



Published in final edited form as:

Obes Rev. 2016 December ; 17(12): 1287–1300. doi:10.1111/obr.12464.

Reporting of Treatment Fidelity in Behavioral Pediatric Obesity Intervention Trials: A Systematic Review

Meghan M JaKa^{1,2,*}, Jacob L Haapala², Erika S Trapl³, Alicia S Kunin-Batson¹, Barbara A Olson-Bullis¹, William J Heerman⁴, Jerica M Berge⁵, Shirley M Moore⁶, Donna Matheson⁷, and Nancy E Sherwood¹

¹HealthPartners Institute

²Division of Epidemiology and Community Health, University of Minnesota

³Department of Epidemiology and Biostatistics, Case Western Reserve University

⁴Department of Pediatrics and Internal Medicine, Vanderbilt University

⁵Family Medicine and Community Health, University of Minnesota Medical School

⁶Frances Payne Bolton School of Nursing, Case Western Reserve University

⁷Department of Pediatrics, Stanford University

Abstract

Behavioral interventions for pediatric obesity are promising, but detailed information on treatment fidelity (i.e., design, training, delivery, receipt, and enactment) is needed to optimize the implementation of more effective interventions. Little is known about current practices for reporting treatment fidelity in pediatric obesity studies. This systematic review, in accordance with PRISMA guidelines, describes the methods used to report treatment fidelity in randomized controlled trials. Treatment fidelity was double-coded using the NIH Fidelity Framework checklist. Three hundred articles (N=193 studies) were included. Mean inter-coder reliability across items was 0.83 (SD=0.09). Reporting of treatment design elements within the field was high (e.g., 77% of studies reported designed length of treatment session), but reporting of other domains was low (e.g., only 7% of studies reported length of treatment sessions delivered). Few reported gold standard methods to evaluate treatment fidelity (e.g., coding treatment content delivered). General study quality was associated with reporting of treatment fidelity ($p<0.01$) as was the number of articles published for a given study ($p<0.01$). The frequency of reporting treatment fidelity components has not improved over time ($p=0.26$). Specific recommendations are made to support pediatric obesity researchers in leading health behavior disciplines toward more rigorous measurement and reporting of treatment fidelity.

*Corresponding Author: 3311 East Old Shakopee Road, Bloomington, MN 55425, Meghan.M.JaKa@HealthPartners.com, Phone: 952-967-5268, Fax: 952-883-5004.

CONFLICTS OF INTEREST STATEMENT

Authors MMJ, JLH, EST, ASK, BAO, WJH, JMB, SMM, DM, and NES have no conflicts of interest to declare.

Key Words (MeSH terms)

Health behavior; pediatric obesity; therapy; systematic review

INTRODUCTION

The prevalence of pediatric obesity remains alarmingly high.^{1–6} Given the known health risks,^{7–13} societal burden,¹⁴ and healthcare costs,¹⁴ managing and preventing this disease is a public health priority. A number of behavioral interventions to address pediatric obesity have been reported, yet interventions that produce reliable, long-term impacts on child weight are rare.^{15–19} Obtaining a better understanding of *why* some of these interventions have not led to desired outcomes is critical for informing the development of more effective interventions in the future. Without information on treatment fidelity (defined as treatment design, provider training, and treatment delivery, receipt and enactment²⁰), it is difficult to interpret null findings and to replicate significant ones. For example, in the absence of information on treatment fidelity, it is impossible to determine whether the treatment itself is not efficacious or whether the intervention was not delivered as intended. Reporting findings from treatment fidelity methods is a critical step in moving the field forward, as an improved focus on fidelity may ultimately lead to enhanced treatment efficacy.²⁰ Treatment fidelity has become even more important with the increasing focus on multi-component behavioral obesity interventions. Without in-depth descriptions of fidelity for each intervention component, it is unclear which components worked and which did not. Clear, detailed, and consistent documentation of treatment fidelity across the field will help researchers improve future interventions.

The first step in improving treatment fidelity is to determine which components are typically measured and reported across the field. Very little is known about current practices in pediatric obesity research.²¹ Authors of multiple systematic reviews have commented on the lack of published information available to evaluate treatment fidelity as a predictor of treatment outcomes, estimating that measurement of any component of fidelity was reported only 5–30% of the time.^{22–26} None of these reviews provide detailed information on what is or is not published. It is not clear which components of treatment fidelity are commonly used; only that use is relatively low across the field. Tools have been developed to assist researchers in standardizing the measurement and reporting of treatment fidelity. The Template for Intervention Description and Replication (TIDieR) checklist gives specific guidance on what to present,²¹ but does not address items specific to behavioral interventions. The Workgroup for Intervention Development and Evaluation Research (WIDER) checklist is another tool promoting a standardized approach to treatment fidelity in randomized controlled trials.²² It includes elements of the TIDieR checklist and adds specific items related to behavioral interventions (e.g., behavioral change techniques). The National Institutes of Health (NIH) Treatment Fidelity Framework, designed by the Behavior Change Consortium (BCC) includes similar items to the TIDieR and WIDER checklists, but further breaks down fidelity into 5 domains: (1) treatment design, (2) provider training, (3) treatment delivery, (4) treatment receipt, and (5) treatment enactment.²⁷ This is the most detailed tool available and has been used to explore the quality of reporting in other

health behavior fields and across behavior change research more broadly.^{28–30} This review is the first to use this tool to explore treatment fidelity in pediatric obesity interventions.

The NIH Treatment Fidelity Framework²⁷ proposes that all 5 domains of treatment fidelity are necessary when reporting the results of intervention trials. (1) Treatment design refers to how an intervention was intended to be delivered and includes theoretical frameworks, intended dose, intended content, and intended qualifications of treatment providers. (2) The provider training component addresses what specific methods will be used to train providers and maintain provider skills throughout the intervention. (3) Treatment delivery corresponds to how well the providers adhere to the intended treatment, and includes information about actual dose and content delivered, as well as the measurement of non-specific factors. (4) Treatment receipt refers to how well the intervention addresses participants' comprehension of and ability to use learned skills during treatment sessions; (5) treatment enactment refers to participants' ability to use these skills outside of formal treatment sessions. Failing to measure and report any one of these components inhibits readers' ability to interpret findings. There is a clear need for improved reporting of treatment fidelity, which will ultimately lead to improved efficacy of future behavioral pediatric obesity interventions.

The current systematic review aims to describe in detail how the childhood obesity prevention and management field reports components of treatment fidelity in randomized controlled trials. Specifically, we aim to (1) identify which domains of treatment fidelity are most commonly used in the field; (2) describe what methods are used to measure these domains; (3) examine associations between treatment fidelity and study quality, number of study articles reviewed, and publication year; and (4) make specific recommendations for more rigorous measurement and reporting of treatment fidelity. Based on previous work, it is hypothesized that treatment design will be the most consistently reported domain and that study quality, number of articles reviewed, and publication year will all be positively associated with treatment fidelity. Throughout this article, the term "treatment" fidelity is used to refer to the fidelity of any interventions aimed at either treating *or* preventing obesity as this is the term is consistently used in other fields. Understanding how treatment fidelity is currently used by researchers in this field can lead to the development of best-practice guidelines and ultimately to more efficacious interventions.

