Citation for published version (APA):
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;39:2089–2095 | DOI: 10.2337/dc16-0346

OBJECTIVE
To investigate the effect of mobile phone applications (apps) on glycemic control (HbA1c) in the self-management of diabetes.

RESEARCH DESIGN AND METHODS
Relevant studies that were published between 1 January 1996 and 1 June 2015 were searched from five databases: Medline, CINAHL, Cochrane Library, Web of Science, and Embase. Randomized controlled trials that evaluated diabetes apps were included. We conducted a systematic review with meta-analysis and GRADE (Grading of Recommendations Assessment, Development and Evaluation) of the evidence.

RESULTS
Participants from 14 studies (n = 1,360) were included and quality assessed. Although there may have been clinical diversity, all type 2 diabetes studies reported a reduction in HbA1c. The mean reduction in participants using an app compared with control was 0.49% (95% CI 0.30, 0.68; I² = 10%), with a moderate GRADE of evidence. Subgroup analyses indicated that younger patients were more likely to benefit from the use of diabetes apps, and the effect size was enhanced with health care professional feedback. There was inadequate data to describe the effectiveness of apps for type 1 diabetes.

CONCLUSIONS
Apps may be an effective component to help control HbA1c and could be considered as an adjuvant intervention to the standard self-management for patients with type 2 diabetes. Given the reported clinical effect, access, and nominal cost of this technology, it is likely to be effective at the population level. The functionality and use of this technology need to be standardized, but policy and guidance are anticipated to improve diabetes self-management care.

The number of patients with diabetes globally is expected to rise to over 500 million by 2030 (1). There is an urgent need for an improved self-management suite of interventions. For self-management to be effective, it needs to be structured and cost-effective (2) and be widely accessible across all health economies, including the developing world (2).
As a newly emerging technology, diabetes mobile phone applications (hereafter referred to as diabetes apps) are a promising tool for self-management. We define diabetes apps as mobile phone software that accepts data (transmitted or manual entry) and provides feedback to patients on improved management (automated or by a health care professional [HCP]). This technology combines the functions of the mobile phone, wireless network for data transmission, and sometimes HCPs for providing feedback. Due to its ubiquitous, low-cost, interactive, and dynamic health promotion, there is potential for diabetes apps to provide an effective intervention in diabetes self-care.

In terms of diabetes self-management, numerous studies have proven the effectiveness of other telemedicine technologies, such as short message service (3), computer-based interventions (4), and web-based interventions (3,5). Compared with these telemedicine interventions, diabetes apps are advantageous in that they are global, cheaper, convenient, and more interactive. There is, however, current uncertainty on the clinical effectiveness of diabetes apps in diabetes self-management (6–9).

RESEARCH DESIGN AND METHODS

Data Sources and Search Strategy

The PRISMA statement and checklist was followed. Five electronic databases were searched (Medline, CINAHL, Cochrane Library, Web of Science, and Embase) for studies published between 1 January 1996 and 1 June 2015. The references of the included studies were hand searched to identify any additional articles. The following terms and the medical subject headings (MeSH) were used during the search: (mobile OR mHealth OR cell phone OR MeSH “Cellular Phone” OR MeSH “Smartphone” OR app OR MeSH “Mobile Applications”) AND (MeSH “Diabetes Mellitus” OR diabetes* OR T2DM OR T1DM OR IDDM OR NIDDM).

Inclusion and Exclusion Criteria

The inclusion criteria were as follows: the participants were over 18 years old and had type 1 or type 2 diabetes, the studies were randomized controlled trials (RCTs), the control group in the study received usual diabetes care without any telehealth programs, and baseline and follow-up mean for HbA1c were reported (or could be calculated). Exclusion criteria were as follows: simulated or self-reported HbA1c data, computer or other mobile terminal–based diabetes apps, diabetes apps were exclusively designed for HCPs, and diabetes apps were exclusively designed for providing general education or allowing communication between patients and HCPs.

Two reviewers (C.H. and T.F.) searched the literature and assessed the studies independently. Any disagreements were resolved through discussion with a third reviewer (B.C.). No language restrictions were applied.

