Challenges of Trustable AI and Added-Value on Health
B. Séroussi et al. (Eds.)

doi:10.3233/SHTI220620

Key Elements for the Evaluation of mHealth Applications: Results from a Delphi Survey

Felix HOLL[a,b,1] and Walter SWOBODA[a]
[a] *DigiHealth Institute, Neu-Ulm University of Applied Sciences, Neu-Ulm, Germany*
[b] *IBE, Ludwig Maximilian University of Munich, Munich, Germany*

Abstract. Mobile Health (mHealth) applications have seen strong growth in recent years, but they are often not systematically evaluated. A Delphi survey was conducted to identify key elements for the evaluation of mHealth applications. Sixteen experts participated in the study, and the study yielded a list of 79 key elements with expert consensus. Thirty-two elements were in the category of structure quality, 29 in process quality, and 18 in outcome quality. The number of key elements highlights the complexity of conducting systematic evaluations of mHealth applications.

Keywords. Mobile Health, Evaluation, Delphi Survey

1. Introduction

Mobile health (mHealth) applications have experienced strong growth in recent years because of their potential to improve patient care [1]. New policies that introduce the reimbursement for mHealth application through publicly financed healthcare systems, such as in Germany or France, are increasing the availability of mHealth applications [2,3]. However, the number of scientific studies to prove this hypothesis is still too small [4]. There is also a lack of a holistic, standardized, and comprehensive evaluation system, which would increase the validity of studies [5].

Several evaluation tools for medical informatics projects and in general and mobile health applications have been developed and implemented in recent years. Some of these have been published in textbooks, and some have been used in reviews [6–9]. The Mobile App Rating Scale, validated and translated in multiple languages, is a widely used academic evaluation tool for mHealth applications [10–12]. Recommendations for reporting evaluation results are made by STARE-HI [13]. However, despite the breadth of evaluation tools available, most evaluations that are performed are not systematical and only focus on single aspects. As a result, the new evaluation tools for eHealth and mHealth should be developed [14,15].

The goal of the study is to define key elements for the evaluation of mHealth applications through expert consensus as a basis for the development of evaluation tools.

[1] Corresponding Author, Felix Holl, DigiHealth Institute, Neu-Ulm University of Applied Sciences, 89231 Neu-Ulm, Germany; E-mail: felix.holl@hnu.de.

2. Methods

We conducted a web-based Delphi survey from April 08th, 2021, until June 20th, 2021. The survey consisted of three rounds (14 days each) and a break of 2 weeks in-between rounds that were used for data analysis and preparation of the subsequent stages. Participants were recruited through a mailing list of the American Medical Informatics Association and a post in a professional social network. Participation was voluntary, and informed consent for participation and data storage was obtained. The responses of the participants were recorded anonymously. E-mail addresses collected to disseminate the invitations for the subsequent survey rounds were stored in a separate database. Participants had the opportunity to withdraw their consent and stop their participation at any point in the study.

2.1. Eligibility criteria

Subject Matter Experts (SME): were eligible to participate if they met the following requirements:

- Five years or more experience in informatics or medicine or health sciences
- Experience in health IT with evaluation methods or having conducted an evaluation before
- Age 18 years or older, able to read and write English, and access to a computer with internet access.

2.2. Study setup and instruments

The first round of the survey was an open idea generation phase using the leading research question "What are important aspects when evaluating mHealth applications?" The idea generation was followed by two rounds of consensus building. (Visualized in **Figure 1**).

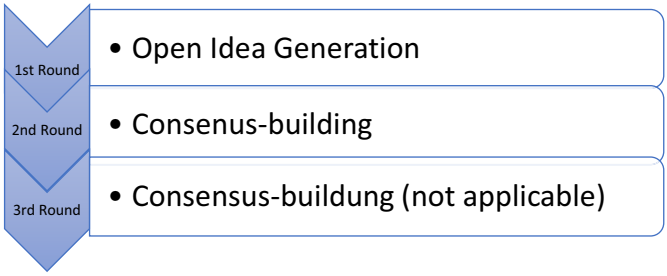

Figure 1. Overview of the study setup

In the open idea generation phase (first round), SMEs were asked to list their ideas on important indicators for the evaluation of mobile health applications that need to be included in a list. The survey consisted of free texts fields, where participants could enter their responses. In addition, demographic information of the participants was collected.

