External Validity Reporting in Behavioral Treatment of Childhood Obesity

Lisa M. Klesges  
*The University of Memphis*, lisa.klesges@memphis.edu

Natalie A. Williams  
*University of Nebraska-Lincoln*, nwilliams17@unl.edu

Kara S. Davis  
*Medical University of South Carolina*

Joanna Buscemi  
*Medical University of South Carolina*

Katherine M. Kitzmann  
*Medical University of South Carolina*

Follow this and additional works at: [http://digitalcommons.unl.edu/famconfacpub](http://digitalcommons.unl.edu/famconfacpub)

Part of the Developmental Psychology Commons, Family, Life Course, and Society Commons, Other Psychology Commons, and the Other Sociology Commons

Klesges, Lisa M.; Williams, Natalie A.; Davis, Kara S.; Buscemi, Joanna; and Kitzmann, Katherine M., "External Validity Reporting in Behavioral Treatment of Childhood Obesity" (2012). Faculty Publications, Department of Child, Youth, and Family Studies. 148.  
[http://digitalcommons.unl.edu/famconfacpub/148](http://digitalcommons.unl.edu/famconfacpub/148)

This Article is brought to you for free and open access by the Child, Youth, and Family Studies, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications, Department of Child, Youth, and Family Studies by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.
External Validity Reporting in Behavioral Treatment of Childhood Obesity

Lisa M. Klesges, PhD, Natalie A. Williams, PhD, Kara S. Davis, MD, Joanna Buscemi, PhD, and Katherine M. Kitzmann, PhD
School of Public Health (Klesges, Williams), the Department of Psychology (Buscemi, Kitzmann), Memphis, Tennessee; Medical University of South Carolina (Davis, Buscemi), Charleston, South Carolina

Abstract

Context—To aid translation of childhood obesity research interventions evidence into practice, research studies must report results in a way that better supports pragmatic decision making. The current review evaluated the extent to which information on key external validity dimensions, participants, settings, interventions, outcomes, and maintenance of effects, was included in research studies on behavioral treatments for childhood obesity.

Evidence acquisition—Peer-reviewed studies of behavioral childhood obesity treatments published between 1980 and 2008 were identified from: (1) electronic searches of social science and medical databases, (2) research reviews of childhood obesity interventions, and (3) reference lists cited in these reviews. Included studies: reported on a controlled obesity intervention trial, targeted overweight or obese children aged 2–18 years, included a primary or secondary anthropometric outcome, and targeted change in dietary intake or physical activity behaviors.

Evidence synthesis—1071 publications were identified and 77 met selection criteria. Studies were coded on established review criteria for external validity elements. All studies lacked full reporting of generalizability elements. Across criteria, the average reporting was 23.9% (range: 0% to 100%). Infrequently reported were setting-level selection criteria and representativeness, characteristics regarding intervention staff, implementation of the intervention content, costs, and program sustainability.

Conclusions—Enhanced reporting of relevant and pragmatic information in behavioral investigations of childhood obesity interventions is needed to improve the ability to evaluate the applicability of results to practice implementation. Such evidence would improve translation of research to practice, provide additional explanation for variability in intervention outcomes, and provide insights into successful adaptations of interventions to local conditions.

Address correspondence to: Lisa M. Klesges, PhD, Dean and Professor, School of Public Health, The University of Memphis, 236A Robison Hall, Memphis TN 38152, lisa.klesges@memphis.edu.

No financial disclosures were reported by the authors of this paper.
Context

Pragmatic intervention evidence is needed to identify behavioral approaches that can be applied in clinical and community practice to reduce the prevalence of childhood overweight and obesity, which has doubled among children and tripled among adolescents since 1980.\(^1,^2\) Recent estimates indicate that nearly 19% of children aged 6–11 years are overweight.\(^2\) Although studies of lifestyle interventions for childhood obesity have proliferated in recent years, there has been little guidance available to support consistent reporting of key elements related to translation of results.

