Methods and Evaluation Criteria for Apps and Digital Interventions for Diabetes Self-Management: Systematic Review

Dillys Larbi¹, MPH, MSc; Pietro Randine¹,², MCS; Eirik Årsand¹,², PhD; Konstantinos Antypas¹,³, PhD; Meghan Bradway¹,⁴, MBA; Elia Gabarron¹, PhD

¹Norwegian Centre for E-health Research, University Hospital of North Norway, Tromsø, Norway
²Department of Computer Science, Faculty of Science and Technology, UiT The Arctic University of Norway, Tromsø, Norway
³SINTEF Digital, Oslo, Norway
⁴Department of Clinical Medicine, Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway

Corresponding Author:
Meghan Bradway, MBA
Norwegian Centre for E-health Research
University Hospital of North Norway
Forskningsparken i Breivika 3rd Fl
Sykehusveien 23
Tromsø, 9019
Norway
Phone: 47 91193393
Email: meghan.bradway@ehealthresearch.no

Abstract

Background: There is growing evidence that apps and digital interventions have a positive impact on diabetes self-management. Standard self-management for patients with diabetes could therefore be supplemented by apps and digital interventions to increase patients’ skills. Several initiatives, models, and frameworks suggest how health apps and digital interventions could be evaluated, but there are few standards for this. And although there are many methods for evaluating apps and digital interventions, a more specific approach might be needed for assessing digital diabetes self-management interventions.

Objective: This review aims to identify which methods and criteria are used to evaluate apps and digital interventions for diabetes self-management, and to describe how patients were involved in these evaluations.

Methods: We searched CINAHL, EMBASE, MEDLINE, and Web of Science for articles published from 2015 that referred to the evaluation of apps and digital interventions for diabetes self-management and involved patients in the evaluation. We then conducted a narrative qualitative synthesis of the findings, structured around the included studies’ quality, methods of evaluation, and evaluation criteria.

Results: Of 1681 articles identified, 31 fulfilled the inclusion criteria. A total of 7 articles were considered of high confidence in the evidence. Apps were the most commonly used platform for diabetes self-management (18/31, 58%), and type 2 diabetes (T2D) was the targeted health condition most studies focused on (12/31, 38%). Questionnaires, interviews, and user-group meetings were the most common methods of evaluation. Furthermore, the most evaluated criteria for apps and digital diabetes self-management interventions were cognitive impact, clinical impact, and usability. Feasibility and security and privacy were not evaluated by studies considered of high confidence in the evidence.

Conclusions: There were few studies with high confidence in the evidence that involved patients in the evaluation of apps and digital interventions for diabetes self-management. Additional evaluation criteria, such as sustainability and interoperability, should be focused on more in future studies to provide a better understanding of the effects and potential of apps and digital interventions for diabetes self-management.

(J Med Internet Res 2020;22(7):e18480) doi: 10.2196/18480

KEYWORDS
self-management; diabetes mellitus; mobile applications; computer communication networks; mHealth; eHealth; health care evaluation mechanisms
**Introduction**

As the number of people with diabetes continues to rise worldwide [1], the need to increase patients’ self-management skills is crucial to improve clinical outcomes and reduce health-related costs [2,3]. There is growing evidence that apps and digital interventions such as websites (web), social media, and other online services have a positive impact on diabetes self-management [4-12], suggesting that standard self-management could be supplemented by digital interventions to aid and improve patients’ skills [4-12]. While some apps and digital interventions have benefited patients, not all of them seem to be based on research, and some of these digital interventions could even compromise the safety of patients with diabetes [13].

To improve diabetes self-management with apps and digital interventions, the World Health Organization and the European Commission [14,15] deem it necessary that the available apps and digital interventions are accurate and reliable. Several initiatives, models, and frameworks suggest how some of these apps and digital interventions could be evaluated [16-19]. These approaches commonly name background information, privacy and security, evidence on the provided information, ease of use, or interoperability as issues that need to be addressed [16-18].

Regarding how to evaluate these criteria, several methods of different complexity have been proposed. These include simple questions to be answered by health care professionals (HCPs) and patients, whereas more complex methodology approaches, such as laboratory-based testing, field testing, and N-of-1 design, are used by researchers [18,20]. Although the aforementioned issues are relevant for diabetes self-management apps and digital interventions, a more specific approach is needed for assessing the growing number and rapidly changing functionalities of these digital diabetes self-management interventions.

Another relevant issue is who should be involved in these evaluations. As patients are often required to make critical decisions based on their own generated health information [21], people with diabetes should be involved in these evaluations. However, a previous assessment of digital health interventions demonstrated limited consideration of user perceptions, and also that of health care personnel [22].

In this systematic review, we identify the specific methods and evaluation criteria that were used to assess apps and digital interventions for diabetes self-management. We also report how patients were involved in these assessments.

**Methods**

This review followed the PRISMA approach [23], and its systematic review protocol is registered in PROSPERO (Registration number: CRD42018115246).

**Data Sources and Search Strategy**

We performed a single data search in June 2018. The search strategy covered all studies that assessed diabetes self-management apps and digital interventions, involved patients, and were published in English after 2015. We chose a short search period to get a rapid overview of the most recent methods and evaluation criteria. The search strategy covered the following databases: CINAHL, EMBASE, MEDLINE, and Web of Science. The full search strategy is available in Multimedia Appendix 1.

**Inclusion and Exclusion Criteria**

We included articles for review if they were (1) primary studies referring to the evaluation of apps or digital interventions for diabetes self-management; and (2) involved patients in the evaluation.

Article were excluded if (1) the evaluation only measured medical values (ie, weight, glycated hemoglobin [HbA1c], blood glucose); (2) it was not a primary study; (3) it did not focus on apps or digital interventions for diabetes self-management; (4) the full-text was not available; (5) it was not a peer-reviewed publication; (6) it was not in English; or (7) it was published before 2015.

