretations, an explanation of the questions is added in appendix 2. (26-29)

APPLICATION OF THE TEMPLATE FOR THE ARTICLES

The articles that were finally included were then scored using the final template to the respective studies. This was done independently by the 2 researchers (BV and WB). The results were compared and discussed during a new consensus meeting (BV, WB and MVN) until an agreement was reached and a conclusive score was given to all template questions for the articles.

RESULTS

LITERATURE SEARCH

The search took place between 09/09/2016 and 18/09/2016 for W.B. and on 04/10/2016 for B.V. The search of B.V. retrieved 565 articles in Pubmed, 4 articles in Cochrane, 669 articles in Embase and 160 articles in Psycinfo. The search of W.B. resulted in somewhat less articles as it was performed a few weeks earlier.

Two different strategies to select the articles were used.

The first author pre-selected (Pubmed 137; Embase 162; Cochrane 0; Psycinfo 5) articles based on the title and made a second selection based on the abstract. Articles that only dealt with one app or one specialization were then excluded, except if they explicitly postulated that they used a new evaluation scale for apps (Pubmed 59; Embase 64; Psycinfo 5). A final selection excluded duplicates, articles without available full texts and framing articles (Pubmed 41; Embase 11; Psycinfo 0). The second author selected the articles immediately on title and abstract based on the inclusion criteria (Pubmed 29; Psycinfo 5; Embase 18). One article was excluded based on language (Spanish), three because they had no full text available and 10 because they were duplicates (Pubmed 28; Embase 10; Psycinfo 0).

A second consensus meeting (with BV and WB) was held to compare the selection results (level of abstract) and to discuss the differences.

* The search of the first author yielded 52 articles of whom 21 were also found in the search of the second author. 31 articles were not withheld by the second author, but after discussion 8 of them were selected for analysis anyway and 23 articles were deleted from the selection.

* The search of the second author yielded 38 results, of whom 21 were also found in the search of the first author. 8 of the remaining results were found in an earlier selection of the first author, of which 4 articles were withheld and 4 articles deleted after discussion. Of the 9 articles that were only found by the second author, one was included and 8 articles were deleted from the selection.

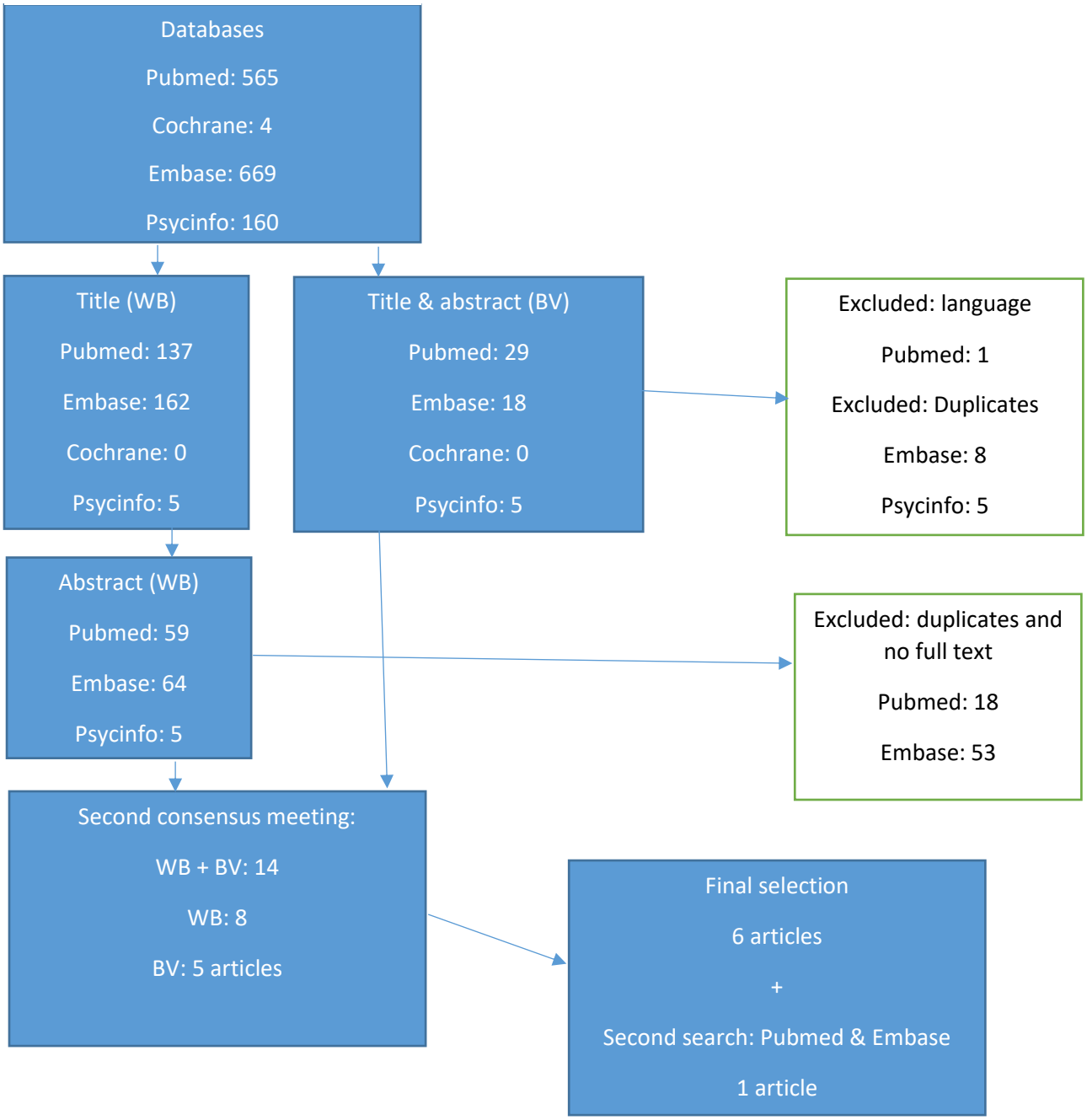
* Of the 21 results found by both, 7 articles were excluded after discussion. As such, a total of 27 articles (21+8+4+1-7) were withheld for full analysis.

The withheld articles were independently appraised for inclusion by reading the whole article. In a third consensus meeting (with BV, WB and MVN) a final agreement on the articles to include was reached. Only 6 articles were finally retained as they sufficiently incorporated evaluation criteria for mobile health apps. (20, 30-34)

As the field of mHealth is rapidly evolving, a search update was executed shortly before finalizing this review. Both BV and WB searched the same databases on 05/02/2017 for new articles, released between 01/09/2016 and 05/02/2017. This yielded 138 results in Pubmed, 155 results in Embase and 98 results in Cochrane. The Psycinfo database was not searched as it was no longer at our disposal. However, we don't expect that this biases the analysis, as the Psycinfo database did not yield any included studies not retrieved in the other databases during the first search.

After independent selection and discussion, five more articles were withheld from Pubmed and one from Embase for full text reading. Of them, only one was added to the final selection. (35)

Figure 1: Literature search



EVALUATION OF THE INCLUDED ARTICLES

The objective of the study of Stoyanov (20) was to develop a reliable measure for rating the quality of mobile health apps. The scale was validated with the use of a sample of mental apps. The consensus meeting after quality assessments of this study retained the following remarks:

- There is a connection to the body of knowledge but the available evidence is so scarce that it is difficult to score in this section.
- The main sources for the criteria were articles concerning website quality criteria and not app quality criteria. Moreover, the MARS-tool was only tested for mental apps, so we should be cautious to accept this tool for all kinds of apps. These critics are written down in the limitations of the study itself.
- Blinding of the investigators (not useful), control for confounding (not appropriate) and reflexivity on the account of the researchers (almost never done) is absent in all of the studies discussed here.
- Doctors or behaviour change experts should preferably be part of the advisory team for the development of a tool. It is however not sure whether this was the case, given the vagueness of the words 'professionals involved in the development of mHealth apps worked together'.

