Journal of Public Health Sciences
Vol. 5, No. 03, pp. 871-890
journal.iistr.org/index.php/JPHS
DOI: 10.56741/IISTR.jphs.002285

Effectiveness of mHealth Applications with Self-Monitoring Features on Glycemic Control in Type 2 Diabetes Mellitus

A Systematic Review with Synthesis Without Meta-analysis (SWiM)

[1]Agung Rahmat Fauzi, [1]Mahendro Prasetyo Kusumo*
Corresponding Author: * mahendroprasetyo@umy.ac.id
[1] Universitas Muhammadiyah Yogyakarta, Yogyakarta, Indonesia

ARTICLE INFO

Article history
Received 6 May 2026
Revised 11 September 2026
Accepted 24 September 2026

ABSTRACT

Diabetes mellitus is a growing global health concern, and inadequate glycemic control increases the risk of long-term complications. This review evaluated the effectiveness of mobile health (mHealth) applications incorporating self-monitoring features on glycated hemoglobin (HbA1c) and secondary outcomes among adults with type 2 diabetes receiving facility-based care. PubMed and ProQuest were searched for studies published from January 2021 to December 2025, supplemented by backward citation searching and verification on August 15, 2026. Completed quantitative studies were analyzed separately from ongoing trial protocols. Ten records were eligible: eight completed studies involving 1,207 participants across seven countries and two protocols. Owing to heterogeneous comparators, follow-up periods, and outcome reporting, together with missing dispersion estimates in five studies, findings were synthesized using Synthesis Without Meta-analysis (SWiM) methods and an effect-direction plot. Six studies reported significant between-group HbA1c reductions ranging from 0.49 to 1.16 percentage points. One study found a within-group reduction without a between-group difference, while another reported no meaningful change (−0.11%, 95% CI −0.24 to 0.02). Larger improvements were observed descriptively in interventions combining applications with scheduled human support. Some studies also reported improvements in self-management behaviors, weight, blood pressure, and satisfaction, although measures varied. GRADE certainty was low for HbA1c and low to very low for secondary outcomes, primarily due to lack of blinding,

attrition, and imprecision. Self-monitoring mHealth applications may modestly improve glycemic control when integrated into supported care, but evidence remains insufficient to identify the most effective intervention components.

Keywords
Diabetes Mellitus
Glycemic Control
HbA1c
mHealth Application
Self-monitoring

This is an open-access article under the CC–BY-SA license.

Introduction

Diabetes mellitus (DM) affects more than 537 million adults worldwide, and the International Diabetes Federation projects that this figure will reach 783 million by 2045 [11]; Global Burden of Disease estimates point in the same direction, with prevalence projected to keep rising through 2050 [12]. Persistent hyperglycemia, whether it arises from insufficient insulin secretion or from impaired insulin action, causes cumulative damage to the cardiovascular, renal, ophthalmic and peripheral nervous systems. Glycated hemoglobin (HbA1c) reflects average glucose exposure over the preceding two to three months and remains the reference measure against which control is judged [13]. Type 2 diabetes mellitus (T2DM) accounts for the large majority of cases and is the population addressed here. Throughout this article the abbreviation T2DM is used whenever a statement is specific to type 2 disease, and DM is reserved for statements that apply to diabetes in general.

Despite the availability of pharmacological therapy and evidence-based clinical guidelines, many patients with T2DM treated in healthcare facilities do not attain their glycemic targets. The gap is repeatedly attributed to limited self-management: irregular self-monitoring of blood glucose (SMBG), inconsistent medication adherence, suboptimal dietary practice and physical inactivity [14]. Structural constraints compound the problem, since short consultations, thin health-literacy support and infrequent follow-up all limit how much individualized care a clinic can deliver.

mHealth applications have been proposed as scalable digital adjuncts to these gaps. By recording SMBG values, tracking diet and activity, prompting medication use and opening a channel for clinician communication, such applications extend structured diabetes care beyond the consultation room. More advanced implementations link self-monitoring data to electronic medical records (EMRs) or to continuous glucose sensors so that clinical decisions can be made on patient-generated data [3], [8], [9]. The terms EMR and electronic health record (EHR) are used in this article as the original authors used them, and are treated as functionally equivalent for the purpose of this synthesis.

A knowledge gap nevertheless remains. Existing reviews have not evaluated the specific contribution of self-monitoring features in mHealth applications deployed inside healthcare facilities. Most prior syntheses pool heterogeneous digital interventions, including telemonitoring, text messaging and web portals, without isolating structured self-monitoring [15], [16]. Evidence from Asian and Southeast Asian health systems is also thinly represented. Together these limitations constrain the ability of clinicians and policymakers to judge whether self-monitoring-enabled applications are worth adopting in facility-based care.

This review therefore synthesis the available evidence on the effectiveness of mHealth applications with self-monitoring features on HbA1c and on related diabetes management outcomes among adults managed in healthcare facilities. In response to the heterogeneity encountered, the synthesis follows the Synthesis Without Meta-analysis (SWiM) reporting guideline [17], presents effect estimates with confidence intervals where these could be obtained, displays results in an effect-direction plot [18], reports domain-level risk of bias, and rates the certainty of the evidence using GRADE [19].

Literature Review

Diabetes self-management education and support (DSMES) is a cornerstone of effective T2DM care. International guidelines recommend structured education covering nutrition, physical activity, medication adherence, glucose monitoring and psychosocial support [20]. Face-to-face DSMES nevertheless scales poorly: it depends on provider availability, requires patients to attend scheduled appointments and offers little continuous reinforcement between visits. These constraints help explain why glycemic targets are missed even in well-resourced systems, and they motivate digital strategies that carry DSMES into daily life.

mHealth applications draw on established behavioral theories, among them the Health Belief Model, Social Cognitive Theory and the Transtheoretical Model, to sustain self-management behavior. Goal-setting, automated reminders, self-monitoring feedback loops and gamification map onto these frameworks by reinforcing self-efficacy and motivation [4], [10]. Theory-driven design on its own, however, appears insufficient. Recent pooled analyses suggest that applications achieving the strongest behavioral outcomes are those that combine automated feedback with human interaction, whether through clinician messaging, coaching calls or family involvement [15]. Work in Indonesian primary care points the same way: a smartphone application improved knowledge, attitudes and self-care behavior among people with diabetes, but the effect depended on how the application was introduced and supported by staff [21].