METHODS

Protocol and registration

This review was conducted in accordance with PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) guidelines.³¹ All methods were specified in advance and documented in a protocol. The protocol was registered on PROSPERO (*Registration #CRD42016036124, date registered March 11, 2016*) and can be accessed here: www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42016036124.

Information sources and search strategy

A trained health sciences librarian with experience in conducting and documenting searches for systematic reviews performed an extensive search of the literature to identify

intervention studies or randomized controlled trials (RCTs) on the management or prevention of pediatric obesity published in the English language. The PubMed (Web-based), Cumulative Index to Nursing and Allied Health Literature (CINAHL—EBSCO platform), PsycINFO (Ovid platform), and EMBASE (Ovid platform) databases were used in this systematic review. Dissertations, books, book chapters, and conference proceedings/abstracts were excluded. In PubMed, the medical subject headings (MeSH) terms defined the concepts of obesity, overweight, or body mass index; treatment, therapy, or prevention; children, childhood, adolescents, or pediatric (under 18 years of age); and RCTs or intervention studies. For optimal retrieval, all terms were supplemented with relevant title and text words. Full PubMed search parameters are available on PROSPERO. The search strategies for CINAHL, PsycINFO, and EMBASE were adjusted for the syntax appropriate for each database using a combination of thesauri and text words. Published reports in the peer-reviewed literature from January 1990 to March 2014 were identified. If an article for a given study met the inclusion criteria and other ancillary study articles were referenced but not identified in the original search (e.g., published after March 2014), these articles were identified and included. Study authors were not contacted to identify additional information, as the primary purpose of this systematic review was to evaluate what is reported in the available literature. Finally, bibliographies from selected key systematic review articles were scanned to identify additional publications.

Study selection process

Articles were independently evaluated for selection in a two-step process by a group of 8 coders. First, titles and abstracts of all identified articles were reviewed by two independent reviewers to make initial exclusions. Exclusion reasons were recorded. All discrepancies were adjudicated by the lead reviewer (MMJ) and discussed with the secondary reviewer when necessary. Then, two reviewers independently read full texts of articles that were not excluded above to determine final selection for inclusion. Studies that did not meet inclusion criteria were removed at that time and reasons for exclusion were documented. Differences were again adjudicated by the lead reviewer (MMJ) and discussed with secondary reviewers when needed. Though all records were reviewed by two reviewers and differences were adjudicated, inter-coder agreement for this step was evaluated. Articles from a single study were then combined into a single record and all study records were imported into REDCap.³² Included studies were published randomized trials testing behavior change interventions to impact weight status of children between the ages of 2–18 at the time of randomization. Table 1 includes detailed information about inclusion and exclusion criteria with reference to PICOS (participants, interventions, comparisons, outcomes, and study design).³¹ This review was designed to evaluate the reporting of treatment fidelity within high quality studies to minimize the potential effect of reporting bias. Thus, only randomized controlled trials (thought to be the gold standard study design) were included. However, no selection criteria were specified for comparison groups as treatment fidelity was only evaluated for a single treatment group within each trial. No between-group comparisons were made. If a trial contained more than one active intervention arm, only one was selected for review using the following criteria: (1) in-person or individually delivered intervention arms were selected over other modes or formats, (2) enhanced or multi-component intervention arms were selected over standard or single-component, and (3)

parent and child intervention arms were selected over parent-only or child-only. If multiple articles for a given study were identified based on the selection criteria in Table 1 (i.e., “Study articles reporting intervention descriptions...”) then all of these articles were used for extraction. An intentionally broad range of interventions was also chosen for this project, including studies targeting participants from any country and participants with chronic or mental health conditions.

Data extraction process and elements

After studies were identified for inclusion, basic information about the study population, study design, and selected intervention arm was extracted. Each study was also coded for study quality as measured by the Delphi checklist (a 9-item checklist with yes/no response options developed to evaluate quality assessment in RCTs).³³ Treatment fidelity was evaluated using a modified version of the NIH Treatment Fidelity Framework.²⁷ For training purposes, a random sample of 10 studies were identified and coded by all three coders. To complete certification of coding, it was intended that the two secondary coders would complete sets of 10 additional studies along with the lead coder. Inter-coder agreement would be calculated after each set of 10 articles was completed until reliability (as measured by prevalence- and bias-adjusted Kappa, PABAK³⁴) was above 0.80 for all items. Due to the large variability in reporting across the studies, the decision was later made to have two coders independently code all studies. Coders met bi-weekly throughout the coding process to refine operational definitions and adjudicate differences. The author of the NIH Treatment Fidelity Framework was contacted when additional clarification on items was needed. Three measures of inter-coder reliability (PABAK, standard Kappa, and percent agreement) were used to calculate reliability of all items in the tools below.

Treatment Fidelity (NIH Treatment Fidelity Framework)—Treatment fidelity was coded using a modified version the National Institutes of Health (NIH) Treatment Fidelity Framework. The framework includes 5 primary domains of fidelity: treatment design, provider training, treatment delivery, treatment receipt, and treatment enactment. Items within each domain were coded as present or absent. If an intervention was multicomponent, a fidelity indicator was marked present if it was described for at least one component. Modifications to the original framework are presented in Table 2. Specifically, 1 item (treatment design information about intended content) was split into 3 items to provide more detail. Because consensus on the discrete components of intervention content does not yet exist, we built on work in other fields to specify 3 components:^{35, 36} behavior change techniques, target behaviors, and therapeutic alliance. Next, the original item related to dose delivered was expanded to mirror the 3 dose intended items, putting increased emphasis on this component. Lastly, a single item was added to evaluate content delivered. This item was not split into the 3 content components as was done above, due to the very limited reporting of these items. Three items were excluded from the framework for this synthesis. The first, regarding the comparison arm, was excluded as only a single intervention group was evaluated. The remaining two were excluded as a reliable operational definition could not be reached. Domain summary scores were calculated for each study by summing the number of items coded as present.

Study Quality (Delphi checklist)—The Delphi checklist³³ was used to measure study quality. This checklist was designed to evaluate the quality assessment of randomized controlled trials (RCTs). It includes 9 items with yes/no response options developed using a Delphi consensus procedure. The final items included: (1) was a method of randomization performed, (2) was treatment allocation concealed, (3) were randomized groups similar at baseline, (4) were eligibility criteria specified, (5) was the outcome assessor blinded, (6) were participants blinded, (7) were treatment providers blinded, (8) were point estimates and measures of variability reported for the primary outcome, and (9) was an intent-to-treat analysis included. A summary score was calculated by summing the number of yes's coded for a single study.