**Eligibility and Data Collection Procedure**

We uploaded all references captured by the search strategy to Rayyan and EndNote and removed duplicates. The eligibility of the articles was assessed in two stages. In the first stage, 2 independent reviewers (PR and EG) examined all titles and abstracts. Eligibility doubts were discussed and agreed with a third and fourth reviewer (KA and EA). In the second stage, the full texts of the selected articles were carefully examined by 2 independent reviewers (PR and EG) to confirm their eligibility.

Two reviewers (PR and MB) independently extracted and recorded the data from these articles on an Excel spreadsheet (Microsoft). We extracted the following information from each article: type of platform, targeted health condition, study population, methods of evaluation, and evaluation criteria. Incongruences with the data extraction were discussed among the research group.

**Confidence in the Evidence and Risk of Bias Assessment**

Two reviewers (EG and KA) assessed the confidence in the evidence and risk of bias of the articles. We used an approach based on the CERQual guidelines [24] to assess the confidence in the evidence of the qualitative primary studies, by evaluating their methodological limitations, relevance, and adequacy. We followed the GRADE guidelines [25] to assess mixed-methods studies, quantitative studies, and randomized trials.

**Strategy for Data Synthesis**

We provide a narrative qualitative synthesis of the findings from the included articles, structured around confidence in the evidence and risk of bias; type of platform (apps, web, or multiplatform [ie, ≥2 types of platform delivering the same intervention in a study]); targeted health condition (type 1 diabetes [T1D], T2D, gestational diabetes mellitus, both T1D and T2D, and unspecified diabetes type); methods of evaluation (questionnaires, interviews, user-group meetings, health measures, system usage analysis, or other); and evaluation criteria (usability, clinical impact, cognitive impact, behavioral impact, feasibility, engagement, acceptability and acceptance, or security and privacy).
Results

Identified and Included Studies
The search strategy resulted in 1681 articles. After removing duplicates, 967 articles remained. In the abstract screening, we excluded 910 articles in accordance with one or more of the exclusion criteria. A total of 57 articles were eligible for full-text screening, 26 of which were excluded (see Multimedia Appendix 2). A total of 31 articles were eventually included in the review [26-56] (see Multimedia Appendix 3). The PRISMA diagram in Figure 1 summarizes the selection process. The confidence in the evidence was considered high in 7 articles [27,33,36,43,51,52,54]; moderate to high in 1 [56]; moderate in 17 [26,28-32,35,37,39,41,42,45,46,48,49,53,55], and low in 6 [34,38,40,44,47,50].

Figure 1. PRISMA flowchart of the selection procedure.

Study Population
The 31 articles in this review included evaluations from 3689 participants. The number of participants in each study ranged from 7 [41,50] to 1041 [43]. In addition to including patients with diabetes in their evaluations, some of the studies expanded the participant group to include HCPs (8/31, 26%) [26,27,29,37,42,48,53,56], researchers (3/31, 10%) [29,38,44], informal caregivers (eg, parents, family members) (4/31, 13%) [29,44,53,56], and others (including experts and other unspecified individuals) (8/31, 26%) [26,29,32,37,38,48,55,56].

Type of Platform and Targeted Health Condition
Most of the 31 included studies evaluated interventions delivered via apps (18/31, 58%) [29-34,36-39,41,46-49,52,54,56], followed by web (9/31, 29%) [27,28,35,43,45,50,51,53,55] and multiplatform (4/31, 13%) [26,40,42,44]. In the studies that conducted a randomized controlled trial, the self-management platform was the main mode of intervention compared with a standard paper diary [33], the intervention plus counseling via
telephone call [54], and a plain text version of the web intervention [43] as opposed to an interactive version. In addition, the same intervention was referred to by some of the studies: Young with Diabetes app [29,36], My Diabetes My Way [30,55], and WellDoc [31,50]. The evaluated digital self-management interventions targeted mostly T2D (12/31, 38%) [27,31,32,34,43,46-51,54], followed by T1D (7/31, 23%) [29,33,36,40,41,44,56], unspecified diabetes type (5/31, 16%) [26,37,39,53,55], gestational diabetes mellitus (4/31, 13%) [35,38,45,52], and T1D and T2D (3/31, 10%) [28,30,42] (Figure 2).

Figure 2. Distribution of types of platform and targeted health conditions among included articles (n=31).

Identified Evaluation Methods
The methods of evaluation were grouped into 6 categories: questionnaires, interviews, user-group meetings, health-related measures, system usage analysis, and other measurements. We also identified 20 specific methods that were either used once or multiple times by the studies during the evaluation process.

The interrater agreement for the methods of evaluation was found to be $\kappa=0.550$, which represents a moderate agreement [57]. A summary of the specific methods of evaluation and studies that used them is presented in Table 1.

Questionnaires were the most common method used to evaluate diabetes self-management apps and digital interventions (21/31 studies, 68%) [29-40,42,43,45,46,50,51,54-56]. Standardized questionnaires were the most frequently used: 16 in total, each used one or multiple times among 13 studies [31,33,36,38,40,42,43,45,50,51,54]. The second most common method of evaluation was interviews (13/31 studies, 42%) [28,29,31,32,35,36,38,45,48,49,52,53,56], mainly semistructured interviews, which were used 14 times in 11 studies [28,31,32,35,36,45,48,49,52,53,56]. Other methods of evaluation that were identified in the included studies were user-group meetings (11/31, 35%) [26,27,29,37,40-42,45,47,48,56], health-related measures (9/31, 29%) [31,33,36,42,45,49,51,54,55], system usage analysis (8/31, 26%) [29,34,35,42,43,45,51,54], and other measurements (7/31, 23%) [26,37,38,44,48,51,55]. Table 1 summarizes the specific methods of evaluation, the number of times these methods were used, and the number of studies that employed these methods.

