Published in final edited form as:

Obes Rev. 2016 July ; 17(7): 636–644. doi:10.1111/obr.12412.

Methodological quality of behavioural weight loss studies: a systematic review

S. C. Lemon[1], **M. L. Wang**[2], **C. F. Haughton**[1], **D. P. Estabrook**[1], **C. F. Frisard**[1], and **S. L. Pagoto**[1]

[1]UMass Worcester Prevention Research Center, Division of Preventive and Behavioral Medicine, University of Massachusetts Medical School, Worcester, MA, USA

[2]Department of Community Health Sciences, Boston University School of Public Health, Boston, MA, USA

Summary

This systematic review assessed the methodological quality of behavioural weight loss intervention studies conducted among adults and associations between quality and statistically significant weight loss outcome, strength of intervention effectiveness and sample size. Searches for trials published between January, 2009 and December, 2014 were conducted using PUBMED, MEDLINE and PSYCINFO and identified ninety studies. Methodological quality indicators included study design, anthropometric measurement approach, sample size calculations, intent-to-treat (ITT) analysis, loss to follow-up rate, missing data strategy, sampling strategy, report of treatment receipt and report of intervention fidelity (mean = 6.3). Indicators most commonly utilized included randomized design (100%), objectively measured anthropometrics (96.7%), ITT analysis (86.7%) and reporting treatment adherence (76.7%). Most studies (62.2%) had a follow-up rate >75% and reported a loss to follow-up analytic strategy or minimal missing data (69.9%). Describing intervention fidelity (34.4%) and sampling from a known population (41.1%) were least common. Methodological quality was not associated with reporting a statistically significant result, effect size or sample size. This review found the published literature of behavioural weight loss trials to be of high quality for specific indicators, including study design and measurement. Identified for improvement include utilization of more rigorous statistical approaches to loss to follow up and better fidelity reporting.

Keywords

Intervention; quality; review; weight loss

Address for correspondence: Professor S. C. Lemon, UMass Worcester Prevention Research Center, Division of Preventive and Behavioral Medicine, University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, MA 01655, USA., Stephenie.Lemon@umassmed.edu.

Conflict of interest statement

The authors have no conflicts of interest to declare

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web site.

Introduction

Scientific articles and popular press publications have revealed broad public skepticism about scientific findings (1). This attitude persists despite efforts to make research more accessible and immediate. The quality of behavioural research specifically, including its reproducibility (2), has also been called into question.

The efficacy of behavioural weight loss interventions in particular has been questioned. Benchmark studies such as the Diabetes Prevention Program (3), the Finnish Diabetes Prevention Study (4), the Da Qing Study (5) and the Look Ahead study (6) have found lifestyle interventions to be efficacious and associated with numerous long-term benefits, including sustained weight loss, improvements in metabolic measures such as blood pressure, fitness, glycemic control and cholesterol levels (6), and decreased incidence of type 2 diabetes. A large number of efficacy and effectiveness studies have attempted to replicate and/or adapt the interventions used in these studies across diverse settings and populations and using a range of different intervention modalities (e.g. individual counselling, group counselling, web-based, multi-method approaches), number of visits and treatment durations (7–10).

Although not all studies have observed statistically significant or clinically meaningful weight loss, a systematic review and meta-analysis systematic review of 80 studies with a minimum 1-year follow-up found that behavioural weight loss interventions aimed at reducing caloric intake and increasing caloric expenditure produce modest weight loss (7). The US Preventive Services Task Force recommends that clinicians screen all adults for obesity and offer or refer patients with body mass indexes (BMI) of 30 or higher to intensive behavioural treatment (Grade B recommendation), a recommendation that has been adopted by the American Academy of Family Physicians (11). Despite this evidence, behavioural weight loss interventions have been referred to as ineffective because of the high prevalence of weight regain and lacking methodological rigour in trials (12–15). In depth examination of the methodological quality of trials in this area might increase the veracity of findings within science and in public opinion.

Given the importance of quality in informing practice, policies and public perceptions, evaluation of the body of literature testing behavioural weight loss interventions from a quality perspective is critical to advance the field. The goal of this systematic review was to assess the methodological quality, of published studies testing the effectiveness of behavioural interventions for weight loss among adults and to examine whether quality was associated with report of a statistically significant weight loss outcome, strength of intervention effectiveness and sample size.

Methods

Search strategy

Literature searches for behavioural weight loss trials published between January 1, 2009 and December 31, 2014 were conducted using PUBMED, MEDLINE and PSYCINFO search engines. Search terms included: ‘weight’, ‘loss’, ‘overweight’, ‘obese’, ‘intervention’ and

'trial'. The reference lists of studies that met inclusion criteria (described below) were searched to identify additional articles.