The U.S. Preventive Services Task Force (USPSTF) and the IOM have identified the lack of consistent reporting of generalizability elements as a limiting factor in their ability to make recommendations for interventions targeting obese children and adolescents.\(^3^-^5\) Providing information within published research on external validity and practical elements could lead to large improvements in the ability of decisions makers to evaluate and synthesize the future evidence base for many interventions, including childhood obesity.\(^4,^9\)

To identify the current status of the literature, and augment a previous review\(^9\), the current review assesses the extent to which the evidence base of behavioral childhood obesity treatment has included key dimensions related to participants, settings, interventions, outcomes, and maintenance of effects. While not all behavioral childhood obesity treatments are expected to be designed as pragmatic clinical trials, elements related to the practical implications of research can be included to support the evaluation of the applicability of study results to local conditions. Behavioral treatment for childhood obesity was generally an intervention that targeted dietary intake, physical activity, or both behaviors for weight loss or to prevent further weight gain. Studies typically included individual, family or environmental strategies to influence changes in behavior and excluded drug or surgical treatments.

Evidence Acquisition

Literature Search and Study Selection

Keyword searches in medical and social databases (i.e., PsycInfo, PubMed) were conducted using Boolean keyword combinations of terms related to intervention, overweight, behavior, and age, resulting in the identification of more than 39,000 abstracts. Database searches were augmented by examination of reference lists for more than 60 qualitative and quantitative reviews of pediatric overweight treatment literature. A total of 1071 studies were individually inspected to determine if they met inclusionary criteria for the current review.

English-language studies were initially evaluated for selection based on inclusion and exclusion criteria used in a recent meta-analytic review of lifestyle interventions targeting overweight youth.\(^10\) These criteria are summarized in Table 1. The present review was conducted in 2009 and included treatment studies published in peer-reviewed journals between 1980 and 2008: (1) with anthropometric measurement (e.g., BMI, body fat) as a primary or secondary outcome; (2) designed as a controlled trial having randomized or
nonrandomized comparison conditions; (3) including at least one behavioral target of either dietary intake or physical activity; and (4) of children aged ≤8 years or in Grade 12 or less in school. Studies that were expressly designed as pilot or feasibility studies were excluded. After applying these criteria, 77 studies were identified for inclusion (Appendix A, available online at www.ajpmonline.org). Although not exhaustive, the current review provides a reasonable representation of studies reporting on behavioral childhood obesity treatments that have the potential for translation and dissemination.

Data Abstraction

A coding manual was developed to guide coding of variables of interest for the present review. The specific dimensions used for coding each investigation are outlined by Green and Glasgow and are considered key quality-rating criteria for external validity by journal editors who publish studies reporting on clinical and community health interventions. Major external validity concepts include: (1) reach (e.g., participation rates of individuals, adoption by clinicians and community settings, as well as representativeness of individuals included); (2) program or policy implementation (e.g., levels of interventionist expertise and training, consistency of delivery, adaptation of an approach to local circumstances); (3) outcomes (e.g., impact on costs, quality of life and adverse consequences, change in BMI); (4) maintenance and sustainability (e.g., which components are institutionalized or modified over time).

All articles were coded by at least two trained raters who independently evaluated each for whether it reported information on the external validity criteria described above. Initial coding resulted in high concordance with 90%–100% agreement on each of the dimensions. When discrepancies occurred, they were resolved by discussion and clarification. Kappa coefficients were not calculated for this study since many categories had true zero cells and this distribution can be problematic when correcting for chance agreement. Results are reported as percentages of papers that reported on the respective external validity criteria. Data on the range and average of reported process and outcome effects are provided as appropriate.

Evidence Synthesis

Description of Studies

The majority of obesity treatment interventions (98.7%) targeted child and/or parent behaviors related to either physical activity or dietary intake and most were designed as RCTs (54 of 77). Treatment outcomes were assessed in terms of body weight, BMI, and/or percentage overweight or percent body fat in 90% of studies. Of 47 studies describing the intervention delivery setting, a medical clinic was most common (n = 28) followed by school-based programs (n = 14). Table 2 summarizes the percentage of studies reporting on various external validity dimensions for the entire sample and by decade of study publication.

In all time periods, studies lacked full reporting on key generalizability and dissemination elements; the most infrequent being setting-level criteria (e.g., methods to identify setting for inclusion, participation rates, and representativeness of settings), participation
percentages and representativeness of individual participants, participation rates of intervention staff, inclusion/exclusion criteria used to select staff, implementation of intervention content, costs, and program sustainability. Across all criteria reported in Table 2, the average reporting was 23.9%, with a range of 0% to 100.0%.