Summary measures and analysis

Descriptive statistics including counts, frequencies, means, and standard deviations were calculated for all items when appropriate. Three measures were used to calculate inter-coder agreement as recommended by the Guidelines for Reporting Reliability and Agreement Studies.³⁷ The PABAK was calculated in addition to traditional measures (Cohen's kappa and percent agreement) as it may be more appropriate when the prevalence of an endorsed item is low.³⁴ A study quality score (Delphi index) and treatment fidelity scores (overall and by domain) were calculated for each study by summing the total number of items present for a given study. Associations between the outcome of overall treatment fidelity (possible range: 0 to 26) and predictors (1) study quality (possible range: 0–9), (2) publication year (in quartiles), and (3) number of articles reviewed per study (1 vs. >1) were evaluated using general linear regression. Statistical significance was considered a p-value less than 0.05. All analyses were conducted using SAS 9.3

RESULTS

Study selection and data extraction process

Figure 1 shows the flow diagram of the process for this review. Overall, 3455 studies were excluded after abstract and title review with high inter-coder agreement (PABAK= 0.85, kappa = 0.77, percent agreement= 0.92) and 814 articles were excluded after full text review with adequate inter-coder agreement (PABAK= 0.78, kappa = 0.75, percent agreement= 0.89).

Descriptive characteristics of included studies

Descriptive information for the 193 included studies is presented in Table 3. Most of the studies were individually randomized trials, conducted in the United States, and published after 2008. Average study quality, as measured by the Delphi checklist, was 4.7 (SD=1.4) out of 9 possible items. The sample for this systematic review was limited to RCTs, thus all of the included studies reported some method of randomization. Due to the type of interventions included in this review, none had providers who were blinded to treatment condition, and very few had participants who were blinded to treatment condition (n=5, 3%). Almost all studies reported point estimates and measures of variability for one or more weight outcomes (n= 178, 92%) and almost all specified eligibility criteria (n = 168, 88%). More variability was seen in the remaining items. The inter-coder agreement for the Delphi

study quality items was high, with a mean (SD) for PABAK, Kappa, and percent agreement of 0.85 (0.12), 0.53 (0.31), and 0.92 (0.06) respectively. Few studies included children in the normal weight range. Most studies evaluated in-person interventions delivered in university or clinic settings. The selected interventions were evenly split between group and individual treatment, or the combination. Most intervened on both parents and children and targeted diet and physical activity combined.

Reporting of Treatment Fidelity in Included Studies

The overall inter-coder agreement for this measure was high with a mean PABAK across all items of 0.83 (SD=0.09). Coders had near-perfect agreement (0.80 to 1.00) for 79% of items, substantial agreement (0.60 to 0.79) for 17% of items, and moderate agreement (0.40 to 0.59) for only 4% of items.³⁸ The mean Kappa was 0.55 (SD=0.23) and the mean percent agreement was 0.92 (SD=0.05) for these same items. Complete results for inter-coder agreement of each item can be found in the Supporting Documents (Table S1). The results from the NIH Treatment Fidelity Framework are presented in Figure 2. There was large variability in reporting of treatment fidelity across the 5 domains; treatment design elements were reported with the highest frequency. Proportion of specific items ranged from 4% (Treatment Enactment: reporting method to assess participant comprehension) to 99% (Treatment Design: reporting target behaviors). Eighty-seven percent of studies (N= 168) reported less than half the items. Individual study results are provided in the Supporting Documents (Table S2). The specific methods used to report selected constructs within each domain are reported in Table 4 and described below.

Treatment Design—The percent of studies reporting some information on the intended dose was high (ranging from 77% to 94%), although the methods of the reporting varied largely as shown in Table 4. Almost all studies reported at least some information about the behavior change techniques used and the targeted weight-related behaviors, but again the variability in the methods with which these were reported was high. Very few used standard definitions of behavior change techniques such as the Behavior Change Technique Taxonomy (BCTTv1).³⁹ Information about the approach to addressing therapeutic alliance was not commonly reported. The most common method was to use a consistent therapist to deliver all sessions for a given participant. Provider credentials were reported more commonly and often included profession, education, or both. Some gave provider characteristics in addition to credentials (e.g., gender, race/ethnicity) which were not captured by this tool. Over half the studies mentioned some theoretical framework or clinical guideline. Common theoretical models were social-ecological model or social-cognitive theory; common types of therapy were motivational interviewing or cognitive-behavioral therapy.

Provider Training—Few studies provided details on the training protocols for their providers. Some authors mentioned that providers were trained or reported the content or length of the training. These instances are not captured by this tool. The reporting of how provider skills were maintained over the course of the intervention was more common. The methods for maintaining skills most often included ongoing supervision or booster trainings.

Treatment Delivery—Overall reporting of treatment delivery was lower compared to the treatment design domain. Information about the dose or content delivered was not often reported, with the exception of number of contact sessions (74%). The most common way to report number of sessions delivered was to report the number of participants with session attendance in categories (e.g., N participants completed the intervention, N participants were high attenders, or N participants completed at least half of the intended sessions). Of those who reported the amount of content delivered during sessions (N = 17), 3 (18%) reported specific behavior change techniques and 4 (24%) reported the weight-related behaviors targeted.

Treatment Receipt and Enactment—Methods to assess participant receipt during treatment sessions were the least likely to be reported relative to the other four domains. Some examples of the methods used to assess participant comprehension of content were in-session quizzes or asking participants to recall treatment messages. Methods to assess participants' use of intervention skills were most common in group physical activity programs, and included the accelerometers or heart rate monitors to determine if children were able to maintain a certain intensity of activity. Measurement of participant enactment of skills outside of treatment sessions was more commonly assessed than receipt of these skills. The most common example of this was the provider review of self-monitoring logs.

Treatment Fidelity and Study Quality, Number of Included Articles, and Publication Year

The summary scores for the NIH Treatment Fidelity Framework are provided in Table 5. There was a statistically significant positive association between study quality as measured by the Delphi checklist and treatment fidelity ($\beta = 0.61$, $p < 0.01$). There was also a statistically significant association between the number of articles included for a given study and treatment fidelity, with higher fidelity reporting by studies with more than one included article ($\beta = 1.8$, $p < 0.01$). There was no relationship between publication year and treatment fidelity ($\beta = 0.20$, $p = 0.26$).

DISCUSSION

The primary aim of this systematic review was to understand how behavioral pediatric obesity interventions report treatment fidelity within randomized controlled trials. Historically, the “treatment fidelity” framework has been used in clinical fields (e.g., psychotherapeutic treatment studies). With this review, we have shown that treatment fidelity is highly relevant, but not fully reported, in behavioral intervention studies designed to prevent and manage obesity. Across this field, researchers were largely successful in reporting treatment design, including identification of theoretical frameworks guiding the intervention development. This is consistent with findings from those who have reviewed literature using the NIH Fidelity Framework in other fields.^{27, 29, 30} Within treatment design, reporting of elements related to therapeutic alliance was relatively low. This item was added by the current research team under the treatment design domain (not within non-specific factors under treatment delivery), as studies in clinical psychotherapy settings have consistently found it to be an essential component of participant engagement and treatment outcomes.³⁶ While the term “therapeutic alliance” (i.e., the quality of the therapist-client

relationship) has its origins in clinical psychotherapy settings, the concept is broadly applicable to any person-to-person intervention in which there is a relationship between the participant and interventionist (e.g., cognitive, behavioral, psychoeducational). Therapeutic alliance has been widely studied in both intervention and prevention trials (e.g., family-based substance abuse prevention, family-based interventions for at risk youth, relapse prevention) and has been consistently shown to be a major variable in explaining both drop-out rates and treatment efficacy.^{40, 41} Thus, it should be adequately addressed during the design phase. Our findings also indicate that some aspects of treatment delivery, including number of sessions and participant satisfaction with the intervention are reported at a high frequency. Other aspects of treatment delivery, such as length of session and content delivered, are reported very infrequently.