Inclusion and exclusion criteria

Study inclusion criteria included: (i) weight or BMI as a primary or secondary outcome; (ii) intervention that tested a behavioural approach to dietary and/or physical activity change for the purpose of weight loss; (iii) inclusion of an adult (age 18+), non-hospitalized or institutionalized sample, (iv) a minimum sample size of 100 total participants at baseline; (v) a minimum of 6 month follow-up period and (vi) publication in English in a peer-reviewed journal. Exclusion criteria included: (i) study aim was to evaluate efficacy of a pharmacological or surgical intervention; (ii) basic science study; (iii) policy and/or environmental intervention only; (iv) evaluation of an already implemented program; (v) survey development study; (vi) weight maintenance or weight gain prevention was the outcome and (vii) secondary/mediation analysis of a trial. For interventions that had multiple publications meeting inclusion criteria, only the article with the final data assessment point was reviewed.

Initial article review process

To determine whether articles met the study inclusion and exclusion criteria, an initial title review of all articles generated from the search was performed by one author. Two authors (SL and CH) then independently reviewed the abstracts of articles that passed initial review. Inter-rater reliability was calculated (93.5%). The authors met to discuss articles with discrepant codings to achieve consensus. Full text articles were pulled for those deemed to meet inclusion criteria based upon abstract review.

Coding

Three authors (SL, CH, MW) developed the initial coding criteria and coding form. This form was pilot tested on five articles by four authors (SL, CH, MW, DE). These rating forms were reviewed by the study team item by item. Items that yielded any discrepancies were discussed to clarify meaning and wording. This process was repeated on a new set of five articles until 100% agreement was achieved on all data elements. A protocol document that described coding rules for each item was then created for use by coders. Each article had two reviewers. The reviewers met to compare ratings, code inter-rater agreement (IRA) for each item and discuss inconsistencies to come to consensus.

Coding criteria

Study elements abstracted included intervention characteristics, study design characteristics and methodological quality indicators.

Intervention characteristics—assessed included the intervention delivery setting (health care, community organization, worksite, church, college, neighbourhood, academic/lab, combination) (IRA = 86.0%), behavioural target(s) of the intervention (diet, exercise or both) (IRA = 96.5%), intervention modality (group or individual; in-person, phone, web,

other technology) (IRA = 100%) and number of intervention sessions offered/dose (IRA = 100%).

Study design characteristics—Factors related to study methodology included number of study conditions (IRA = 100%), comparison condition (IRA = 98.8%), level of randomization (IRA = 100%), sample size at baseline (IRA = 100%) and final follow-up assessment time period (IRA = 100%), and final follow-up assessment time period in months (IRA = 98.8%). Weight-related outcomes used in the study included change in weight (in both lbs. and kgs.), change in BMI, percent of weight change and achieving a specific weight loss percentage (typically 5% of baseline weight) (IRA = 86.2%).

Study effectiveness—was assessed in two ways. First, we abstracted whether or not a statistically significant result ($p < .05$) was reported for any weight-related outcome at the final assessment point (IRA = 98.8%). Second, within each intervention condition, baseline values and results at the final assessment point were abstracted for each weight-related outcome that was reported (IRA = 100%). Types of outcomes included were change in weight (in either pounds or kilogrammes), change in BMI, change in percentage weight loss and the percentage of the sample that achieved a pre-defined weight loss threshold (e.g. 5% of baseline weight). For intervention conditions that included a given outcome, we calculated the median and interquartile range across those results. Then, within each intervention condition that included change in weight or BMI, a Cohen's D statistic was computed in order to standardize estimates across the multiple ways in which weight-related outcomes were assessed. Cohen's D was calculated as the standardized mean change from baseline in weight or BMI. Mean change in weight or BMI at the last follow-up visit was divided by the standard deviation of the baseline measure. The effect size was very similar for studies reporting both change in weight and change in BMI. Because of this, when both were reported, we used effect size for change in weight in our analyses as this was the most commonly reported outcome measure.

Methodological quality—We included ten indicators reflecting key components of methodological and/or reporting quality (16). Quality indicators were an amalgam of known standards for rating intervention studies that were adapted by the study team to be specific for behavioural weight loss interventions (17–19). Individual indicators, rather than a summary score, were used so that specific elements could be assessed (17,18). The quality indicators selected therefore covered these two areas and were as follows:

1. *Study design* was assessed (randomized controlled trial (RCT), non-randomized design with control group, single arm design). RCTs were considered the gold standard. Level of randomization (i.e. individual participant, provider, site) was also assessed (IRA = 100%).
2. We assessed the *measurement approach* used to assess weight and/or height data. Options included: measured by research staff, obtained from medical or other records, self-report or a combination of these (IRA = 100%).
3. Whether or not *power/sample size calculations* were reported was assessed (yes/no) (IRA = 95.3%).

4. We defined '*intent-to-treat' (ITT)* analysis as whether or not participants' data were included in the intervention condition to which they were randomized in the analysis, regardless of whether or not they participated in/attended that intervention/program/ treatment. ITT was considered independently of how missing data at follow-up was handled. Because of discrepancies in how investigators define ITT (20–23), text was reviewed for a clear definition of ITT consistent with the one used in this study and CONSORT-type diagrams and denominators in tables were reviewed to establish whether the ITT criteria was met. ITT was classified as yes or no (IRA = 95.3%).