Among the 7 studies considered of high confidence in the evidence, the evaluations of the apps and digital diabetes self-management interventions were based mostly on standardized questionnaires [33,36,43,51,54], medical tests [33,36,51,54], and usage log analysis [43,51,54], followed by author-created questionnaires [43,51], semistructured interviews [36,52], focus groups [27], self-reported health measures [33], self-reported usage [43], alpha testing [51], and other oral and written feedback [51].
Table 1. Specific methods of evaluation and studies that used them.

| Method of evaluation, specific type (n=times used), and details | Reference(s) |
|---------------------------------------------------------------|---------------|
| **Standardized questionnaires (n=26)**                       |               |
| Block Food Frequency Assessment                               | [45]          |
| Dietary Knowledge, Attitude, and Behavior Questionnaire       | [51]          |
| Health Care Climate Questionnaire                             | [36]          |
| Paffenbarger Questionnaire                                    | [45]          |
| Patient Enablement Instrument                                 | [43]          |
| Patient Health Questionnaire-9                                | [50]          |
| Patient Reported Diabetes Symptoms Scale                       | [50]          |
| Perceived Competence in Diabetes                              | [36]          |
| Problem Areas in Diabetes-5                                   | [31]          |
| Problem Areas in Diabetes                                     | [33,36,40,42] |
| RAND 36-Item Health Survey 1.0                                | [33]          |
| Self-Efficacy for Diabetes Scale                              | [50]          |
| System Usability Scale                                        | [33-35,38,40,42]|
| The Health Education Impact Questionnaire                     | [54]          |
| The Service User Technology Acceptability Questionnaire       | [54]          |
| 36-Item Short Form Survey                                     | [50]          |
| **Author-created questionnaires (n=20)**                      |               |
| N/A                                                           | [29,30,37-39,42,43,45,46,51,55,56] |
| **User-group meetings**                                       |               |
| Focus groups (n=9)                                            |               |
| N/A                                                           | [27,37,40-42,45,47] |
| Workshops (n=7)                                               |               |
| N/A                                                           | [26,29,48,56] |
| **Interviews**                                                |               |
| Semistructured interviews (n=14)                              |               |
| N/A                                                           | [28,31,32,35,36,45,48,49,52,53,56] |
| Unspecified interview format (n=3)                            |               |
| N/A                                                           | [29,38]       |
| **System usage analysis**                                     |               |
| Usage log analysis (n=8)                                      |               |
| N/A                                                           | [34,35,43,45,51,54] |
| Self-reported usage (n=2)                                     |               |
| N/A                                                           | [42,43]       |
| Think-aloud protocol (n=1)                                    |               |
| N/A                                                           | [29]          |
| **Health-related measures**                                   |               |
| Medical tests (n=8)                                           |               |
The evaluated criteria were grouped into 8 categories: usability, clinical impact, cognitive impact, behavioral impact, feasibility, engagement, acceptability and acceptance, and security and privacy. The included studies evaluated one or several of these identified criteria. The interrater agreement (κ) for the evaluation criteria was found to be 0.563, which represents a moderate agreement [57].

Among the 7 studies considered of high confidence in the evidence, the most commonly evaluated criteria were clinical impact [33,36,51,54], cognitive impact [33,36,43,54], and engagement [43,51,54], followed by usability [33,51], behavioral impact [33,51], and acceptability and acceptance [51,54]. None of these studies considered of high confidence evaluated feasibility or security and privacy.

Qualitative and mixed-method studies that used thematic analysis in their evaluation focused mostly on usability as an evaluation criterion. Three of the studies considered of high confidence in evidence were qualitative and mixed-method studies. Of these, 2 evaluated cognitive impact [36,52] and usability [27,52], and 1 evaluated engagement [36].

Figure 3 shows the number of studies that used each of the specific methods to evaluate the identified criteria. It illustrates that several methods were used to evaluate one criterion in a single study. Likewise, some studies evaluated several criteria using one or more of the identified methods of evaluation. For example, of the 31 included studies, 9 [31,33,36,40,42,43,45,50,54] evaluated cognitive impact using standardized questionnaires.

---

| Method of evaluation, specific type (n=times used), and details | Reference(s) |
|---------------------------------------------------------------|---------------|
| HbA1c                                                         | [31,33,36,45,49,51,54,55] |
| Fasting blood glucose                                         | [51]          |
| Blood pressure and cholesterol                                | [55]          |
| Gestational weight gain                                       | [45]          |
| **Self-reported health measures (n=5)**                       |               |
| Self-reported blood glucose                                   | [33,42]       |
| Self-reported physical activity and nutritional habits         | [42]          |
| **Other measurements**                                        |               |
| Security assessment (n=1)                                     | [44]          |
| N/A                                                           |               |
| Scenarios (n=2)                                               | [37,38]       |
| N/A                                                           |               |
| Cost-effectiveness (n=1)                                      | [55]          |
| N/A                                                           |               |
| Alpha testing (n=1)                                           | [51]          |
| N/A                                                           |               |
| Observation (n=2)                                             | [48]          |
| N/A                                                           |               |
| Rating system (n=1)                                           | [26]          |
| Star rating                                                   |               |
| Heuristics method (n=1)                                       | [37]          |
| Bertini’s mobile tool                                         |               |
| Anecdotal feedback (n=1)                                      | [26]          |
| Open text review                                              |               |
| Other oral and written feedback (n=2)                         | [44,51]       |
| N/A                                                           |               |

aN/A: not applicable.
Figure 3. Number of studies using the various methods of evaluation and evaluation criteria. Blank boxes (NaN): No studies within this category.

| Method of Evaluation | Specific evaluation method | Evaluation criteria |
|----------------------|---------------------------|---------------------|
|                      |                           | Useability | Feasibility | Acceptability and acceptance | Behavioral Impact | Cognitive Impact | Engagement | Clinical Impact | Security and privacy |
| Questionnaires        | Standardized              | 6         | 1          | 2          | 9          | 2         |
|                      | Author-created            | 8         | 5          | 5          | 2          | 2         |
| User-group meetings   | Focus groups              | 5         | 2          | 2          | 2          | 1         | 1         |
|                      | Workshops                 | 4         | 2          |             |             |           | 1         |
| Interviews            | Semistructured            | 7         | 3          | 1          | 5          | 5         | 3         | 1         |
|                      | Unspecified format        | 2         |             |             |             |           | 1         |
| System usage analysis | Usage log analysis        | 1         | 2          |             |             |           | 3         |
|                      | Patient-reported usage    |           |             | 2          |             |           |           |           |
|                      | Think-aloud test          |           |             |             |             |           | 1         |
| Health-related measures | Medical tests           |           |             |             |             |           | 8         |
|                      | Self-reported health measures |       |             |             |             |           |           |
| Other measurements    | Security assessment       |           |             |             |             |           | 1         |
|                      | Scenarios                 |           |             |             |             |           | 2         |
|                      | Cost-effectiveness        |           |             |             |             |           | 1         |
|                      | Alpha testing             |           |             |             |             |           | 1         |
|                      | Observation               |           |             |             |             |           | 1         |
|                      | Rating system             |           |             |             |             |           | 1         |
|                      | Heuristics method         |           |             |             |             |           | 1         |
|                      | Anecdotal feedback        |           |             |             |             |           | 1         | 1         |
|                      | Other oral and written feedback |       |             |             |             |           | 2         |