The indicators entered by the SMEs were be grouped based on the Donabedian model of quality (structure, process, and outcome), and duplicates removed before the consensus-building rounds [16]. The Donabedian model of quality was used as a theoretical basis to structure the results of the survey and improve the organization of the

results with the lens of quality of care and categorize elements of app quality according to their influence on quality of care. Following the idea generation phase, two rounds of consensus building took place in which the participants were asked to rate each of the indicators from the idea generation phase using a 4-point Likert scale (Strongly agree, agree, disagree, strongly disagree) whether they should be included in the list of key elements. In addition to the rating, participants could suggest additional elements or rewording of elements. Consensus was considered if at least 75% of the participants rated an element with "agree" or "strongly agree". Elements with less than 75% agreement were dropped. Studies have found that 75% is a common level of agreement in Delphi surveys [17]. Because no additional elements were suggested in the first consensus-building cycle, the study was stopped after the 2nd round. Data cleaning was done by one author and validated by a second author. Data analysis was done using Microsoft Excel Version 16.

3. Results

3.1. Expert demographics

A total of 16 experts participated in the web-based Delphi survey. All participants met the experience criteria for participating in the study. All participants fulfilled the eligibility criteria of at least five years of professional experience. Participants came from several different professional backgrounds.

- User Adoption
- Epidemiology & Public Health
- Public health
- Medical informatics
- Health information management
- Law; Health technology regulation

Three participants had 5 to 7 years professional experience, one 8 to 10, four 11 to 15, four 16 to 20 and another four more than 20 years.

3.2. Results of the open idea generation phase (Round 1)

The idea generation in round 1 yielded a total of 82 elements after the removal of duplicates. Of the 82 elements, 34 were categorized as structure quality, 30 as process quality, and 18 as outcome quality.

3.3. Results of the consensus-building

The first cycle of consensus-building (round 2) resulted in 79 key elements that achieved at least 75% consensus (rated with "agree" or "strongly agree" on the 4-point Likert scale). These 79 elements were divided into 32 elements in the category of structure quality, 29 in the category of process quality, and 18 in the category of outcome quality. Three elements from the open idea generation phase did not achieve consensus. As mentioned, no additional elements were suggested in the second round, and the Delphi survey was stopped after the second round.

Table 1 give an overview of the number of key elements in each quality category and subcategories. The complete list of key elements can be accessed here [18].

Table 1. Structure quality

Structure Quality (n=32)
Data Quality and Interoperability (n=4)
Privacy (n=5)
Security (n=7)
Funding/Cost (n=3)
Access (n=3)
Functionality (n=9)
Certification (n=1)
Process Quality (n=29)
UX (n=9)
Usability (n=3)
Quality improvement (n=4)
Content (n=7)
Usage (n=6)
Outcome Quality (n=18)
Health outcomes (n=9)
Economic outcomes (n=2)
Care process outcomes (n=1)
Patient-reported outcomes (n=5)
Other outcomes (n=1)

4. Discussion

Seventy-nine key elements achieved expert consensus in the Delphi survey. The elements are spread out across all three of the Donabedian model of quality. Structure quality has the most elements while outcome quality has the least, as the outcome quality can be assessed with few indicators. In contrast, the structure quality is determined by several elements. The usual usability studies on mHealth apps [19] should be followed in the future by content studies based on key elements such as those shown here.

5. Conclusions

The high number of elements highlights the complexity of conducting holistic, standardized, and comprehensive evaluations of mHealth applications. The results from the Delphi survey can serve as a basis for the development of a comprehensive evaluation framework for mHealth applications. The next step is to search for validated measurement tools for each key element and create new tools for those where no validated ones exist. Afterward, the framework will be developed and piloted.

References

[1] Steinhubl SR, Muse ED, Topol EJ. The emerging field of mobile health. Sci Transl Med. 2015 Apr 15;7(283):283rv3. doi: 10.1126/scitranslmed.aaa3487. PMID: 25877894; PMCID: PMC4748838.

[2] Gerke S, Stern AD, Minssen T. Germany's digital health reforms in the COVID-19 era: lessons and opportunities for other countries. NPJ Digit Med. 2020 Jul 10;3:94. doi: 10.1038/s41746-020-0306-7. PMID: 32685700; PMCID: PMC7351985.