Data Extraction

Participant demographics, study design considerations, and context were extracted from the included studies. Two reviewers independently carried out the data extraction (C.H. and T.F.). Study authors were contacted to provide additional data, and missing SDs were estimated by calculation (10).

Quality Assessment

The quality assessment was conducted by two reviewers independently (C.H. and T.F.), using the quality rating tool proposed by the U.S. Preventive Services Task Force (11). Seven criteria were used to assess quality: baseline comparability of the groups, the maintenance of comparability of the groups, differential or high loss to follow-up, reliable and valid measurement, clear definition of the intervention, consideration of important outcomes, and an intention-to-treat analysis. The quality of each study was graded as good, fair, or poor. To be rated as good, studies needed to meet all the criteria. A study was rated as poor if one (or more) domain was assessed as having a serious flaw. Studies that met some but not all of the criteria were rated as fair quality.

Data Analysis

Changes in HbA1c or HbA1c at follow-up, were compared between groups using a mean difference and were presented with an associated 95% CI. When studies investigated interventions and contexts that were both deemed clinically similar and free from statistical heterogeneity, pooling was carried out using an inverse variance random-effects model (12). Meta-analyses were conducted using the Comprehensive Meta-Analysis Software (version 2.2). The level of evidence was applied to the GRADE (Grading of Recommendations Assessment, Development and Evaluation) criteria and reported.

RESULTS

Identified and Included Studies

Searches identified 5,209 articles; 4,238 were screened after removing duplicate records and 4,178 were excluded. Sixty studies were eligible for full text review and 42 were excluded (Fig. 1), resulting in 14 included studies. Four studies examined type 1 diabetes and 10 studies examined type 2 diabetes.

Characteristics of the Included Studies and Quality Assessment

In the 14 studies, there were 1,360 participants: 509 and 851 with type 1 and type 2 diabetes, respectively (Supplementary Table 1). In the type 1 diabetes studies, the mean age of participants ranged from 34 (13) to 36 years old (14), and the mean duration of diabetes ranged from 16 (13–15) to 19 years (15). Two studies were undertaken in Europe (13,14), one in Australia (15), and one was multinational (16). In the type 2 diabetes studies, the mean age of the participants was much higher, ranging from 51 (17) to 62 years old (18), and the mean duration of diabetes ranged from
Identification

5,209 records identified through database searching

2 additional records identified through manual-searching

Screening

4,238 records after duplicates removed

4,238 records screened

4,178 records excluded

Eligibility

60 full-text articles assessed for eligibility

42 full-text articles excluded:
1. Other telehealth interventions (n=7)
2. No HbA1c data (n=1)
3. Not an RCT (n=19)
4. Participants aged <18 (n=2)
5. Irrelevant control group (n=2)
6. Ongoing study (n=1)
7. Conventional diabetes education program (n=1)
8. Conference abstracts (n=8)

Included

14 studies (18 papers) included in qualitative synthesis

14 studies (18 papers) included in quantitative synthesis (meta-analysis)

Type 1 Diabetes Apps

Three apps were used for participants with type 1 diabetes and aimed to help patients to calculate the most appropriate insulin bolus on the basis of patient blood glucose (BG) levels, food intake, and physical activity. Data for all three apps were manually entered. One study reported that there was little impact of the app on the total time spent on face-to-face or telephone follow-up and concluded that the software did not require more time for patients to manage their diabetes (13). A further study estimated the average cost to patients and educators’ time was £38 per patient, attributed to the app over a 9-month period (15). HCP feedback was provided in all apps, with a frequency ranging from every week to every 3 weeks (Supplementary Table 4).

Type 2 Diabetes Apps

Nine apps were used for participants with type 2 diabetes. The apps were designed to improve patient self-management by providing personalized feedback on self-monitoring data, such as BG, food intake, and physical activity. In eight of the apps, BG was automatically transferred and other data was manually entered, with one exception where blood pressure, body weight, and pedometer were also automatically transferred (25). Quinn et al. (17) reported that the app was associated with shorter consultation times. Among seven apps with HCP feedback, three provided feedback when needed (e.g., patient data were considered abnormal). In the other apps, the frequency of feedback ranged from once a week to once every 3 months (Supplementary Table 4).