5. *Loss to follow-up rate* was assessed by determining the percentage of individuals at baseline who completed assessments at the final follow-up point. We used the most conservative approach, not taking into consideration specific study designed exclusion criteria (e.g. some studies excluded participants who became pregnant or began taking weight gain medications), we used this approach to apply a consistent definition across studies (IRA = 93.5%). While there is no universal consensus on what loss follow-up rate results in biassed study results, we considered less than 25% to indicate high quality (24), as suggested by the Methodological Quality Rating Scale.

6. *Missing data strategy* refers to the approach(es) used to handle missing data at follow-up assessments. Classifications were based on standards described in 2010 by the National Research Council (25) and included: no or minimal (>95% of the baseline sample included in the final follow-up time point) missing data (26), multiple imputation, generalized estimating equations or other non-parametric model-based approaches, inverse probability weighting, maximum likelihood/parametric models, imputed weight gain carried forward; single imputation (i.e. last or baseline carried forward) and complete case analysis with and without comparisons to baseline sample). We additionally assessed whether sensitivity analyses were performed using more than one method (IRA = 96.5%).

7. The *sampling strategy* used to enrol participants was classified as being from a known population (i.e. having a denominator) or not (IRA = 91.9%). Among those studies with a known population, we assessed whether comparisons were made between the included sample and the known population (IRA = 100%).

8. We assessed whether data were reported in participant *treatment receipt or adherence* was reported (IRA = 97.7%).

9. Last, we assessed whether attempts to ensure *intervention fidelity* were described or reported (IRA = 95.3%). We additionally assessed the reporting/methods used for fidelity which was categorized as none stated, followed a standardized protocol, or quality review by study staff.

Analysis

The distribution of intervention characteristics, study design characteristics and each of the nine methodological quality indicators were computed among the included studies. To determine whether or not methodological quality was associated with intervention

effectiveness, bivariate comparisons were made to assess the associations of the sum of quality indicators with whether or not the study reported a statistically significant outcome at the final assessment point (*t*-test) and the strength of association/Cohen's D statistic (using Pearson's correlation coefficient). The latter was computed only among those studies that included change in weight (lbs. or kgs.) or BMI as an outcome. The bivariate association of the study sample size and the sum of number of quality indicators was also assessed (one-way ANOVA).

Results

Study selection

A total of 5,499 non-duplicate articles were retrieved from the initial search strategy (Fig. 1). Of these, 4,765 were excluded based on title review; 734 abstracts were reviewed. Of these, 543 were excluded and 191 articles were selected for full review. An additional six articles were identified from reference review for full review, resulting in 197 articles for full review. Of these, 107 were deemed ineligible. The most common reasons for ineligibility included: follow-up period of less than 6 months ($n = 30$), results presented in another publication ($n = 17$) and methods paper ($n = 9$). A total of 90 articles were included in this review; 156 unique interventions were tested in these 90 studies. A list of these articles is included in the online Appendix.

Study characteristics

Table 1 describes the intervention characteristics of the included studies. Most targeted both diet and physical activity (93.3%). The most common primary intervention delivery settings were health care (38.9%) and academic research labs (36.7%). A range of intervention delivery modalities were used, with many studies reporting multiple modalities within a single intervention condition and/or across intervention conditions. More than half of the interventions (51.3%) offered more than 20 sessions.

Table 2 presents a summary of the design characteristics and outcomes of the included studies. The majority included two treatment conditions (66.7%). A wide variety of comparison conditions were used, with comparative effectiveness (i.e. comparing one intervention approach to other) (48.9%) (either with or without a no intervention control group), usual care (22.2%) and materials only (20.0%) being the most common. Most studies included more than one outcome measure. The majority of the 90 studies included change in weight and/or BMI as outcomes (87.7%). Among those that did not include such a measure, outcomes included proportion of participants who reached a certain threshold of weight loss (typically 5%) (1.1%); percentage of weight loss achieved (8.9%); or both a threshold of weight loss and weight loss percentage (2.2%).

One hundred forty of the 156 unique interventions included in the 90 studies presented change in weight as an outcome (in pounds or kilogrammes). Converting each of these to pounds, median weight loss in the 140 interventions was −5.13 (interquartile range (IQR) = −10.14 to −2.55). The median change in BMI, among the 86 interventions that assessed it, was −.80 (IQR = −1.8 to −.30). With respect to change in percentage weight loss (included

in 46 interventions), the median was −4.7% (IQR = −1.68% to −6.4%). Among studies that included proportion achieving a specified weight loss threshold, a median of 36.0% of participants did so (IQR = 21% – 49%). Cohen's D could be calculated for 76 of the 79 studies that assessed change in weight or BMI; these included 133 unique weight loss interventions. Mean Cohen's D statistic was −.26 (sd = .24; median = −.21). Almost half (47.4%) demonstrated no effect; 36.1% a small effect; 12.8% a medium effect and 3.8% a large effect.