**Discussion**

**Summary of the Findings**

This review aimed to identify the existing methods and criteria used to assess apps and digital diabetes self-management interventions that involved patients in their evaluations. A total of 31 articles were included in the review, 7 of which were considered of high confidence in the evidence [27,33,36,43,51,52,54]. More than half of the studies (18/31, 58%) focused on the evaluation of apps for diabetes self-management, and 12 of the 31 studies addressed T2D. The most commonly used methods of evaluation were questionnaires, interviews, and user-group meetings. The most used evaluation criteria to assess apps and digital interventions for diabetes self-management were cognitive impact, clinical impact, and usability.

**Specific Evaluation Criteria and Diabetes Patients’ Assessment**

In our review, we have found that studies dealing specifically with apps and digital interventions for diabetes self-management focus on the evaluation of more technology-related and users’ interaction aspects (ie, acceptability and acceptance, and engagement). In addition, these studies focus on the impact that these digital self-management interventions have on the individual. Behavioral impact, cognitive impact, and clinical impact were used as relevant criteria for assessing all types of digital interventions for diabetes self-management. It is vital to measure the interventions’ impact on their users because those that have shown benefits related to behavioral, cognitive, and clinical impact could reduce health-related costs [2,3].

Evidence shows that involving individuals in the assessment of different health interventions has a positive impact on health...
Involving patients with diabetes in assessing apps and digital self-management interventions, and obtaining their feedback regarding additional evaluation criteria could also increase our knowledge about the features that support engagement with these technologies. This could also help create better digital health interventions that encourage more continuous and effective use [59]. The most common methods of evaluation with the patients were questionnaires, interviews, and user-group meetings. Simple methods such as these elicit the opinion and perceptions of users, as well as encourage them to critically analyze self-management apps and digital interventions. Therefore, such methods should be used in conjunction with complex methods used by researchers and developers [18,20], especially to measure the same criterion.

**Improving Reported Evaluations of Digital Interventions for Diabetes Self-Management**

Apps and digital health interventions have evolved quickly. Yet, compared with other sectors, the health industry seems to be behind with regard to digitalization [60]. Currently, most apps and digital interventions for self-management are not recommended as part of the treatment plan, maybe because their design and development do not take into consideration sustainability [61]. In fact, digital health interventions rarely advance beyond a pilot phase [62,63], or the duration of an intervention study.

In 2016, the mobile health (mHealth) evidence reporting and assessment checklist was developed by the World Health Organization to help with reporting evidence of the effectiveness of mHealth interventions [64]. The checklist recommended reporting on items that touch on sustainability, scalability, and transparency, such as infrastructure, interoperability, contextual adaptability, and replicability, which we still see are not much focused on in today’s studies. Future studies should also consider these evaluation criteria in addition to gender and equity issues associated with the use of apps and digital interventions for diabetes self-management.

Evaluation reports for apps and digital interventions for diabetes self-management must be standardized, as recommended by the CONSORT-EHEALTH guidelines for reporting digital health interventions [65]. The lack of standardization made it challenging to compare studies as different authors used different terminologies to describe the same evaluation criterion. For example, one study [37] used the term *heuristics evaluation*, which was grouped under usability because it evaluated measures such as the visibility of app status, ease of input, and readability. Likewise, another study [32] evaluated *satisfaction*, which falls under usability because it evaluated among others, visual attractiveness and ease of use.

As electronic health (eHealth) research is a multidisciplinary field, we assume that the authors chose these terms based on the various educational or professional backgrounds and the corresponding target audiences. By following the World Health Organization classification of digital health interventions [66], terminologies related to the evaluation of apps and digital interventions for diabetes self-management could be standardized to facilitate straightforward interpretation and aggregation of research evidence.

**Association Between Methods Used and Criteria Evaluated**

In our review we have found that there was an almost even split of studies that used standardized questionnaires, author-created questionnaires, and semistructured interviews to evaluate usability. Our results are to some extent in line with the findings of a previous review that found that usability was mainly assessed though polls and questionnaires [67]. The usability of a digital self-management intervention is crucial to its successful adoption, its acceptance, and the individual’s engagement with it. In addition, we found that cognitive impact was often assessed not only through standardized questionnaires, but also through semistructured interviews.

Comparing the methods for the evaluation of usability with those for the evaluation of cognitive impact, we identified that it was more common to use author-created questionnaire for usability. A possible explanation might be the wide variety of intervention delivery platforms (eg, different types of apps and online resources) that might create different evaluation needs not captured in existing standardized usability questionnaires. Another explanation might be the different research traditions in different disciplines. Usability might be more often a concern of computer science researchers, whereas cognitive impact a concern of health researchers and professionals.

Finally, health outcomes were almost exclusively evaluated by medical tests, showing the preference of health researchers and professionals in using standardized tests to determine the impact of digital interventions. Several other methods can be used to evaluate multiple criteria; however, depending on the aim and the type of study, researchers must endeavor to exhaust all available methods to ensure consistency of results.

