The effectiveness of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes: A systematic review and meta-analysis

Karoline Villinger\(^1\) | Deborah R. Wahl\(^1\) | Heiner Boeing\(^2\) | Harald T. Schupp\(^3\) | Britta Renner\(^1\)

\(^1\)Department of Psychology, Psychological Assessment and Health Psychology, University of Konstanz, Konstanz, Germany
\(^2\)Department of Epidemiology, German Institute of Human Nutrition, Nuthetal, Germany
\(^3\)Department of Psychology, General and Biological Psychology, University of Konstanz, Konstanz, Germany

Correspondence
Britta Renner, Department of Psychology, Psychological Assessment and Health Psychology, University of Konstanz, PO Box 47, Konstanz D-78457, Germany.
Email: britta.renner@uni-konstanz.de

Funding information
German Research Foundation, Grant/Award Number: 2374; Federal Ministry of Education and Research, Grant/Award Number: 01EL1820A

Summary
A systematic review and meta-analysis were conducted to assess the effectiveness of app-based mobile interventions for improving nutrition behaviours and nutrition-related health outcomes, including obesity indices (eg, body mass index [BMI]) and clinical parameters (eg, blood lipids). Seven databases were searched for studies published between 2006 and 2017. Forty-one of 10 132 identified records were included, comprising 6348 participants and 373 outcomes with sample sizes ranging from 10 to 833, including 27 randomized controlled trials (RCTs). A beneficial effect of app-based mobile interventions was identified for improving nutrition behaviours (\(g = 0.19; CI, 0.06-0.32, P = .004\)) and nutrition-related health outcomes (\(g = 0.23; CI, 0.11-0.36, P < .001\)), including positive effects on obesity indices (\(g = 0.30; CI, 0.15-0.45, P < .001\)), blood pressure (\(g = 0.21; CI, 0.01-0.42, P = .043\)), and blood lipids (\(g = 0.15; CI, 0.03-0.28, P = .018\)). Most interventions were composed of four behaviour change technique (BCT) clusters, namely, “goals/planning,” “feedback/monitoring,” “shaping knowledge,” and “social support.” Moderating effects including study design, type of app (commercial/research app), sample characteristics (clinical/non-clinical sample), and intervention characteristics were not statistically significant. The inclusion of additional treatment components besides the app or the number or type of BCTs implemented did not moderate the observed effectiveness, which underscores the potential of app-based mobile interventions for implementing effective and feasible interventions operating at scale for fighting the obesity epidemic in a broad spectrum of the population.

KEYWORDS
BCT, diet, intervention, m-Health, mobile apps, nutrition behaviour, nutritional outcomes, obesity

Abbreviations: App, application; BCT, behaviour change technique; BMI, body mass index; CONSORT, Consolidated Standards of Reporting Trials; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; RCT, randomized controlled trial

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2019 The Authors. Obesity Reviews published by John Wiley & Sons Ltd on behalf of World Obesity Federation
1 | INTRODUCTION

There were 2.1 billion people worldwide classified as overweight or obese in 2013, which equates to 27.5% of all adults. Since being overweight or obese is associated with both physical and mental health consequences and huge economic costs, it is one of today’s most crucial health issues. Since nutrition-related behaviours are well established as major risk factors in becoming overweight, accounting for a considerable percentage of global disability-adjusted life years, preventing obesity, and being overweight are not only of peripheral matters but also of social, societal, and governmental interest.

To fight the obesity epidemic, it is important to scale effective, feasible, and affordable interventions that address a broad spectrum of the population. New intervention delivery modes such as e-Health (web based) and m-Health (mobile) technologies are booming and have even evoked a “self-track” trend. The term “m-Health” refers to the concept of using mobile devices, such as mobile phones, personal digital assistants (PDAs), tablets, wireless devices, and smartphones, in medicine and public health. Early functionalities of these devices were text messaging (SMS), paging, and voice communication. However, with the emergence of smartphones, more advanced functionalities including fully automated applications (apps) were developed. App-based mobile health interventions are a particularly promising method of changing nutrition behaviours and nutrition-related health outcomes due to the high level of global smartphone penetration and the ease of installing apps in all kinds of mobile devices. The advantages of app-based mobile interventions are numerous, including the possibility of intervening in “real life” and “real time” while also offering interactivity, the ability to tailor interventions to personal needs, and the potential to provide effective and feasible interventions to different target groups. So far, however, most users and patients have relied on commercially available app-based mobile health interventions that have not been empirically evaluated and rarely include evidence-based strategies for behaviour change. The effectiveness of app-based mobile health interventions must be determined to enable evidence-based decisions and to evaluate the potential contribution of app-based mobile interventions as large-scale health prevention measures.

While an increasing number of systematic reviews have examined technology-based interventions, a considerable gap exists in the research since most of these reviews have examined combined intervention delivery modes using various e-Health and m-Health technologies simultaneously. In addition, they have a specific focus on study selection, target population, and outcome measure. More specifically, most reviews combine different target behaviours (eg, diet and physical activity); they focus on specific audiences (eg, adults with overweight or obesity or with diabetes or patients with cancer) or research designs (eg, randomized controlled trials [RCTs]) or examine single nutrition-related health outcomes (eg, weight loss). The few reviews that specifically look at app-based mobile interventions are also heterogeneous in scope; some concentrate on single nutrition-related health outcomes (eg, weight loss, diabetes, or glycaemic control) or a specific target group (eg, healthy adults), while others combine multiple health behaviours (eg, diet and physical activity). Almost all previous reviews analysed the studies narratively, and the number of intervention studies that they included was rather limited, ranging from four to 27. Not surprisingly, this leads to findings that are divergent, difficult to integrate, and limited in their ability to provide quantified effect sizes on the effectiveness of app-based mobile health interventions. Only six previous meta-analyses, which included between seven and 22 studies, revealed that app-based mobile interventions were associated with significant weight loss and improved diabetes indicators. On the basis of these findings, app-based mobile interventions might be a promising approach for combating obesity and nutrition-related diseases, but to our knowledge, no conclusions about the effects of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes have yet been drawn, and there have been no attempts to quantify the effects. Addressing both nutrition behaviours and nutrition-related health outcomes provides a more comprehensive picture of the effectiveness of mobile interventions using a fully automated mobile dietary application, as intervention studies often target multiple outcomes within the same study.

Therefore, the aim of this systematic review and meta-analysis was to address these research gaps by evaluating the effectiveness of mobile interventions using a fully automated mobile dietary application on nutritional outcomes in both healthy and clinical audiences. Nutritional outcomes included (a) nutrition behaviours (primary outcomes) such as nutrition scores and calorie intake and (b) consequent nutrition-related health outcomes (secondary outcomes) such as obesity indices (eg, body mass index [BMI]) and clinical metabolic parameters (eg, blood lipids). Unlike most previous reviews that have narratively summarized empirical evidence, we add to the literature by conducting both a systematic review and a meta-analysis using a random effects model for quantifying intervention effects. In addition, we extended the scope of previous reviews by examining relevant moderator effects such as sample characteristics and the duration of the intervention and assessing the “building blocks” of app-based mobile interventions that target nutrition behaviours and nutrition-related health outcomes by coding the intervention characteristics according to implemented behaviour change techniques (BCTs).