More advanced configurations connect self-monitoring data to clinical workflows. EMR-linked applications let clinicians review patient-generated data between appointments, which supports timely medication adjustment and personalized counselling [6], [9]. Hardware

innovations such as implantable continuous glucose sensors paired with an application supply resolution that passive self-reporting cannot reach [8]. Triadic models involving patients, family members and pharmacists or nurses have shown promise among older adults and lower-literacy populations [10].

The evidence base nonetheless remains heterogeneous in design, population, intervention components and follow-up duration. Two recent meta-analyses illustrate both the promise and the imprecision of the field: pooled digital interventions reduced HbA1c by 0.31% (95% CI −0.45 to −0.16) compared with usual care [15], while a synthesis restricted to digital management with self-monitoring reported a larger pooled reduction of 0.52% (95% CI −0.63 to −0.42) [16]. Neither isolated the incremental contribution of the self-monitoring component within facility-delivered programmes, which is the gap the present review addresses.

Materials and Methods

A. Study Design and Reporting Guidance

This is a systematic review conducted in line with the PRISMA 2020 statement [22]. The PICO framework (Table 1) defined the research question and guided the search, screening and data extraction. Because the anticipated heterogeneity made statistical pooling questionable from the outset, the synthesis was planned and is reported according to the SWiM reporting guideline [17], and the certainty of the evidence was rated with GRADE [19]. The review was not registered prospectively, which is acknowledged as a limitation.

Table 1. PICO framework used in this systematic review

Component	Description
P – Population	Adults (≥18 years) with type 2 diabetes mellitus (T2DM) receiving care in healthcare facilities (hospitals, primary care clinics, community health centres or community pharmacies). Studies of type 1 diabetes alone were not eligible; mixed populations were eligible only when T2DM results were reported separately.
I – Intervention	mHealth application with at least one dedicated self-monitoring feature (blood glucose recording, dietary log, physical activity tracking, medication adherence prompt, or a feedback module built on self-monitored data)
C – Comparison	Usual care (standard outpatient diabetes management) or an alternative intervention (telephone follow-up, paper-based education) without a dedicated mHealth self-monitoring component
O – Outcome	Primary: change in HbA1c. Secondary: fasting blood glucose, body weight or BMI, blood pressure, diabetes self-management behavior, medication adherence, patient satisfaction and quality of life
Evidence sets	Effectiveness set: completed studies reporting at least one quantitative outcome. Contextual set: published protocols of registered ongoing trials, used only to describe the research pipeline and excluded from every effect estimate.

B. Search Strategy

A structured search was performed in PubMed and ProQuest for articles published between January 2021 and December 2025. Search terms were combined with Boolean

operators: ("mobile health" OR "mHealth" OR "smartphone application" OR "mobile application") AND ("self-monitoring" OR "self-management") AND ("diabetes mellitus" OR "type 2 diabetes" OR "T2DM") AND ("HbA1c" OR "glycemic control" OR "blood glucose"). The PubMed string was: (("mobile health"[tiab] OR "mHealth"[tiab] OR "smartphone application"[tiab] OR "mobile application"[tiab]) AND ("self-monitoring"[tiab] OR "self-management"[tiab]) AND ("diabetes mellitus"[MeSH Terms] OR "type 2 diabetes"[tiab] OR "T2DM"[tiab]) AND ("HbA1c"[tiab] OR "glycemic control"[tiab] OR "blood glucose"[tiab])), filtered to 2021/01/01–2025/12/31. An equivalent string was run in ProQuest through its Advanced Search interface using subject headings and thesaurus terms mapped to the same concepts. No language filter was applied at the identification stage, but only English-language reports were retained because translation was not feasible.

Two databases were selected deliberately rather than by convenience. PubMed indexes the MEDLINE corpus and therefore captures the clinical and digital-health journals in which trials of facility-based diabetes applications are overwhelmingly published, while ProQuest adds multidisciplinary and grey coverage, including dissertations and regional health-science titles from Southeast Asia that MEDLINE indexes incompletely. Scopus, Web of Science, Embase, CINAHL and IEEE Xplore were not available through the authors' institutional subscription at the time of the review.

To reduce the risk that eligible studies were missed, three verification steps were added after the reviewers' comments. First, the reference lists of all included studies were screened (backward citation searching). Second, the reference lists of three recent systematic reviews and meta-analyses of digital diabetes interventions [15], [16], [14] were screened in the same way. Third, a cross-check of the first 200 hits of an equivalent Google Scholar query was performed. The final search across all sources was run on 15 August 2026, and none of these steps produced an additional study that met the eligibility criteria. The possibility that studies indexed only in Embase, CINAHL or IEEE Xplore were missed cannot be excluded and is stated among the limitations.

C. Eligibility Criteria and the Two Evidence Sets

Records had to satisfy the following criteria: (a) written in English; (b) adults aged 18 years or older with T2DM; (c) an intervention using a mobile or smartphone application with at least one dedicated self-monitoring feature; (d) delivery within a healthcare facility; and (e) at least one glycemic or self-management outcome reported or, for protocols, pre-specified. Editorials, narrative reviews without primary data, conference abstracts without a full text, and studies confined to non-T2DM populations were excluded, as were secondary analyses of a trial already represented in the review.

The handling of study protocols was specified before synthesis began, and this replaces the single undifferentiated inclusion list used in the earlier version of this manuscript. Two mutually exclusive evidence sets were defined. The effectiveness set comprises completed studies that report at least one quantitative outcome; only these studies contribute to effect estimates, to the subgroup synthesis, to the risk-of-bias assessment and to the GRADE ratings. The contextual set comprises published protocols of registered ongoing trials that met every population, intervention and setting criterion and that were conducted in Asian health systems identified a priori as under-represented among completed trials; these protocols are described separately in Section 4.8 and are used only to map the research pipeline. Protocols that did not meet the contextual-set criteria were excluded at full-text screening (n = 7) and appear as such in Fig. 1. Consequently the review reports eight effectiveness studies plus two ongoing-trial protocols, rather than ten equivalent included studies.

D. Study Selection and Data Extraction

Records were screened sequentially by title, abstract and full text, with duplicates removed beforehand. Two reviewers assessed eligibility independently and resolved disagreements by discussion. Extracted data covered authors and year, country, study design, analyzed sample size, population characteristics, intervention details including whether scheduled human support was provided and by which professional, duration, comparator, primary outcome (change in HbA1c), secondary outcomes and conclusions. For every effectiveness study the extraction was verified a second time against the published full text, and the effect estimate, its comparison type (between-group or within-group), the timepoint to which it refers, and any reported 95% confidence interval were recorded explicitly. Where an interval was not reported but could be derived from group-specific estimates given by the authors.