**Reach, Representativeness, and Adoption**

All studies included inclusionary criteria for participant enrollment but only half reported specific descriptions of the larger intended target audience for the intervention. Fewer than six studies reported participation rates or examined the representativeness of individual participants on social and demographic factors. Regarding setting-level adoption, 20.8% of studies described the target settings for adoption. Similarly, key setting-level information such as selection criteria, recruitment methods, or participation of eligible settings were rarely described.

**Implementation and Adaptation**

Adaptation of existing programs was documented in about half of studies but few reported alterations during the study period. All studies described the intended intervention content; only 5.0% reported the extent to which this content was delivered. Roughly half of studies described characteristics of intervention delivery staff but only two evaluated variability of intervention delivery among staff members.

**Outcomes**

Very few studies reported intervention effects in comparison to national benchmarks such as the percentage of children classified as overweight or obese, reported on indices of participant quality of life, or examined any potential adverse outcomes. Subgroup effects or potential interactions between treatment outcome and relevant participant characteristics (e.g., age, gender) were included in less than one third of studies; even fewer reporting differential intervention effects by intervention agent or setting-level characteristics (e.g., size of organization). Participant attrition was available in most studies (>80%) with 27.0% reporting on the representativeness of demographic characteristics between those dropping out and those remaining. Few studies (16.9%) tested for differences in dropout between treatment conditions. Although highly useful for policy planning purposes, only five studies reported a cost estimate and none reported on the total amount of time required to deliver the intervention.

**Maintenance and Sustainability**

Just over half of studies (58.4%) reported the length of follow-up after completion of the intervention, with 28.6% examining follow-up ≥12 months. No studies described whether further modifications to the intervention were considered or examined whether the program was sustained after the research phase was completed.

**Discussion**

The current review evaluated the extent to which key external validity dimensions related to translation of behavioral childhood obesity treatments were reported in papers published
during a period of nearly 3 decades. Although 77 controlled studies were identified, most research studies failed to include key generalizability elements that could support translation of results to other practice and community settings. These findings underscore recent concerns over the lack of evidence available to support actionable steps to address childhood obesity⁴,⁹ and general observations of a lack of reporting of contextual factors in controlled intervention trials.⁹,¹¹–¹⁷

In general, elements associated with internal validity were more commonly reported than those focusing on external validity. This could be due to the influence of reporting guidelines such as CONSORT that have historically emphasized the quality of internal validity reporting. There were changes in reporting over time with an increase in the percentage of studies describing their intended target population, intervention implementation, differential attrition, and representativeness of dropouts. Several reporting guidelines¹⁸–²⁰ published since 1980 focused on these elements so changes could reflect that emphasis. Decreases in the percentage of studies reporting long-term follow-up after intervention completion were seen. Perhaps with growing awareness of the need for research evidence regarding childhood obesity treatments during this time, authors submitted more papers without long-term follow-up and editors accepted these publications in a more timely fashion. Although difficult to verify, it is plausible given the contemporaneous reviews highlighting the need for additional evidence in the field.²¹–²⁵

**Implications and Recommendations**

In parallel to the idea that “all politics is local,” dissemination decisions must consider local conditions in determining whether and how best to implement future evidence-based programs. If reported research does not include context-specific information to evaluate whether an intervention produces a generalized effect or is variable on localized conditions across previous implementations, these decisions are impossible to achieve. From the current review, contextual elements needing most attention included setting-level criteria (e.g., selection criteria, participation rates, and representativeness of organizations), implementation information (e.g., program adaptation, variable delivery by staff), outcomes reporting (e.g., no comparison to a population benchmark, variable outcome by setting, or time and cost of program delivery), and sustainability or institutionalization of program content. Lack of available evidence clearly limits insight into what adaptations to local practice would be needed for translation.

Applying more consistent criteria would improve the ability to move evidence from discovery to delivery generally²⁸,³² and in particular to behavioral treatments for childhood obesity.⁴ For example, understanding whether a program has broad reach or draws only a segment of the targeted population would be essential to understand future recruitment and retention efforts in implementing a local program. Similarly, differences in participation of various settings and characteristics related to program sustainability would be important to consider in comparison to local organizations.