**Feasibility of Using Digital Self-Management Interventions in Clinical Workflow**

Although most apps and digital health interventions are intended for self-management, some of them also provide access to the health care system, such as communication with HCPs and electronic health journals. The reviewed studies consistently reported that this is in response to patients’ interest in being able to contact their HCPs or share results (eg, their blood glucose results with their health care team). This was the case not only within our review [35,36,42,44,47,48,56] but also by industry research groups [68,69]. This implies the potential and expectation for further involvement of HCPs in patients’ use of apps and digital interventions for diabetes self-management.
Several studies, including many in this review, have shown that involving HCPs in digital interventions is associated with improved self-management of diabetes and the success of these interventions [31,48,49,52,70-72]. Therefore, studies focusing on apps and digital interventions for diabetes self-management should evaluate the possibilities of effortlessly integrating these interventions in the workflow of HCPs—the connection and interaction with electronic health journals and other existing health information systems. Such an integration can be achieved by evaluating the infrastructure needed for digital self-management interventions [64].

Limitations and Strengths
The search for articles covered a short period (2015-2018) and focused on articles published in the English language. Therefore, we may have missed relevant studies that reported additional evaluation methods or evaluation criteria. Our interrater agreement of the data extraction was only moderate; however, all incongruences were discussed among the research group. Our findings have provided a useful overview of the recent evaluation methods and criteria that researchers are using to assess current apps and digital interventions for diabetes self-management. Furthermore, our review included both quantitative and qualitative studies which provided a better characterization of different evaluation methods and criteria that are being used to assess digital diabetes self-management interventions.

Conclusions
There are only few studies that involved patients in the evaluation of apps and digital interventions for diabetes self-management, and even fewer still considered of high confidence in the evidence. The most common evaluation methods were questionnaires, interviews, and user-group meetings, whereas evaluation criteria were cognitive impact, clinical impact, and usability. Studies with high confidence in the evidence did not evaluate feasibility or security and privacy, neither were patients involved in evaluating the latter criterion which was evaluated in only 2 [29,44] of the included studies.

It is important to the successful implementation and continuous use of apps and digital interventions for diabetes self-management that patients are involved in evaluating every criteria. In that way, they can contribute to the development and modification of these digital interventions to better meet their specific self-management needs. Furthermore, the methods and criteria evaluated in digital diabetes self-management interventions should be expanded to assess and ensure sustainability and interoperability. In addition, studies should evaluate the association between cognitive, clinical, and behavioral impact of these apps and digital interventions, and health-related costs for individuals with diabetes. This could help improve health care associated with the management of diabetes and promote the incorporation of apps and digital interventions for self-management in the services provided at health care facilities.

Acknowledgments
This project is funded by Helse Nord (HNF1425-18). The coauthors acknowledge the advice of the project’s Advisory board: Professor Gunnar Hartvigsen, Anne Grethe Olsen MD, and Dr. Med. Anne Helen Hansen, and also the support and contributions of Per Erlend Hasvold in his role as an “internal reviewer.” Furthermore, we thank Dr. Steven Bradway for his assistance with the coarse data extraction and organization at the start of this review. The publication charges for this article have been funded by a grant from the publication fund of UiT The Arctic University of Norway.

Authors’ Contributions
KA was responsible for database searching; EG, EÅ, KA, and PR were responsible for title, abstract, and full-text screening; MB and PR performed independent data extraction; EG and KA evaluated risk of bias; and DL, EG, EÅ, MB, KA, and PR performed data analysis and interpretation. All the coauthors contributed to drafting and revising the review. All coauthors approved the final version of the manuscript.

Conflicts of Interest
None declared.

Multimedia Appendix 1
Search strategy (search date: June 21, 2018).
[DOC File , 28 KB-Multimedia Appendix 1]

Multimedia Appendix 2
List of rejected articles after full-text review (n=26).
[DOC File , 51 KB-Multimedia Appendix 2]

Multimedia Appendix 3
Articles included in qualitative synthesis (n=31).
[DOC File , 267 KB-Multimedia Appendix 3]
Multimedia Appendix 4

PRISMA checklist.