E. Risk of Bias and Quality Appraisal

Risk of bias was assessed independently by two reviewers. For randomized trials the revised Cochrane tool (RoB 2) was applied across five domains: randomization process, deviations from intended interventions, missing outcome data, measurement of the outcome, and selection of the reported result [23]. For non-randomized studies ROBINS-I was used [24]. Disagreements were settled by discussion.

F. Data Synthesis

A random-effects meta-analysis was considered and rejected for three reasons. First, the comparators differ in kind: usual care [1], [5], [8], [9], [10], a paper-based diary [3], telephone follow-up as a third arm [2], and standard care within a single-blind design [4]. Second, the timepoint at which the primary glycemic endpoint was measured ranges from 4 weeks [8] to 9 months [10], and two studies report attenuation between an interim and a final measurement [4], [9], so a single pooled estimate would obscure rather than summarize. Third,

and decisively, five of the eight studies report neither a standard deviation of change nor a confidence interval for the between-group difference, so the variance required for inverse-variance weighting is unavailable and would have to be imputed for the majority of the evidence. Pooling under those conditions would create a precision the primary literature does not support.

The synthesis therefore follows SWiM [17]. Studies were grouped for synthesis by intervention characteristic rather than by outcome alone, and four groupings were specified: presence or absence of scheduled human support; integration or non-integration with an EMR or EHR; intervention length of 12 weeks or less versus longer; and explicit cultural or linguistic adaptation versus none. Within each grouping the synthesis reports the number of studies, the available between-group effect estimates and their range, and the number of studies whose result favored the mHealth arm. The standardized metric for synthesis is the absolute between-group difference in HbA1c expressed in percentage points; where only a within-group change was reported, this is stated and the study is not counted as showing a between-group effect. Vote counting based on statistical significance alone was avoided as the primary method, and results are presented as an effect-direction plot in which symbol area is proportional to sample size and overall risk of bias is shown alongside each study [18]. With eight outcome studies, formal meta-regression was not attempted.

G. Certainty of Evidence

Certainty was rated for HbA1c and for each major secondary outcome using GRADE [19]. Bodies of evidence from randomized trials started at high certainty and were downgraded for risk of bias, inconsistency, indirectness, imprecision and publication bias. Publication bias could not be assessed statistically because fewer than ten studies contributed to any outcome, and it is reported as undetected rather than absent. Ratings and the reasons for each downgrade are shown in Table 5.

H. Ethical Consideration

This review uses previously published literature, so no participants were involved and separate ethical approval was not required. All included studies reported approval from their respective institutional review boards.

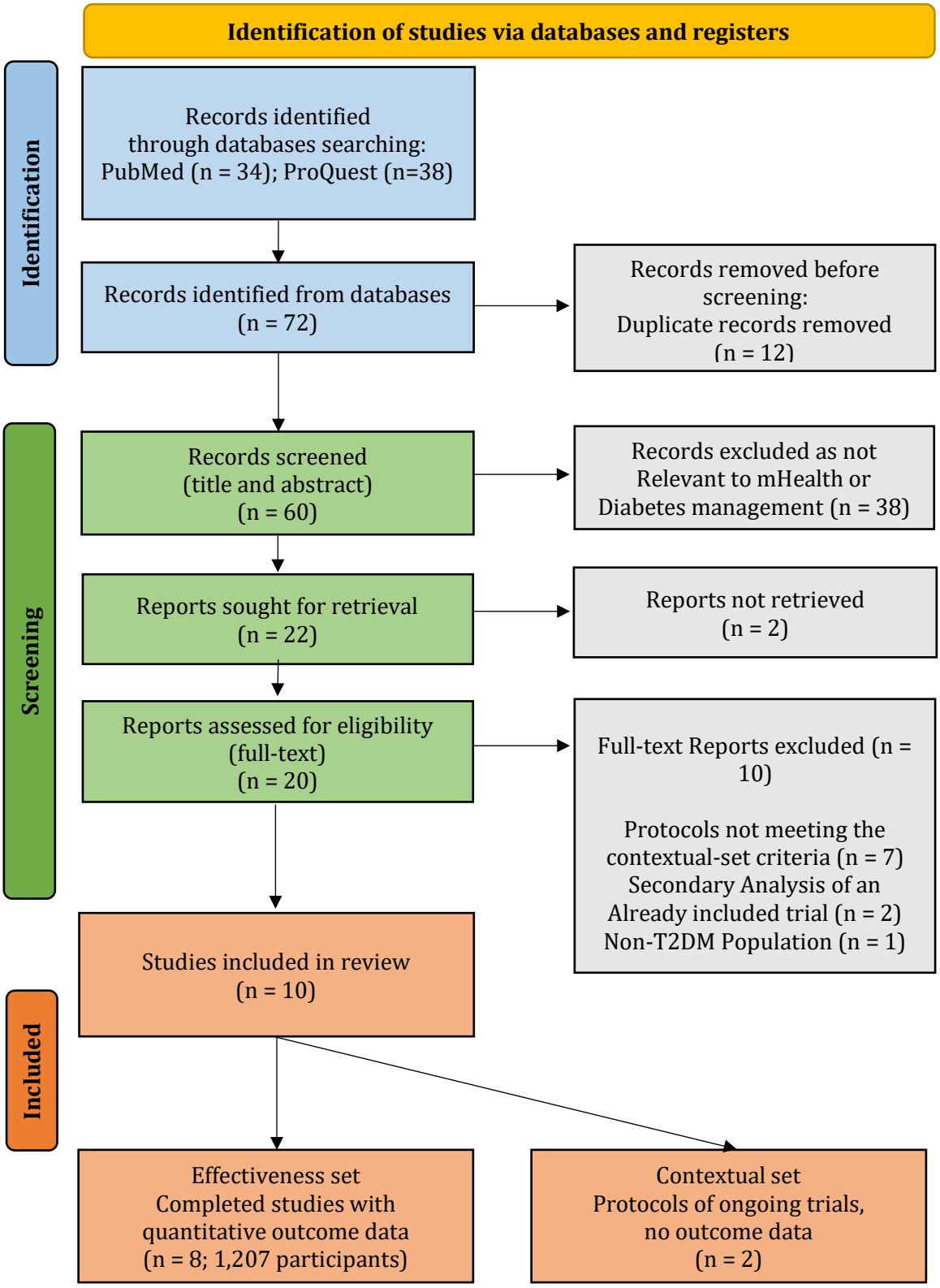

Fig. 1. PRISMA 2020 flow diagram of study selection

Results

A. Study Selection

Database searching identified 72 records (PubMed, n = 34; ProQuest, n = 38). After 12 duplicates were removed, 60 records were screened by title and abstract and 38 were excluded as unrelated to mHealth or diabetes management. Of the 22 reports sought, 2 could not be retrieved, leaving 20 full texts assessed for eligibility. Ten were excluded: protocols that did not meet the contextual-set criteria (n = 7), secondary analyses of an already included trial (n = 2) and a non-T2DM population (n = 1). Ten records were included. Eight of these forms the effectiveness set and contribute 1,207 participants across seven countries; the remaining two are ongoing-trial protocols in the contextual set and contribute no outcome data. The selection process is shown in Fig. 1.