Contrasting intervention outcomes with national benchmarks and consideration of possible adverse effects (e.g., diminished quality of life) are critical for determining the potential significance of childhood obesity treatment programs. Consideration of a range of negative
outcomes (e.g., disordered eating and weight concerns, lowered self-esteem) is needed to comprehensively assess whether an intervention might have future negative impact on recipients and what adaptations might be necessary for local implementation. Finally, ease of program delivery and estimates of the cost of interventions are essential data for financial and policy planning in applied venues.

Coordinated efforts among researchers, funders, editors, grant reviewers, and evidence review consensus panels are needed to address the challenges of expanding the availability of external validity research evidence, in particular to improve translation of childhood obesity interventions. The IOM expert panel for bridging this evidence gap in childhood obesity research recommends that government, foundations, professional organizations and research institutions should establish and support specific guidance on common standards to evaluate evidence quality and that research funders, researchers and publishers consider allowing for generalizability of findings and related implementation information at every stage of research development. This could be accomplished by adding generalizability criteria to proposal review procedures and training reviewers to evaluate generalizability, and journal editors providing guidelines and space for authors to give greater depth of description to clarify generalizability. Editors should encourage reporting of generalizability among studies at all points in the research continuum from pilot investigations to effectiveness trials to dissemination studies.

Implementing these recommendations will require concerted effort to overcome actual and perceived barriers. In particular, editors and researchers may be apprehensive that submissions will be greatly lengthened if expanded external validity elements are included. While extensive descriptions of all elements may not be feasible, it is possible to address a majority of elements in a single publication. Professional meetings to build consensus among researchers, editors, practitioners, and policymakers have resulted in a number of journal editors agreeing to adopt external validity reporting standards. Coordination of efforts in this area has expanded at NIH with annual conferences on the science of dissemination and implementation. Change among reviewers and review criteria is emerging with a new chartered study section at the Center for Scientific Review at NIH formed to meet the needs of targeted grant submissions in dissemination and implementation research.

Conclusion

As national prevention strategies emphasize the need to expand the evidence base of effective interventions and improve the dissemination of research results into community benefit, the consequences of an inability to translate evidence into actionable change in practice becomes ever more poignant. The current review of behavioral childhood obesity treatment research highlights the need for a concerted effort to increase the reporting of external validity and contextual elements that will allow evaluation of generalizability to inform future implementation decisions. It is hoped that with the more recent focus on improving the design and reporting of pragmatic trials and emphasis on external validity reporting criteria, enhancements will be seen in the relevance of future investigations, in particular for research related to childhood obesity. Such changes are
needed and would greatly improve decision makers’ ability to evaluate the applicability of results to current practice and community settings to reduce the burden of childhood obesity in local communities.

The goal of the current review was to highlight the continued need to focus attention on increasing the availability of evidence that reports on key external validity dimensions, and the particular need in the area of childhood obesity interventions, in order to improve the ability to summarize research evidence for literature reviews, guideline development, and implementation and dissemination decisions. Further improvement appears well justified given the current results, the increasing need for evidence to support health promotion, and population trends that make the issue particular pressing in the area of childhood obesity.

**Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

**Acknowledgments**

We acknowledge Dr. Kenneth Ward for his helpful comments in revision of this paper.

**References**

3. IOM. Progress in preventing childhood obesity: How do we measure up? 2006
4. IOM. Bridging the evidence gap in obesity prevention: A framework to inform decision making. 2010


**Table 1**

Summary of study inclusion and exclusion criteria

<table>
<thead>
<tr>
<th>Intervention</th>
<th>Goal was to produce weight loss or prevent weight gain in already overweight children</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment of overweight as primary target</td>
</tr>
<tr>
<td></td>
<td>Focused on change in weight-related health behaviors (e.g., not surgery)</td>
</tr>
<tr>
<td></td>
<td>Primarily an evaluation of intervention efficacy or effectiveness related to weight outcomes</td>
</tr>
<tr>
<td>Sample</td>
<td>Children and/or adolescents aged ≤18 years or in Grade 12 or less in school</td>
</tr>
<tr>
<td></td>
<td>Identified as overweight or obese</td>
</tr>
<tr>
<td></td>
<td>Included in study based on weight status rather than another medical condition (e.g., diabetes)</td>
</tr>
<tr>
<td>Study Design</td>
<td>Between groups with participants in both groups overweight</td>
</tr>
<tr>
<td></td>
<td>Groups were tests of alternate treatments or treatment versus no treatment</td>
</tr>
<tr>
<td>Outcomes</td>
<td>Assessment based on weight parameter was used (e.g., BMI)</td>
</tr>
<tr>
<td></td>
<td>Assessment conducted at equivalent time points in all groups</td>
</tr>
<tr>
<td>Research Study</td>
<td>Reported empirical data</td>
</tr>
<tr>
<td></td>
<td>Provided enough information to compute an effect size</td>
</tr>
<tr>
<td></td>
<td>Published between 1980 and 2008</td>
</tr>
<tr>
<td></td>
<td>Reported in English language</td>
</tr>
</tbody>
</table>
**Table 2**