References

1. World Health Organization. Diabetes. 2019. URL: https://www.who.int/news-room fact-sheets/detail/diabetes [accessed 2019-12-01]
2. Barker I, Steventon A, Williamson R, Deeny SR. Self-management capability in patients with long-term conditions is associated with reduced healthcare utilisation across a whole health economy: cross-sectional analysis of electronic health records. BMJ Qual Saf 2018 Dec;27(12):989-999 [FREE Full text] [doi: 10.1136/bmjqs-2017-007635] [Medline: 30139822]
3. Grady PA, Gough LL. Self-management: a comprehensive approach to management of chronic conditions. Am J Public Health 2014 Aug;104(8):e25-e31. [doi: 10.2105/AJPH.2014.302041] [Medline: 24922170]
4. Willhide ICC, Peeples MM, Kouraté RCA. Evidence-Based mHealth Chronic Disease Mobile App Intervention Design: Development of a Framework. JMIR Res Protoc 2016;5(1):e25 [FREE Full text] [doi: 10.2196/resprot.4838] [Medline: 26883135]
5. Marcolino MS, Oliveira JAQ, D’Agostino M, Ribeiro AL, Alkmim MBM, Novillo-Ortiz D. The Impact of mHealth Interventions: Systematic Review of Systematic Reviews. JMIR Mhealth Uhealth 2018 Jan 17;6(1):e23 [FREE Full text] [doi: 10.2196/mhealth.8873] [Medline: 29343463]
6. Cotter AP, Durant N, Agne AA, Cherrington AL. Internet interventions to support lifestyle modification for diabetes management: a systematic review of the evidence. J Diabetes Complications 2014;28(2):243-251 [FREE Full text] [doi: 10.1016/j.jdiacomp.2013.07.003] [Medline: 24332469]
7. Gabarron E, Årsand E, Wynn R. Social Media Use in Interventions for Diabetes: Rapid Evidence-Based Review. J Med Internet Res 2018 Aug 10;20(8):e10303 [FREE Full text] [doi: 10.2196/jmir.10303] [Medline: 30097421]
8. Holster J, Carter B, Hewitt J, Francisca T, Mayor S. Do Mobile Phone Applications Improve Glycemic Control (HbA1c) in the Self-management of Diabetes? A Systematic Review, Meta-analysis, and GRADE of 14 Randomized Trials. Diabetes Care 2016 Nov;39(11):2089-2095. [doi: 10.2337/dci16-0346] [Medline: 27926892]
9. Hamada A, Quek KF, Chan CKY, Oldenburg B. Web-based interventions for the management of type 2 diabetes mellitus: a systematic review of recent evidence. Int J Med Inform 2011 Jun;80(6):389-405. [doi: 10.1016/j.ijmedinf.2011.02.002] [Medline: 21481632]
10. Shan R, Sarkar S, Martin SS. Digital health technology and mobile devices for the management of diabetes mellitus: state of the art. Diabetologia 2019 Jun;62(6):877-887. [doi: 10.1007/s00125-019-4864-7] [Medline: 30963188]
11. Greenwood DA, Gee PM, Fatkin KJ, Peeples M. A Systematic Review of Reviews Evaluating Technology-Enabled Diabetes Self-Management Education and Support. J Diabetes Sci Technol 2017 Sep;11(5):1015-1027. [doi: 10.1177/1932296817713506] [Medline: 28560898]
12. Huang Z, Soljak M, Boehm BO, Car J. Clinical relevance of smartphone apps for diabetes management: A global overview. Diabetes Metab Res Rev 2015;31(4):299. [doi: 10.1002/dmrr.2990] [Medline: 29431916]
13. Huckvale K, Adomaviciute S, Prieto JT, Leow MK, Car J. Smartphone apps for calculating insulin dose: a systematic assessment. BMC Med 2015;13:106 [FREE Full text] [doi: 10.1186/s12916-015-0314-7] [Medline: 25943590]
14. European Commission. Report of the Working Group on mHealth Assessment Guidelines. 2017 Jun. URL: https://ec.europa.eu/digital-single-market/en/news/report-working-group-mhealth-assessment-guidelines [accessed 2019-12-01]
15. World Health Organization. mHealth. New Horizons for Health Through Mobile Technologies. 2011. URL: https://www.who.int/goe/publications/goe_mhealth_web.pdf [accessed 2019-12-01]
16. Henson P, David G, Albright K, Torous J. Deriving a practical framework for the evaluation of health apps. Lancet Digital Health 2019 Jun;1(2):e52-e54. [doi: 10.1016/S2468-2667(19)30013-5] [Medline: 31333133]
17. Health on the net. HON Code. 2019. URL: https://www.hon.ch/en/ [accessed 2020-01-01]
18. American Psychiatric Association. App Evaluation Model. 2020. URL: https://www.psychiatry.org/psychiatrists/practice/mental-health-apps/app-evaluation-model [accessed 2020-01-01]
19. National Institute for Health and Care Excellence. Evidence Standards Framework for Digital Health Technologies. London, UK: National Institute for Health and Care Excellence; 2019. URL: https://www.nice.org.uk/Media/Default/About/what-we-do/our-programmes/evidence-standards-framework/digital-evidence-standards-framework.pdf [accessed 2020-02-26]
20. Luke-Schoffman DE, Sillec VF, Waring ME, Boudreaux ED, Sasadivam RS, Mullen SP, et al. Methods for Evaluating the Content, Usability, and Efficacy of Commercial Mobile Health Apps. JMIR Mhealth Uhealth 2017 Dec 18;5(12):e190 [FREE Full text] [doi: 10.2196/mhealth.8757] [Medline: 29524914]
21. D'Angelo L, Wyatt JC. mHealth and mobile medical Apps: a framework to assess risk and promote safer use. J Med Internet Res 2014 Sep 15;16(9):e210 [FREE Full text] [doi: 10.2196/jmir.3133] [Medline: 25223398]
22. Lewis TL, Barron C, Vallespin B, Saadatford O, Puigdomenech E, Espallargues M, et al. mHealth Assessment: Conceptualization of a Global Framework. JMIR Mhealth Uhealth 2017 May 02;5(5):e60 [FREE Full text] [doi: 10.2196/mhealth.7291] [Medline: 28465282]
23. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 2009 Aug 18;151(4):264-9, W64. [Medline: 19622511]

24. Levin S, Bohren M, Rashidian A, Munthe-Kaas H, Glenton C, Colvin CJ, et al. Applying GRADE-CERQual to qualitative evidence synthesis findings-paper 2: how to make an overall CERQual assessment of confidence and create a Summary of Qualitative Findings table. Implement Sci 2018 Jan 25;13(Suppl 1):10 [FREE Full text] [doi: 10.1186/s13012-017-0689-2] [Medline: 29384082]

25. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011 Apr;64(4):401-406. [Medline: 21208779]

26. Ashurst EJ, Jones RB. Is the Health App Challenge approach of patient-led application conception, development, and review worthwhile? Health Policy Technol 2017 Mar;6(1):83-92. [Medline: 10.1016/j.hpt.2016.12.001]

27. Bernhard G, Mahler C, Seidlung HM, Stützle M, Ose D, Baudendistel I, et al. Developing a Shared Patient-Centered, Web-Based Medication Platform for Type 2 Diabetes Patients and Their Health Care Providers: Qualitative Study on User Requirements. J Med Internet Res 2018 Mar 27;20(3):e105 [FREE Full text] [doi: 10.2196/jmir.8666] [Medline: 29588269]

28. Brady E, Segar J, Sanders C. Accessing support and empowerment online: The experiences of individuals with diabetes. Health Expect 2017 Oct;20(5):1088-1095 [FREE Full text] [doi: 10.1111/hex.12552] [Medline: 28718928]

29. Castensøe-Seidenfaden P, Husted GR, Teilmann G, Hommel E, Olsen BS, Kensing F. Designing a Self-Management App for Young People With Type 1 Diabetes: Methodological Challenges, Experiences, and Recommendations. JMIR Mhealth Uhealth 2017 Oct 23;5(10):e124 [FREE Full text] [doi: 10.2196/mhealth.8137] [Medline: 29061552]