B. Characteristic of Included Studies

The eight completed studies were conducted in seven countries: Vietnam [1], Thailand [2], [10], Japan [3], Spain [4], Germany [5], China [8] and South Korea [9]. Seven were randomized trials and one used a quasi-experimental design [1]. Analyzed sample sizes ranged from 68 [8] to 269 [9], giving 1,207 participants in total. Intervention length ranged from 4 weeks [8] to 9 months [10]. All participants were adults with T2DM managed in hospital outpatient departments, primary care clinics, community health centres or community pharmacies. Characteristics are summarized in Table 2. The two protocols [6], [7] are described separately in Section 4.8 and are deliberately absent from Table 2, from Table 3 and from every effect statement below.

Table 2. Characteristics of the eight studies in the effectiveness set

No	Author	Year	Country	Study Design	Analyzed (n)	Application and support model	Comparator	Duration
1	Nguyen et al. [1]	2025	Vietnam	Quasi-experimental, Multicentre, open-label	163 (81/82)	diaB: 26-lesson DSMES curriculum with twice-weekly group video sessions led by physicians, dietitians and health coaches, plus messaging-based coaching	Usual care	10 weeks
2	Riangkam et al. [2]	2022	Thailand	3-arm RCT	129 (43/43/ 43)	MHDSMP: self-monitoring application with motivational text messages and nurse-delivered telephone coaching	Telephone follow-up; usual care	12 weeks
3	Leung et al. [3]	2024	Japan	2-arm RCT, open-label	149	SMC-01: glucose, diet and activity recording with guideline-based automated feedback and no scheduled human contact	Paper-based self-monitoring diary	24 weeks

No	Author	Year	Country	Study Design	Analyzed (n)	Application and support model	Comparator	Duration
4	Ruiz-Leon et al. [4]	2025	Spain	Single-blind RCT	123	Greenhabit: gamified daily challenges covering diet, physical activity and psychological wellbeing	Standard care	12 weeks
5	Bretschneider et al. [5]	2025	Germany	Multicenter RCT	149 (73/76)	Vitadio: structured digital therapeutic with dietary and activity self-monitoring and remote coach feedback	Standard care	6 months
6	Guo et al. [8]	2021	China	Pilot RCT	68 (34/34)	Implantable continuous glucose sensor paired with a mobile application, with general-practitioner guidance delivered through the application	Usual care	4 weeks
7	Lee et al. [9]	2022	South Korea	3-arm RCT	269	iCareD: EMR-integrated self-care application, with and without scheduled physician feedback	Usual care	26 weeks
8	Poonprapai et al. [10]	2022	Thailand	2-arm RCT	157 (78/79)	Pharmacist-delivered family-support model in which family members received education and prompts through the application and reinforced self-monitoring at home	Usual care	9 months

Note: RCT (randomized controlled trial); DSMES (diabetes self-management education and support); EMR (electronic medical record. Study identifiers follow the IEEE reference numbers used throughout this article)

C. Risk of Bias

No randomized trial was judged at low risk of bias overall. Five were rated as raising some concerns [2], [3], [4], [5], [10] and two were rated at high risk [8], [9]. The most common source of concern was the second domain: participants and, in most cases, the professionals delivering care could not be blinded to an application-based intervention, and adherence was usually self-reported. Missing outcome data drove the high rating for the 26-week trial [9], where attrition rose steadily over follow-up. For the pilot trial [8] the concern is one of outcome measurement rather than attrition: HbA1c integrates glucose exposure over approximately two to three months, so a change measured after a 4-week intervention cannot be interpreted in the same way as a change measured at 12 weeks or later, and the reported reduction of 1.16 percentage points should be treated with corresponding caution. The single non-randomized study [1] was assessed with ROBINS-I and rated at serious risk overall, driven by confounding, since allocation followed site rather than randomization and no propensity adjustment was reported, with moderate concerns for participant selection, deviations from intended interventions and selection of the reported result. Domain-level judgements for the randomized trials are shown in Fig. 2.

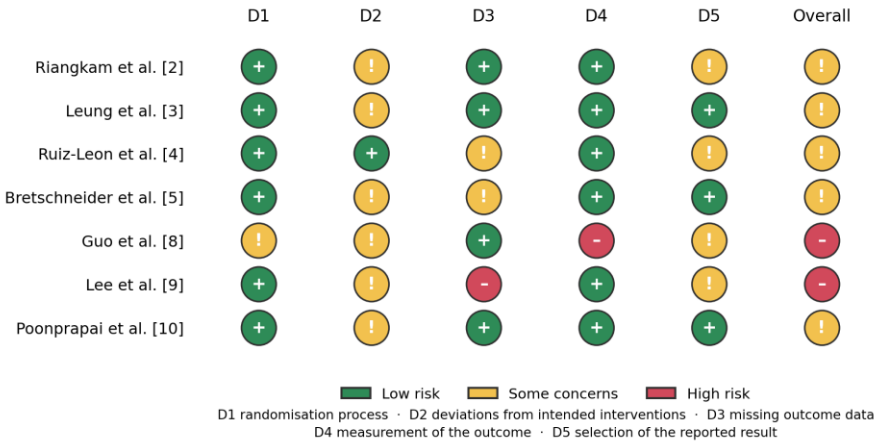

Fig. 2. Domain-level risk-of-bias judgements (RoB 2) for the seven randomized trials in the effectiveness set. The single non-randomized study [1] was appraised with ROBINS-I and rated at serious risk of bias overall

D. Effect on Glycemic Control (HbA1c)

Effect estimates, the comparison from which each was taken, the timepoint to which it refers and the corresponding confidence interval are set out in Table 3. Six of the eight studies reported a statistically significant between-group reduction in HbA1c [1], [2], [5], [8], [9], [10], with estimates ranging from 0.49 to 1.16 percentage points. One study reported a within-group reduction of 0.4 percentage points in the intervention arm, but the control arm also improved and no between-group difference was demonstrated [4]. One study found no difference (−0.11 percentage points, 95% CI −0.24 to 0.02) [3]. A reduction of at least 0.5 percentage points is conventionally treated as clinically meaningful in T2DM management [13], and five of the six between-group estimates meet or approach that threshold.