2 | METHODS

The systematic review of the literature and quantitative meta-analysis to evaluate the effectiveness of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes were conducted according to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines (see Data S1 for the completed PRISMA checklist).
2.1 | Search strategy and selection criteria

To investigate the effectiveness of app-based mobile dietary interventions for both clinical and healthy samples in changing nutrition behaviours and nutrition-related health outcomes, studies were eligible if (a) they included a mobile intervention using a fully automated mobile application, (b) the app assessed any kind of nutrition behaviour, (c) the measured outcome was nutrition-related, including either nutrition behaviours or nutrition-related health outcomes, and (d) they targeted adolescents or adults.

Any intervention study design or study population from clinical and healthy audiences was considered for inclusion. All types and units of measurement for nutrition behaviours and nutrition-related health outcomes were acceptable (eg, self-report, objective measures, calories, and kilograms). The app intervention could be a stand-alone intervention using apps only or a multicomponent intervention where the use of an app was one of several intervention components (eg, face-to-face counselling and dietary education). RCTs and pre-post studies were included if they encompassed an intervention that was delivered through an app that targeted nutrition behaviours or nutrition-related health outcomes. We considered both between (control-intervention) and within (pre-post) comparisons for the quantitative meta-analyses. We looked at studies between 2006 and June 2017, since smartphones and apps are a recent development over the last 10 years. Studies were excluded if (a) they were not available in English language, (b) they were published before 2006, (c) they were targeting children (12 y of age or younger), (d) they were papers, study protocols, or conference presentations that did not report empirical data, or (e) the app did not assess any kind of nutrition behaviour.

To ensure that the literature search was multidisciplinary and comprehensive, we searched databases from the fields of medicine, nutrition, and sport (MEDLINE, PubMed, PsycInfo, PsycIndex, PsyArticle, SPORTDiscus, and Web of Science) using a predefined systematic database search protocol developed in cooperation with a scientific literature specialist from the university library. The search strategy incorporated both keywords and a controlled vocabulary (eg, Medical Subject Headings and MeSH terms) and free-text search terms. The complete search terms as well as a specification of the search strategy are provided in Data S2. We also searched reference lists of relevant review papers and retraced study protocols and conference presentations to identify other potentially eligible studies.

After duplicates are removed, articles were selected in a three-step process (see Figure 1). Firstly, trained reviewers independently screened titles and abstracts according to the four eligibility criteria, with no reviewer screening both the title and abstract of the same paper. The full-text articles that remained after this first selection were then screened independently by two authors (K.V./D.W.), who also screened a random sample of 25 studies to ensure that the four eligibility criteria were being applied consistently. Disagreements were resolved through discussion by the authors until a consensus was reached.

2.2 | Data extraction

A standardized manual for data extraction was developed by the authors, and two authors (K.V./D.W.) independently extracted key study and sample characteristics (including study design, number and type of participants, and dropout rate), intervention characteristics (including type of intervention, duration, BCTs, and type of app used), nutrition assessment (including quantity and type of food assessed and assessment method), and prespecified outcomes. The information was then entered into a customized Excel database and discussed by the authors. The coders were trained in recognizing BCTs as defined by the v1 taxonomy and reference list. The v1 taxonomy includes a total of 93 BCTs such as “goal setting for behaviours” (BCT 1.1), which describes a behaviour goal such as eating five pieces of fruit per day, or “goal setting for outcome” (BCT 1.3), representing a positive outcome of a wanted behaviour, eg, a weight loss goal of 0.5 kg over 1 week as an outcome of a changed eating pattern. Both BCTs belong to the BCT cluster “goals and planning” (BCT 1). In total, the v1 taxonomy includes 93 BCTs, which are grouped into 16 BCT clusters.

![FIGURE 1 Study selection process](image-url)
The coders independently extracted and coded the 93 BCTs and 16 BCT clusters, and disagreements were resolved through discussions between the authors until a consensus was reached. Since missing details in the descriptions of the interventions made it impractical to only analyse BCTs that were implemented in the app, additional treatments were also considered when classifying the implemented intervention strategies. In order to retrieve as much detail and information about the interventions and BCTs as possible, we used three sources: (a) the description and information provided in the retrieved articles, (b) the original apps used in the included studies (which we downloaded for data extraction where possible), and (c) in case of incomplete reports, we contacted the authors and asked for the missing information. Hence, we took a great deal of effort to retrieve as much information as possible, so we could make objective and non-contentious decisions. Neither authorship, publication journal, nor study results were blinded for data extraction.

2.3 Outcomes

Our primary outcomes were changes in nutrition behaviours, including changes in overall nutrition (eg, healthy eating score), the consumption of specific foods (eg, fruit and vegetables), nutrient intake (eg, vitamin C), and caloric intake. Secondary outcomes were nutrition-related health outcomes including obesity indices (eg, body weight and BMI) and clinical parameters (eg, blood lipids and blood pressure). See also Data S8 for the identified outcomes included in the present meta-analysis (three right-hand columns).

2.4 Study quality assessment

The quality of the included studies was independently evaluated by two authors (K.V./D.W.) according to the 25 criteria outlined by the Consolidated Standards of Reporting Trials (CONSORT) 2010 checklist. Criteria and respective items are related to the background and objectives, methods (including participant selection and outcome measures), study analysis and results, and potential selection bias or bias from funding (see Figure 2 and Data S4 for a definition of all 25 criteria). While the CONSORT checklist is intended for controlled trials, most criteria are applicable to other study designs, and the weaker study designs justifiably received a lower score than studies using a controlled trial design. This approach has been used in other reviews.1,5 Each item of the 25 criteria was rated as 1 (fulfilled), 0 (only partially fulfilled), 0 (not fulfilled), or not applicable to the study design. For example, criterion 2, “background and objectives,” entails two items with Item 2a, “scientific background and explanation of rationale,” and item 2b, “specific objectives or hypotheses.” Items were averaged per criterion. Adapted from previous reviews,1,5,12 study quality was classified as “high,” “fair,” and “low” based on an “overall study quality score” (sum of points). Nonapplicable criteria were discounted from the “overall study quality score,” and as a result, the highest attainable quality score was not 25 for all studies. The study quality score for each study was divided by the highest attainable score and multiplied by 100 to give a percentage of fulfilled criteria, with more than 66.6% = high, 50% to 66.6% = fair, and less than 50% = low study quality (Figure 2). In addition, two authors (K.V./D.W.) independently assessed the risk of bias according to the International Cochrane Collaboration criteria. All study criteria were dual coded, and any discrepancies were resolved by consensus.