Findings from this review highlight additional areas for improvement with respect to treatment fidelity reporting. Components in the domains of provider training, treatment receipt, and treatment enactment were infrequently reported. This is in contrast to reviews utilizing the NIH Fidelity Framework in other areas that found higher reporting of enactment.^{42, 43} It should be noted that one enactment item (“Reporting use of a strategy to improve subject performance of intervention skills”) was excluded from this review, as adequate reliability could not be obtained. This is likely due to the low prevalence and high variability in methods of this item across this field. The importance of provider training was highlighted in recent research of Brose and colleagues (2015) that found availability and use of a training manual was associated with better outcomes for smoking cessation interventions.⁴⁴ Further, using more robust methods to evaluate participant receipt and enactment can highlight potential breakdowns in the pathway from treatment design to participant outcomes. As shown in Table 4, there is room for improvement in the quality of methods selected to measure fidelity within each domain. Within the field, emphasis has been placed on moving toward objective, valid, and reliable methods to measure study *outcomes* (e.g., physical activity). This same emphasis should be placed on using rigorous methodology to measure treatment fidelity. For example, only one of the 44 studies reporting a method to assess treatment enactment used an objective measure (specifically, observation of home environment or behaviors).

In examining specific studies for completeness, only a single study (of 193 evaluated) included at least one item from each of the 5 domains, and this study (the HIKCUPS study) was tied for the highest proportion of reported items at 71%. This is notable in that Borrelli and colleagues (2005) highlight the mutually exclusive nature of the five domains.²⁷ Inattention to one domain could threaten the internal validity of a study. It is important to mention that reporting on treatment fidelity may be diminished due to space limitations in journals as opposed to what was actually implemented. Again, this may reflect prioritizing outcome evaluation over treatment fidelity evaluation. It is necessary to report information about each of these domains to fully understand the quality of the intervention and nuances of implementation. In addition, this would allow standardized comparisons between interventions so that the specific components that produce behavior changes may be identified. To address space limitations, authors should consider publishing separate articles addressing treatment fidelity. We found that the number of published articles included for a given study was associated with reporting of treatment fidelity. The aforementioned

HIKCUPS study had 9 published articles included in this review (the most of any included study). Two of these articles focused heavily on treatment fidelity, one on study design⁴⁵ and one on process evaluation.⁴⁶ This multi-paper approach allowed the authors to dedicate significant space to the description of treatment fidelity components and may have contributed to their success in addressing all five domains.

The current review also sought to examine associations of treatment fidelity and study quality. It could be posited that those designing more rigorous studies may also design more rigorous treatment fidelity methods, thus having better reporting. In this review, there was a positive and significant association between study quality and treatment fidelity. Still, within the highest quality studies (6 or more items on the Delphi checklist, N = 49 studies) an average of only 37% (SD=11%) of treatment fidelity items were reported. Interestingly, this study found no association between treatment fidelity reporting and year of publication, contrary to the findings of interventions to address secondhand smoke.⁴³ This is further evidence that efforts are needed to help researchers improve their published descriptions of treatment fidelity. As a first step, detailed recommendations for improving treatment fidelity reporting are presented in Table 6. Additionally, we urge researchers within the field to continue refining the available treatment fidelity reporting tools and to develop and publish rigorous methods for reporting specific items within the treatment fidelity domains.

Funding agencies and professional organizations can also help obesity researchers to become leaders in treatment fidelity. Specifically, funders could require applicants to describe methods to assess treatment fidelity within grant applications. As mentioned previously, treatment design elements were the most commonly reported fidelity domains in this review, which is likely due to the inclusion of these items in the research plan methodology of grant applications. Requiring investigators to include all components of treatment fidelity may motivate researchers to more intentionally measure the multiple domains of treatment fidelity. However, it is important to note the additional costs associated with this level of methodological rigor and will have budget implications which should be anticipated by funding agencies. Professional organizations and associated journals can also aid in moving forward the science by requiring the use of existing checklists (e.g., WIDER, TiDIER, or NIH Fidelity Framework checklists) by authors during submission and during the peer review process. To further encourage work in this area, obesity researchers should propose symposia focusing on treatment fidelity at national obesity meetings and those who sit on editorial boards should bring these issues forward. Finally, journals could offer submission categories for treatment fidelity articles, analogous to the “Study Design, Statistical Design, Study Protocols” option offered by Contemporary Clinical Trials.

This systematic review has a number of strengths that bolster the findings. Inter-coder reliability was high and comparable to other systematic reviews using this measure.^{27, 42, 43, 47} Further, a thorough approach to coding was taken, such that two reviewers independently coded all studies and adjudicated differences. This project undertook an expansive review with broad inclusion criteria spanning a twenty-four year period to fully capture the state of the field. Third, strong methodology was employed, i.e., following PRISMA guidelines, following reliability reporting recommendations to include multiple measures of reliability at each step and for each item (Supporting Documents, Table

S1), registering the review on PROSPERO along with publication of the search strategy, and including the entire data table in the Supporting Documents (Table S2). Finally, this work went beyond looking simply at the presence of treatment fidelity items to characterizing the specific methods used by research teams, which has been a noted limitation of previous reviews.

There are limitations that should be noted. Incomplete descriptions of intervention components within a treatment arm made evaluating fidelity by treatment component (e.g., home visits vs. group classes) impossible. To address this in the future, authors should be clear in defining the designed setting, participants, and mode. Similarly, multiple interventions are often compared in randomized controlled trials; however, this project focused exclusively on fidelity reporting in one intervention arm, selected by the reviewers. Authors should be urged to report detailed treatment fidelity for each component of all intervention groups whether or not they are the primary intervention of interest. This would allow future systematic reviews to compare fidelity between study arms. Lastly, the current review did not evaluate the level of fidelity within a study (e.g., adequacy of the reported dose or content delivered). Future systematic reviews can use treatment fidelity to answer important questions, such as determining the optimal treatment dose for behavioral pediatric obesity interventions. Researchers have begun to examine associations between intervention content and outcomes.^{48–51} However, this work is hampered by low levels of fidelity reporting across the field. By giving more attention to the reporting of all components of treatment fidelity, from design to enactment, it is likely that these components will be better implemented and the effectiveness of pediatric obesity interventions will improve. Obesity researchers have the opportunity to be in the vanguard of behavioral research and can lead the discipline forward by improving standards for reporting within our field, ultimately leading to more effective interventions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Special thanks goes to additional coders, Sarah Toov, Karen Omlung, and Erin Schwartz, and to Eli Poe for help in formatting references. This research was supported by the National Heart, Lung, and Blood Institute, the Eunice Kennedy Shriver National Institute of Child Health and Development, and the NIH Office of Behavioral and Social Sciences Research (Grant Numbers U01HL103561, U01HL103620, U01HL103622, U01HL103629, and U01HD068890).