Two qualifications matter for interpretation. The estimate in Lee et al. [9] refers to week 12; by week 26 the difference was no longer statistically significant, which the authors attribute to declining engagement. The estimate in Guo et al. [8] is the largest in the set but comes from the smallest sample, the shortest intervention and the study at highest risk of bias. Read together, the estimates that survive the most scrutiny are the more modest ones: 0.53 percentage points in the multicentre trial with the most complete reporting [5] and 0.49 percentage points in the largest trial [9]. These sit close to the pooled estimates reported in the wider literature [15], [16].

Table 3. Structured evidence table for HbA1c

Study	Comparison from which the estimate is taken	Time-point	Effect estimate (HbA1c, percentage points)	95% CI	P	Direct-ion
Nguyen et al. [1]	Adjusted between-group difference (regression coefficient)	12 weeks	-0.72	-1.14 to -0.31	0.001	Favours mHealth

Study	Comparison from which the estimate is taken	Time-point	Effect estimate (HbA1c, percentage points)	95% CI	P	Direct-ion
Riangkam et al. [2]	Within-group change in the mHealth arm; between-group comparison by repeated-measures analysis of variance	12 weeks	-0.63	Not reported	<0.001 within-group; F = 12.28, p<0.001 between groups	Favours mHealth
Leung et al. [3]	Between-group difference in change, derived from the group-specific changes reported by the authors	12 weeks	-0.11	-0.24 to 0.03	0.11	No difference
Ruiz-Leon et al. [4]	Within-group change in the mHealth arm; the control arm also improved significantly	12 weeks	-0.40 (within-group)	Not reported	<0.001 within-group; between-group difference not demonstrated	Uncertain
Bretschneider et al. [5]	Adjusted between-group difference	6 months	-0.53	Not reported	<0.001	Favours mHealth
Guo et al. [8]	Between-group difference	4 weeks	-1.16	Not reported	0.014	Favours mHealth
Lee et al. [9]	Between-group difference	12 weeks; 26 weeks	-0.49 at 12 weeks; not significant at 26 weeks	Not reported	0.02 at 12 weeks	Favours mHealth at 12 weeks only
Poonprapai et al. [10]	Between-group difference (−0.97 in the mHealth arm versus −0.12 in the control arm)	9 months	-0.85	Not reported	0.001	Favours mHealth

Note: A negative estimate favors the mHealth arm. Confidence intervals are reported as published; the interval for study [3] was derived from the group-specific mean changes and intervals given by the authors. The absence of dispersion estimates in five of eight studies is the principal reason a pooled estimate was not calculated

Fig. 3 displays the direction of effect across all four outcome domains, with symbol area proportional to sample size and overall risk of bias shown in the final column. The pattern in the HbA1c column is consistent rather than uniform: improvement in most studies, no change in one, and an unresolved result in one. The plot also makes visible that the two studies with the darkest risk-of-bias markers sit at opposite ends of the effect range, which argues against reading the spread of estimates as a genuine gradient of effectiveness.

E. Structured Subgroup Synthesis

The four pre-specified groupings are summarized in Table 4. Studies that added scheduled human support to the application reported larger and more consistent between-group reductions than those relying on automated feedback alone, and this is the clearest pattern in the data. The comparison is nonetheless confounded, because the two applications without scheduled human support [3], [4] also differ from the others in comparator, in population and

in the density of behavior-change techniques, so human support cannot be isolated as the operative ingredient. The grouping by health-system integration is the weakest: only one completed study featured EMR integration [9], so no comparison is possible, and the claim that integration is an independent contributor to effectiveness is not supported by the present evidence. Shorter interventions were more often positive than longer ones, which is compatible either with a novelty effect or with the attenuation observed directly in the two studies that measured more than one timepoint [4], [9].

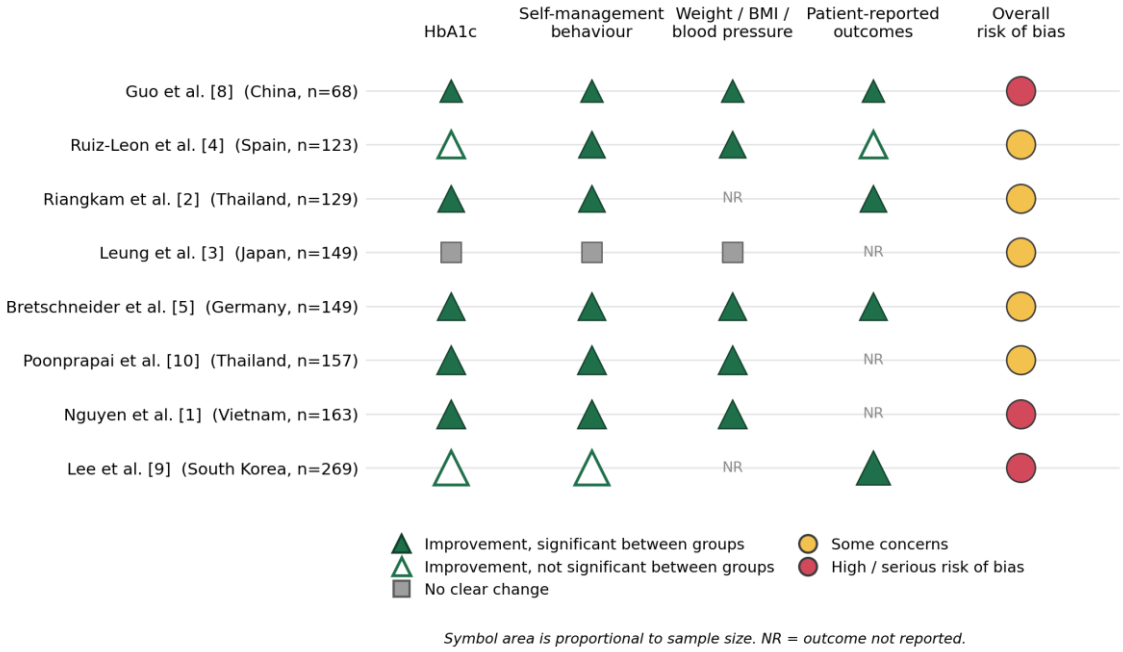

Fig. 3. Effect-direction plot for the effectiveness set across four outcome domains

Cultural or linguistic adaptation was described in the three Southeast Asian studies, all of which favoured the mHealth arm, but adaptation is entangled with region, delivery model and comparator, and with three studies no conclusion about adaptation as such can be drawn.