2.5 Data analysis and synthesis

Standardized effect sizes were calculated to conduct quantitative meta-analyses. Effect sizes were calculated based on two approaches to account for between group (control-intervention) and within group (pre-post) comparisons.54 Effect sizes were calculated separately for each measured primary and secondary outcome. Referring to Higgins and Green,55 outcomes reported for subsamples (eg, gender) and for studies with more than two groups were pooled to create single pairwise comparisons to address the unit-of-analysis error due to “double counts.” Specifically, we combined all relevant intervention groups of a study into a single group and combined all relevant control groups into a single control group. Cohen’s d was calculated to provide a standardized effect size and converted into Hedges’ g to correct for the slight upward sample bias.57 Hedges’ g is a variation of Cohen’s d that corrects for biases because of small sample sizes. However, the effect sizes calculated in Cohen’s d (see Data S11) or Hedges’ g (see Data S12) were highly comparable and did not show marked statistical changes. The magnitude of Hedges’ g is interpreted using Cohen’s convention as small (0.2), medium (0.5), and large (0.8). The standardized mean difference was prioritized for between-group comparisons, and the standardized mean change was preferred for within-group comparisons. Reported effect sizes (eg, odds ratios) were transformed into Cohen’s d. Furthermore, where possible, effect sizes were calculated from the statistics provided.58,59 Positive effect sizes refer to intended changes by the intervention, indicating, for example, weight loss or an increase in fruit consumption.

The standardized effect sizes were synthesized using a meta-analysis model with random effects.50 The choice of a fixed effect or a random effects statistical model affects the method used to calculate the total overall estimate and hence the interpretation of the summary estimates. A fixed-effect meta-analysis assumes that all studies are estimating the same (fixed) intervention effect, whereas a random effects meta-analysis allows for differences in the intervention effect from study to study (see, for example, Riley et al52 and Borenstein et al60). Thus, random effects meta-analysis models address heterogeneity in the interventions effects caused by differences in study populations, interventions received, follow-up length, and other factors. However, random effects commonly yield a wider scatter of effect estimates and substantially wider confidence intervals than fixed effect models because each effect size has two components of variation: one due to sampling error and one from the underlying distribution. In one study,51 the within-group comparison effect sizes for weight (d = 46.57) and BMI (d = 15.00) deviated markedly from other studies, and these two outcomes effects were therefore excluded from the
analyses. To address within-study dependencies, effect sizes of studies measuring multiple outcomes were aggregated using the univariate procedure developed by Borenstein, Hedges, Higgins, and Rothstein \(^6^0\) (BHHR), which is considered as the most precise and least biased. \(^6^2\) Heterogeneity was investigated by \(Q\)-statistics and Higgins \(I^2\).

The meta-analysis was conducted in a three-step process to account for the heterogeneity of study designs and outcome variables: (a) firstly, an all-encompassing data set was created to calculate an overall effect size across all 373 identified primary and secondary outcomes, including both between-group and within-group comparisons. Sensitivity analyses were then computed following the recommendation of Viechtbauer and Cheung \(^6^3\) for outlier and influence diagnostic procedures to determine the stability of conclusions. Publication bias was assessed by a visual inspection of the funnel plot and Egger’s test. \(^6^4\) Subgroup analyses were conducted by type of comparison (between or within group). (b) Secondly, an adjusted data set was created, which prioritized between-group comparisons over within-group comparisons. \(^5^4,^6^5\) If a primary or secondary outcome was measured as both between-group and within-group comparisons, the between-group comparison was prioritized and included in the data set. Within-group comparisons were only included when no between-group comparisons were provided for the outcome. Follow-up intervals from the baseline were calculated and grouped into short-term (less than 3 mo), intermediate-term (3-6 mo), and long-term (more than 6 mo) effects. \(^5^5\) In cases where an outcome (eg, BMI) was measured multiple times within a follow-up interval, the shortest in duration within the respective follow-up interval was

![FIGURE 2 Heat map visualizing the assessment of the 25 Consolidated Standards of Reporting Trials (CONSORT) criteria of study quality for the 41 studies included. Colours range from dark blue (item fulfilled) to light blue (item not fulfilled or unclear) and grey (not applicable). *Wharton et al (2014): semirandomized trial](image-url)
prioritized. Subgroup analyses were performed for the three different follow-up intervals. Meta-regressions were also conducted for a priori identified moderators, including study design (RCT vs non-RCT), sample size and characteristic (clinical vs non-clinical), study and intervention duration, dropout rate, and the number of included outcomes, as well as intervention characteristics including the type of app (commercial vs research app), the inclusion of treatment components in addition to the app (stand-alone app [app only] vs app combined with another intervention [app+]), and the number of BCTs implemented as intervention strategies. Furthermore, we conducted meta-regressions for samples including adolescents vs adults and the quality of the studies according to the 25 CONSORT criteria (high vs fair). We conducted additional moderation analysis to get more insights into the question of whether the presence or absence of each BCT identified in the included intervention studies impacts on the effect size estimates (see Goodwin, Ostuzzi, Khan, Hotopf, and Moss-Morris, Tang, Smith, McSharry, Hann, and French, and Williams and French for a similar approach). (c) Thirdly, to get more detailed insights into intervention effects, primary and secondary outcomes were analysed separately and further subdivided into short-term, intermediate-term, and long-term follow-up intervals. In addition, constituent outcomes of nutrition behaviours and nutrition-related health outcomes were analysed separately (given a sufficient number of studies was available for the respective outcome), including caloric and fruit/vegetable intake as well as obesity indices (including body weight, BMI, body fat, hip, waist, and arm circumference), blood pressure, blood lipids (cholesterol, LDL, HDL, and triglyceride), and blood sugar (including fasting [plasma] glucose, glucose, HbA1c, and glucose/HbA1c). All analyses were conducted in Statistical Package for the Social Sciences (SPSS) (version 24) and R using the packages compute.es, metafor, MAD, and altmeta.

3 RESULTS

The search identified a total of 11,707 electronic records. After duplicates are removed, 10,132 titles were screened, and the full text of 101 potentially eligible articles was retrieved (see Figure 1). Studies were excluded for multiple reasons, such as not having an intervention component (see Data S3 for a list of the excluded studies). In total, 41 studies had investigated the effectiveness of app-based mobile interventions for improving nutrition behaviours or nutrition-related health outcomes and met all four eligibility criteria. These 41 studies were then included in the systematic review and quantitative meta-analysis.

3.1 Study quality

A detailed summary of quality assessments of the studies included according to the CONSORT 2010 checklist is presented in Figure 2 and depicted as a heat map. The 25 CONSORT criteria show a considerable variation across the 41 studies (see Data S4 for criteria definitions and details). Overall, study quality ranged from high (29 studies) to fair (12 studies) (see Figure 2 for study details). While the study quality of 89% of the 27 RCT intervention studies was classified as "high," 43% of the non-RCT intervention studies were classified as being of "high" quality. On average, the studies included fulfilled 74% of the quality assessment criteria (range: 51%-92%). Most studies (at least 85%) met the CONSORT criteria requirements of providing a clear scientific rationale and describing their scientific background and objectives (criterion 2), delivered interventions (criterion 5), statistical methods (criterion 12), sample characteristics (criterion 15), and number analysed (criterion 16); they also considered limitations (criterion 20) and provided a consistent and balanced interpretation of their results (criterion 22). A majority of the studies (at least 63%) reported a detailed and complete description of the study design, participants, and outcomes as specified by the CONSORT criteria (criteria 3, 4, 6, 13, 14, and 17). Fewer studies reported sample size calculations (criterion 7) and included randomization (criterion 8-10). Most of the studies did not include blinding procedures in their study design (criterion 11). Therefore, the results of the evaluation of risk of bias according to the six International Cochrane Collaboration criteria indicate a high or unclear risk of bias for the two blinding procedures criteria and the allocation concealment criterion. Risk of bias due to incomplete data or random sequence generation was low (36 studies and 27 studies, respectively). See Data S5 and S6.