References

1. Bethell C, Simpson L, Stumbo S, Carle AC, Gombojav N. National, state and local disparities in childhood obesity. *Health Aff (Millwood)*. 2010; 29:347–56. [PubMed: 20194972]
2. Wilson DK. New perspectives on health disparities and obesity interventions in youth. *J Pediatr Psychol*. 2009; 34:231–44. [PubMed: 19223277]
3. Ogden C, Lamb M, Carroll M, Flegal K. Obesity and socioeconomic status in children and adolescents: United States, 2005–2008. *NCHS Data Brief*. 2010
4. Ogden C, Carroll M, Curtin L, Lamb M, Flegal K. Prevalence of high body mass index in US children and adolescents, 2007–2008. *J Am Med Assoc*. 2010; 303:8.

5. Wang YC, Orleans CT, Gortmaker SL. Reaching the healthy people goals for reducing childhood obesity: Closing the energy gap. *Am J Prev Med.* 2012; 42:437–44. [PubMed: 22516482]
6. Orsi C, Hale D, Lynch J. Pediatric obesity epidemiology. *Curr Opin Endocrinol Diabetes Obes.* 2011; 18:14–22. [PubMed: 21157323]
7. Popkin BM. Understanding global nutrition dynamics as a step towards controlling cancer incidence. *Nat Rev Cancer.* 2007; 7:61–67. [PubMed: 17186019]
8. Whitaker RC, Wright JA, Pepe MS, Seidel KD, Dietz WH. Predicting obesity in young adulthood from childhood and parental obesity. *N Engl J Med.* 1997; 337:869–73. [PubMed: 9302300]
9. Daniels SR. The consequences of childhood overweight and obesity. *Future Child.* 2006; 16:47–67. [PubMed: 16532658]
10. Merten MJ. Weight status continuity and change from adolescence to young adulthood: Examining disease and health risk conditions. *Obesity (Silver Spring).* 2010; 18:1423–8. [PubMed: 19851300]
11. Gordon-Larsen P, The NS, Adair LS. Longitudinal trends in obesity in the United States from adolescence to the third decade of life. *Obesity.* 2009
12. Pi-Sunyer FX. The obesity epidemic: Pathophysiology and consequences of obesity. *Obes Res.* 2002; 10(Suppl 2):97S–104S. [PubMed: 12490658]
13. Stovitz S, Hannan P, Lytle L, Demerath E, Pereira M, Himes J. Child height and the risk of young-adult obesity. *Am J Prev Med.* 2010; 38:74–77. [PubMed: 20117560]
14. Finkelstein EA, Trogdon JG, Cohen JW, Dietz W. Annual medical spending attributable to obesity: payer-and service-specific estimates. *Health Aff (Millwood).* 2009; 28:w822–31. [PubMed: 19635784]
15. Boon CS, Clydesdale FM. A review of childhood and adolescent obesity interventions. *Crit Rev Food Sci Nutr.* 2005; 45:511–25. [PubMed: 16371325]
16. Berge JM. A review of familial correlates of child and adolescent obesity: what has the 21st century taught us so far? *Int J Adolesc Med Health.* 2009; 21:457–83. [PubMed: 20306760]
17. Young KM, Northern JJ, Lister KM, Drummond JA, O'Brien WH. A meta-analysis of family-behavioral weight-loss treatments for children. *Clin Psychol Rev.* 2007; 27:240–49. [PubMed: 17070638]
18. Seo DC, Sa J. A meta-analysis of obesity interventions among U.S. minority children. *J Adolesc Health.* 2010; 46:309–23. [PubMed: 20307819]
19. Stice E, Shaw H, Marti CN. A meta-analytic review of obesity prevention programs for children and adolescents: The skinny on interventions that work. *Psychol Bull.* 2006; 132:667–91. [PubMed: 16910747]
20. Bellg AJ, Borrelli B, Resnick B, et al. Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychol.* 2004; 23:443. [PubMed: 15367063]
21. Hoffmann TC, Glasziou PP, Boutron I, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ.* 2014; 348:g1687. [PubMed: 24609605]
22. Albrecht L, Archibald M, Arseneau D, Scott SD. Development of a checklist to assess the quality of reporting of knowledge translation interventions using the Workgroup for Intervention Development and Evaluation Research (WIDER) recommendations. *Implement Sci.* 2013; 8:52. [PubMed: 23680355]
23. Dane AV, Schneider BH. Program integrity in primary and early secondary prevention: are implementation effects out of control? *Clin Psychol Rev.* 1998; 18:23–45. [PubMed: 9455622]
24. Gresham FM, Gansle KA, Noell GH. Treatment integrity in applied behavior analysis with children. *J Appl Behav Anal.* 1993; 26:257–63. [PubMed: 8331022]
25. Moncher FJ, Prinz RJ. Treatment fidelity in outcome studies. *Clin Psychol Rev.* 1991; 11:247–66.
26. Odom SL, Brown WH, Frey T, Karasu N, Smith-Canter LL, Strain PS. Evidence-based practices for young children with autism contributions for single-subject design research. *Focus Autism Other Dev Disabl.* 2003; 18:166–75.

27. Borrelli B, Sepinwall D, Ernst D, et al. A new tool to assess treatment fidelity and evaluation of treatment fidelity across 10 years of health behavior research. *J Consult Clin Psychol*. 2005; 73:852. [PubMed: 16287385]
28. Slaughter SE, Hill JN, Snelgrove-Clarke E. What is the extent and quality of documentation and reporting of fidelity to implementation strategies: a scoping review. *Implementation Science*. 2015; 10:129. [PubMed: 26345357]
29. Borrelli B, Tooley EM, Scott-Sheldon LA. Motivational interviewing for parent-child health interventions: a systematic review and meta-analysis. *Pediatr Dent*. 2015; 37:254–65. [PubMed: 26063554]
30. Schober I, Sharpe H, Schmidt U. The reporting of fidelity measures in primary prevention programmes for eating disorders in schools. *European Eating Disorders Review*. 2013; 21:374–81. [PubMed: 23881537]
31. Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *International journal of surgery*. 2010; 8:336–41. [PubMed: 20171303]
32. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics*. 2009; 42:377–81. [PubMed: 18929686]
33. Verhagen AP, de Vet HC, de Bie RA, et al. The Delphi list: a criteria list for quality assessment of randomized clinical trials for conducting systematic reviews developed by Delphi consensus. *J Clin Epidemiol*. 1998; 51:1235–41. [PubMed: 10086815]
34. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol*. 1993; 46:423–29. [PubMed: 8501467]
35. Michie, S.; Atkins, L.; West, R. *The behaviour change wheel: a guide to designing interventions*. London: Silverback; 2014.
36. Martin DJ, Garske JP, Davis MK. Relation of the therapeutic alliance with outcome and other variables: a meta-analytic review. *J Consult Clin Psychol*. 2000; 68:438. [PubMed: 10883561]
37. Kottner J, Audigé L, Brorson S, et al. Guidelines for reporting reliability and agreement studies (GRRAS) were proposed. *Int J Nurs Stud*. 2011; 48:661–71. [PubMed: 21514934]
38. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977:159–74. [PubMed: 843571]
39. Michie S, Richardson M, Johnston M, et al. The behavior change technique taxonomy (v1) of 93 hierarchically clustered techniques: Building an international consensus for the reporting of behavior change interventions. *Ann Behav Med*. 2013; 46:81–95. [PubMed: 23512568]
40. Faw L, Hogue A, Johnson S, Diamond GM, Liddle HA. The Adolescent Therapeutic Alliance Scale (ATAS): Initial psychometrics and prediction of outcome in family-based substance abuse prevention counseling. *Psychotherapy Research*. 2005; 15:141–54.
41. Elvins R, Green J. The conceptualization and measurement of therapeutic alliance: An empirical review. *Clin Psychol Rev*. 2008; 28:1167–87. [PubMed: 18538907]
42. Reiser RP, Milne DL. A systematic review and reformulation of outcome evaluation in clinical supervision: Applying the fidelity framework. *Training and Education in Professional Psychology*. 2014; 8:149.
43. Johnson-Kozlow M, Hovell MF, Rovniak LS, Sirikulvadhana L, Wahlgren DR, Zakarian JM. Fidelity issues in secondhand smoking interventions for children. *Nicotine & tobacco research*. 2008; 10:1677–90. [PubMed: 19023822]
44. Brose LS, McEwen A, Michie S, West R, Chew XY, Lorencatto F. Treatment manuals, training and successful provision of stop smoking behavioural support. *Behav Res Ther*. 2015; 71:34–39. [PubMed: 26057438]
45. Jones RA, Okely AD, Collins CE, et al. The HIKCUPS trial: a multi-site randomized controlled trial of a combined physical activity skill-development and dietary modification program in overweight and obese children. *BMC Public Health*. 2007; 7:15. [PubMed: 17263896]