Table 4. Structured subgroup synthesis of between-group HbA1c effects

Grouping	Category	Studies (n)	Between-group estimation available	Range of estimates (percentage points)	Studies favouring the mHealth arm
Scheduled human support	Application plus scheduled human support ([1], [2], [5], [8], [9], [10])	6	5	-1.16 to -0.49	6 of 6
	Application with automated feedback only ([3], [4])	2	1	-0.11	0 of 2
EMR or EHR integration	Integrated ([9])	1	1	-0.49	1 of 1 (12 weeks only)
	Not integrated ([1]–[5], [8], [10])	7	5	-1.16 to -0.11	5 of 7

Grouping	Category	Studies (n)	Between-group estimation available	Range of estimates (percentage points)	Studies favouring the mHealth arm
Intervention length	12 weeks or less ([1], [2], [4], [8])	4	3	-1.16 to -0.63	3 of 4
	Longer than 12 weeks ([3], [5], [9], [10])	4	3	-0.85 to -0.11	2 of 4 at the final timepoint
Cultural or linguistic adaptation	Locally developed or explicitly adapted ([1], [2], [10])	3	2	-0.85 to -0.72	3 of 3
	Not described ([3], [4], [5], [8], [9])	5	4	-1.16 to -0.11	3 of 5

Note: These groupings are descriptive. They summarize how effects are distributed across study characteristics; they are not a formal test of moderation, and no causal claim about any single characteristic follows from them.

F. Secondary Outcomes

Beyond glycemic control, the interventions produced a broadly favorable but unevenly measured set of secondary results. Diabetes self-management behavior, covering dietary adherence, physical activity, medication use and regular SMBG, improved in the studies that included structured behavioral modules [1], [2], [5], [8], [10]. These behavioral changes matter mechanistically, since they represent the pathway through which self-monitoring would be expected to affect HbA1c. They were, however, measured with different instruments across studies and were self-reported in every case.

Cardiometabolic outcomes improved in several studies. Reductions in body weight, BMI and systolic blood pressure were reported for the Vitadio [5] and Greenhabit [4] applications, both of which paired dietary and activity monitoring with glycemic self-monitoring, and blood pressure also improved in the pharmacist-delivered family-support trial [10]. Patient-reported outcomes, including satisfaction, self-efficacy and diabetes distress, improved in most intervention arms; the EMR-integrated application recorded an 87.2% satisfaction rate [9] and the Thai programme showed significant gains in self-management competency [2]. Because these outcomes were reported by four or fewer studies each, and because satisfaction was in some cases measured only in the intervention arm, they are rated at low to very low certainty in Table 5.

G. Certainty of the Evidence

Table 5 sets out the GRADE assessment. Certainty was rated low for HbA1c: the body of evidence was downgraded once for risk of bias, since no study was at low risk overall and two were at high risk, and once for imprecision, because sample sizes were small and confidence intervals were unavailable for most estimates. Inconsistency was not downgraded, since the direction of effect is stable across studies and the variation in magnitude is largely explained by the subgroup structure in Table 4. Secondary outcomes were rated low to very low,

reflecting fewer contributing studies, self-reported measurement and, for satisfaction, instruments that were not consistently comparable. Publication bias could not be assessed statistically and is reported as undetected rather than absent.

Table 5. GRADE summary of findings for the effectiveness set

Outcome	Studies (participant)	Risk of bias	Inconsistency	Indirect-ness	Impre-cision	Certainty	Interpre-tation
HbA1c	8 (1,207)	Serious	Not serious	Not serious	Serious	⊕⊕◯◯ Low	mHealth applications with self-monitoring may produce a small reduction in HbA1c; the size of the effect is uncertain
Diabetes self-management behavior	5 (666)	Serious	Not serious	Not serious	Serious	⊕⊕◯◯ Low	These interventions may improve self-management behavior, measured by self-report
Body weight or BMI	4 (503)	Serious	Serious	Not serious	Serious	⊕◯◯◯ Very low	The effect on body weight is very uncertain
Systolic blood pressure	3 (429)	Serious	Serious	Not serious	Serious	⊕◯◯◯ Very low	The effect on systolic blood pressure is very uncertain
Patient satis-faction and diabetes distress	4 (615)	Serious	Not serious	Serious	Serious	⊕◯◯◯ Very low	The effect on patient-reported outcomes is very uncertain; instruments differed across studies

H. Ongoing Trials: The Contextual Set

Two published protocols met the contextual-set criteria and are reported here in isolation from every effectiveness statement above. The first describes a primary care-based, multicentre, cluster-randomized pragmatic trial in South Korea evaluating the WellCheck application linked to electronic health records, with a planned enrolment of 480 participants and change in HbA1c at 24 weeks as the primary endpoint [6]. Its comparison of application-plus-physician-support against application-only addresses precisely the question that the completed evidence cannot answer, since only one completed study featured EMR integration. The second is an exploratory sequential mixed-methods protocol from Yogyakarta, Indonesia, in which an Integrated Diabetes Self-Management application is developed from qualitative needs assessment, usability tested, and then evaluated over three months using a non-equivalent control group design [7]. Neither protocol reports outcome data, and no statement about the effectiveness of mHealth applications in this review rests on either of them. They are retained because they document the trajectory of facility-based mHealth research in two Asian health systems where completed trials are scarce.

Discussion

A. Patterns Associated with Greater Effectiveness

The synthesis identifies patterns associated with greater effectiveness rather than determinants of it, and the distinction is not merely semantic. With eight outcome studies and no formal moderator analysis, the groupings in Table 4 describe how effects are distributed across study characteristics; they cannot establish that any characteristic causes the difference.

The most consistent pattern concerns human support. Interventions that paired the application with scheduled contact from a professional or a trained family member reported larger between-group reductions than those relying on automated feedback alone: pharmacist-coordinated family support [10], physician feedback on EMR-linked data [9], group video sessions with clinicians and dietitians [1], nurse-delivered telephone coaching [2] and remote coach feedback [5] all fall on the favorable side, while the application offering recording and automated guidance without scheduled contact [3] did not differ from a paper diary. This is congruent with the stratified pooled analysis of Kerr et al., in which higher-intensity coaching was associated with a larger HbA1c reduction than lower-intensity delivery [15]. The reasonable reading is that self-monitoring functionality is necessary but not sufficient, and that it works when embedded in a supported care pathway. The reading that the present evidence cannot support is that human support has been shown to be the active ingredient, since it co-varies with comparator, population and behavioral content.