The study of Loy (30) aimed at developing a quality assessment tool to evaluate apps that deal with medication-related problems (MRP). Different types of medication-related problem apps and the trends in the evaluation are discussed. The consensus meeting after quality assessments of this study retained the following remarks:

- We could only approve partially with the design of the study, because the only focus of the evaluation tool is apps for MRP. This article provides a ranking list of apps for MRP and the comparison of different aspects or types of MRP apps seems more important than creating and more thoroughly validating the tool. The asset of the design is that the authors first created the tool, then selected apps and finally used the tool to test the apps in a well-structured way. Regrettably, this test was done to evaluate the available apps in the app store and not to validate the tool. The qualitative part of this study is therefore less interesting for our systematic review.
- Information on how the criteria were extracted from the 70 studies is vague.
- It remains unclear who, or how many people developed the tool, selected the apps and carried out the quality assessment.
- There is no verification or validation of the tool.

- There are no verification procedures or control for confounding for the results of the study. However, an estimate of variance is noted.

The article of Anderson (35) tries to find and evaluate apps dealing with chronic disease self-management. It shortlists apps for evaluation and proposes the App Chronic Disease Checklist (ACDC) for quality and reliability. By doing so, the article wishes to indicate the best available apps. The consensus meeting after quality assessments of this study retained the following remarks:

- This study has a very good introduction and connection to the wider body of knowledge.
- The items for the evaluation tool were mainly based on usability checklists in the literature. The articles at the basis of the tool are comprehensively described, but there is no information on how these articles were retrieved. How the elements in the article were adapted and filtered to form the proposed checklist also remains unclear.
- The validation of the tool was only done by the three authors and there was no real testing of the tool (no quantitative appreciation in our study). It is mentioned how apps can be tested, but this proposed method was not applied to this study. As long as this tool has not been tested, it seems exaggerated to call the tool comprehensive and evidence-based.
- The study admits to be mainly based on the MARS-tool (20), but claims to add some extra items. The criticism that MARS is based on quality criteria for websites is correct. However, as the ACDC-checklist was mainly deducted from MARS, this commentary is backfiring at the ACDC-tool. The remark that MARS is only made for mental apps, whereas the ACDC is for all chronic disease apps should be considered. However, it seems like the MARS was created for all apps and tested on mental apps, whereas the ACDC was not tested at all. The authors' comment that the ease of use- category (in the ACDC) is better than the aesthetics (in the MARS) seems reasonable.
- The mentioned self-reflective limitations in this study are -if any- very limited.

The article of Yasini (31) proposes the implementation of a process to seal good quality apps and provide their logo when they seem to be trustable. The aim of the study is to automate app evaluation using the use-cases described in their previous research. The consensus meeting after quality assessments of this study retained the following remarks:

- In general the methods for this study were unclear and poor.

- The literature search was not described and only three articles were used. Moreover, the content of these articles was not discussed.
- The validation procedure by the expert panel is unclear.
- The actual quality questionnaire is not included in the article. There is a link to the site, but this site is only available in French and the questionnaire cannot be found on the site without an account. The proposed quality criteria remain unavailable.
- There is a strong emphasis in this study on the use-cases (based on their previous research), and without acceptance of this system, apps cannot be evaluated by this tool. It feels like the primary aim of the author is to implement this system and spread the mHealth Quality seal. Private certification has however proven not to be a reliable system for the approval of apps so far. (12)
- The usability of the questionnaire is tested only by target users from their own evaluation community and the security check is outsourced with the consequence that the used security criteria are not mentioned.
- There is no testing of the presented tool, so we were not able to give a quantitative appreciation.

Developing and testing a new evaluation tool for healthcare smartphone apps was the intention of the study of Jin (33). The consensus meeting after quality assessments of this study retained the following remarks:

- The design of this study is good, as it tries to do exactly what we are looking for: creating a new tool and testing its validity. The methods are however not as performant, both for the creation as for the testing of the tool. In the construction phase, the sampling of criteria is not satisfactory as it is based on only four studies listing quality criteria. The search strategy and criteria to select these studies is lacking. Also, the method to select the studies' criteria for the tool was not described. Further on the assumption is made that apps are a downscaled version of websites and therefore need to correspond to the same criteria. This is not an obvious assumption. In the discussion, the authors admit that different aspects from the tool should be re-evaluated for the distinct characteristics of smartphone apps.
- There is a validation of the tool by five experts, based on tool analysis and an unknown number of closed-ended questions. More information concerning this evaluation is not available. As a result of the validation, different items were removed. Even though it was proven to be

important for use of apps, community aspects were deleted from the final tool without a clear argumentation.

- The testing of the tool does not add a lot of credibility to the tool. A convenience cohort of students in medicine and nursing were asked what they thought about the tool by answering a questionnaire. There is no information on how the questionnaire was constructed. The outcome measure of this test remains rather unclear. There is also no discussion or statistical analysis of these results.
- In their conclusion, the authors argue that specific apps should be used to test the evaluation tool, but such testing was not done by this study.
- The limitations are too restricted.

The study of Chyjek (34) tries to identify the best available pregnancy wheel apps on the market by using a newly created tool that can be used for all healthcare apps. The consensus meeting after quality assessments of this study retained the following remarks:

- The connection to a wider body of knowledge is poor and only focused on obstetrics and gynaecology.
- The focus of this study is an overview of the available apps with pregnancy wheels and quoting these apps rather than developing a new validated, evidence based tool for app evaluation.
- Information on how the tool was build is not available, only the references are given. There is no proof that the tool analyses apps thoroughly. Some important evaluation elements might be missing. Nothing is mentioned on the data collection methods of the sources for tool criteria, nor on the sampling or the data analysis. The methods of the tool creation are therefore rather weak. Furthermore, the validation of the tool is not satisfying: an objective component reporting error rate is given, without sufficient additional explanation to it.
- The methods concerning the exploration of pregnancy wheel apps are more adequate. Sampling, data collection methods and data analysis are described and appropriate. Bearing in mind the weak construction of the tool, the results of the appliance of the tool are only as strong as the tool itself.
- There is no further analysis of the results, no estimate of variance is available (not possible given the methods) and no control for confounding. Only the total score for the APPLICATIONS-tool is given.
- In the conclusion of this article, the authors claim that the tool could be used for apps in different specializations. The tool itself might not seem to be specific for pregnancy wheel apps

and the claim might be true, but the authors should be careful with these kinds of claims, bearing in mind the above-mentioned criticism to the tool.

The primary objective of the study by Mc Millan (32) was to create a specific tool for apps that aim at health behaviour change. In addition to this, the study scanned behaviour change apps from the NHS Apps Library to the items of this tool. The consensus meeting after quality assessments of this study retained the following remarks:

- The design of the study is acceptable, but the tool only aims at assessing the quality of apps that deal with behaviour change. Its applicability for other types of apps should not be taken for granted.
- The criteria are only based on the NICE behaviour change guide 2014, which might be considered as a gold standard for behaviour change, but not for quality of apps in general. The selection of criteria out of the guideline and their conversion into questions was done by only one author.
- The verification of this tool immediately serves as testing of the tool. The design of this testing is weak, which is mainly reflected in the quantitative part of our appreciation of this study. Only an interrater agreement percentage for each question was calculated between the two evaluators. No other analytic methods or validation procedures were applied. The value of this percentage is doubtful.
- The tool is only tested on apps that were found in the NHS Library (and deal with behaviour change). This might represent a huge bias for this study as this sample does not represent the average app available on the market. However, many health and lifestyle apps try to effectuate behaviour changes and as a result, maybe a lot of apps could benefit from compliance to quality criteria for behaviour change apps.

Qualitative		Stoyanov		Loy		Andersson		Yasini		Jin		Chijek		McMillan	
1	Objective	2		2		2		1		1		2		2	
2	Design	2		1		2		1		2		1		1	
3	Context	2		2		2		2		2		2		2	
4	Framework	1		2		2		0		0		1		2	
5	Sampling	1		2		0		0		0		0		1	
6	Data collection	2		2		0		1		0		0		-	
7	Data analysis	2		1		1		0		1		0		1	
8	Verification	2		0		0		1		2		1		1	
9	Conclusions	2		2		1		0		0		1		1	
10	Limitations	2		2		1		0		1		1		2	
11	Reflexivity	1		1		1		0		1		0		1	
12	Conflict of Interest	2		2		2		0		2		2		0	
13	Funding	2		2		1		0		2		0		2	
14	Country	2		1		0		1		2		2		2	
15	List of apps	2		2		-		0		-		2		0	
Quantitative															
1	Objective	-		-						-		-		-	
2	Design	-		-						-		-		-	
3	Group selection	2		2						0		2		2	
4	Characteristics	2		2						2		2		1	
5	Blinding	-		-						-		-		-	
6	Outcome	2		2						0		0		1	
7	Size	2		2						2		0		1	
8	Analytic	2		2						2		-		1	
9	Estimate/ variance	2		2						2		0		0	
10	Confounding	0		0						0		0		0	
11	Conflict of Interest	-		-						-		-		-	
12	Funding	-		-						-		-		-	
13	Results	2		2						1		2		2	
14	Conclusions	2		2						0		1		1	
15	Limitations	-		-						-		-		-	
16	Downloaded	2		2						-		2		2	
17	Country	-		-						-		-		-	
18	List of apps	-		-						-		-		-	

Table 2: Scores (after consensus) for the included articles on the template for quality assessment and risk of bias

COMPARISON OF SCALES

The template (see Table 2) is not intended to calculate a sum, but should be seen as a tool to describe the strengths and weaknesses of the studies.