Methodological quality

A summary of the methodological quality indicators is presented in Table 3. Quality indicators that were most commonly utilized included a randomized controlled trial design (100% (individual level randomization 95.6%, site level randomization 4.4%), objectively measured weight/height (96.7%) and description of an ITT analytic approach (86.7%). Most studies did not describe fidelity assessments (65.6%), with 15.5% reporting following a standardized protocol and 18.9% reporting quality review by the study team. Sixty-two percent (62.2%) of studies had a follow-up retention rate of more than 25%. The majority of studies reported using a strategy to address loss to follow-up (either primary or sensitivity analysis) or having minimal missing follow-up data (69.9%). A variety of approaches to loss to follow-up were used; the most common were last/baseline value carried forward (35.6%) and multiple imputation (24.5%), either alone or as a sensitivity analysis to a primary analysis that used a complete case approach. Overall, the mean number of quality indicators achieved was 6.3, with a median of 6 (range 2–9).

No differences in methodological quality were observed between studies that reported a statistically significant reduction in a primary outcome at final assessment point (mean number of quality indicators = 6.3 (SD = 1.5)) and those that did not (mean = 6.0 (SD = 1.4); $p = .31$). We also found no association between methodological quality and effect size (Pearson's correlation coefficient = 0.07) among those studies for which a Cohen's D statistic could be calculated. No differences in quality indicators were observed by study sample size (mean number of quality indicators for sample size 0–199 = 6.1 (SD = 1.4); 200–399 = 6.3 (SD = 1.5); 400 or greater = 6.2 (SD = 1.4); $p = 0.88$).

Discussion

In this study, the methodological quality of contemporary trials testing the efficacy and effectiveness of behavioural weight loss interventions among adults was examined, with an emphasis on key indicators of internal and external validity. A total of 90 articles using a wide range of intervention modalities, comparison conditions and conducted in a variety of settings were included. Overall, the studies were of moderate to high quality (median of six out of nine quality indicators), with considerable variation across indicators.

Internal validity indicators that were assessed included key features that have potential to bias the results of weight loss trials. Most studies used a protocol that required weight and height to be objectively measured, a positive finding in light of the recognized bias in the use of self-report or other methods in attaining these measures (27,28). Most studies additionally clearly stated that ITT analysis was completed, that is they were included in the analysis in

the intervention group to which they were assigned regardless of their participation in or adherence to the intervention. Studies that did not use an ITT approach typically excluded participants who did not meet a specified intervention participation benchmark, which has potential to overestimate intervention effects.

While the majority of studies used a strategy to address loss to follow-up, more than one-fifth did not, instead using a complete case analysis approach. This can lead to reduced power and selection bias, particularly if loss to follow-up is differential across conditions or related to weight loss. Traditional clinical trial methodology additionally call for the most conservative approach to addressing loss to follow-up be used in order to decrease the probability of committing a type 1 or 2 error. In this review, we found that single imputation methods, such as last observation carried forward and baseline observation carried forward were most commonly used, which has been true in other areas. Such approaches, however, are typically not recommended as a primary approach for handling missing data because they can introduce bias (25). This may be particularly true with weight-related outcomes. On average, American adults are estimated to gain 1 to 2 pounds per year (29). Thus, baseline and last value carried forward approaches, the mostly commonly used in the studies included in this review and the easiest to implement statistically, may not be the most conservative approach. Another potential loss to follow up strategy is to impute an average weight gain to coincide with estimated population weight gain. However, only one study included in this review used this approach. More advanced statistical approaches such as multiple imputation, model-based approaches, while more computationally intensive, are increasingly considered a gold standard. In a meta analysis of pharmaceutical randomized controlled trials published between 2000 and 2006 that assessed weight loss or weight gain, Elobeid and colleagues compared various loss to follow-up analysis approaches on trial results. The authors concluded that multiple imputation and model based approaches are the most appropriate for weight loss trials (30). Less than 30% of studies included in this review used such a strategy.

There have been calls within the field of behavioural medicine and related disciplines to increase reporting of indicators that relate to project implementation and external validity in an effort to better promote the translation of research into practice and enhance the public health impact of research-developed interventions (31–33). In this review, only 32.7% of studies reported on intervention fidelity, despite these calls to action and the availability of frameworks and standards for fidelity reporting (34–36). Only 17.3% including objective reviews such as reviewing sessions in person or via audiotape for protocol adherence; other methods reported included describing training of study staff, use of standardized protocols, selecting intervention staff with specialized degrees and supervision of interventionists. It is possible that quality control and fidelity checks were in place, but not reported in many studies, particular given journal restrictions. Intervention adherence rates were reported by 76.7% of studies. We found that only 41.1% of studies samples from a known population, with the majority enrolling convenience samples. Of these, only one study made comparisons between the known population and the included sample. Thus, most studies were not designed to assess reach, and among those that were, only one attempted to describe the success in achieving a representative population. The results of these three indictors (fidelity, participation and reach/representativeness) are similar to those reported in

a systematic review of childhood obesity prevention interventions published from 1980 to 2008 (37).