Two patterns that the earlier version of this manuscript presented as independent factors do not survive structured examination. Health-system integration cannot be evaluated at all, because only one completed study featured EMR linkage [9]; the claim that integration modulates effectiveness rests on a single study whose own effect attenuated by 26 weeks. Cultural adaptation is described only in the three Southeast Asian studies [1], [2], [10], all of which are also the studies with the most intensive human support, so adaptation and support cannot be separated. Behavioral theory-driven design was likewise inconsistently reported across the primary papers and could not be coded reliably. These are hypotheses generated by the synthesis, not findings established by it.

Duration deserves separate comment because the evidence here is direct rather than cross-sectional. Two studies measured HbA1c at more than one timepoint, and both showed attenuation: the EMR-integrated application produced a significant reduction at 12 weeks that was no longer significant at 26 weeks [9], and the gamified application produced within-group improvement in both arms by 12 weeks [4]. Attenuation observed within a study is stronger evidence about durability than the apparent advantage of short trials in Table 4, which may simply reflect which studies chose short endpoints.

B. How Risk of Bias Shapes These Conclusions

The conclusions of this review are driven by studies at moderate rather than low risk of bias, and this should temper their strength. No randomized trial in the effectiveness set was rated at low risk overall. Five raised some concerns and two were rated high. The relationship between methodological quality and effect size runs in the expected direction: the largest reported reduction comes from the study at highest risk of bias, with the smallest sample and an intervention shorter than the biological interval that HbA1c reflects [8], whereas the trials with the most complete reporting and the largest samples report reductions of 0.53 and 0.49 percentage points [5], [9]. If the high-risk studies are set aside, the central estimate of benefit moves toward roughly half a percentage point, which is close to the pooled figures in the wider literature [15], [16] and remains at the boundary of what is conventionally considered clinically meaningful.

The single non-randomized study [1] was rated at serious risk of bias under ROBINS-I, primarily because of confounding, and it contributes one of the larger adjusted estimates. Excluding it would not change the direction of the synthesis, but it would remove one of the two estimates that carry a published confidence interval. This illustrates a broader reporting problem in the field: the studies with the most complete statistical reporting are not always the ones with the strongest designs, and the absence of dispersion estimates in five of eight papers is itself an obstacle to cumulative synthesis.

C. Implications for Practice and Research

For health systems considering adoption, the practical implication is narrower than a general endorsement of mHealth. The evidence is most favourable for applications that sit inside a care pathway with a named professional responsible for reviewing patient-generated data, and it is weakest for standalone applications offering recording and automated feedback. Because certainty is low, adoption decisions should be framed as implementation with evaluation rather than as deployment of an established intervention. For researchers, the most useful contributions would be trials that report between-group differences with confidence intervals as a matter of course, that measure HbA1c at intervals appropriate to its biology, that follow participants for at least twelve months, and that report engagement with a standardized metric so that attenuation can be interpreted. The two ongoing trials described in Section 4.8 are positioned to answer part of this, particularly the Korean trial that compares application-plus-physician-support against application-only within a single design [6].

Conclusion

This review synthesized eight completed studies with 1,207 participants and two protocols of ongoing trials. Among adults with T2DM managed in healthcare facilities, mHealth applications with self-monitoring features were associated with reductions in HbA1c in six of

eight studies, with between-group estimates from 0.49 to 1.16 percentage points; one study showed improvement in both arms without a between-group difference and one showed no change. Certainty of evidence was low for HbA1c and low to very low for secondary outcomes. Larger reductions were associated with interventions that added scheduled human support to the application, but with eight studies and no formal moderator analysis this remains a pattern rather than an established determinant. The evidence does not support treating self-monitoring functionality on its own as sufficient, nor does it yet identify which design feature carries the benefit. Health systems adopting these tools should embed them within supported care pathways and should evaluate them locally, since the present evidence base is too imprecise to underwrite confident predictions of effect size.

Limitations and Suggestion

Several limitations qualify these findings. First, and most consequentially, statistical pooling was not possible: comparators, endpoints and reporting formats differed, and five of eight studies reported no dispersion estimate for the between-group difference. The SWiM synthesis reported here is a defensible response to that constraint, but it does not deliver the precision of a meta-analysis, and the ranges in Table 4 should not be read as pooled estimates. Second, the search covered two databases and English-language reports only. The rationale for PubMed and ProQuest is given in Section 3.2, and backward citation searching together with a Google Scholar cross-check on 15 August 2026 identified no additional eligible studies, but studies indexed only in Embase, CINAHL, Web of Science or IEEE Xplore may have been missed, and non-English trials from high-burden settings are systematically absent. Third, no study was at low risk of bias overall, and two of eight were at high risk, so the certainty of every conclusion is correspondingly limited. Fourth, only one completed study evaluated an EMR-integrated application, which leaves one of the more clinically interesting configurations essentially untested. Fifth, most included studies addressed T2DM in outpatient settings, which constrains generalisability to inpatient care, to type 1 diabetes and to resource-limited community settings. Sixth, the review was not registered prospectively, so protection against selective reporting at the review level rests on the a priori specifications.

Future work should prioritize multicentre randomized trials with pre-registered protocols, follow-up of at least twelve months, standardized engagement metrics and prospective cost-effectiveness analysis. Trials should report between-group differences with confidence intervals rather than within-group changes with p-values alone, since the latter cannot be cumulated. Implementation science frameworks would help explain context-specific determinants of adoption and sustained use, particularly in low- and middle-income countries where the diabetes burden is rising fastest.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] P. V. N. Nguyen et al., "Effectiveness of a digital diabetes self-management education and support program in Vietnamese adults with type 2 diabetes," Sci. Rep., vol. 15, no. 1, art. no. 41369, 2025, doi: 10.1038/s41598-025-25245-4.

[2] C. Riangkam, A. Sriyuktasuth, K. Pongthavornkamol, W. Kusakunniran, and A. Sriwijitkamol, "Effects of a mobile health diabetes self-management program on HbA1c, self-management and patient satisfaction in adults with uncontrolled type 2 diabetes: a randomized controlled trial," J. Health Res., vol. 36, no. 5, pp. 878–888, 2022, doi: 10.1108/JHR-02-2021-0126.

[3] N. Leung et al., "Efficacy of Save Medical Corporation (SMC)-01, a smartphone app designed to support type 2 diabetes self-management based on established guidelines: randomized controlled trial," J. Med. Internet Res., vol. 26, no. 1, art. no. e53740, 2024, doi: 10.2196/53740.