Several questions (Kmet 1, Kmet 2, Kmet 6, SRQR 19, SRQR 20, SRQR 21, BH 6, BH 8) are asked in both the qualitative and quantitative part and do not address different issues. To limit their impact, we entered N/A in the quantitative part.

Only a few articles had a connection to a theoretical framework which might not be that surprising as a relative new domain of research can be regarded as having less theoretical and evidence based backgrounds.

None of the seven articles had a sufficient description of 'the reflexivity of the account'. The likely impact of the authors' characteristics and methods used on the obtained data is not assessed. We also found little added value of mentioning the country where the study was performed.

As the studies are based on apps, blinding is hard to achieve or not appropriate. Also, no article controlled for confounding. As a result, these items were not discriminating between different studies in our quality assessment.

CRITERIA USED IN THE TOOLS

The evaluation criteria used in six of the seven studies were mentioned in the respective articles. Despite our search for the exact criteria of the study of Yasini (31) on the website, no criteria were found.

As we compare the different criteria, some are more important. Since applications are using sensitive data of patients, privacy and security is one of the key criteria to rate. The usability and aesthetics are important, in order for patients to keep using the app. These can be found in all evaluation tools. Next to security and aesthetics, the information mentioned in the application needs to be evidence-or guideline-based. Misleading or false information can potentially harm the health of the user.

The criteria of Stoyanov (20) seem to have an appropriate amount (23) of questions. This evaluation tool is the only one with a general part. This general part consists of the focus of the app, the theoretical background, age group and technical aspects about security and privacy (sharing, password, reminders, and offline function). The rating part of the tool includes six components. Every question has a five-points scale (or N/A) and for every score a full explanation is given. Fun to use, appropriateness and customization is mandatory in order to reach a better compliance. The second part relates to the functionality and ease of use and learn. Aesthetics represents a smaller

component (three questions). The biggest part focuses on the information provided in the app: the accuracy, goals, quality and quantity of information, credibility and source are discussed. Unfortunately, the part about evidence based information is excluded in the testing of the tool, since no apps were tested yet in clinical trials. As mHealth is being used more frequently, one can expect that evaluating apps in the setting of clinical trials is also likely to increase. The final two components relate to the subjective quality and perceived impact of the app. This is a relevant issue as this can be related to patients compliance, which itself is probably related to the expected health care gains.

The study of Loy (30) has an extensive evaluation tool of 86 questions. Different scores are used with explanation added. The first section is extensive on the features of the apps. They discuss the monitoring, medication interaction checker, dose calculator, medication information and medication recording features. This final feature also consists of a part about passwords and data storage. The next part, dealing with reliability, is really complete. It discusses the audience, source, design, disclaimer, funding, conflict of interest, EBM and updates. The third component discusses the usability. The part about privacy is narrow but interesting, since they are discussing the possibility of allowing users to choose which information will be shared with whom. Questions about conflict of interest, confounding and evidence of bias are only present in this tool. They seem interesting to consider.

The criteria of Anderson (35) are based on those of Stoyanov (20). They added a part about positive behaviour change, which can be useful since a large part of mHealth apps are ought to help adopting a healthy lifestyle. Most components (11/24) are similar to the ones used in the tool of Stoyanov. Aesthetics is lowered in importance and the subjective part is deleted. Also questions about user profile setup, offline mode, automation, medical and technological jargon and privacy and data security are added. Data security and privacy seems to be too limited. They worked with a three-point scale with explanation for every score.

The study of Jin (33) consists of a zero to three scale with 23 questions. The first three parts deal with the content use, the sources and the information. The aesthetics are relatively more important, this contains the consistency and suitability of design, interface and accuracy of wording. The last section discusses the security and privacy but does not evaluate an adequate privacy policy.

The study of Chyjek (34) has the smallest evaluation tool, including only 14 questions. Explanations for the scores are added. Only one question asks about the literature (references used or not). Most questions address interface and use. A component about privacy and security is not available. An interesting question is about the price of the app, since patients need to be convinced to buy the app. Still this part can be questionable because paying is only needed once and other treatments can be

expensive too. Other lines about advertisements, presents of images and videos and text search field are doubtful.

The study of Mc Millan (32) includes an extensive list of 62 questions. However, only yes and no can be answered, which of course limits the answering options. The first part of the questionnaire assesses the app purpose and is similar to the general part of the study by Stoyanov (20). The second part is about the development of the app. The third component is a small (2 questions) part discussing the usability of the app. Theme four to six are only applicable for behaviour change apps. These are derived from the NICE Guideline of behaviour change. The seventh part is extensive and includes 18 questions on privacy and data collection. Some of these questions are however very similar and should have been bundled. Only one final question evaluates data protection. The last theme discusses the evidence based information used in the app.

DISCUSSION

We conducted a systematic literature search to identify articles that contain evaluation criteria for mobile health apps. Seven articles were withheld and their content was evaluated by a template that focused first on tool development and secondly on testing of this tool. Apart from that, the available criteria in the retrieved tools were analysed and compared with each other.

Our main conclusion is that the evidence base for health app evaluation is very limited. Apps are still a relatively new technology and a firm evidence base is known to grow slowly. This contrasts with the fast development of the technology. This problem is further aggravated by the complexity of evaluating a health tool as multi-layered and diversified as mobile apps can be. Bearing in mind that personal medical information is very vulnerable and financial interests could be very big, there is a clear need for more security. The necessity for good quality evaluation criteria is therefore unquestionable. There is a clear need for more research and for a vision to implement health apps in our health society. A suggestion on how we could organize app evaluation is given in Appendix 3.

The scales we found were mostly created by the author or by an expert panel and meaningful validation of the tools was scarce. If a tool was tested, it was always done with a sample of apps that address one single disease or purpose, even if the scale was meant to evaluate all sorts of health apps.

The tool that is used most and is referred to most frequently is the MARS-tool. (20) This is the only tool that was found to be used by multiple articles to rate different kinds of apps. It is the most generally accepted tool in the literature we found. All other articles had methodological issues. A short summary of the article methodologies and reflection on the criteria used is presented here.

The article of Stoyanov (20) has the best design, very little methodological issues and the best tool validation of all studies we retrieved. The main limitation of this evaluation tool is the concise discussion of the privacy and security section. The tool still needs to be validated by other studies with other healthcare problems. As far as we know (and based on our literature search) we are not aware of any new studies with the MARS-tool that could contribute to its validity.

Loy (30) made a well-constructed study, but the tool development stays unclear and the testing is uninteresting as it is not a true validation. The tool itself is only meant for medication-related problem apps and the main objective of the study was to rank these apps in order to recommend specific apps. The criteria are too extensive (86 questions) to be generally used, but it has an easy procedure. The criteria are based on medication-related problems and cannot be used for all health apps. The privacy section is relatively less important.

Andersson (35) is claiming their ACDC-tool is better than the MARS, but this claim cannot be confirmed by our quality assessment. Although the tool itself seems very usable, the criticism to the MARS-tool is poorly founded and there is no validation or testing of the ACDC-tool. Also, no information on the creation of this tool is available. The tool used 24 questions, mainly based on the study of Stoyanov. Only one quarter of the questionnaire addresses the quality of the information, privacy and security, while three quarters assess the design, usability and interactivity of the app.

The study of Yasini (31) should not have been included as the tool itself is unavailable, even on the linked website. It is a very deficient study and a lot of essential information on tool creation and validation is missing.

Jin (33) created a study with a good design, but the tool is based on a limited number of sources and has a poor validation with a strong possibility of bias. As we compare the evaluation criteria the amount of questions is more appropriate, but too extensive on aesthetics and too concise on privacy and security. It is warranted to validate this tool in a new study as it could be a useful tool.