We found that methodological quality was not associated with whether or not a statistically significant weight loss result was reported, effect size, or sample size. This indicates that differences in methodological quality are not driving inconsistencies in translation of efficacy findings to effectiveness studies. Instead, such differences are likely a function of intervention design, intensity and/or delivery and the associated challenges of achieving behaviour change across diverse populations and settings. Mean weight loss in the studies included was 6.79 pounds (sd = 6.70) relative to 14.33 (sd = 10.36) in the Diabetes Prevention Program and 9.23 (sd = 11.24) in the Finnish Diabetes Prevention Program, both of which found significant associated reductions in diabetes risk of 58% (38). While weight losses in effectiveness studies were 47–74% of those of these larger efficacy trials and the degree to which this magnitude of weight loss affects diabetes risk is unknown, it would seem premature to dismiss these interventions as ineffective. Two main criticisms have been launched against behavioural weight loss interventions. The first is that weight regain is extremely common and is an indicator of treatment failure (13,14). Although weight regain is common over the long-term, the Diabetes Prevention Program found that 10 years after the lifestyle intervention, in spite of average weight regain of 71%, reduction in incidence of diabetes was 34% relative to control (39). The Da Qing Diabetes Study found a reduction in diabetes incidence of 43% relative to control at 20 years although did not report weight regain rates (40). Weight loss may not be the best measure of impact of lifestyle interventions, although it is the most feasible measure of impact in shorter term effectiveness studies, which poses a challenge to this field. It should also be noted that termination of treatment for many conditions results in a return of symptoms. Lifestyle interventions may not be a permanent cure for obesity, but this is an extraordinarily high bar for any preventive intervention (41). Chronic care models are certainly needed. The second criticism of weight loss interventions is that they lack methodological rigour (12). Our findings show that recent trials are of moderate to high quality.

This study must be viewed in the context of its limitations. Only studies published in English were included. Publication bias is a well-known limiting factor in synthesizing research evidence that could impact the conclusions of this study (42). Studies with positive or significant results are more likely to be published in the scientific literature, and selected outcomes with null findings are more likely to be intentionally excluded from published reports. We additionally restricted articles to those with at least 100 people and six months or more follow-up time. The former was done to eliminate pilot type studies which typically are not designed with maximal methodological rigour and are not powered to find statistically significant results. The later restriction was improved because we were interested in studies designed to assess weight loss beyond the short-term impact of the intervention. We examined whether studies were statistically significant, rather than using an indicator of whether or not clinically meaningful weight loss was achieved. This was done because vast heterogeneity in how results were reported made it impossible to operationalize clinically meaningful results. Only a select number of quality indicators were included, indicators were limited to those deemed by the study team to be most likely to impact study bias. Lastly, we were only able to assess what was reported in the included studies. It is

possible that a given study could have included/addressed a given quality indicator and the investigators chose not to report it for reasons such as journal word count restrictions, response to journal editors and/or reviewers and the investigators' simply not perceiving it to be important.

In summary, this systematic review of behavioural trials testing weight loss interventions found this body of literature to be of high quality for specific quality indicators, including study design and measurement. However, specific areas in need of improvement were identified. This review identified a need for utilization of more methodologically rigorous statistical approaches to address loss to follow-up in such trials, more representative sampling approaches and better reporting of fidelity, in particular. These are potential areas of improvement for this field of research that are relevant to investigators, journal editors and reviewers.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Drs. Lemon and Pagoto receive funding from Grant # 5 U48 DP005031; Dr. Pagoto receives funding from Grant # K24 HL124366; Dr. Wang receives funding from Grant # K01 DK102447; and Christina Haughton is supported by Grant # T32 HL120823.

References

1. Ioannidis JP. Why most published research findings are false. PLoS Med. 2005; 2:e124. [PubMed: 16060722]
2. PSYCHOLOGY. Estimating the reproducibility of psychological science. Science. 2015; 349:aac4716. [PubMed: 26315443]
3. Knowler WC, Barrett-Connor E, Fowler SE, et al. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med. 2002; 346:393–403. [PubMed: 11832527]
4. Lindstrom J, Louheranta A, Mannelin M, et al. The Finnish Diabetes Prevention Study (DPS): lifestyle intervention and 3-year results on diet and physical activity. Diabetes Care. 2003; 26:3230–3236. [PubMed: 14633807]
5. Pan XR, Li GW, Hu YH, et al. Effects of diet and exercise in preventing NIDDM in people with impaired glucose tolerance. The Da Qing IGT and Diabetes Study. Diabetes Care. 1997; 20:537–544. [PubMed: 9096977]
6. Wing RR. Long-term effects of a lifestyle intervention on weight and cardiovascular risk factors in individuals with type 2 diabetes mellitus: four-year results of the Look AHEAD trial. Arch Intern Med. 2010; 170:1566–1575. [PubMed: 20876408]
7. Franz MJ, VanWormer JJ, Crain AL, et al. Weight-loss outcomes: a systematic review and meta-analysis of weight-loss clinical trials with a minimum 1-year follow-up. J Am Diet Assoc. 2007; 107:1755–1767. [PubMed: 17904936]
8. Williams RL, Wood LG, Collins CE, Callister R. Effectiveness of weight loss interventions—is there a difference between men and women: a systematic review. Obes Rev. 2015; 16:171–186. [PubMed: 25494712]
9. Franz MJ, Boucher JL, Rutten-Ramos S, VanWormer JJ. Lifestyle weight-loss intervention outcomes in overweight and obese adults with type 2 diabetes: a systematic review and meta-analysis of randomized clinical trials. J Acad Nutr Diet. 2015; 115:1447–1463. [PubMed: 25935570]