30. Conway N, Campbell I, Forbes P, Cunningham S, Wake D. mHealth applications for diabetes: User preference and implications for app development. Health Informatics J 2016 Dec;22(4):1111-1120. [doi: 10.1177/1460458215616265] [Medline: 26635324]

31. Desveaux L, Shaw J, Saragosa M, Soobiah C, Marani H, Hensel J, et al. A Mobile App to Improve Self-Management of Individuals With Type 2 Diabetes: Qualitative Realist Evaluation. J Med Internet Res 2018 Mar 16;20(3):e81 [FREE Full text] [doi: 10.2196/jmir.8712] [Medline: 29549070]

32. Dewi DS, Irfoni AR, Rahman A. Kansesi Engineering Approach for Designing a Self-monitoring Blood Glucose Application. IJTech 2017 Apr 29;8(2):272-282. [doi: 10.14716/ijtech.v8i2.6144]

33. Drion I, Pameijer LR, van Dijk PR, Groenier KH, Kleefstra N, Bilo HJG. The Effects of a Mobile Phone Application on Quality of Life in Patients With Type 1 Diabetes Mellitus: A Randomized Controlled Trial. J Diabetes Sci Technol 2015 May 11;9(5):1086-1091 [FREE Full text] [doi: 10.1177/1932296815585871] [Medline: 25963412]

34. Georgsson M, Staggers N. Quantifying usability: an evaluation of a diabetes mHealth system on effectiveness, efficiency, and satisfaction metrics with associated user characteristics. J Am Med Inform Assoc 2016 Jun 23;23(1):5-11 [FREE Full text] [doi: 10.1093/jamia/ocv099] [Medline: 26377990]

35. Gianfrancesco C, Darwin Z, McGowan L, Smith DM, Haddrill R, Carter M, et al. Exploring the Feasibility of Use of An Online Dietary Assessment Tool (myfood24) in Women with Gestational Diabetes. Nutrients 2018 Aug 23;10(9):1147 [doi: 10.3390/nu10091147]

36. Georgsson M, Staggers N. Quantifying usability: an evaluation of a diabetes mHealth system on effectiveness, efficiency, and satisfaction metrics with associated user characteristics. J Am Med Inform Assoc 2016 Jun 23;23(1):5-11 [FREE Full text] [doi: 10.1093/jamia/ocv099] [Medline: 26377990]

37. Gianfrancesco C, Darwin Z, McGowan L, Smith DM, Haddrill R, Carter M, et al. Exploring the Feasibility of Use of An Online Dietary Assessment Tool (myfood24) in Women with Gestational Diabetes. Nutrients 2018 Aug 23;10(9):1147 [doi: 10.3390/nu10091147] [Medline: 26377990]

38. Geiger C, Darwiczy S, Deckers R, van der Burg G, van der Vleuten CMC, et al. Coaching and Gamification Platform for Young Diabetes Patients. Sensors (Basel) 2018 Jan 30;18(2):402 [FREE Full text] [doi: 10.3390/s18020402]

39. Gianfrancesco C, Darwin Z, McGowan L, Smith DM, Haddrill R, Carter M, et al. Exploring the Feasibility of Use of An Online Dietary Assessment Tool (myfood24) in Women with Gestational Diabetes. Nutrients 2018 Aug 23;10(9):1147 [doi: 10.3390/nu10091147] [Medline: 26377990]

40. Jo S, Park H. Development of the IMB Model and an Evidence-Based Diabetes Self-management Mobile Application. Healhc Inform Res 2018 Apr;24(2):125-138 [FREE Full text] [doi: 10.4258/hir.2018.24.2.125] [Medline: 29700246]

41. Jo S, Park H. Development of the IMB Model and an Evidence-Based Diabetes Self-management Mobile Application. Healhc Inform Res 2018 Apr;24(2):125-138 [FREE Full text] [doi: 10.4258/hir.2018.24.2.125] [Medline: 29700246]

42. Keria X, Larbi L, Larbi N, Larbi M, et al. Qualitative Findings table. Implement Sci 2018 Jan 25;13(Suppl 1):10 [FREE Full text] [doi: 10.1186/s13012-017-0689-2] [Medline: 29384082]

43. Keria X, Larbi L, Larbi N, Larbi M, et al. Qualitative Findings table. Implement Sci 2018 Jan 25;13(Suppl 1):10 [FREE Full text] [doi: 10.1186/s13012-017-0689-2] [Medline: 29384082]

44. Keria X, Larbi L, Larbi N, Larbi M, et al. Qualitative Findings table. Implement Sci 2018 Jan 25;13(Suppl 1):10 [FREE Full text] [doi: 10.1186/s13012-017-0689-2] [Medline: 29384082]

45. Keria X, Larbi L, Larbi N, Larbi M, et al. Qualitative Findings table. Implement Sci 2018 Jan 25;13(Suppl 1):10 [FREE Full text] [doi: 10.1186/s13012-017-0689-2] [Medline: 29384082]
44. Neinstein A, Wong J, Look H, Arbiter B, Quirk K, McCanne S, et al. A case study in open source innovation: developing the Tidepool Platform for interoperability in type 1 diabetes management. J Am Med Inform Assoc 2016 Mar;23(2):324-332 [FREE Full text] [doi: 10.1093/jamia/ocv104] [Medline: 26338218]

45. Nicholson WK, Beckham AJ, Hatley K, Diamond M, Johnson L, Green SL, et al. The Gestational Diabetes Management System (GooDMoms): development, feasibility and lessons learned from a patient-informed, web-based pregnancy and postpartum lifestyle intervention. BMC Pregnancy Childbirth 2016 Sep 21;16(1):277 [FREE Full text] [doi: 10.1186/s12884-016-1064-2] [Medline: 27654119]

46. Park S, Burford S, Nolan C, Hanlen L. The Role of Digital Engagement in the Self-Management of Type 2 Diabetes. Health Commun 2016 Dec;31(12):1557-1565. [doi: 10.1080/10410236.2015.1089468] [Medline: 27124817]