Effectiveness of the Apps

Type 1 Diabetes

There were mixed results from the type 1 diabetes studies. Two studies (14,16) found no difference between the intervention group and the control group and two studies (13,15) reported statistically significant results that favored the apps. There was a statistically insignificant difference in HbA1c between the apps and control group of -0.36% (95% CI -0.87, 0.14; P = 0.16; I² = 87%) (Fig. 2). No subgroup analyses were reported.

Type 2 Diabetes

All 10 studies of type 2 diabetes reported a reduction of HbA1c in participants using an app, with a median reduction of 0.55% (range 0.15–1.87). After pooling, the mean reduction in HbA1c was 0.49% (95% CI 0.30, 0.68; P < 0.01; I² = 10%) (Fig. 3). These results exhibited consistent findings with no heterogeneity. One study reported a reduction larger than clinically anticipated, which raised debate over the legitimacy of their findings (26). After excluding the subgroup of studies that were assessed as poor quality, we found a mean reduction of 0.41% (95% CI 0.22, 0.61; P < 0.001; I² = 0%) (Fig. 3). The level of evidence by GRADE was...
moderate due to the findings being downgraded due to quality.

Type 2 Diabetes Subgroup Analyses
The subgroup analysis by follow-up duration showed that five studies with a shorter follow-up duration (<6 months) displayed a larger (but nonsignificant) HbA1c reduction than those with a longer duration (>6 months), 0.62 vs. 0.40% (P = 0.33), respectively. There was no difference in the reduction of HbA1c in three studies with a mean diabetes duration of ≤9 years (0.53%) compared with those with a duration ≥9 years (0.55%; P = 0.93). Studies of younger participants with a mean age of ≤55 years reported a larger and clinically significant reduction in HbA1c level of 1.03% compared with those with an average age of >55 years, but the result was not found to be statistically significant (P = 0.10).

In the subgroup analysis by number of self-monitoring tasks, six diabetes apps supported at most three self-monitoring tasks and had results similar to the studies with more than three self-monitoring tasks (mean reduction of 0.44 vs. 0.58%; P = 0.56). Two studies of diabetes apps with only automated feedback had a small and statistically nonsignificant change in HbA1c of −0.26% (95% CI −0.62, 0.09). When the diabetes apps that included HCP feedback were pooled, eight studies reported a reduction of 0.56% (95% CI 0.35, 0.78). There was no statistically significant difference between HCP versus automatic feedback subgroup (P = 0.16).

Four sensitivity analyses were undertaken to test the robustness of the results. Removing three studies (20, 23, 26) with poor quality reported a mean reduction of 0.41% (95% CI 0.22, 0.61) (Fig. 3). The removal of one study (17) with incomplete statistical information was associated with a mean reduction of 0.48% (95% CI 0.28, 0.67), and the exclusion of one study (20) conducted on mixed participants with type 1 and type 2 diabetes had an attendant mean reduction of 0.48% (95% CI 0.27, 0.69). Finally, the exclusion of two studies (17, 23) with baseline HbA1c levels >9.0% was associated with a mean reduction of 0.47% (95% CI 0.25, 0.69).

**CONCLUSIONS**

Ten studies were included for type 2 diabetes, predominantly of fair quality. The results of these studies indicated a consistent reduction in HbA1c of 0.5%. Although there was no indication of heterogeneity, the study conducted by Takenga et al. (26) introduced a large effect that was likely to be caused by poor study quality (high attrition rate, differential loss to follow-up, and high baseline HbA1c level). Thus, studies were stratified into subgroups determined by their quality assessment (27). No differences were found between the subgroups, and the studies of poor quality were included for completeness and to highlight the challenges in study design.

Five subgroup analyses showed that the effect did not differ significantly by follow-up duration, mean diabetes duration of participants, mean age of participants, number of self-monitoring tasks supported by the diabetes apps, or types of feedback. Compared with studies that investigated the effectiveness of alternative interventions such
as text messaging, mobile device use, and computer-based and conventional self-management, we have found that apps offer promising results and reinforce the message argued by other authors (3,4,28–30). The evidence for this finding by GRADE was moderate, after downgrading due to quality.