The study of Chyjek (34) is a poorly designed study with unclear tool construction and no validation. Probably some evaluation items will be missed or forgotten. The APPLICATIONS-tool therefore seems to give an incomplete evaluation. It is a small scoring system without a privacy and security part and a quality of information component that is too limited.

Finally, the study of Mc Millan (32) is a well-designed study but has an unusable tool as it can only be specifically applied for behaviour change apps. Moreover, the validation of the tool is weak. The tool consists of an extended list, mainly dealing with behaviour change and only a small component on ease of use, aesthetics, privacy and security. It is questionable whether this scale can be used for other health apps. It should however be considered that behaviour change is an essential purpose of mobile health apps and therefore elements of this tool could be used in other health app evaluation tools.

LIMITATIONS

- A possible limitation of our review may lay in the literature search strategy for evaluation tools for apps. Some useful scales or checklists may have been missed due to our search with only one search string, the way we selected articles for full text reading or due to the search in only 4 databases. We believe that most of the articles using a new evaluation tool would however put one of the following words in their title: “regulation, evaluat*, rating, scale, asses*, quality, criteria or certification”. We based our search string on the keywords of an exploratory literature search. Because of excluding articles dealing with only one disease or problem, articles may have been missed. However, the broad first selection for reading the entire abstract made it possible to find more tools that explicitly announce a new evaluation tool (which is reflected in the titles of the selected studies).

  Furthermore, we performed a second literature search before finishing our review which found one new article. Further research could use the same (or an adapted) search string to find useful articles in the future.
- Another limitation may be the assessment of the articles with the template we used. A search for other templates yielded just a few articles and a perfect template to evaluate all aspects of the studies was not found. Instead we tried to keep the template in the study of Kmet intact, as a basis for our evaluation of the articles and added only the few questions we thought to add extra value. (28)
- This research aimed to find evaluation tools for health and lifestyle apps. Even though we never excluded evaluation tools for not being aimed at these types of apps, we are not sure the tools we analysed can also be used for other types of apps.
- It was the first time for both evaluators to perform a systematic review, which might be a negative point in our study. However, the support of an experienced promotor makes this limitation less relevant.
- Evaluating quality may suffer from reviewers’ bias. Although we tried to minimize this by conducting most of the phases of the review process independently and by discussing any discrepancies until agreement was reached, we cannot fully guarantee the absence of reviewer bias.
- The division between the qualitative and quantitative part of the studies might be artificial, but was needed to evaluate both the tool construction and the testing of the tool separately.
- The references within the retained articles were not searched for other relevant studies (snowballing).

CONCLUSION

We can conclude that there is a big need for evidence-based health app evaluation criteria. The currently available evidence is scarce. The most frequently used tool with a credible evidence base for health app evaluation is the MARS-tool, developed by Stoyanov et al. (20). The ACDC-tool (35) added some interesting aspects but needs further testing and validation. Further research should be done to validate or refine these tools, or to create an additional evaluation tool that will be universally accepted.

There was no funding for this study and no conflicts of interest.

REFERENCES

1. Bijma B. Zorg per smartphone : 10 apps voor patiënt en dokter. LVH De Dokter. 2014 13th of March 2014.
2. Key M, Santos, J., Takane, M. mHealth : New horizons for health through mobile technologies: second global survey on eHealth. WHO Library Cataloguing-in-Publication Data. 2011.
3. Mehl G, Labrique, A. Prioritizing integrated mHealth strategies for universal health coverage. 2016.
4. Labrique A, Vasudevan, L., Kochi, E. Fabricant, R., Mehld, G. mHealth innovations as health system strengthening tools: 12 common applications and a visual framework. 2013.
5. FAGG R. Actiepunt 19: Mobile Health. eGezondheid. 2015.
6. Geudens L. Mobile Health maakt geïntegreerde zorg mogelijk. Huisarts Nu. 2015;44(5):206-10.
7. Van Emelen J, Staelraeve, S., Van Herck, P. mHealth als sleutel tot kwaliteit en betaalbaarheid van zorg. Whitepaper Voka Health Community mobile Health. 2015.
8. Digital Agenda for Europe: mHealth. 2011.
9. Stakeholder meeting on Quality and reliability of mobile health applications. Paper gepresenteerd op: Stakeholder meeting mobile health applications (European Commission). 2015.
10. Lewis TL, Wyatt JC. mHealth and mobile medical Apps: a framework to assess risk and promote safer use. J Med Internet Res. 2014;16(9):e210.
11. IT Glossary 2016. Available from: http://www.gartner.com/it-glossary/big-data/.
12. Misra S. Happtique's recent setback shows that health app certification is a flawed proposition. iMedicalApps. 2014.
13. Sunyaev A, Dehling, T, Taylor, P., et al. Availability of mobile health app privacy policies. J Am Med Inform Assoc. Sunyaev, A., Dehling, T, Taylor, P., et al;22:e28-33.
14. Nicholas J, Larsen ME, Proudfoot J, Christensen H. Mobile Apps for Bipolar Disorder: A Systematic Review of Features and Content Quality. J Med Internet Res. 2015;17(8):e198.
15. Majeed-Ariss R, Baildarm, E., Campbell, M., et al. Apps and adolescents: a systematic review of adolescents' use of mobile phone and tablet apps that support personal management of their chronic of long-term physical conditions. J Med Internet Res. 2015;17(12):e287.
16. Huckvale KC, M., Morrison, C., et al. Apps for asthma self- management: a systematic assessment of content and tools. BMC Medicine. 2012;10:144.
17. Huckvale K, Adomaviciute, S., Prieto, J., et al. Smartphone apps for calculating insuline dose : a systematic assessment. BMC Medicine. 2015;13.
18. Dehling T, Gao F, Schneider S, Sunyaev A. Exploring the Far Side of Mobile Health: Information Security and Privacy of Mobile Health Apps on iOS and Android. JMIR Mhealth Uhealth. 2015;3(1):e8.
19. Becker S, Miron-Shatz T, Schumacher N, Krocza J, Diamantidis C, Albrecht UV. mHealth 2.0: Experiences, Possibilities, and Perspectives. JMIR Mhealth Uhealth. 2014;2(2):e24.
20. Stoyanov SR, Hides L, Kavanagh DJ, Zelenko O, Tjondronegoro D, Mani M. Mobile app rating scale: a new tool for assessing the quality of health mobile apps. JMIR Mhealth Uhealth. 2015;3(1):e27.
21. Powell AC, Torous J, Chan S, Raynor GS, Shwarts E, Shanahan M, et al. Interrater Reliability of mHealth App Rating Measures: Analysis of Top Depression and Smoking Cessation Apps. JMIR Mhealth Uhealth. 2016;4(1):e15.
22. Grundy QH, Wang, Z., Bero, L. A. Challenges in Assessing Mobile Health App Quality. A Systematic Review of Prevalent and Innovative Methods. Am J Prev Med. 2016;51(6):1051-9.
23. Wicks P, Chiauzzi E. 'Trust but verify'--five approaches to ensure safe medical apps. BMC Med. 2015;13:205.
24. Boulos MNK, Brewer, A.C, Karimkhani, C, Buller, D.B, Dellavalle, R.P. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. Online J Public Health Inform. 2014;5(3):229.
25. Boudreaux ED, Waring ME, Hayes RB, Sadasivam RS, Mullen S, Pagoto S. Evaluating and selecting mobile health apps: strategies for healthcare providers and healthcare organizations. Transl Behav Med. 2014;4(4):363-71.