3.2 Study and sample characteristics

Overall, the 41 studies contained a total of 6348 participants, yielded 373 primary and secondary outcomes, and were published between 2006 and 2017 (80.5% in 2014 or later). Of these, 27 studies were RCTs with either a two-arm (19 studies), three-arm (seven studies), or four-arm (one study) design. The remaining 14 studies were either single-arm, pre-post studies (eight studies), or studies with different control group designs (six studies). Sample size in the studies ranged from 10 to 833 (M = 154.83; SD = 177.80), with an average attrition rate of 18.7% (SD = 16.27; range: 0%-72%). Sample descriptions were provided for 5701 participants. The mean age was 41.51 years (SD = 13.44; range: 14-68 y), and 3678 (64.5%) of the sample population were women. Thirty-eight studies focussed on adults, and three included adolescents aged between 13 to 19 years of age and 13 to 17 years. The average BMI of the sample population was 30.77 kg/m² (SD = 3.73; range: 22.25-36.40 kg/m²). Most of the studies focussed on clinical samples, with 16 studies including participants classified as being overweight or obese and eight including patients diagnosed with diabetes or prediabetes. In addition, one study focussed on survivors of endometrial and breast cancer and one on smokers. The remaining 15 studies focussed on non-clinical, generally healthy samples. Study durations ranged from 20 days to 24 months with an average duration of 24 weeks (SD = 21.71). Intervention duration ranged from 2 weeks to 96 weeks (M = 21.05; SD = 21.17).
A total of 30 different smartphone apps were used across the studies, mostly running on Android (nine studies), iOS (eight studies), or both (13 studies). To implement the intervention, 17 studies developed their own research app, while 15 used pre-existing commercial apps. Two studies modified a pre-existing app to fit their study purposes, and seven did not provide respective information. Eighteen studies included an exclusively app-based intervention, and 23 studies included additional treatment components, which were delivered to participants either before (13 studies), during (23 studies), and/or after (two studies) the app-based intervention.

3.3 Classification of implemented BCTs

The classification of implemented intervention strategies for achieving outcome changes according to the BCT taxonomy\(^4^8\) showed that in total, nine of the 16 different BCT clusters with an average of 3.88 (SD = 1.29, range: 1-6) were employed across the 41 studies. Figure 3 depicts the frequency of BCT clusters implemented across the 41 studies and Table 1 the frequency of inclusion of single BCTs. A detailed summary of BCTs for each study is provided in Data S7.

As Figure 3 illustrates, the BCT intervention strategy clusters “feedback and monitoring” (BCT 2, 41 studies), “goals and planning” (BCT 1, 31 studies), “social support” (BCT 3, 28 studies), and “shaping knowledge” (BCT 4, 25 studies) were implemented in the majority of the 41 studies, while “associations” (BCT 7, 17 studies), “reward and threat” (BCT 10, nine studies), and “comparison of behaviour” (BCT 6, six studies) were less prominent, and the BCT clusters “antecedents” (BCT 12) and “self-belief” (BCT 15) were only implemented in one study each.

In addition to the 16 main BCT clusters of the BCT taxonomy, intervention strategies can be further classified into 93 BCTs. The number of implemented BCTs across the 41 studies ranged from two to 11, with an average of 6.9 (SD = 2.46). For example, the frequently implemented BCT cluster “goals and planning” (BCT 1) includes nine different BCTs, of which seven were actually implemented across the 41 studies with “goal setting for behaviours” (BCT 1.1, 28 studies), “goal setting for outcome” (BCT 1.3, 14 studies), and “review behaviour goal(s)” (BCT 1.5, 20 studies) being realized the most frequently. The intervention strategies “problem solving” (BCT 1.2, six studies), “action planning” (BCT 1.4, two studies), and “review outcome goal(s)” (BCT 1.7, eight studies) were implemented less often, and strategies such as “discrepancy between current behaviour and goal” (BCT 1.6), “behavioural contract” (BCT 1.8), and “commitment” (BCT 1.9) were implemented in none of the studies. The BCT cluster “feedback and monitoring” includes seven different BCTs, of which five were actually implemented across the 41 studies. The most frequently implemented strategies were “feedback on behaviour” (BCT 2.2, 37 studies), “self-monitoring of behaviour” (BCT 2.3, 41 studies), “self-monitoring of outcome(s) of behaviour” (BCT 2.4, 17 studies), and “feedback on outcome(s) of behaviour” (BCT 2.7, 11 studies), while the three remaining strategies were implemented either infrequently or not at all. The third main intervention strategy cluster “social support,” used in 28 studies, encompasses the three BCTs: “unspecified social support” (BCT 3.1, 12 studies), “practical social support” (BCT 3.2, 10 studies), and “emotional social support” (BCT 3.3, 12 studies).

3.4 Assessment of nutrition behaviour

Of the 41 studies, 29 included an app-based assessment of the total nutrition intake\(^6^9\)-\(^7^2\), \(^7^4\)-\(^7^6\), \(^7^8\)-\(^8^0\), \(^8^3\)-\(^8^5\), \(^8^7\)-\(^8^9\), \(^9^2\)-\(^9^6\), \(^9^9\), \(^1^0^1\)-\(^1^0^6\), \(^1^0^8\) and nine an assessment of specific foods\(^6^1\), \(^7^7\), \(^7^8\), \(^8^4\), \(^9^3\), \(^9^5\), \(^1^0^0\), \(^1^0^7\) such as vegetables.
or specific food consumption patterns (eg, adherence to specific guideline-based recommendations), one assessed meal replacements,
and two did not provide further information. Twenty one of the 32

TABLE 1 Comparisons between effect sizes, according to whether
specific BCT clusters and single BCTs were present or absent in the
intervention study

<table>
<thead>
<tr>
<th>Moderator</th>
<th>Present</th>
<th>Absent</th>
<th>Estimate</th>
<th>SE</th>
<th>P</th>
<th>CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>BCT 1</td>
<td>31</td>
<td>10</td>
<td>0.21</td>
<td>0.06</td>
<td>0.13</td>
<td>.663</td>
</tr>
<tr>
<td>BCT 1.1</td>
<td>28</td>
<td>13</td>
<td>0.25</td>
<td>0.02</td>
<td>0.12</td>
<td>.875</td>
</tr>
<tr>
<td>BCT 1.2</td>
<td>6</td>
<td>0.30</td>
<td>0.25</td>
<td>0.04</td>
<td>0.15</td>
<td>.803</td>
</tr>
<tr>
<td>BCT 1.3</td>
<td>14</td>
<td>27</td>
<td>0.29</td>
<td>−0.10</td>
<td>0.11</td>
<td>.396</td>
</tr>
<tr>
<td>BCT 1.4</td>
<td>2</td>
<td>39</td>
<td>0.26</td>
<td>0.05</td>
<td>0.27</td>
<td>.854</td>
</tr>
<tr>
<td>BCT 1.5</td>
<td>20</td>
<td>21</td>
<td>0.26</td>
<td>0.00</td>
<td>0.11</td>
<td>.973</td>
</tr>
<tr>
<td>BCT 1.7</td>
<td>8</td>
<td>16</td>
<td>0.32</td>
<td>−0.11</td>
<td>0.13</td>
<td>.391</td>
</tr>
</tbody>
</table>

Note. BCT cluster and single BCTs as classified in the BCT taxonomy. Comparisons were only made if the BCT cluster or single BCT was at least included in two intervention studies; k = number of studies; g = Hedges’ g. Abbreviation: BCT, behaviour change technique.