The subgroup analysis by follow-up duration suggested that the effect of diabetes apps on BG control may attenuate over time. A possible rationale for this subgroup effect is a lack in user friendliness, a lack in perceived additional benefits, and a lack of use of gamification elements, resulting in a lack of efficacy following use (31). The subgroup analysis by mean age of participants indicated that younger patients were more likely to benefit from the use of the diabetes apps. It may be speculated that younger patients are more amenable to new technologies and more familiar with the use of mobile phones. The subgroup analysis by personalized feedback system highlighted the gap between automated feedback and HCP feedback. Although automated feedback has the advantage of being interactive and dynamic, there is a limit to presupposed scenarios, whereas feedback provided by HCPs was more individual, especially in emergency situations. Feedback options ranged widely between the apps, but it is postulated that it was the feedback that triggered improved lifestyle choices, which in turn lowered HbA1c. None of the five sensitivity analyses changed the overall effect size significantly, which suggests that the findings are not sensitive to these scenarios. The results of our meta-analysis lend support to the use of diabetes apps in diabetes self-management, especially for type 2 diabetes. However, we have highlighted a number of limitations of current diabetes apps.

For type 1 diabetes, there was little difference in HbA1c between intervention and control groups and the results were associated with considerable heterogeneity. The level of evidence by GRADE was downgraded to very low due to study quality, inconsistency, and uncertainty, so the findings should be interpreted as very uncertain and likely to change after future research. Furthermore, none of the apps in the included type 1 diabetes studies had

<table>
<thead>
<tr>
<th>Study or subgroup</th>
<th>Intervention</th>
<th>Control</th>
<th>Difference in means</th>
<th>Forest Plot</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean (%)</td>
<td>SD (%)</td>
<td>N</td>
<td>Weight (95% CI)</td>
</tr>
<tr>
<td>Orsama et al. 2013a</td>
<td>-0.37</td>
<td>0.83</td>
<td>23</td>
<td>16.31%</td>
</tr>
<tr>
<td>Holmen et al. 2014a</td>
<td>-0.31</td>
<td>1.10</td>
<td>39</td>
<td>15.95%</td>
</tr>
<tr>
<td>Faridi et al. 2008a</td>
<td>-0.10</td>
<td>0.30</td>
<td>15</td>
<td>13.16%</td>
</tr>
<tr>
<td>Waki et al. 2014b</td>
<td>6.70</td>
<td>0.70</td>
<td>27</td>
<td>15.19%</td>
</tr>
<tr>
<td>Nagrebetsky et al. 2013a</td>
<td>6.90</td>
<td>0.70</td>
<td>7</td>
<td>2.67%</td>
</tr>
<tr>
<td>Yoo et al. 2009b</td>
<td>7.10</td>
<td>0.80</td>
<td>57</td>
<td>32.54%</td>
</tr>
<tr>
<td>Quinn et al. 2008b</td>
<td>7.48</td>
<td>1.10</td>
<td>13</td>
<td>4.17%</td>
</tr>
<tr>
<td><strong>Subtotal</strong></td>
<td>181</td>
<td>180</td>
<td>100%</td>
<td>-0.41 (-0.61, -0.22)</td>
</tr>
<tr>
<td><strong>Z = 4.24, P &lt; 0.001</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Heterogeneity: Tau² = 0, Q-value = 2.53, df = 6, P = 0.86, I² = 0%</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2. studies with poor quality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Isteponian et al. 2009b</td>
</tr>
<tr>
<td>Takenga et al. 2014a</td>
</tr>
<tr>
<td>Quinn et al. 2011b</td>
</tr>
<tr>
<td><strong>Subtotal</strong></td>
</tr>
<tr>
<td><strong>Z = 2.70, P = 0.01</strong></td>
</tr>
<tr>
<td><strong>Heterogeneity: Tau² = 0.19, Q-value = 4.54, df = 2, P = 0.10, I² = 55.91%</strong></td>
</tr>
<tr>
<td><strong>Test for differences: Q-value = 2.05, df = 1, P = 0.15</strong></td>
</tr>
<tr>
<td><strong>Total</strong></td>
</tr>
<tr>
<td><strong>Z = 5.07, P &lt; 0.01</strong></td>
</tr>
<tr>
<td><strong>Heterogeneity: Tau² = 0.01, Q-value = 10.04, df = 9, P = 0.35, I² = 10.39%</strong></td>
</tr>
</tbody>
</table>