[3] Yaghobian S, Ohannessian R, Duong TA, Medeiros de Bustos E, Le Douarin YM, Moulin T. France extends its tele-expertise funding model nationally after COVID-19. J Telemed Telecare. 2022 Apr;28(3):233-235. doi: 10.1177/1357633X211067067. Epub 2021 Dec 21. PMID: 34931877.
[4] Rowland SP, Fitzgerald JE, Holme T, Powell J, McGregor A. What is the clinical value of mHealth for patients? NPJ Digit Med. 2020 Jan 13;3:4. doi: 10.1038/s41746-019-0206-x. PMID: 31970289; PMCID: PMC6957674.
[5] Holl F, Kircher J, Swoboda WJ, Schobel J. Methods Used to Evaluate mHealth Applications for Cardiovascular Disease: A Quasi-Systematic Scoping Review. Int J Environ Res Public Health. 2021 Nov 23;18(23):12315. doi: 10.3390/ijerph182312315. PMID: 34886039; PMCID: PMC8656469.
[6] Friedman CP, Wyatt J. Evaluation methods in biomedical informatics. Springer Nature; 2022.
[7] Brender J. Handbook of evaluation methods for health informatics. London: Elsevier Academic Press; 2006.
[8] Eslami Andargoli A, Scheepers H, Rajendran D, Sohal A. Health information systems evaluation frameworks: A systematic review. Int J Med Inform. 2017 Jan;97:195-209. doi: 10.1016/j.ijmedinf.2016.10.008. Epub 2016 Oct 15. PMID: 27919378.
[9] Fanta G, Pretorius L, Erasmus L. An evaluation of ehealth systems implementation frameworks for sustainability in resource constrained environments: A literature review. InInternational Association for Management of Technology (IAMOT) 2015 Conference Proceedings 2015 Jun 8 (pp. 1046-1063).
[10] 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 Mar 11;3(1):e27. doi: 10.2196/mhealth.3422. PMID: 25760773; PMCID: PMC4376132.
[11] Terhorst Y, Philippi P, Sander LB, Schultchen D, Paganini S, Bardus M, Santo K, Knitza J, Machado GC, Schoeppe S, Bauereiß N, Portenhauser A, Domhardt M, Walter B, Krusche M, Baumeister H, Messner EM. Validation of the Mobile Application Rating Scale (MARS). PLoS One. 2020 Nov 2;15(11):e0241480. doi: 10.1371/journal.pone.0241480. PMID: 33137123; PMCID: PMC7605637.
[12] Knitza J, Tascilar K, Messner EM, Meyer M, Vossen D, Pulla A, Bosch P, Kittler J, Kleyer A, Sewerin P, Mucke J, Haase I, Simon D, Krusche M. German Mobile Apps in Rheumatology: Review and Analysis Using the Mobile Application Rating Scale (MARS). JMIR Mhealth Uhealth. 2019 Aug 5;7(8):e14991. doi: 10.2196/14991. PMID: 31381501; PMCID: PMC6699116.
[13] Talmon J, Ammenwerth E, Brender J, de Keizer N, Nykänen P, Rigby M. STARE-HI--Statement on reporting of evaluation studies in Health Informatics. Int J Med Inform. 2009 Jan;78(1):1-9. doi: 10.1016/j.ijmedinf.2008.09.002. Epub 2008 Oct 18. PMID: 18930696.
[14] Catwell L, Sheikh A. Evaluating eHealth interventions: the need for continuous systemic evaluation. PLoS Med. 2009 Aug;6(8):e1000126. doi: 10.1371/journal.pmed.1000126. Epub 2009 Aug 18. PMID: 19688038; PMCID: PMC2719100.
[15] Enam A, Torres-Bonilla J, Eriksson H. Evidence-Based Evaluation of eHealth Interventions: Systematic Literature Review. J Med Internet Res. 2018 Nov 23;20(11):e10971. doi: 10.2196/10971. PMID: 30470678; PMCID: PMC6286426.
[16] Donabedian A. The quality of care. How can it be assessed? JAMA. 1988 Sep 23-30;260(12):1743-8. doi: 10.1001/jama.260.12.1743. PMID: 3045356.
[17] Foth T, Efstathiou N, Vanderspank-Wright B, Ufholz LA, Dütthorn N, Zimansky M, Humphrey-Murto S. The use of Delphi and Nominal Group Technique in nursing education: A review. Int J Nurs Stud. 2016 Aug;60:112-20. doi: 10.1016/j.ijnurstu.2016.04.015. Epub 2016 May 1. PMID: 27297373.
[18] Holl F. Key Elements for the Evaluation of mHealth Applications, (2022). doi:doi/10.7910/DVN/D5XR0Q.
[19] Reza Soroushmehr SM, Davuluri P, Molaei S, Hargraves RH, Tang Y, Cockrell CH, Ward K, Najarian K. Spleen segmentation and assessment in CT images for traumatic abdominal injuries. Journal of medical systems. 2015 Sep;39(9):1-1.