10. Raaijmakers LC, Pouwels S, Berghuis KA, Nienhuijs SW. Technology-based interventions in the treatment of overweight and obesity: a systematic review. Appetite. 2015; 95:138–151. [PubMed: 26165415]
11. American Academy of Family Physicians. [accessed October 27, 2015] Clinical Preventive Services Recommendation: Obesity. 2015. [WWW document]. URL http://aafp.org/patient-care/clinical-recommendations/all/obesity.html
12. Mann T, Tomiyama AJ, Westling E, Lew AM, Samuels B, Chatman J. Medicare's search for effective obesity treatments: diets are not the answer. American Psychol. 2007; 62:220–233.
13. Bacon L, Aphramor L. Weight science: evaluating the evidence for a paradigm shift. Nutr J. 2011; 10:9. [PubMed: 21261939]
14. Taubes G. The science of obesity: what do we really know about what makes us fat? An essay by Gary Taubes. BMJ. 2013; 346:f1050. [PubMed: 23596125]
15. Aphramor L. Validity of claims made in weight management research: a narrative review of dietetic articles. Nutr J. 2010; 9:30. [PubMed: 20646282]
16. Higgins, JPT., Green, S. [accessed October 5, 2015] Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0. The Cochrane Collaboration. (updated March 2011) [WWW document]. www.cochrane.org/resources/handbook
17. Miller WR, Wilbourne PL. Mesa Grande: a methodological analysis of clinical trials of treatments for alcohol use disorders. Addiction. 2002; 97:265–277. [PubMed: 11964100]
18. Community Preventive Services Task Force. [October 21, 2015] Systematic review methods. (last updated June 2014). [WWW document]. URL http://www.thecommunityguide.org/about/methods.html
19. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Int J Surg. 2010; 8:336–341. [PubMed: 20171303]
20. Abraha I, Montedori A. Modified intention to treat reporting in randomised controlled trials: systematic review. BMJ. 2010; 340:c2697. [PubMed: 20547685]
21. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. BMJ. 1999; 319:670–674. [PubMed: 10480822]
22. Abraha I, Cherubini A, Cozzolino F, et al. Deviation from intention to treat analysis in randomised trials and treatment effect estimates: meta-epidemiological study. BMJ. 2015; 350:h2445. [PubMed: 26016488]
23. Alshurafa M, Briel M, Akl EA, et al. Inconsistent definitions for intention-to-treat in relation to missing outcome data: systematic review of the methods literature. PLoS One. 2012; 7:e49163. [PubMed: 23166608]
24. Miller, WR., Simpson, TL. What works? A methodological analysis of the alcohol treatment outcome literature. In: Hester, RK., editor. Handbook of Alcoholism Treatment Approaches: Effective Alternatives. Allyn & Bacon; Needham Heights, MA: 1995. p. 12-44.
25. National Research Council. The Prevention and Treatment of Missing Data in Clinical Trials. Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics. Division of Behavioral and Social Sciences and Education. The National Academic Press; Washington, DC: 2010.
26. Fewtrell MS, Kennedy K, Singhal A, et al. How much loss to follow-up is acceptable in long-term randomised trials and prospective studies? Arch Dis Child. 2008; 93:458–461. [PubMed: 18495909]
27. Merrill RM, Richardson JS. Validity of self-reported height, weight, and body mass index: findings from the National Health and Nutrition Examination Survey, 2001–2006. Prev Chronic Dis. 2009; 6:A121. [PubMed: 19754997]
28. Preston SH, Fishman E, Stokes A. Effects of categorization and self-report bias on estimates of the association between obesity and mortality. Ann Epidemiol. 2015; 25:907–911. e2. [PubMed: 26385831]
29. Hill JO, Wyatt HR, Reed GW, Peters JC. Obesity and the environment: where do we go from here? Science. 2003; 299:853–855. [PubMed: 12574618]

30. Elobeid MA, Padilla MA, McVie T, et al. Missing data in randomized clinical trials for weight loss: scope of the problem, state of the field, and performance of statistical methods. PLoS One. 2009; 4:e6624. [PubMed: 19675667]

31. Glasgow RE, Green LW, Klesges LM, et al. External validity: we need to do more. Ann Behav Med. 2006; 31:105–108. [PubMed: 16542124]

32. Green LW, Glasgow RE. Evaluating the relevance, generalization, and applicability of research: issues in external validation and translation methodology. Eval Health Prof. 2006; 29:126–153. [PubMed: 16510882]

33. Allen, JD., Linnan, LA., Emmons, KM. Fidelity and its relationship to implementation effectiveness, adaptation and dissemination. In: Brownson, RC., Coldtz, GA., editors. Dissemination and Implementation Research in Health: Translating Science to Practice. Oxford University Press; New York, NY: 2012.