Mean values reported as:

a: change in mean HbA1c from baseline to follow-up
b: follow-up mean HbA1c

**Figure 3**—Pooled type 2 diabetes studies of HbA1c comparison of apps vs. control.
an automatic data uploading functionality. In future studies for type 1 diabetes, we encourage investigators to include apps with this functionality, not only for the purpose of being user friendly but also for safety concerns by reducing the risk of data entry errors.

Two studies reported on the cost-effectiveness of the apps for type 1 diabetes with inconclusive findings (15,16). Of three studies on type 2 diabetes that discussed compliance, two highlighted poor compliance, with only 35% of patients being recorded as regular app users (21,24). One study (25) reported a decline in patient use over time, from 70% in the 1st week to 50% in the last 2 weeks. Four studies tried to explore the mechanisms behind the effects, but the conclusions were inconsistent (15,17,21,24). We postulate that diabetes apps influence lifestyle choice, but how this occurs is unclear. One hypothesis is that the reminder and feedback features of diabetes apps can lead to improvement in health beliefs, self-efficacy, and social support (32).

By the end of the decade, worldwide mobile phone usage is anticipated to exceed 5 billion (33). Therefore apps may be able to offer an affordable and widely available adjunct to diabetes self-management. We have included studies across a variety of health care systems, from both the developed and developing world, so we argue that the apps are currently available and could form the basis of improved health promotion on diabetes education and self-management.

This study had several limitations. Since this review was restricted to published studies, publication bias cannot be ruled out, as highlighted by other investigators (30). All included study designs were not blinded and so were downgraded in the quality assessment tool (highlighting the increased risk of ascertainment bias). Furthermore, patient-important outcomes and behavioral mechanisms were not considered and highlighted as a clear gap to be addressed in future studies. A further weakness is that some of the effect attributed to the apps could be explained by health care providers. Finally, there is no clear definition of diabetes apps, and study authors defined their interventions in different ways as a result. In this review, we defined diabetes apps as software that is designed for use on a mobile phone allowing patients to enter data into the app and receive feedback. The implications for future research include establishing a common standardized platform of functionality. Investigators of future studies need to consider adequately powered pragmatic RCTs with secure sequence generation, concealed allocation, use of an active control app, and comparable access to HCPs. Features such as these might reduce the impact of ascertainment bias, and effects due to HCPs. RCTs with longer duration of follow-up (>6 months) using standardized app technology may well demonstrate beneficial clinical effects in type 2 diabetes. Furthermore, there is significant scope for research in the use of apps in other areas of self-management, such as increasing physical activity, weight loss, and smoking cessation.

In a clinical context, we recommend that HCP feedback should be central in all future app designs and supplemented with dynamic automated feedback. Future technology should also be underpinned by behavior change theories and gamification elements to achieve a larger effect on BG control and improve compliance of patients in using diabetes apps. Finally, future technology should also consider the needs of older patients.

Acknowledgments. The authors acknowledge Dr. Haroon Ahmed, of the Division of Population Medicine, Cardiff University School of Medicine, and the editorial base at Diabetes Care for their clinical and methodological guidance that strengthened the manuscript.

This article is an honest, accurate, and transparent account of the study being reported; no important aspects of the study have been omitted, and any discrepancies from the study as planned (and, if relevant, registered) have been explained.

Duality of Interest. No potential conflicts of interest relevant to this article were reported.

Author Contributions. C.H. designed the protocol, searched the literature, extracted the data, conducted the analysis, and drafted the manuscript. B.C. interpreted the results and drafted the manuscript. J.H. reviewed the manuscript and advised on the clinical context of the review. T.F. searched the literature and extracted the data. S.M. designed the protocol, interpreted the results, and contributed to the manuscript. B.C. is the guarantor of this work and, as such, had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

References