46. Jones RA, Warren JM, Okely AD, et al. Process evaluation of the Hunter Illawarra Kids Challenge Using Parent Support study: a multisite randomized controlled trial for the management of child obesity. *Health Promot Pract.* 2010; 11:917–27. [PubMed: 19158237]
47. Corley NA, Kim I. An assessment of intervention fidelity in published social work intervention research studies. *Research on Social Work Practice.* 2016; 26:53–60.
48. Hendrie GA, Brindal E, Corsini N, Gardner C, Baird D, Golley RK. Combined home and school obesity prevention interventions for children: What behavior change strategies and intervention characteristics are associated with effectiveness? *Health Educ Behav.* 2012; 39:159–71. [PubMed: 21984691]
49. Golley R, Hendrie G, Slater A, Corsini N. Interventions that involve parents to improve children's weight-related nutrition intake and activity patterns—what nutrition and activity targets and behaviour change techniques are associated with intervention effectiveness? *Obes Rev.* 2011; 12:114–30. [PubMed: 20406416]
50. Nixon C, Moore H, Douthwaite W, et al. Identifying effective behavioural models and behaviour change strategies underpinning preschool-and school-based obesity prevention interventions aimed at 4–6-year-olds: a systematic review. *Obes Rev.* 2012; 13:106–17.
51. Martin J, Chater A, Lorencatto F. Effective behaviour change techniques in the prevention and management of childhood obesity. *Int J Obes.* 2013; 37:1287–94.

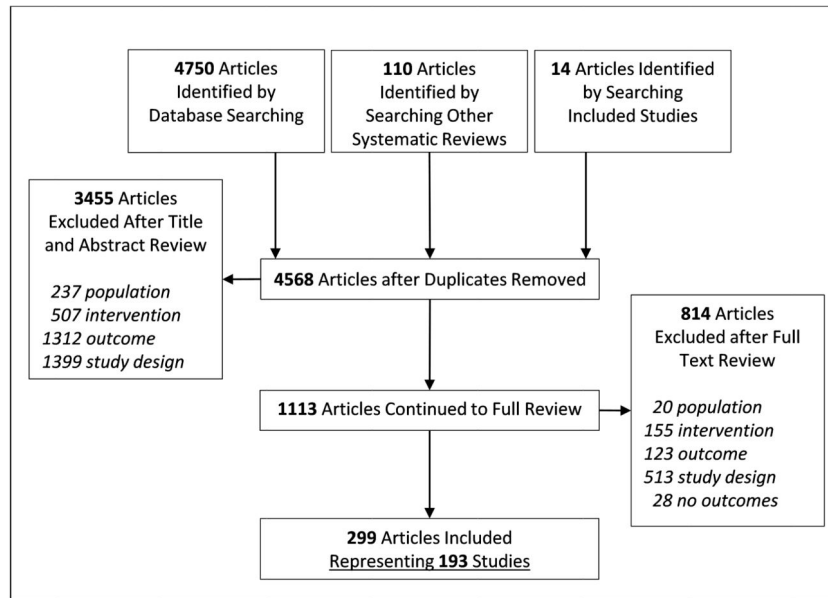


Figure 1.
Flow of articles through the search and selection process.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

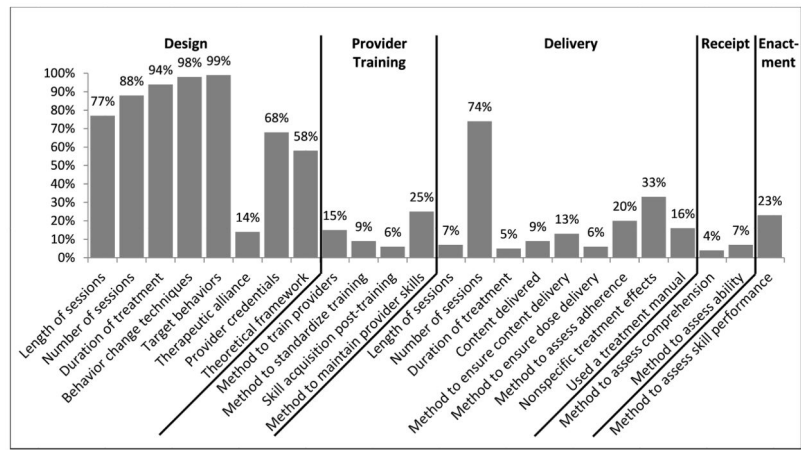


Figure 2. Reporting of treatment fidelity, N = 193 studies.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

Study selection criteria following the PICOS guidelines.