Three hundred seventy-three outcome effect sizes were reported in the 41 studies, covering a broad range of different primary and secondary outcomes (see Data S8, right-hand columns, as well as Data S12). The primary outcome, nutrition behaviours, was assessed through both general and specific nutrition scores (eg, Healthy Eating Index [HEI] as a measure of diet quality, which reflects the concordance of dietary pattern with key dietary recommendations from the Dietary Guidelines for Americans109), total caloric intake, the consumption of specific foods (eg, fruits and low-fat milk), consumption of meal types (eg, take-away meals), or the intake of specific nutrients (eg, sodium). Secondary outcomes also included a broad range of indicators ranging from obesity indices (eg, body weight, BMI, body fat, and waist circumference) to clinical metabolic parameters (eg, blood pressure, glucose, or triglyceride).

The random effects meta-analyses based on the all-encompassing data set including all 373 primary and secondary outcomes effect sizes across the 41 studies showed an overall significant small-to-medium positive effect for app-based mobile interventions, with overall Hedges’ g = 0.33 (CI, 0.21−0.44, P < .001).

To assess the robustness of the effect, we conducted a sensitivity analysis including and excluding outliers.63 One study was identified as an outlier64 (Hedges’ g = 1.80). However, removing this study did not appreciably change the overall effect size (Hedges’ g = 0.27; CI, 0.20−0.35). Therefore, we kept the study in the analysis, given our interest in providing a comprehensive assessment of the overall impact of app-based mobile interventions on nutrition behaviours.

The investigation of publication bias, analysed by a visual examination of the funnel plot with the observed effect size g on the horizontal axis plotted against the standard error, revealed no asymmetry (see Data S9). In addition, Egger’s regression coefficient64 did not suggest a publication bias (z = 0.47, P = .636). However, since the Q-statistics for the overall effect across the 373 effect sizes were significant, indicating considerable heterogeneity between the 41 studies with Q(40) = 312.19, P < .001, and I² = 86.79%, the type of comparison (between group or within group) was further examined as a moderator. Analyses indicated a significant moderating effect, Q(1) = 13.13, P < .001, explaining R² = 3.63% of the among-study heterogeneity. Subsequent separate subgroup analyses yielded a significant small effect size of Hedges’ g = 0.22 (CI, 0.08−0.36, P = .002, Q(27) = 200.61, P < .001, I² = 83.61%) for between-group comparisons (k = 28, outcome n = 190) and a significant medium-to-large effect size with Hedges’ g = 0.47 (CI, 0.29−0.65, P < .001, Q(33) = 390.52, P < .001, I² = 93.61%) for within-group comparisons (k = 34, outcome n = 183). As the type of comparison (between group or within group) was identified as a significant moderator, further analyses were conducted with the adjusted data set, which prioritized between-group effects.

The random effects meta-analyses based on the adjusted data set across the 41 studies including 224 effect sizes revealed an overall significant small but positive effect of Hedges’ g = 0.26 (CI, 0.15−0.36, P < .001). The Q and I² statistics indicated considerable heterogeneity.
across studies with $Q(40) = 220.98$, $P < .001$, and $I^2 = 80.94\%$ (see Figure 4). Separate meta-regressions were conducted to identify moderating effects of study design (RCT vs non-RCT), study quality (high vs fair), sample size and characteristic (non-clinical vs clinical sample; adolescents vs adults), study and intervention duration, dropout rate, and the number of included outcomes, along with intervention characteristics including the type of the app, the inclusion of treatment components in addition to the app, and the number of BCTs implemented as intervention strategies. The meta-regressions revealed no significant effects, with $.875 \leq P \leq .120$ (see Data S10). In addition to testing whether or not the number of BCTs implemented had a moderating effect, it was additionally tested whether specific BCTs were a predictor of the pooled effect size (see, for a similar approach, Goodwin et al., Tang et al., and Williams and French). The separate meta-regressions for the implemented BCT clusters and single BCTs (technique present vs not present in the intervention) yielded no significant effect, with $.973 \leq P \leq .074$ (see Table 1).

### 3.6 Primary outcomes

Twenty one of the 41 studies assessed at least one primary outcome, addressing the effect of app-based mobile interventions on nutrition behaviours, resulting in 24 analysed behavioural outcomes (see Figure 5). Analysing the effect of app-based mobile interventions on behavioural outcomes ($k = 21$, $n = 24$) showed a small significant effect size, Hedges’ $g = 0.19$ (CI, 0.06-0.32, $P = .004$), with considerable heterogeneity of $Q(20) = 57.41$, $P < .001$, and $I^2 = 62.96\%$. On the basis of the number of available studies, behavioural outcomes were further separated into calorie ($k = 8$, $n = 9$) and fruit and vegetable intake ($k = 8$, $n = 27$). For a detailed summary, see Data S12 and S13. While both outcomes yielded overall positive effects, only the effect for fruit and vegetable intake reached statistical significance with Hedges’ $g = 0.32$ (CI, 0.15-0.50, $P < .001$, $Q(7) = 11.83$, $P = .106$, $I^2 = 24.19\%$).

### 3.7 Secondary outcomes

Thirty four of the 41 studies included at least one secondary, nutrition-related health outcome, resulting in 42 analysed outcomes (see Figure 6). The effect size for nutrition-related health outcomes was small to medium, with Hedges’ $g = 0.23$ (CI, 0.11-0.36, $P < .001$). Effect sizes showed considerable heterogeneity between studies, $Q(33) = 202.39$, $P < .001$, and $I^2 = 84.15\%$. Furthermore, based on the number of available studies, assessed nutrition-related health outcomes were divided into obesity indices (eg, body weight and BMI), blood pressure, blood lipids, and blood sugar (see also Data S13). The strongest effect was found for obesity indices with Hedges’ $g = 0.30$ (CI, 0.15-0.45, $P < .001$, $Q(31) = 230.49$, $P < .001$, $I^2 = 87.88\%$, $k = 32$, $n = 76$). Separate analyses for body weight and BMI also revealed comparable effect sizes with an effect of Hedges’ $g = 0.27$ for body weight (CI, 0.13-0.41, $P < .001$, $Q(30) = 125.59$, $P < .001$, $I^2 = 81.93\%$, $k = 31$, $n = 39$) and an effect of Hedges’ $g = 0.37$ for BMI (CI, 0.18-0.55, $P < .001$, $Q(16) = 66.35$, $P < .001$, $I^2 = 81.55\%$, $k = 17$, $n = 21$). The effects for the other health indicators were also significantly positive but less pronounced, with blood pressure showing an overall effect of Hedges’ $g = 0.21$ (CI, 0.01-0.42, $P = .043$, $Q(6) = 20.99$, $P = .002$, $I^2 = 73.81\%$, $k = 7$, $n = 19$) and blood lipids of Hedges’ $g = 0.15$ (CI, 0.03-0.28, $P = .018$, $Q(4) = 1.67$, $P = .797$, $I^2 = 0.00\%$, $k = 5$, $n = 22$). For cholesterol, a significant positive overall effect was found with Hedges’ $g = 0.37$ (CI, 0.04-0.71, $P = .031$, $Q(4) = 21.99$, $P < .001$, $I^2 = 72.81\%$, $k = 5$, $n = 7$). The effect for blood sugar was also positive but not statistically significant, Hedges’ $g = 0.18$, $P = .429$ ($k = 7$, $n = 10$). We did not conduct separate meta-analyses for the remaining single health outcomes, since the numbers of studies and outcomes were too small.