[4] A. M. Ruiz-Leon et al., "Efficacy of a mobile health–based behavioral treatment for lifestyle modification in type 2 diabetes self-management: Greenhabit randomized controlled trial," J. Med. Internet Res., vol. 27, no. 1, art. no. e58319, 2025, doi: 10.2196/58319.

[5] M. P. Bretschneider et al., "Evaluation of the impact of mobile health app Vitadio in patients with type 2 diabetes: randomized controlled trial," J. Med. Internet Res., vol. 27, no. 1, art. no. e68648, 2025, doi: 10.2196/68648.

[6] H. Sang et al., "Impact of the WellCheck smartphone app linked to electronic health records on clinical outcomes in patients with type 2 diabetes: study protocol for primary care-based, prospective, multicenter, cluster-randomized, pragmatic clinical trials," PLoS ONE, vol. 20, no. 8, art. no. e0329003, 2025, doi: 10.1371/journal.pone.0329003.

[7] D. M. P. Putri, Y. Suhoyo, A. A. P. Pertiwi, and C. Effendy, "Integrated Diabetes Self-Management (IDSM) mobile application to improve self-management and glycemic control among patients with type 2 diabetes mellitus in Indonesia: a mixed methods study protocol," PLoS ONE, vol. 17, no. 11, art. no. e0277127, 2022, doi: 10.1371/journal.pone.0277127.

[8] M. Guo et al., "Effectiveness of mHealth management with an implantable glucose sensor and a mobile application among Chinese adults with type 2 diabetes," J. Telemed. Telecare, vol. 29, no. 8, pp. 632–640, 2023, doi: 10.1177/1357633X211020261.

[9] E. Y. Lee et al., "Efficacy of personalized diabetes self-care using an electronic medical record–integrated mobile app in patients with type 2 diabetes: 6-month randomized controlled trial," J. Med. Internet Res., vol. 24, no. 7, art. no. e37430, 2022, doi: 10.2196/37430.

[10] P. Poonprapai, S. Lerkiatbundit, and W. Saengcharoen, "Family support-based intervention using a mobile application provided by pharmacists for older adults with diabetes to improve glycaemic control: a randomised controlled trial," Int. J. Clin. Pharm., vol. 44, no. 3, pp. 680–688, 2022, doi: 10.1007/s11096-022-01389-5.

[11] International Diabetes Federation, IDF Diabetes Atlas, 10th ed. Brussels, Belgium: International Diabetes Federation, 2021. [Online]. Available: https://www.diabetesatlas.org

[12] GBD 2021 Diabetes Collaborators, "Global, regional, and national burden of diabetes from 1990 to 2021, with projections of prevalence to 2050: a systematic analysis for the Global Burden of Disease Study 2021," Lancet, vol. 402, no. 10397, pp. 203–234, 2023, doi: 10.1016/S0140-6736(23)01301-6.

[13] American Diabetes Association Professional Practice Committee, "6. Glycemic goals and hypoglycemia: Standards of Care in Diabetes—2024," Diabetes Care, vol. 47, suppl. 1, pp. S111–S125, 2024, doi: 10.2337/dc24-S006.

[14] K. Pal et al., "Computer-based diabetes self-management interventions for adults with type 2 diabetes mellitus," Cochrane Database Syst. Rev., vol. 2013, no. 3, art. no. CD008776, 2013, doi: 10.1002/14651858.CD008776.pub2.

[15] D. Kerr, D. Ahn, K. Waki, J. Wang, B. Breznen, and D. C. Klonoff, "Digital interventions for self-management of type 2 diabetes mellitus: systematic literature review and meta-analysis," J. Med. Internet Res., vol. 26, no. 1, art. no. e55757, 2024, doi: 10.2196/55757.

[16] Y. Xiao et al., "Effectiveness of digital diabetes management technology on blood glucose in patients with type 2 diabetes at home: systematic review and meta-analysis," J. Med. Internet Res., vol. 27, no. 1, art. no. e66441, 2025, doi: 10.2196/66441.

[17] M. Campbell et al., "Synthesis without meta-analysis (SWiM) in systematic reviews: reporting guideline," BMJ, vol. 368, art. no. l6890, 2020, doi: 10.1136/bmj.l6890.
[18] M. Hilton Boon and H. Thomson, "The effect direction plot revisited: application of the 2019 Cochrane Handbook guidance on alternative synthesis methods," Res. Synth. Methods, vol. 12, no. 1, pp. 29–33, 2021, doi: 10.1002/jrsm.1458.
[19] G. H. Guyatt et al., "GRADE: an emerging consensus on rating quality of evidence and strength of recommendations," BMJ, vol. 336, no. 7650, pp. 924–926, 2008, doi: 10.1136/bmj.39489.470347.AD.
[20] American Diabetes Association Professional Practice Committee, "5. Facilitating positive health behaviors and well-being to improve health outcomes: Standards of Care in Diabetes—2024," Diabetes Care, vol. 47, suppl. 1, pp. S77–S110, 2024, doi: 10.2337/dc24-S005.
[21] N. Budhiwardoyo and M. P. Kusumo, "Transforming diabetes management: a study on the effectiveness of smartphone applications in enhancing knowledge, attitudes, and behavior of people with diabetes mellitus," Open Access Maced. J. Med. Sci., vol. 11, no. E, pp. 388–393, 2023, doi: 10.3889/oamjms.2023.11601.
[22] M. J. Page et al., "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," BMJ, vol. 372, art. no. n71, 2021, doi: 10.1136/bmj.n71.
[23] J. A. C. Sterne et al., "RoB 2: a revised tool for assessing risk of bias in randomised trials," BMJ, vol. 366, art. no. l4898, 2019, doi: 10.1136/bmj.l4898.
[24] J. A. C. Sterne et al., "ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions," BMJ, vol. 355, art. no. i4919, 2016, doi: 10.1136/bmj.i4919.

Authors



Agung Rahmat Fauzi is a medical doctor and graduate student in the Master of Hospital Administration, Postgraduate Program, Universitas Muhammadiyah Yogyakarta, Indonesia. His research interests include digital health, mobile health applications, non-communicable disease management, and healthcare quality improvement. (email: rfagung@gmail.com).



Dr. dr. Mahendro Prasetyo Kusumo, MMR., FISPH., FISCM., AIFO-K, FRSPH is a medical doctor and lecturer in the Master of Hospital Administration, Postgraduate Program, Universitas Muhammadiyah Yogyakarta, Indonesia. His research interests include telehealth, IoT-based monitoring systems, non-communicable disease management, and hospital administration. (email: mahendro_prasetyo@umy.ac.id).