	Inclusion Criteria	Exclusion Criteria
<i>Population</i>	<ul style="list-style-type: none"> • Studies targeting children ages 2 to 18 	<ul style="list-style-type: none"> • Studies targeting pregnant women, infants, adults • Studies targeting children who are underweight • Studies targeting participants who are hospitalized, in residential overnight camps, or are in assisted living
<i>Intervention</i>	<ul style="list-style-type: none"> • Behavior change intervention aiming to impact child, parent, or family weight-related behaviors • Interventions delivered in any modality (e.g., in person, phone, or web), format (e.g., individual/family, or group), or setting (e.g., school, clinic, or community) 	<ul style="list-style-type: none"> • Pharmacologic or surgical interventions • Breastfeeding interventions • Prescribed diet or exercise or environmental interventions with no behavioral component • Interventions delivered entirely during the school day or community-wide interventions where fidelity cannot be feasibly measured at the individual or family level
<i>Outcome</i>	<ul style="list-style-type: none"> • Objective measure of child weight outcome (e.g., BMIz or percent overweight) 	
<i>Study Design</i>	<ul style="list-style-type: none"> • Individual or group randomized controlled trials published as original articles • Study articles published after 1990 • Study articles published in English • Study articles reporting intervention descriptions or fidelity data for the primary intervention (e.g., design papers, process evaluations, cost-effectiveness analyses) 	<ul style="list-style-type: none"> • Observational studies (e.g., developing but not testing interventions, developing theoretical frameworks, validating outcome measures, correlates of obesity, case studies, cohort studies, or case-control studies) • Quasi-experimental studies (e.g., pre/post studies, or non-randomized control group) • Cross-sectional (as opposed to cohort) group randomized trials • Secondary syntheses of RCTs not likely to report on treatment fidelity (e.g., study recruitment or study measure development) • Secondary syntheses of existing literature (e.g., systematic reviews, commentaries, or book chapters) • Unpublished works (e.g., dissertations or abstracts)

PICOS- participants, interventions, comparisons, outcomes, and study design.

Table 2

Specific items coded using the NIH Treatment Fidelity Framework.

Borelli 2005 Item	Adapted Operational Definition
<i>Provided information about treatment dose intended</i>	
1. Length of contact sessions	Must report entire session length, not only the length of an activity within a session.
2. Number of contact sessions	Must report for at least one component.
3. Duration of treatment over time	Must report treatment duration specifically, not evaluation period.
<i>Provided information about treatment content intended</i>	
Treatment Design	The next 3 items were added for this review.
4. Behavior change techniques	Must report at least one strategy to induce behavior change (e.g., goal setting).
5. Target behaviors	Must report targeting at least one weight-related behavior (e.g., physical activity).
6. Therapeutic alliance	Must report a design element addressing the participant/provider relationship.
X Provided information about the comparison treatment	This item was not coded for this review.
7. Mention of provider credentials	Includes credentials for any provider, as originally designed or otherwise.
8. Mention a theoretical model or clinical guidelines	May be referenced anywhere in the manuscript.
Provider Training	
1. Description of how providers were trained	Must report the method of training (e.g., role play); training content not included.
2. Standardized provider training	Must report how training was similar for all providers; training length not included.
3. Measured provider skill acquisition post-training	Must be reported for at least one component and can include accreditation.
4. Described how provider skills maintained over time	Must be reported for at least one component (e.g., supervision or ongoing training).
<i>Provided information about dose delivered</i>	
The next 3 items were added for this review.	
1. Length of contact sessions	Must report session length either delivered or received (e.g., average session length).
2. Number of contact sessions	Must report the number of participants completing some portion of the treatment.
3. Duration of treatment	Must report the average duration actually delivered or received across participants.
4. Provided information about content delivered	This item was added for this review. Must report actual content delivered during treatment sessions.
5. Method to ensure that content was delivered as specified	Must report method used during sessions with intent of improving content delivery.
6. Method to ensure that dose was delivered as specified	Must report method used during sessions with the intent of improving dose delivery.
7. Method to assess if the provider adhered to the content	Must report method used to evaluate and report actual content delivered.
8. Assessed nonspecific treatment effects	May include participant satisfaction; therapeutic alliance methods not included.
9. Used a treatment manual	Must state "treatment manual" or report written instructions given to providers.
Treatment Receipt	
1. Assessed subject comprehension of the intervention during the intervention period	Must be a method used by the treatment team to assess skills learned during the intervention period; methods used by the evaluation team are not included.

Borelli 2005 Item	Adapted Operational Definition
<p>X Strategy to improve subject comprehension of the intervention</p> <p>2. Assessed subject ability to use the intervention skills during the intervention period</p> <p>X Strategy to improve subject performance of intervention skills</p>	<p>This item was not coded for this review.</p> <p>Must be assessment of skills during the treatment session (e.g. evaluation role play) by the treatment team; methods used by the evaluation team are not included.</p> <p>This item was not coded for this review.</p>
<p>1. Assessed subject performance of the intervention skills in settings which the intervention might be applied</p> <p>X Strategy to improve subject performance of intervention skills in settings which the intervention might be applied</p>	<p>Must mention assessment of learned skills performed outside treatment sessions by the treatment team; methods used by the evaluation team are not included.</p> <p>This item was not coded for this review.</p>

Treatment Enactment

Table 3Summary characteristics of included studies, N = 193.^A

		N or Mean	% or SD	
	Studies with more than one article included	64	33%	
	Study length (months)	13.0	10.4	
	Study participants (N)	153	206	
<i>Study Design</i>	GRT	33	17%	
	RCT	160	83%	
	United States	106	55%	
<i>Study Country</i>	Europe, Australia, or Canada	64	34%	
	Other	23	11%	
	1990–2007	52	27%	
<i>Study Year</i>	2008–2010	55	29%	
	2011–2012	51	26%	
	2013–2014	35	18%	
	Method of randomization used	193	100%	
<i>Study Quality</i>	Treatment allocation concealed	57	30%	
	Groups similar at baseline	123	64%	
	Eligibility criteria specified	169	88%	
	Outcome assessor blinded	50	26%	
	Care provider blinded	0	0%	
	Participant blinded	5	3%	
	Point estimates and variability presented	178	92%	
	Intention-to-treat analysis included	122	63%	
	<i>Participant Age</i>	Ages 2–11	79	41%
		Ages 12–18	84	44%
Ages 2–11 + 12–18		29	15%	
<i>Participant Weight Status</i>	Overweight/obese	144	75%	
	Normal + overweight/obese	49	25%	
<i>Intervention Mode</i>	In-Person	130	67%	
	Phone/web/email/text	12	6%	
	In-person + phone/web/email/text	50	26%	
	Home	15	8%	
<i>Intervention Setting</i>	School/community	36	19%	
	Clinic/university	65	34%	
	Home + school/community/clinic/university	51	26%	
	School/community + clinic/university	7	4%	
<i>Intervention Format</i>	Individual	64	33%	
	Group	70	36%	
	Individual + group	57	30%	
<i>Intervention Participants</i>	Parent	20	10%	
	Child	20	10%	

	N or Mean	% or SD
Parent + child	153	80%
PA/TV	14	7%
<i>Intervention Target Behavior</i> Diet	12	6%
PA/TV + diet	166	86%

^ANot all categories add up to 193 as some studies did not report certain intervention information.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4

Methods used to measure treatment fidelity, N = 193 studies.