26. Tountas Y, Dimitrakaki, C., Bollars, C., Kok, H., Van den Broucke, S., Molleman, G. European Quality instrument for health promotion : The EQUIHP. Archives of Hellenic Medicine. 2007;24(1):75-91.
27. O'Brien BC, Harris, I. B., Beckman, T. J., Reed, D. A., Cook, D. A. . Standards for reporting qualitative research: a synthesis of recommendations. Acad Med. 2014;89(9):1245-51.
28. Kmet LM, Lee, R. C., Cook, L. S. Standard quality assessment criteria for evaluating primary research papers from a variety of fields. HTA initiative. 2004;13.
29. BinDhim NF, Hawkey A, Trevena L. A systematic review of quality assessment methods for smartphone health apps. Telemed J E Health. 2015;21(2):97-104.
30. Loy JS, Ali EE, Yap KY. Quality Assessment of Medical Apps that Target Medication-Related Problems. J Manag Care Spec Pharm. 2016;22(10):1124-40.
31. Yasini M, Beranger J, Desmarais P, Perez L, Marchand G. mHealth Quality: A Process to Seal the Qualified Mobile Health Apps. Stud Health Technol Inform. 2016;228:205-9.
32. McMillan B, Hickey E, Patel MG, Mitchell C. Quality assessment of a sample of mobile app-based health behavior change interventions using a tool based on the National Institute of Health and Care Excellence behavior change guidance. Patient Educ Couns. 2016;99(3):429-35.
33. Jin M, Kim J. Development and Evaluation of an Evaluation Tool for Healthcare Smartphone Applications. Telemed J E Health. 2015;21(10):831-7.
34. Chyjek K, Farag S, Chen KT. Rating Pregnancy Wheel Applications Using the APPLICATIONS Scoring System. Obstet Gynecol. 2015;125(6):1478-83.
35. Anderson K, Burford O, Emmerton L. App Chronic Disease Checklist: Protocol to Evaluate Mobile Apps for Chronic Disease Self-Management. JMIR Res Protoc. 2016;5(4):e204.

APPENDICES

APPENDIX 1: Construction of the template

The following templates were retrieved in our search for templates, but not used in the construction of the final template after discussion: COREQ; ENTREQ; GRADE; PRISMA; ROBINS-I; Quality indicators for health promotion programs; Cochrane Collaboration's tool for assessing risk of bias in randomized trials.

The following templates were retrieved in our search for templates and were used in the construction of the final template after discussion: Standards for Reporting Qualitative Research (SRQR) (27); Systematic review of quality assessment methods for smartphone health apps (29); Standard Quality Assessment Criteria for evaluating primary research papers from a variety of fields (28); European Quality Instrument for Health Promotion (EQUIHP) (26). This last one is different to the other templates as it analyses the quality of a health promotion initiative rather than analysing an article. This was not compatible with the other templates and therefore only 2 items were adopted to fit in the final template.

The fundament of our final template was the template of Kmet (28). This template consists of 10 criteria for qualitative studies (Q1-10) and of 14 criteria for quantitative studies (K1-14). We made the same division between a qualitative and a quantitative checklist, because of the important differences in analysing these two types of research. In the appliance of the template, many articles discussing quality criteria for apps had both a qualitative and a quantitative part and we divided the template accordingly for the evaluation.

In the qualitative part, we maintained all questions from the Kmet qualitative checklist. Most of the SRQR criteria (27) were deleted because they resembled to questions already included by Kmet. Three criteria from SRQR were added (limitations, conflict of interest and funding). From EQUIHP (26) two items seemed necessary to adapt and fit in (aim of the tool and target population). We classified these two items as a part of the first question of Kmet (Q1). The BinDhim criteria (29) mostly cover questions on the recording of the study rather than on the intrinsic quality of the study. Many criteria leaned close to existing questions in the template and were therefore inserted or added to these questions: The first BinDhim criterion (BH1) was inserted into question Q6; the fifth BinDhim criterion (BH5) was inserted into question Q5. BH7 was inserted into question Q1. BinDhim criterion 6 and 8 were added as separate questions to the final template.

In the quantitative part, we maintained 12 out of 14 questions from the Kmet quantitative checklist. K5 and K7 were eliminated as they were not appropriate. Blinding of the subjects was never used

because the subjects were the apps. Random allocation was unnecessary because of the nature of the articles. The same SRQR and EQUIHP criteria were added as in the qualitative list. Also the BinDhim criteria added were similar to those added in the qualitative list, except for the extra addition of BH2 (Does the study specifically mention that the app was downloaded and analysed based on its content?), which is an interesting question for the quantitative part.

APPENDIX 2: Manual for the template

Manual for scoring of qualitative part of the articles

1. Question / objective clearly described? (28)

Yes: Research question or objective is clear by the end of the research process (if not at the outset).

Partial: Research question or objective is vaguely/incompletely reported.

No: Question or objective is not reported, or is incomprehensible.

+ Aims & objectives (26)

The aims and objectives of the project are clearly defined. The following indicators will help you to assess the adequacy of the aims and objectives.

Is there a written specification of the project aims and objectives? Is there a distinction between short-term (immediate), medium-term (intermediate) and long-term (ultimate) objectives? Is the formulation of the objectives SMART? (Specific, Measurable, Acceptable, Realistic, Time-framed) Are the objectives matched to the analysis of the problem and determinants?

+ Target group (26)

The group of people the project intends to influence is clearly defined. The following indicators will help you to assess the adequacy of the strategy and method: Have the relevant demographic features of the target group been identified? Have the priority needs, wishes and social norms of the target group been identified? Has the target group been segmented with regard to motivational phase and possibilities for change? Has the size of the target group been established (number of subjects)? Is it clear how the target group will be reached?

+ Does the study clearly identify the targeted group of users the app is intended for (consumers or healthcare professionals)? Identifying the intended audience for apps is important to allow for the comparison of app quality within targeted groups. (29)

2. Design evident and appropriate to answer study question? (If the study question is not clearly identified, infer appropriateness from results/conclusions.) (28)

Yes: Design is easily identified. Design is the general description of the type of research and is appropriate to address the study question.

Partial: Design is not clearly identified, but gross inappropriateness is not evident; or design is easily identified but a different method would have been more appropriate.

No: Design used is not appropriate to the study question (e.g. a causal hypothesis is tested using qualitative methods); or design cannot be identified.

3. Context for the study is clear? (28)

Yes: The context/setting is adequately described, permitting the reader to relate the findings to other settings.

Partial: The context/setting is partially described.

No: The context/setting is not described.

4. Connection to a theoretical framework / wider body of knowledge? (28)

Yes: The theoretical framework/wider body of knowledge informing the study and the methods used is sufficiently described and justified.

Partial: The theoretical framework/wider body of knowledge is not well described or justified; link to the study methods is not clear.

No: Theoretical framework/wider body of knowledge is not discussed.

5. Sampling strategy described, relevant and justified? (28)

Yes: The sampling strategy is clearly described and justified. The sample includes the full range of relevant, possible cases/settings (i.e., more than simple convenience sampling), permitting conceptual (rather than statistical) generalizations.

Partial: The sampling strategy is not completely described, or is not fully justified. Or the sample does not include the full range of relevant, possible cases/settings (i.e., includes a convenience sample only).

No: Sampling strategy is not described.

+ Does the study clearly mention the inclusion and exclusion criteria for app selection? Providing an explanation for why some apps were included and others excluded increases the likelihood of reproducible results and further specifies the type of apps included. (29)

6. Data collection methods clearly described and systematic? (28)

Yes: The data collection procedures are systematic, and clearly described, permitting an “audit trail” such that the procedures could be replicated.

Partial: Data collection procedures are not clearly described; difficult to determine if systematic or replicable.

No: Data collection procedures are not described.

+ Has the study clearly mentioned the data collection time frame? New apps are frequently released and updated. The period in which data collection has taken place is important for further research, which may assess the quality of a specific group of health-related apps in relation to time. (29)

7. Data analysis clearly described, complete and systematic? (28)

Yes: Systematic analytic methods are clearly described, permitting an “audit trail” such that the procedures could be replicated. The iteration between the data and the explanations for the data (i.e., the theory) is clear – it is apparent how early, simple classifications evolved into more sophisticated coding structures which then evolved into clearly defined concepts/explanations for the data). Sufficient data is provided to allow the reader to judge whether the interpretation offered is adequately supported by the data.

Partial: Analytic methods are not fully described. Or the iterative link between data and theory is not clear.

No: The analytic methods are not described. Or it is not apparent that a link to theory informs the analysis.

8. Use of verification procedure(s) to establish credibility of the study? (28)

Yes: One or more verification procedures were used to help establish credibility/ trustworthiness of the study (e.g., prolonged engagement in the field, triangulation, peer review or debriefing, negative case analysis, member checks, external audits/interrater reliability, “batch” analysis).

No: Verification procedure(s) not evident.

9. Conclusions supported by the results? (28)

Yes: Sufficient original evidence supports the conclusions. A link to theory informs any claims of generalizability.

Partial: The conclusions are only partly supported by the data. Or claims of generalizability are not supported.

No: The conclusions are not supported by the data. Or conclusions are absent.

10. Limitations. Trustworthiness and limitations of findings. (27)

11. Reflexivity of the account? (28)

Yes: The researcher explicitly assessed the likely impact of their own personal characteristics (such as age, sex and professional status) and the methods used on the data obtained.