47. Peng W, Yuan S, Holtz BE. Exploring the Challenges and Opportunities of Health Mobile Apps for Individuals with Type 2 Diabetes Living in Rural Communities. Telemed J E Health 2016 Sep;22(9):733-738. [doi: 10.1089/tmj.2015.0180] [Medline: 26982017]

48. Petersen M, Hempler NF. Development and testing of a mobile application to support diabetes self-management for people with newly diagnosed type 2 diabetes: a design thinking case study. BMC Med Inform Decis Mak 2017 Jun 26;17(1):91 [FREE Full text] [doi: 10.1186/s12911-017-0493-6] [Medline: 28651639]

49. Pludwinski S, Ahmad F, Wayne N, Ritvo P. Participant experiences in a smartphone-based health coaching intervention for type 2 diabetes: A qualitative inquiry. J Telemed Telecare 2015 Jul 21;172-178. [doi: 10.1177/1357633X15595178] [Medline: 26199275]

50. Quinn CC, Khokhar B, Weed K, Barr E, Gruber-Baldini AL. Older Adult Self-Efficacy Study of Mobile Phone Diabetes Management. Diabetes Technol Ther 2015 Jul;17(7):455-461 [FREE Full text] [doi: 10.1089/dia.2014.0341] [Medline: 25692373]

51. Ramadas A, Chan CKY, Oldenburg B, Hussen Z, Quek KF. A web-based dietary intervention for people with type 2 diabetes: development, implementation, and evaluation. Int J Behav Med 2015 Jun;22(3):365-373 [FREE Full text] [doi: 10.1007/s12529-014-9445-z] [Medline: 25274015]

52. Skar JB, Garnweidner-Holme LM, Lukasse M, Terragni L. Women's experiences with using a smartphone app (the Pregnant+) to manage gestational diabetes mellitus in a randomised controlled trial. Midwifery 2018 Mar;58:102-108. [doi: 10.1016/j.midw.2017.12.021] [Medline: 29329023]

53. Tieu L, Sarkar U, Schillinger D, Ralston JD, Ratanawongs A, Pasick R, et al. Barriers and Facilitators to Online Portal Use Among Patients and Caregivers in a Safety Net Health Care System: A Qualitative Study. J Med Internet Res 2015;17(12):e275 [FREE Full text] [doi: 10.2196/jmir.4847] [Medline: 26681155]

54. Torbjørnsen A, Småstuen MC, Jenum AK, Årstad E, Ribu L. Acceptability of an mHealth App Intervention for Persons With Type 2 Diabetes and its Associations With Initial Self-Management: Randomized Controlled Trial. JMIR Mhealth Uhealth 2018 May 21;6(5):e125 [FREE Full text] [doi: 10.2196/mhealth.8824] [Medline: 29784635]

55. Wake DJ, He J, Czesak AM, Mughal F, Cunningham SG. MyDiabetesMyWay: An Evolving National Data Driven Diabetes Self-Management Platform. J Diabetes Sci Technol 2016 Dec;10(5):1050-1058 [FREE Full text] [doi: 10.1177/1932296816648168] [Medline: 27126192]

56. Zhang Y, Li X, Luo S, Liu C, Liu F, Zhou Z. Exploration of Users' Perspectives and Needs and Design of a Type 1 Diabetes Management Mobile App: Mixed-Methods Study. JMIR Mhealth Uhealth 2018 Sep 21;6(9):e11400 [FREE Full text] [doi: 10.2196/jmir.11400] [Medline: 30249580]

57. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977 Mar;33(1):159-174. [Medline: 843571]

58. Campbell M, Escobar O, Fenton C, Craig P. The impact of participatory budgeting on health and wellbeing: a scoping review of evaluations. BMC Public Health 2018 Jul 03;18(1):822 [FREE Full text] [doi: 10.1186/s12889-018-5735-8] [Medline: 29970044]

59. Adu MD, Malau-Uh CJ, Calander EJ, Malau-Aduli AE, Malau-Aduli BS. Considerations for the Development of Mobile Phone Apps to Support Diabetes Self-Management: Systematic Review. JMIR Mhealth Uhealth 2018 Jun 21;6(6):e10115 [FREE Full text] [doi: 10.2196/10115] [Medline: 29929949]

60. Azzopardi-Muscat N, Ricciardi W, Odone A, Buttigieg S, Zeegers Paget D. Digitalization: potentials and pitfalls from a public health perspective. Eur J Public Health 2019 Oct 01;29(Supplement_3):1-2 [FREE Full text] [doi: 10.1093/eurpub/ckz169] [Medline: 31738438]

61. Wali S, Keshavjee K, Demers C. Moving towards sustainable electronic health applications. In: Heston TF, editor. eHealth - Making Health Care Smarter. London, UK: IntechOpen; 2018.

62. Huang F, Blaschke S, Lucas H. Beyond pilotitis: taking digital health interventions to the national level in China and Uganda. Global Health 2017 Jul 31;13(1):49 [FREE Full text] [doi: 10.1186/s12929-017-0275-z] [Medline: 28756767]

63. Wilson K, Gertz B, Arenth B, Salisbury N. The Journey to Scale: Moving Together Past Digital Health Pilots. Seattle, WA: PATH; 2014. URL: https://pathazureedge.net/media/documents/TS_dhs_journey_to_scale.pdf [accessed 2020-02-24]

64. Agarwal S, LeFevre AE, Lee J, L'Engle K, MGH, Sinha C, et al. Guidelines for reporting of health interventions using mobile phones: mobile health (mHealth) evidence reporting and assessment (mERA) checklist. BMJ 2016;352:i1174. [Medline: 26988021]
Abbreviations

- **HbA1c**: glycated hemoglobin
- **HCPs**: health care professionals
- **T1D**: type 1 diabetes
- **T2D**: type 2 diabetes

©Dillys Larbi, Pietro Randine, Eirik Årsand, Konstantinos Antypas, Meghan Bradway, Elia Gabarron. Originally published in the Journal of Medical Internet Research (http://www.jmir.org), 06.07.2020. This is an open-access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on http://www.jmir.org/, as well as this copyright and license information must be included.

[URL](https://www.jmir.org/2020/7/e18480)