### 3.8 Follow-up intervals: Short-term, intermediate-term, and long-term effects

Effect sizes for different follow-up intervals were also examined. Assessing short-term, intermediate-term, and long-term effects separately by subgroup analysis revealed positive effects for all follow-up intervals. However, only studies targeting short-term and/or intermediate-term follow-up intervals yielded significant small effect sizes (short-term: Hedges’ $g = 0.27$; CI, 0.12-0.43, $P = .001$, $Q(12) = 22.34$, $P = .034$, $I^2 = 48.53\%$, $k = 13$, $n = 57$; intermediate-term: Hedges’ $g = 0.28$; CI, 0.15-0.40, $P < .001$, $Q(28) = 186.20$, $P < .001$, $I^2 = 83.85\%$, $k = 29$, $n = 136$). The effects for long-term intervals did not reach statistical significance (Hedges’ $g = 0.08$; CI, 0.00-0.20, $P = .404$, $Q(8) = 27.84$, $P = .001$, $I^2 = 72.99\%$, $k = 9$, $n = 31$). In a subsequent step, the effects of different follow-up intervals were examined separately for primary and secondary outcomes (see Figures 5 and 6). For nutrition behaviours (primary outcome), the effect for the intermediate-term follow-up interval was statistically significant with Hedges’ $g = 0.25$ (CI, 0.10-0.39, $P < .001$, $Q(12) = 28.32$, $P < .005$, $I^2 = 52.32\%$, $k = 13$, $n = 49$). The effect for nutrition-related health outcomes (secondary outcome) was also significant and small with Hedges’ $g = 0.26$ (CI, 0.12-0.40, $P < .001$, $Q(24) = 167.52$, $P < .001$, $I^2 = 84.68\%$, $k = 25$, $n = 86$). Effects of short-term and long-term intervals did not reach statistical significance for either outcome. An overview of the different levels of analysis and corresponding statistical characteristics is summarized in Data S12.

### 4 DISCUSSION

This meta-analysis of data from 41 studies, which included more than 6300 participants, showed that app-based mobile interventions can be effective for changing nutrition behaviours and their main nutrition-related health outcomes in a wide range of study settings, with both clinical and generally healthy samples. The study extends previous work by including 373 primary and secondary outcomes addressing...
FIGURE 4  Forest plot showing the effects of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes ($k = 41$, outcome $n = 224$; adjusted data set)
nutrition behaviours, obesity indices, and clinical metabolic parameters, revealing small-to-moderate positive effect sizes.

To make the results as relevant and robust as possible to inform prevention and clinical practice, we prioritized between group effects, which also resulted in a significant positive effect with Hedges’ $g = 0.26$ (CI, 0.15-0.36). We found no evidence that the inclusion of additional treatment components besides the app or the number of BCTs implemented moderated the observed effectiveness, which underscores the potential of app-based mobile interventions for implementing effective and feasible, cost-effective interventions with a high reach of target groups.

**FIGURE 5** Forest plot showing the effects of app-based mobile interventions on nutrition behaviours (primary outcome) for short-term, intermediate-term, and long-term follow-up intervals ($k = 21$, outcome $n = 24$; adjusted data set)
Although the results of studies comparing app-based mobile intervention versus control and pre-post comparisons yielded broadly consistent positive results on their relative effectiveness, a substantial diversity was observed in the range of efficacy across measured outcomes. In view of the rapid increase in the prevalence and disease burden of obesity worldwide, it is particularly encouraging that the results of app-based mobile interventions revealed a positive effect size for changing obesity indices with Hedges’ $g = 0.30$ (CI, 0.15-0.45) pooled across 32 studies. This confirms previous reviews summarizing the effectiveness of various types of mobile interventions.
on obesity indices. This effectiveness was also observed for clinical metabolic parameters including blood lipids and blood pressure, although with smaller effect sizes, while a positive, non-significant effect was observed for blood sugar. This differential efficacy across nutrition-related health outcomes might reflect heterogeneous mechanisms and causes, smaller numbers of studies on clinical metabolic parameters, or different methodological issues affecting studies targeting obesity indices and clinical metabolic parameters. New developments in sensor technology, eg, continuous glucose measurement, seem to offer promising avenues for improving the assessment of the clinical metabolic parameters implicated in body weight regulation.

4.1 Implications for intervention research and clinical practice

While the pooled effect size indicates a significant overall effect on nutrition behaviours and related health outcomes (Hedges’ $g = 0.26$; CI, 0.15–0.36), converting this statistical significance into clinical significance is not unequivocally possible. Nakagawa and Cuthill and other researchers suggested taking multiple criteria into account besides comparisons with benchmark values (eg, Cohen’s classification), such as comparing effect size values of the current work with previous research and practical meaningful measures.

Comparing the effect size values of the current study with previous research shows that the observed effects coincide with m-Health interventions in the same and other domains. For example, Schippers et al. reported a pooled body weight reduction of $d = -0.23$; CI, $-0.38$ to $-0.08$ based on 12 studies, which is comparable with the present pooled obesity indices estimate, which includes 32 studies (Hedges’ $g = 0.30$; CI, 0.15–0.45). The effect size value is also comparable with effect sizes found in previous meta-analyses of weight loss interventions either using mobile phones or not using mobile phones. A comparison with other domains leads to similar conclusions. For example, a recent meta-analysis in the domain of physical activity by Eckerstorfer et al. found a small overall effect of Hedges’ $g = 0.29$ (CI, 0.20–0.37), which agrees with another recent meta-analysis of m-Health physical activity interventions.