Partial: Possible sources of influence on the data obtained were mentioned, but the likely impact of the influence or influences was not discussed.

No: There is no evidence of reflexivity in the study report.

12. Conflicts of interest: Potential sources of influence or perceived influence on study conduct and conclusions; how these were managed. (27)

13. Funding: Sources of funding and other support; role of funders in data collection, interpretation, and reporting. (27)

14. Does the study identify the country in which the search was conducted? (29)
Some app stores such as the Apple app store and Google Play allow developers to restrict the purchasing of apps to specific countries. Therefore, the search results may vary as they are country-dependent.

15. Does the study provide a list of the apps included in the review? (29)

Providing the final apps included in the study will help future researchers to monitor the progress and quality of apps that may have been updated over time.

Manual for scoring of quantitative part of the articles

1. Question or objective sufficiently described? (28)

Yes: Is easily identified in the introductory section (or first paragraph of methods section). Specifies (where applicable, depending on study design) all of the following: purpose, subjects/target population, and the specific intervention(s) /association(s)/descriptive parameter(s) under investigation. A study purpose that only becomes apparent after studying other parts of the paper is not considered sufficiently described.

Partial: Vaguely/incompletely reported (e.g. “describe the effect of” or “examine the role of” or “assess opinion on many issues” or “explore the general attitudes”...); or some information has to be gathered from parts of the paper other than the introduction/background/objective section.

No: Question or objective is not reported, or is incomprehensible.

N/A: Should not be checked for this question.

+ Aims & objectives (26)

The aims and objectives of the project are clearly defined. The following indicators will help you to assess the adequacy of the aims and objectives.

Is there a written specification of the project aims and objectives? Is there a distinction between short-term (immediate), medium-term (intermediate) and long-term (ultimate) objectives? Is the formulation of the objectives SMART? (Specific, Measurable, Acceptable, Realistic, Time-framed) Are the objectives matched to the analysis of the problem and determinants?

+ Target group (26)

The group of people the project intends to influence is clearly defined. The following indicators will help you to assess the adequacy of the strategy and method:

Have the relevant demographic features of the target group been identified? Have the priority needs, wishes and social norms of the target group been identified? Has the target group been segmented with regard to motivational phase and possibilities for change? Has the size of the target group been established (number of subjects)? Is it clear how the target group will be reached?

+ Does the study clearly identify the targeted group of users the app is intended for (consumers or healthcare professionals)? Identifying the intended audience for apps is important to allow for the comparison of app quality within targeted groups. (29)

2. Design evident and appropriate to answer study question? (If the study question is not given, infer from the conclusions). (28)

Yes: Design is easily identified and is appropriate to address the study question / objective.

Partial: Design and /or study question not clearly identified, but gross inappropriateness is not evident; or design is easily identified but only partially addresses the study question.

No: Design used does not answer study question (e.g., a comparison group is required to answer the study question, but none was used); or design cannot be identified.

N/A: Should not be checked for this question.

3. Method of subject selection (and comparison group selection, if applicable) or source of information/input variables (e.g., for decision analysis) is described and appropriate. (28)

Yes: Described and appropriate. Selection strategy designed (i.e., consider sampling frame and strategy) to obtain an unbiased sample of the relevant target population or the entire target population of interest (e.g., consecutive patients for clinical trials, population-based random sample for case-control studies or surveys). Where applicable, inclusion/exclusion criteria are described and defined (e.g., “cancer” -- ICD code or equivalent should be provided). Studies of volunteers: methods and setting of recruitment reported. Surveys: sampling frame/ strategy clearly described and appropriate.

Partial: Selection methods (and inclusion/exclusion criteria, where applicable) are not completely described, but no obvious inappropriateness. Or selection strategy is not ideal (i.e., likely introduced bias) but did not likely seriously distort the results (e.g., telephone survey sampled from listed phone numbers only; hospital based case-control study identified all cases admitted during the study period, but recruited controls admitted during the day/evening only). Any study describing participants only as “volunteers” or “healthy volunteers”. Surveys: target population mentioned but sampling strategy unclear.

No: No information provided. Or obviously inappropriate selection procedures (e.g., inappropriate comparison group if intervention in women is compared to intervention in men). Or presence of selection bias which likely seriously distorted the results (e.g., obvious selection on “exposure” in a case-control study).

N/A: Descriptive case series/reports.

+ Has the study clearly mentioned the data collection time frame? New apps are frequently released and updated. The period in which data collection has taken place is important for further research, which may assess the quality of a specific group of health-related apps in relation to time. (29)

+ Does the study clearly mention the inclusion and exclusion criteria for app selection? Providing an explanation for why some apps were included and others excluded increases the likelihood of reproducible results and further specifies the type of apps included. (29)

4. Subject (and comparison group, if applicable) characteristics or input variables/information (e.g., for decision analyses) sufficiently described? (28)

Yes: Sufficient relevant baseline/demographic information clearly characterizing the participants is provided (or reference to previously published baseline data is provided). Where applicable, reproducible criteria used to describe/categorize the participants are clearly defined (e.g., ever-smokers, depression scores, systolic blood pressure > 140). If “healthy volunteers” are used, age and sex must be reported (at minimum). Decision analyses: baseline estimates for input variables are clearly specified.

Partial: Poorly defined criteria (e.g. “hypertension”, “healthy volunteers”, “and smoking”). Or incomplete relevant baseline / demographic information (e.g., information on likely confounders not reported). Decision analyses: incomplete reporting of baseline estimates for input variables.

No: No baseline / demographic information provided. Decision analyses: baseline estimates of input variables not given.

N/A: Should not be checked for this question.

5. If interventional and blinding of investigators to intervention was possible, is it reported?

Yes: Blinding reported. (28)

Partial: Blinding reported but it is not clear who was blinded.

No: Blinding would have been possible (and was possibly done) but is not reported.

N/A: Observational analytic studies. Uncontrolled experimental studies. Surveys. Descriptive case series / reports. Decision analyses.

6. Outcome and (if applicable) exposure measure(s) well defined and robust to measurement / misclassification bias? Means of assessment reported? (28)

Yes: Defined (or reference to complete definitions is provided) and measured according to reproducible, "objective" criteria (e.g., death, test completion – yes/no, clinical scores). Little or minimal potential for measurement / misclassification errors. Surveys: clear description (or reference to clear description) of questionnaire/interview content and response options. Decision analyses: sources of uncertainty are defined for all input variables.

Partial: Definition of measures leaves room for subjectivity, or not sure (i.e., not reported in detail, but probably acceptable). Or precise definition(s) are missing, but no evidence or problems in the paper that would lead one to assume major problems. Or instrument/mode of assessment(s) not reported. Or misclassification errors may have occurred, but they did not likely seriously distort the results (e.g., slight difficulty with recall of long-ago events; exposure is measured only at baseline in a long cohort study). Surveys: description of questionnaire/interview content incomplete; response options unclear. Decision analyses: sources of uncertainty are defined only for some input variables.

No: Measures not defined, or are inconsistent throughout the paper. Or measures employ only ill-defined, subjective assessments, e.g. "anxiety" or "pain." Or obvious misclassification errors/measurement bias likely seriously distorted the results (e.g., a prospective cohort relies on self-reported outcomes among the "unexposed" but requires clinical assessment of the "exposed"). Surveys: no description of questionnaire/interview content or response options. Decision analyses: sources of uncertainty are not defined for input variables.

N/A: Descriptive case series / reports.

7. Sample size appropriate? (28)

Yes: Seems reasonable with respect to the outcome under study and the study design. When statistically significant results are achieved for major outcomes, appropriate sample size can usually

be assumed, unless large standard errors (SE > 1⁄2 effect size) and/or problems with multiple testing are evident. Decision analyses: size of modelled cohort / number of iterations specified and justified.

Partial: Insufficient data to assess sample size (e.g., sample seems "small" and there is no mention of power/sample size/effect size of interest and/or variance estimates aren't provided). Or some statistically significant results with standard errors > 1⁄2 effect size (i.e., imprecise results). Or some statistically significant results in the absence of variance estimates. Decision analyses: incomplete description or justification of size of modelled cohort / number of iterations.

No: Obviously inadequate (e.g., statistically non-significant results and standard errors > 1⁄2 effect size; or standard deviations > _ of effect size; or statistically non-significant results with no variance estimates and obviously inadequate sample size). Decision analyses: size of modelled cohort / number of iterations not specified.