	N	%
Length of sessions, N = 149		
Approximate length (e.g., about 1 hour)	9	6%
Minimum or maximum length	6	4%
Range of lengths (e.g., 15–30 minutes)	22	15%
Exact length	98	66%
Length using multiple methods	13	9%
Behavior change techniques, N = 190		
Referring to type of therapy (e.g., motivational interviewing)	9	5%
Used standardized definitions (e.g., BCTTv1)	6	3%
Mapped to specific components of a theoretical model	2	1%
Method not specified	173	91%
Treatment Design		
Therapeutic Alliance, N = 27		
The same provider delivered all sessions (i.e., continuity of care)	20	74%
Used rapport-building techniques	6	22%
Used both of the above methods	1	4%
Provider credentials, N = 131		
Education level (e.g., Master's level counselors)	14	11%
Previous profession (e.g., dietician)	77	59%
Reported both of the above methods	40	31%
Theoretical framework, N = 112		
Clinical guidelines	10	9%
Counseling approach or type of therapy (e.g., motivational interviewing)	28	25%
Theoretical model (e.g., Social Ecological Model)	56	50%
Reported more than one of the above methods	18	16%
Standardize provider training, N = 17		
Standardized training materials	8	47%
Providers attend same training	3	18%
Single trainer lead all trainings	1	6%
Accredited or recognized training programs	3	18%
Method not specified	2	12%
Provider Training		
Provider skill acquisition post-training, N = 11		
Paper/pencil skills test	3	27%
Coding of observed certification sessions	3	27%
Method not specified	5	45%
Method to maintain provider skills, N = 48		
Ongoing supervision only	30	63%
Periodic booster training and ongoing supervision	7	15%
Reviewing recorded or observed session and ongoing supervision	10	21%

	N	%
Formal evaluations of adherence to protocol triggering retraining	1	2%
Length of sessions, N = 14		
Number of session delivered in intended length	1	7%
Approximate length of sessions delivered	2	14%
Range of session lengths delivered	2	14%
Average session length	9	64%
Number of sessions, n = 142		
Average number of sessions attended per participant	37	26%
Session attendance by category (e.g., low, medium, or high)	74	52%
Percent of sessions attended (e.g., 80% of sessions attended)	31	22%
Content delivered, N = 17		
Average amount of content delivered per participant	2	12%
Average amount of content delivered per session	3	18%
Number of sessions delivered above a certain threshold	2	12%
Percent of content delivered overall	10	59%
Method to ensure content delivered, N = 25		
Protocols or scripts during intervention sessions	5	20%
Provider checklists during intervention sessions	8	32%
Computer-generated algorithm to determine content delivered	12	48%
Method to assess provider adherence to content, N = 39		
Interventionist-reported	9	23%
Independently coded via observation or session recording	28	72%
Method not specified	2	5%
Non-specific treatment effects, N = 63		
Evaluated a measure of participant satisfaction	61	97%
Evaluated therapist characteristics	1	2%
Evaluated non-intended behavior change techniques	1	2%
Method to assess participant comprehension, N = 7		
Participant quizzes	5	71%
Participant recall of learned concepts	1	14%
Provider evaluation of comprehension	1	14%
Method to assess ability to use intervention skills, N = 13		
Objective measures (e.g., heart rate monitor or accelerometer)	8	62%
Participant self-report	1	8%
Provider direct observation	4	31%
Method to assess performance of intervention skills, N = 44		
Participant report of behavior change techniques used between sessions	5	11%
Provider evaluation of self-monitoring logs	31	70%
Provider evaluation of homework assignments	7	16%
Provider observation of home environment or behaviors	1	2%

Table 5

Summary scores for the NIH Treatment Fidelity Framework, N= 193 studies.

	M	SD	Min	Max
NIH Treatment Fidelity Framework Summary Score (24 items)	8.67	2.63	1	17
Treatment Design Score (8 items)	5.97	1.25	1	8
Provider Training Score (4 items)	0.54	0.89	0	4
Treatment Delivery Score (9 items)	1.82	1.48	0	8
Treatment Receipt Score(2 items)	0.10	0.32	0	2
Treatment Enactment Score (1 item)	0.23	0.42	0	1

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 6

Specific recommendations for future reporting of treatment fidelity in behavioral interventions to manage or prevent pediatric obesity.

Recommendations	
<ul style="list-style-type: none"> • Clearly report the intended mode (e.g., in-person or phone), setting (e.g., school or home), format (e.g., group or individual), and target participant (e.g., parent or child) for each component of the treatment. • Report intended dose in terms of length (e.g., minutes), number, duration (e.g., months), pace (e.g., constant or tapered), and density (e.g., weekly). • Clearly distinguish between treatment duration and study duration, as well as which sessions are intended to deliver treatment and which are intended to collect data for study evaluation. • Use consistent definitions of behavior change techniques (e.g., the Behavior Change Technique Taxonomy, BCTTv1³⁸). • Use theoretical frameworks to guide choice of behavior change techniques (e.g., Intervention Mapping⁵⁰). • Report detailed information about the behaviors targeted within diet and physical activity (e.g., fruit and vegetable intake or portion control). A consistent taxonomy of target behaviors for pediatric obesity interventions should be developed. • Clearly distinguish between intended components (e.g., dose, content, and provider credentials) and what was actually delivered. • Report relevant provider characteristics (e.g., years of experience, gender, and race/ethnicity) in addition to academic or professional credentials. • Report information on how specific constructs from theoretical frameworks were chosen (e.g., APEASE criteria⁵¹). 	
Treatment Design	
<ul style="list-style-type: none"> • Provider training should be tailored based on baseline provider skills assessments or and this information should be reported.⁵² • Provider training should be reported as both what was intended and what was ultimately delivered. • Detailed information about provider training should also include information about the content and dose of the training. • Utilize and report provider certification prior to delivering treatment, with a gold standard method of directly observing and coding standardized certification sessions. • Identify appropriate evidence-based methods to maintain provider skills over time (e.g., clinical supervision).⁴⁰ 	
Provider Training	
<ul style="list-style-type: none"> • Report detailed information on the mode, setting, and format of sessions actually delivered if different from originally designed. • Report detailed information on the dose (e.g., length, number, duration, pace, and density of sessions delivered) and content (e.g., behavior change techniques and target behaviors) actually delivered across all participants randomized to the treatment. • Use direct observation or recordings to code and report content delivered to avoid self-presentation bias that may occur if relying on provider report.⁵³ Conversely, if the purpose of having a provider report treatment delivery is to improve adherence, the method should be completed by the provider during treatment sessions. • Non-specific factors are not systematically measured and the definitions are not clear in the literature.⁵⁴ However, best practice would be to have an independent observer code for the delivery of non-intended behavior change techniques and to have participants report satisfaction. • Publish treatment manuals as supporting documents. 	
Treatment Delivery	
<ul style="list-style-type: none"> • Distinguish between dose/content delivered and dose/content received, which may vary depending on the mode of delivery (e.g., e-mail vs. in-person) or the treatment format (e.g., individual vs. group). 	
Treatment Receipt	

Recommendations

Analogous to treatment delivery, (1) utilize provider methods (e.g., in-session skills assessments) to ensure treatment receipt during intervention sessions and (2) independently evaluate participant receipt for reporting purposes. Validated methods and tools should be developed for use within the field.

Utilize methods to report participant enactment of learned skills during the treatment period (but outside of treatment sessions) and participant enactment after the treatment period has concluded. The latter may be thought of as participant retention of skills.

As with treatment delivery and receipt, use provider-completed methods (e.g., provider observation of participant behaviors *in vivo*) to ensure enactment and independent evaluation of participant enactment for reporting purposes.

Treatment Enactment