As pooled overall estimates are based on standardized effect sizes converted from measures with different units (eg, body weight loss in kg, BMI change, etc), translating them into physiological and clinical significance is not directly possible. However, effect sizes in original units, eg, body weight loss in kilograms as a result of the intervention, can be obtained from reported changes in the intervention arms (see also Schippers et al.). Summarizing the 13 intervention studies reporting changes in BMI (kg/m$^2$) in intervention arms (see Data S14, Table S14.1), the average weighted BMI weight loss was $-0.90$ kg/m$^2$, ranging from $0.24^{49}$ to $-1.41^{83}$ kg/m$^2$. In the 29 studies reporting body weight changes in kilograms (or lbs) in intervention arms (see Data S14, Table S14.2), the reduction in the intervention arm ranged from $0.64^{49}$ to $-9.65^{103}$ kg with an average weight loss of $-2.69$ kg. This weighted effect size for body weight reduction is similar to those reported in previous meta-analysis of mobile or e-Health interventions within adults with chronic diseases ($-2.45$ kg$^{47}$, adults ($-3.1$ kg$^{40}$), and adults with overweight or obesity ($-2.70$ kg$^{20}$). Whether this is or is not enough to result in physiologically relevant effects or clinically meaningful health improvements probably depends on the situation and context. For some individuals with overweight or obesity, this weight loss might be insufficient. However, from a public health perspective, small changes might also be considered as relevant (see Fisher et al. for a discussion).

The overall effectiveness of the interventions did not significantly vary in dependence of whether the interventions were based on apps as stand-alone or they were combined with additional intervention components. Eighteen of the 41 intervention studies (44%) used an app as a stand-alone intervention delivery method, while a larger proportion of interventions used an app in combination with other intervention strategies such as groups sessions, weekly meetings and online tools, coaching calls, text messages, and emails, or face-to-face contact (see Data S8 for a detailed study description). Also, the 11 most successful intervention studies, which included four intervention studies using an app as stand-alone strategy (36%). Comparing the effect sizes between these "success models" shows an overall comparable result pattern for stand-alone interventions (range: Hedges’ $g = 0.58^{100}$ to Hedges’ $g = 0.76^{71}$) compared with combined interventions (range: Hedges’ $g = 0.28^{74}$ to Hedges’ $g = 1.19^{61}$).

A major finding of our analysis is that app-based mobile interventions are effective in changing nutrition behaviours, which is an essential prerequisite for changes in obesity indices and clinical metabolic parameters. However, the number and type of assessed behavioural outcomes varied between studies, which limited the calculation of effect sizes for all specific behavioural outcomes except for caloric intake. A positive effect was found for these two behavioural outcomes, but the pooled effect was only significant for fruit and vegetable intake (Hedges’ $g = 0.32$; CI, 0.15–0.50), which is one of the most commonly measured dietary change indicators.

Studies showed substantial variations in the nutrition assessment methods, which are often study and/or country specific and are not harmonized. This diversity in methods for assessing dietary intake might have contributed to the identified heterogeneity and the smaller effect sizes compared with nutrition-related health outcomes. It is therefore desirable to harmonize and standardize dietary assessment methods to enable better cross-study comparisons.

Behaviour assessments also provide a major input for creating tailored feedback implemented in mobile interventions. So far, however, the full potential of mobile technology for dietary assessment has barely been realized. The focus is often on conventional food frequency questionnaires relying on retrospective self-reports, including memory bias. Mobile devices utilizing image-based methods offer in-the-moment dietary assessments, and it is likely that automated food identification and portion size estimation will allow more a comprehensive and accurate assessment of the quality and quantity of dietary intake.
Considering that on average, only four of the 16 different BCT clusters were implemented in the interventions and that mobile devices impose excessive restrictions on message length and interaction duration,\textsuperscript{121–123} the generally positive effects we observed across nutrition behaviours and major health outcomes were induced by a highly focussed intervention effort. The BCT clusters that were utilized, which form the central “building blocks” of interventions, mainly encompassed goal setting, feedback and self-monitoring, information, and social support provision, which coincide with successful conventional individual and group-based interventions\textsuperscript{124} and reviews on m-Health interventions.\textsuperscript{21,31,40,115,125–128} Setting goals, monitoring behaviour, receiving feedback, and reviewing relevant goals in the light of feedback are central to self-management and behavioural control, as specified by control theories.\textsuperscript{129,130} Considering that app-based mobile interventions can operate at scale and at lower comparable implementation costs than individual and group-based interventions, they offer the potential to be a cost-effective method for improving nutrition behaviours and health indicators. A recent systematic review of economic evaluations of m-Health solutions\textsuperscript{131} found a consistent overall reporting of positive economic outcomes (eg, increase in life-years gained, cost savings, and cost-effectiveness). Of 35 intervention studies using m-Health as primary intervention component, 26 (74%) reported a positive costing outcome. This supports the notion that mobile intervention might be a viable alternative to more cost-intensive face-to-face intervention formats, offering a potentially more effective alternative to common non-mobile interventions.

Although the number of single BCTs implemented varied from two to 11 between studies, the results did not show the association, which some studies have suggested between greater effectiveness and an increasing number of BCTs.\textsuperscript{57,125,127,128,132} We conducted additional moderation analysis to get more insights into the question of whether or not the presence of each of the 24 BCTs identified impacts on the effect size estimates (see Goodwin et al.\textsuperscript{66} Tang et al.\textsuperscript{67} and Williams and French\textsuperscript{68}). While we identified 19 BCTs, which were implemented in two or more intervention studies, we did not find evidence that any of the BCTs predicted the pooled effect size. Hence, there does not appear to be a single effective approach to changing nutrition behaviours and their main nutrition-related health outcomes (see also Goodwin et al\textsuperscript{66} for similar results but see Olander et al.\textsuperscript{133} Tang et al.\textsuperscript{67} and Williams and French\textsuperscript{68} for significant results). However, the effect of a single BCT may generally be very small.\textsuperscript{134} For example, according to control theories,\textsuperscript{129,130} setting and reviewing goals in the light of feedback are central to self-management, behavioural control, and ultimately behaviour changes. Accordingly, techniques revolving around goal setting and reviewing of goals are likely to be more effective if implemented conjointly in an intervention. However, “goal setting” (BCT 1.1 and BCT 1.3) was implemented more commonly than “review of goals” (BCT 1.5 and BCT 1.7).

The BCT taxonomy, while already complex with 16 different theory-based BCT clusters and 93 single BCTs, is mainly focussing on static concepts. However, since mobile interventions distinguish themselves by being interactive, adaptive, time-sensitive, and intrindividually dynamic,\textsuperscript{12} more dynamic concepts including the timing of information provision, feedback, and reminders or tailoring tasks and goals to individual progress and capacities as specified in persuasive technology might be essential ingredients of effective focussed mobile interventions.\textsuperscript{12,13,121,122,135,136} Hence, while mobile phones are a promising platform for accessible and cost-effective interventions, further development and innovation are required to ensure that the medium’s possibilities are leveraged to ensure efficacious changes.\textsuperscript{14,40}

### 4.2 Study strengths and limitations

To the best of our knowledge, this is one of the most comprehensive and up-to-date reviews and meta-analyses to evaluate the effects of app-based mobile interventions on both major nutrition-related health outcomes and intermediate nutrition behaviours across a broad spectrum of the population, combining data from 41 studies, 6348 participants, and 373 outcomes with sample sizes ranging from 10 to 833 participants and including 27 RCTs. The meta-analyses should therefore provide sufficiently reliable estimates of the intervention effects associated with app-based mobile interventions targeting nutrition behaviours and related health outcomes. By analysing the currently available evidence for app-based mobile intervention studies and reporting results separately for study designs and characteristics, we address the observed heterogeneity and provide estimates across the whole range of intervention studies. We report results separately for two outcomes (nutrition behaviours and nutrition-related health outcomes), each according to short-term, intermediate-term, and long-term follow-up intervals. Importantly, results showed consistency between study designs, clinical, and generally healthy samples and between the different forms of analysis.