34. Pinnock H, Epiphaniou E, Sheikh A, et al. Developing standards for reporting implementation studies of complex interventions (StaRI): a systematic review and e-Delphi. Implement Sci. 2015; 10:42. [PubMed: 25888928]

35. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: the RE-AIM framework. Am J Public Health. 1999; 89:1322–1327. [PubMed: 10474547]

36. Tabak RG, Khoong EC, Chambers DA, Brownson RC. Bridging research and practice: models for dissemination and implementation research. Am J Prev Med. 2012; 43:337–350. [PubMed: 22898128]

37. Klesges LM, Williams NA, Davis KS, Buscemi J, Kitzmann KM. External validity reporting in behavioral treatment of childhood obesity: a systematic review. Am J Prev Med. 2012; 42:185–192. [PubMed: 22261216]

38. Lindstrom J, Eriksson JG, Valle TT, et al. Prevention of diabetes mellitus in subjects with impaired glucose tolerance in the Finnish Diabetes Prevention Study: results from a randomized clinical trial. J Am Soc Nephrol. 2003; 14:S108–S113. [PubMed: 12819313]

39. Knowler WC, Fowler SE, Hamman RF, et al. 10-year follow-up of diabetes incidence and weight loss in the Diabetes Prevention Program Outcomes Study. Lancet. 2009; 374:1677–1686. [PubMed: 19878986]

40. Li G, Zhang P, Wang J, et al. The long-term effect of lifestyle interventions to prevent diabetes in the China Da Qing Diabetes Prevention Study: a 20-year follow-up study. Lancet. 2008; 371:1783–1789. [PubMed: 18502303]

41. Pagoto SL, Appelhans BM. A call for an end to the diet debates. JAMA. 2013; 310:687–688. [PubMed: 23989081]

42. Dwan K, Gamble C, Williamson PR, Kirkham JJ. Systematic review of the empirical evidence of study publication bias and outcome reporting bias—an updated review. PLoS One. 2013; 8:e66844. [PubMed: 23861749]