N/A: Most surveys (except surveys comparing responses between groups or change over time). Descriptive case series / reports.

8. Analysis described and appropriate? (28)

Yes: Analytic methods are described (e.g. "chi square"/ "t-tests"/"Kaplan-Meier with log rank tests", etc.) and appropriate.

Partial: Analytic methods are not reported and must be guessed at, but are probably appropriate. Or minor flaws or some tests appropriate, some not (e.g., parametric tests used, but unsure whether appropriate; control group exists but is not used for statistical analysis). Or multiple testing problems not addressed.

No: Analysis methods not described and cannot be determined. Or obviously inappropriate analysis methods (e.g., chi-square tests for continuous data, SE given where normality is highly unlikely, etc.). Or a study with a descriptive goal / objective is over-analysed.

N/A: Descriptive case series / reports.

9. Some estimate of variance (e.g., confidence intervals, standard errors) is reported for the main results/outcomes (i.e., those directly addressing the study question/ objective upon which the conclusions are based)?

Yes: Appropriate variances estimate(s) is/are provided (e.g., range, distribution, confidence intervals, etc.). Decision analyses: sensitivity analysis includes all variables in the model.

Partial: Undefined "+/-" expressions. Or no specific data given, but insufficient power acknowledged as a problem. Or variance estimates not provided for all main results/outcomes. Or inappropriate variance estimates (e.g., a study examining change over time provides a variance around the parameter of interest at "time 1" or "time 2", but does not provide an estimate of the variance around the difference). Decision analyses: sensitivity analysis is limited, including only some variables in the model.

No: No information regarding uncertainty of the estimates. Decision analyses: No sensitivity analysis.

N/A: Descriptive case series / reports. Descriptive surveys collecting information using open-ended questions.

10. Controlled for confounding? (28)

Yes: Randomized study, with comparability of baseline characteristics reported (or non-comparability controlled for in the analysis). Or appropriate control at the design or analysis stage (e.g., matching, subgroup analysis, multivariate models, etc.). Decision analyses: dependencies between variables fully accounted for (e.g., joint variables are considered).

Partial: Incomplete control of confounding. Or control of confounding reportedly done but not completely described. Or randomized study without report of comparability of baseline characteristics. Or confounding not considered, but not likely to have seriously distorted the results. Decision analyses: incomplete consideration of dependencies between variables.

No: Confounding not considered, and may have seriously distorted the results. Decision analyses: dependencies between variables not considered.

N/A: Cross-sectional surveys of a single group (i.e., surveys examining change over time or surveys comparing different groups should address the potential for confounding). Descriptive studies. Studies explicitly stating the analysis is strictly descriptive/exploratory in nature.

11. Conflicts of interest: Potential sources of influence or perceived influence on study conduct and conclusions; how these were managed. (27)

12. Funding: Sources of funding and other support; role of funders in data collection, interpretation, and reporting. (27)

13. Results reported in sufficient detail? (28)

Yes: Results include major outcomes and all mentioned secondary outcomes.

Partial: Quantitative results reported only for some outcomes. Or difficult to assess as study question/objective not fully described (and is not made clear in the methods section), but results seem appropriate.

No: Quantitative results are reported for a subsample only, or "n" changes continually across the denominator (e.g., reported proportions do not account for the entire study sample, but are reported only for those with complete data - i.e., the category of "unknown" is not used where needed). Or results for some major or mentioned secondary outcomes are only qualitatively reported when quantitative reporting would have been possible (e.g., results include vague comments such as "more likely" without quantitative report of actual numbers).

N/A: Should not be checked for this question.

14. Do the results support the conclusions? (28)

Yes: All the conclusions are supported by the data (even if analysis was inappropriate). Conclusions are based on all results relevant to the study question, negative as well as positive ones (e.g., they aren't based on the sole significant finding while ignoring the negative results). Part of the conclusions may expand beyond the results, if made in addition to rather than instead of those strictly supported by data, and if including indicators of their interpretative nature (e.g., "suggesting," "possibly").

Partial: Some of the major conclusions are supported by the data, some are not. Or speculative interpretations are not indicated as such. Or low (or unreported) response rates call into question the validity of generalizing the results to the target population of interest (i.e., the population defined by the sampling frame/strategy).

No: None or a very small minority of the major conclusions are supported by the data. Or negative findings clearly due to low power are reported as definitive evidence against the alternate hypothesis. Or conclusions are missing. Or extremely low response rates invalidate generalizing the results to the target population of interest (i.e., the population defined by the sampling frame/ strategy).

N/A: Should not be checked for this question.

15. Limitations. Trustworthiness and limitations of findings. (27)

16. Does the study specifically mention that the reviewed app was downloaded and analysed based on its content? (29)

Reviewing the app description may not be enough to make a full judgment about the app quality. Reporting how apps were initially reviewed is necessary to understand the depth of analysis. (If it is not mentioned explicitly or only a selected sample is fully downloaded, this criterion will not be met.)

17. Does the study identify the country in which the search was conducted? (29)

Some app stores such as the Apple app store and Google Play allow developers to restrict the purchasing of apps to specific countries. Therefore, the search results may vary as they are country-dependent.

18. Does the study provide a list of the apps included in the review? (29)

Providing the final apps included in the study will help future researchers to monitor the progress and quality of apps that may have been updated over time.

APPENDIX 3: Suggestion

It is unfeasible to evaluate all existing health and lifestyle apps and every update with any evaluation tool. Private organizations and governmental initiatives have failed to certify reliable apps. (12, 31) The different aspects of health apps and the fact that not many people are skilled in all facets will be important hurdles when evaluating apps. User ratings are currently one of the only evaluation parameter an end user gets when downloading from the app store. It is a very robust score, but has the advantage of the mass effect: the more people evaluating the app, the more the rating seems to be trustable (if bots can be eliminate). We suggest a slightly different approach, still using the mass effect to increase trust in certain apps.

The core of our suggestion is the inclusion of different stakeholders focusing on their specific skills in evaluating apps. The result will be presented to the public to make an informed choice about the used apps. A platform with a grid can be used where all aspects of the suggested apps are discussed. This should preferably be controlled by governmental institutions per country.

User ratings and popularity of the apps (popularity index) could be the guidance to start evaluation. End users could also make suggestions to the institution (e.g. about one specific problem that remains underexposed) and app developers should make sure all the information needed is available. End users are in the best position to evaluate the ease of use of use, interested doctors should evaluate the medical content, a team of privacy experts should assess the privacy and security features and the government should check whether the app has all the required certificates (FDA approval or CE-mark). The grid could give information which evaluations have already been done and how many persons evaluated each item.

By doing so, the entire community is activated and the vast weight of this evaluation task can be carried out by different stakeholders at once. Each country should work for its own specific context but it is likely that the same app will be evaluated similarly in similar countries. This could build international support for positively evaluated apps.

This approach could come at a price (paying the platform supervisors, the certificate checkers, the privacy team and potentially a compensation for the doctors), but many believe huge savings could be made once trusted mHealth is fully operational.

APPENDIX 4: Beschrijving van de werkverdeling

In het verkennende literatuuronderzoek werden verschillende aandachtspunten gevestigd door de auteurs. Birgit zocht informatie over de Belgische en Europese context terwijl ik de Amerikaanse context onderzocht en al ging zoeken naar bestaande kwaliteitscriteria.

De beslissing om het verdere onderzoek toe te spitsen op kwaliteitscriteria van apps werd door ons beiden gedragen en werd mee gevoed door de promotor.

Het systematische literatuuronderzoek werd door beide auteurs onafhankelijk van elkaar uitgevoerd. Achteraf bleek dat we een verschillende aanpak hadden gebruikt. Nadat de uiteindelijke artikels geselecteerd waren, vergeleek Birgit de tools uit de studies terwijl ik de kwaliteit van de gebruikte studies bekeek. Verschillende consensusmeetings werden met de twee auteurs gehouden om elke stap voldoende op elkaar af te stemmen. Regelmatig werd ook de promotor hierbij betrokken.

De discussie en beperkingen van de studie werden samen geschreven en ik voegde er nog een suggestie aan toe. Hiermee tracht ik een voorstel te doen waarin het beste van alle voorgaande ideeën om apps te beoordelen geclusterd zit. Het massa-effect blijft behouden voor de items waarvoor dit nuttig kan zijn en iedereen focust zich op dat onderdeel dat hij als stakeholder best kan beoordelen.