However, our review and meta-analysis have limitations. Since we strived to cover all the globally available evidence on app-based mobile interventions targeting nutrition behaviours and nutrition-related health outcomes, considerable heterogeneity in the results could be due to methodological differences restricting the interpretation. Firstly, the systematic review is based on 30 different apps, with 15 studies using commercial apps. Although meta-regression did not yield a significant difference in effectiveness between commercial and research apps, considerable heterogeneity within both types of apps might still be of concern. Secondly, estimated intervention effects were smaller in between-group comparisons than in within-group comparisons, which are associated with less precise estimations due to within-subject correlations and potentially confounding variables.\textsuperscript{54,65} Thirdly, although the pooled effect size did not vary in dependence of the study or intervention duration (see also Schippers et al\textsuperscript{40} for similar findings for weight loss interventions), the effect sizes varied between short-term, intermediate-term, and long-term follow-ups as for the latter, effect sizes were generally smaller and non-significant. There are several possible explanations for this, as many factors may have caused lower effectiveness for long-term outcomes. This differential efficacy might indicate a smaller number of studies with long-term follow-ups, a decrease in the maintenance
of changes, or a general lack of effectiveness. Specifically, systematic reviews and meta-analyses have shown that the efficacy of both traditional and technology-based weight loss interventions is greatest during the first 6 months, which might be due to a well-documented decline in engagement with intervention modalities. Another possible explanation could be bias in conducting, analysing, or reporting. In our analysis, neither dropout rate, sample size, study quality, intervention duration, number or type of included BCTs, nor outcomes were associated with differential effectiveness. Also, the funnel plot did not provide consistent evidence for publication bias (see Data S9). However, such plots are known to be insensitive, and it is possible that a small study bias might have led to an overestimation of the magnitude of the effectiveness.

Considering the BCTs, we have identified four core clusters including "goals/planning," "feedback/monitoring," "shaping knowledge," and "social support," but not all interventions using these strategies were more effective. There are various possible reasons for this. The BCTs may be implemented very differently in one study versus another. Since meta-analyses are reliant on published descriptions of intervention and control conditions, a further "drill down" was limited. We included information from the articles, authors, and apps to decrease the likelihood of missing out on actually implemented BCTs. Since these are different data sources, the surplus of having more detailed information available comes with the downside that the available information differs between studies. As one anonymous reviewer suggested, some apps might, for example, vary the information included depending on how they are used (eg, some interactive apps trigger specific BCTs after repeated and constant use), which might explain why some BCTs are not coded. Methodological limitations may also have contributed. A certain number of studies and a sufficient variation among them in terms of intervention components and effect sizes are needed to have a chance of detecting relevant associations (see also Michie et al for a detailed discussion). Specifically, there was insufficient variation between the 41 available studies for the assessment of the relative effectiveness of BCTs or evaluating different BCT combinations.

The potential for sample selection bias is another concern for the quality of mobile intervention studies. In this respect, most of the studies scored in the moderate to high range on CONSORT criteria and neither study quality (high vs fair quality) nor study design (RCT vs non-RCT) moderated the effects. Moreover, the type of study sample did not reveal a significant moderating effect, indicating that the effects were comparable for studies with both clinical and generally healthy samples and across adolescents and adults. Overall, the findings suggest a high potential for app-based mobile interventions across multiple audiences, offering confidence that app-based mobile interventions are a promising method for fighting the obesity epidemic in a broad spectrum of the population.

5 | CONCLUSION

The present findings represent one of the most comprehensive currently available evidence bases demonstrating that app-based mobile interventions are effective and highly promising for changing nutrition behaviours and nutrition-related health outcomes, which is essential for conquering the obesity epidemic. The overall pooled effect size was positive, and in general, the effects were relatively consistent across different outcomes, which provides some confidence in the conclusion that the effects are likely to be small but, to a certain degree, reliable and of practical relevance. Moreover, the present results do not indicate that this generally positive effect is limited to certain populations (healthy vs clinical samples; adolescents vs adults), intervention strategies (app only vs app+; number or type of BCTs), or type of app (commercial vs research). Considering that app-based mobile interventions can operate at scale and at lower comparable implementation costs than individual and group-based interventions, they offer the potential to be a cost-effective method for improving nutrition behaviours and health indicators in a wide range of target groups. However, long-term follow-up effect sizes were generally smaller and non-significant. Hence, increasing engagement with intervention modalities might be a key avenue for future developments.

Mobile technologies offer the possibility of realizing more dynamic concepts with interactive, adaptive, time-sensitive, and intrindividually dynamic strategies. However, all statements comparing the merits of mobile app-based interventions with other delivery modes must be tempered by the potential limitations of the methodology, the complexity of specific populations, and treatment settings. These findings raise important questions around intervention design. Additional research to unpack the effective ingredients of mobile interventions would be helpful to identify which intervention components might be the most likely to be universally effective and which are more contextually dependent.

ACKNOWLEDGEMENTS

We thank Alexandra Grimm for supporting the data analyses. We also thank Valentina Boos, Marie-Luise Diepolder, Timo Dimitriadis, Desiree Katzenberger, Heike Koch, Laura König, Monika May, Sabrina Nöth, Kristian Sabados, Julia Schuhmann, Tobias Volk, Katrin Ziesemer, and Tony Arthur for their valuable support.

CONFLICT OF INTERESTS

No conflict of interest was declared.

AUTHORS’ CONTRIBUTIONS

Karoline Villinger, Deborah R. Wahl, Harald T. Schupp, and Britta Renner designed the systematic review. Karoline Villinger, Deborah R. Wahl, and Britta Renner did the search. Karoline Villinger and Deborah R. Wahl did the screening and data extraction process, assessed data quality, and analysed data with critical comments from Britta Renner, Harald T. Schupp, and Heiner Boeinge. Karoline Villinger, Deborah R. Wahl, and Britta Renner drafted the manuscript with input from Harald T. Schupp and Heiner Boeinge. All authors read, commented on, and approved the draft and final manuscripts.
No additional data available.

REFERENCES


68. Williams SL, French DP. What are the most effective intervention techniques for changing physical activity self-efficacy and physical activity behaviour—and are they the same? *Health Educ Res*. 2011;26(2):308-322. https://doi.org/10.1093/her/cyr005


SUPPORTING INFORMATION

Additional supporting information may be found in the Supporting Information section at the end of the article.

How to cite this article: Villinger K, Wahl DR, Boeijing H, Schupp HT, Renner B. The effectiveness of app-based mobile interventions on nutrition behaviours and nutrition-related health outcomes: A systematic review and meta-analysis. Obesity Reviews. 2019;20:1465–1484. https://doi.org/10.1111/obr.12903