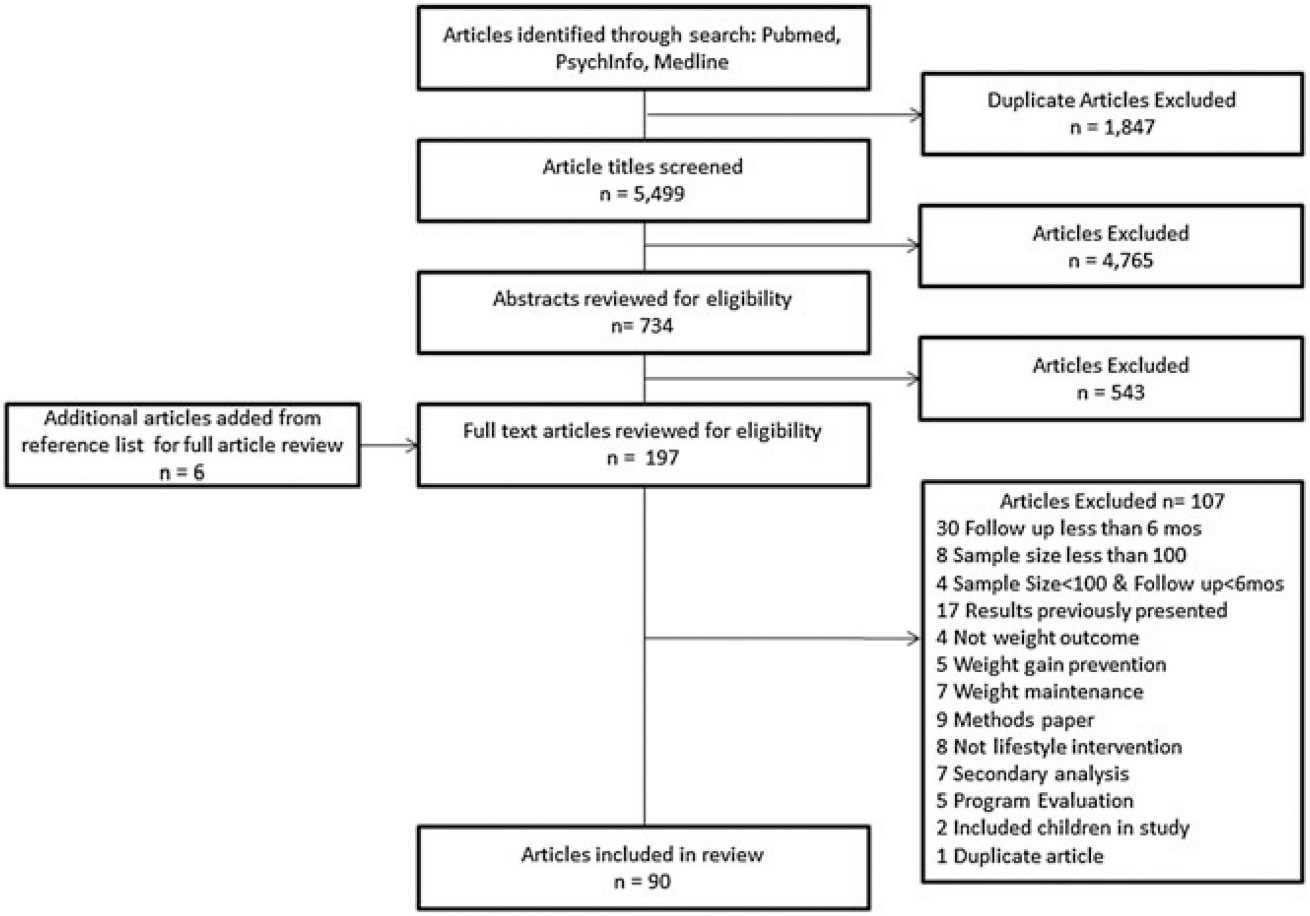

Figure 1.
Diagram of study inclusion.

Table 1

Description of intervention characteristics of included studies ($n = 90$)

	***N* (%)**
Behavioural targets	
Diet only	6 (6.7%)
Diet and physical activity	84 (93.3%)
Delivery setting	
Health care	35 (38.9%)
Academic/research lab	33 (36.7%)
Community organization	8 (8.9%)
Worksite	7 (7.8%)
Commercial weight loss	3 (3.3%)
Healthcare and community	3 (3.3%)
Not described	1 (1.1%)
Intervention modality *	
In-person group	47 (30.1%)
In-person individual counselling	54 (34.6%)
In-person groups and individual counselling	25 (16.0%)
Individually-targeted web or other technology-based individual counselling	21 (13.5%)
Group-targeted web or other technology-based	4 (2.6%)
Web or other technology-based groups combined with individual counselling	5 (3.2%)
Maximum number of intervention sessions *	
1–10	28 (18.0%)
11–20	39 (25.0%)
21–30	37 (23.7%)
31–40	11 (7.1%)
41+	32 (20.5%)
Unknown/web-based continuous access	9 (5.8%)

* Total equals 156 because of studies testing more than one intervention arm.

Table 2

Description of methodological characteristics and results of included studies (n = 90)

	N (%)
Last follow-up period	
6 months	19 (21.1%)
7–12 months	36 (40.0%)
13–24 months	31 (34.4%)
>24 months	4 (4.4%)
Number of treatment arms	
2	60 (66.7%)
3	20 (22.2%)
4+	10 (11.1%)
Sample size	
100–199	28 (31.1%)
200–399	40 (44.4%)
400 or greater	22 (24.4%)
Comparison condition	
Nothing/usual care	20 (22.2%)
Attention control	1 (1.1%)
Materials only	18 (20.0%)
Wait list	6 (6.7%)
Fitness club membership	1 (1.1%)
Other approach/comparative effectiveness/no control	27 (30.0%)
Comparative effectiveness and other control group	17 (18.9%)
Outcome measures *	
Change in weight (lbs. or kg.) and/or BMI only	47 (52.2%)
Change in weight and/or BMI and percent weight loss	10 (11.1%)
Change in weight and/or BMI and proportion achieving a threshold	12 (13.3%)
Change in weight and/or BMI, percent weight loss and proportion achieving a threshold	10 (11.1%)
Percent weight loss only	8 (8.9%)
Proportion achieving a threshold only	1 (1.1%)
Percent weight loss and proportion achieving a threshold	2 (2.2%)
Intervention effect size *	
No effect (Cohen's D < .2)	63 (47.4%)
Small effect (Cohen's D .2–.49)	48 (36.1%)
Moderate effect (Cohen's D .5–.79)	17 (12.8%)
Large effect (Cohen's D ≥ .8)	5 (3.8%)

* Includes 133 interventions conditions from 76 studies with means and standard deviations for weight loss and/or BMI at baseline and final follow-up assessment.

Table 3

Description of methodological quality on included studies (n = 90)

	***N* (%)**
Study design	
Randomized controlled trial	90 (100%)
Objectively measured weight/height	8(96.7%)
Power/sample size calculations presented	59 (65.6%)
Intent-to-treat analysis	78 (86.7%)
Loss to follow-up rate	
<25%	56 (62.2%)
25–49%	31 (34.4%)
≥50%	3 (3.3%)
Missing data/loss to follow-up strategy	
<5% missing data	3 (3.3%)
Multiple imputation	15 (16.7%)
Generalized estimating equations or other non-parametric modelling approaches	1 (1.1%)
Inverse probability weighting or maximum likelihood methods	0
Maximum likelihood/parametric models	3 (3.3%)
Imputed weight gain carry forward	1 (1.1%)
Baseline or last value carried forward only	18 (20.0%)
Complete case analysis with sensitivity analysis-multiple Imputation	7 (7.8%)
Complete case analysis with sensitivity analysis-last value or baseline carried forward	14 (15.6%)
Complete case analysis only	22 (24.4%)
Unknown/not described	6 (6.7%)
Sampled from known population	37 (41.1%)
Comparisons to target population	1 (1.1%)
Treatment receipt/adherence rate reported	69 (76.7%)
Intervention fidelity reported	31 (34.4%)