De samenwerking met promotor en coauteur verliep uitermate vlot van het begin tot op het einde.

APPENDIX 5: Advies Ethisch Comité (OBC)

Gunstig advies van OBC

Bekijk status geschiedenis

2016-06-13 17:37:53

Student(en):

Wouter Billiet

Birgit Vanden Bûssche

Promotor:

Marc Van Nuland (marc.vannuland@med.kuleuven.be - u0040434 - Acad. Centr. voor Huisartsgeneeskunde)

Faculteit:

Geneeskunde

Opleiding:

Master in de huisartsgeneeskunde (Leuven e.a.)

Het onderzoek is:

multicentrisch

De opdrachtgever is:

academisch (universiteit,...)

Het onderzoek is:

Het onderzoek is (gedeeltelijk) prospectief.

Het onderzoek is:

Prospectief interventioneel onderzoek

Gebruikte methode:

zonder geneesmiddelen

BIJKOMENDE GEGEVENS:

Achtergrond:

De laatste jaren wordt er steeds meer verwacht van m-Health als katalysator voor een efficiëntere en goedkopere gezondheidzorg. (1) M-Health is een onderdeel van e-Health (wat op zich een verzamelterm is voor elektronische overdracht van gezondheidszorggegevens) en staat voor het gebruik van mobiele apparaten om de medische praktijk en in ruimere zin de volksgezondheid te ondersteunen. (2) Voorbeelden van m-Health zijn: elektronische gezondheidsdata, apps voor zorgverleners (voor bv hulp bij diagnostiek), apps voor patiënten (voorlichtingsapps, levensstijlapps, ziektespecifieke apps) of telegeneeskunde (zonder fysische aanwezigheid van arts en patiënt op dezelfde plaats). (3) E-Health zou zorgen voor meer beschikbaarheid en toegankelijkheid van gegevens en kan meer innovatie en kosten-effectiviteit bewerkstelligen (2). Deze nieuwe technieken evolueren zeer snel en kunnen aldus een belangrijke rol spelen in een meer kwaliteitsvolle gezondheidszorg waarbij de patiënten zelf het initiatief in handen kunnen nemen (patient empowerment). (4)
Er bestaan reeds erg veel apps voor gezondheidsthema's en deze aantallen groeien exponentieel, maar de kwaliteit van deze apps is erg variabel. Het is dan ook interessant om die apps te kunnen selecteren die voldoende aandacht schenken aan evidentie, veiligheid en privacy. (4,5) Enkele ondernemingen hebben geprobeerd om via een certificatie-programma geld te slaan uit de kwaliteitsbeoordeling van apps, (6,7) maar gezien de economische drijfveren zijn deze programma's niet altijd even betrouwbaar of even clear-cut. (8) Om dit probleem te kunnen omzeilen werden ook overheden ingeschakeld om een kwaliteitscontrole voor m-Health uit te voeren, maar ook dit loopt niet van een leien dakje gezien dit erg arbeidsintensief werk is. (3,6,7)
Voor onze masterproef zullen we het thema apps in de huisartspraktijk uitdiepen en zullen we trachten te kijken naar wat de belangrijkste criteria zijn voor kwaliteitsvolle m-Health apps. Daarna zullen wij deze criteria toepassen op 2 geselecteerde apps. De app 'Slaap Lekker' werkt rond slaapstoornissen en is gemaakt door VGZ, de app 'Versterk je enkel' is gemaakt door VeiligheidNL. We ontvingen van beide producenten goedkeuring voor gebruik van hun apps in onze studie.

Vraagstelling:

- Vraag Birgit: Wat zijn de voor- en nadelen, de uitdagingen en mogelijkheden van gezondheidsapps in de huisartsenpraktijk? Bieden dergelijke apps een duidelijke meerwaarde voor de gezondheidszorg?
- Vraag Wouter: Bestaan er reeds kwaliteitscriteria en zijn deze relevant voor arts en patiënt bij gebruik van apps in de huisartspraktijk? Vallen gezondheidsapps wel op deze manier te kwalificeren?
- Vraag voor beiden: Hoe verloopt de toepassing van deze nieuwe vragenlijst in de praktijk? Dit wordt door elk afzonderlijk toegepast via een eigen studie met een eigen app.

Methodologie:

Vooreerst wordt er een uitgebreide literatuurstudie verricht waarin er door Birgit op zoek gegaan wordt naar de uitdagingen en mogelijkheden, voor- en nadelen van en een ethisch-juridisch kader voor gezondheidsapplicaties. Terzelfdertijd zal Wouter de wetenschappelijke literatuur en aanbevelingen vanuit onder meer West-Europese en Noord-Amerikaanse beleidsinstanties bestuderen om de basis voor de kwaliteitscriteria op te staven en deze te vergelijken met commerciële initiatieven rond kwaliteitscriteria. Mogelijke oplossingen worden dan door beiden gesuggereerd rond kwaliteitscriteria, regularisatie en appwinkels. Na de samenstelling van een lijst met de belangrijkste criteria wordt deze door allebei toegepast, zowel op de 'Slaap Lekker' en 'Versterk je enkel' app en dit gebeurt blind van elkaar. Tijdens een consensusvergadering met de promotor en co-promotor worden beide evaluaties van deze apps vergeleken en besproken. Hieruit worden de moeilijkheden en mogelijkheden, voor- en nadelen voor de arts gegenereerd. Uiteindelijk wordt de 'Versterk je enkel' app getest door Birgit in een solopraktijk en de 'Slaap Lekker' door Wouter in een duopraktijk met een pilootstudie, dit om de haalbaarheid en bruikbaarheid van deze kwaliteitscriteria en de relevantie voor patiënt te onderzoeken bij tweemaal een dertigtal personen. Via een affiche in de wachtzaal en persoonlijk contact bij een gerelateerd probleem worden de deelnemers gerecruteerd. Een korte vragenlijst wordt voor de start van het gebruik aan de patiënten voorgelegd, dit gebeurt na het lezen en ondertekenen van de informed consent. De app wordt één maand getest door telkens een dertigtal patiënten. Zij worden na één maand opnieuw gecontacteerd om een tweede uitgebreidere vragenlijst in te vullen. Hier wordt de nadruk gelegd op de volgens ons belangrijkste criteria. Na analyse van de gegevens worden de haalbaarheid, bruikbaarheid en relevantie van de criteria voor de patiënt besproken en neergeschreven.

Referenties:

(1) Geudens, L. Mobile Health maakt geïntegreerde zorg mogelijk. Huisarts Nu. Oktober 2015. 44 (5) : 206-210.
(2) mHealth: New horizons for health through mobile technologies: second global survey on eHealth. 2011. URL : www.who.int.
(3) Boulos, M.N.K, Brewer, A.C, Karimkhani, C, Buller, D.B, Dellavalle, R.P. Mobile medical and health apps: state of the art, concerns, regulatory control and certification. Online J Public Health Inform. 2014; 5(3): 229.
(4) Digital Agenda for Europe: mHealth. November 2011 URL: https://ec.europa.eu/digital-agenda/en/mhealth.
(5) Buijink, A.W.G., Visser, B.J., Marshall L. Medical apps for smartphones: lack of evidence undermines quality and safety. Eivdence-Based Medicine. 2013; 18 (3): 90-92.
(6) Stakeholder meeting on Quality and reliability of mobile health applications. Paper gepresenteerd op: Stakeholder meeting mobile health applications (European Commission) 6 Juli 2015; Brussel. URL: file:///C:/Users/gebruiker/Downloads/Summaryofthemeeting.pdf
(7) Misra, S. Happtique's recent setback shows that health app certification is a flawed proposition [Internet]. Baltimore: iMedicalApps; 8 Januari 2014. URL: http://www.imedicalapps.com/2014/01/happtiques-setback-future-app-certification/
(8) IT Glossary [Internet]. Gartner; 2016. URL: http://www.gartner.com/it-glossary/big-data/

89246_1456732326_6301_affichewachtzaalenkel_pdf.pdf
20309_1456732556_6974_informed_consent_pdf.pdf
73227_1456732556_8688_inleidingsbrief_pdf.pdf
90542_1456732557_084_vragenlijst_slaaplekker_pdf.pdf
58542_1456732557_298_vragenlijst_enkel_pdf.pdf
21456_1461686291_8807_affichewachtzaalslaaplekker_pdf.pdf