Percentages of studies reporting external validity dimensions\(^a\) overall (N = 77) and by decade

<table>
<thead>
<tr>
<th>External validity dimension(^b)</th>
<th>% reporting, all studies reviewed (N = 77)</th>
<th>1980-1989 (n = 22)</th>
<th>1990-1999 (n = 20)</th>
<th>2000-2008 (n = 35)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>REACH AND REPRESENTATIVENESS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual participants</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Target audience description</td>
<td>59.7</td>
<td>54.6</td>
<td>45.0</td>
<td>71.4</td>
</tr>
<tr>
<td>Participant inclusion/exclusion criteria</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
</tr>
<tr>
<td>Participation rate</td>
<td>3.9</td>
<td>9.1</td>
<td>0.0</td>
<td>2.9</td>
</tr>
<tr>
<td>Representativeness of participants</td>
<td>7.8</td>
<td>0.0</td>
<td>15.0</td>
<td>8.6</td>
</tr>
<tr>
<td>Setting level</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Description of included settings</td>
<td>20.8</td>
<td>18.2</td>
<td>30.0</td>
<td>17.1</td>
</tr>
<tr>
<td>Setting inclusion/exclusion criteria</td>
<td>6.6</td>
<td>0.0</td>
<td>15.0</td>
<td>5.7</td>
</tr>
<tr>
<td>Participation rate</td>
<td>1.3</td>
<td>0.0</td>
<td>5.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Representativeness of settings</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td><strong>IMPLEMENTATION AND ADAPTATION</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consistent implementation of program</td>
<td>15.6</td>
<td>0.0</td>
<td>5.0</td>
<td>31.4</td>
</tr>
<tr>
<td>Staff expertise and training</td>
<td>59.7</td>
<td>50.0</td>
<td>55.0</td>
<td>68.6</td>
</tr>
<tr>
<td>Implementation differed by staff</td>
<td>2.6</td>
<td>0.0</td>
<td>0.0</td>
<td>5.7</td>
</tr>
<tr>
<td>Customization of existing program</td>
<td>57.1</td>
<td>68.2</td>
<td>50.0</td>
<td>54.3</td>
</tr>
<tr>
<td>Program adaptation during active intervention</td>
<td>6.5</td>
<td>0.0</td>
<td>10.0</td>
<td>8.6</td>
</tr>
<tr>
<td><strong>Outcomes of decision making</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcomes compared to standard goal</td>
<td>6.5</td>
<td>0.0</td>
<td>15.0</td>
<td>5.7</td>
</tr>
<tr>
<td>Adverse consequences</td>
<td>16.9</td>
<td>13.6</td>
<td>15.0</td>
<td>20.0</td>
</tr>
<tr>
<td>Effect moderator by participant characteristic(s)</td>
<td>28.6</td>
<td>31.8</td>
<td>15.0</td>
<td>34.3</td>
</tr>
<tr>
<td>Effect moderator by staff/setting</td>
<td>5.2</td>
<td>9.1</td>
<td>0.0</td>
<td>5.7</td>
</tr>
<tr>
<td>Time needed to deliver interventions</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Costs</td>
<td>6.5</td>
<td>0.0</td>
<td>0.0</td>
<td>14.3</td>
</tr>
<tr>
<td><strong>Maintenance and institutionalization</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Program sustainability</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Conducted follow-up after intervention completion</td>
<td>58.4</td>
<td>77.3</td>
<td>60.0</td>
<td>45.7</td>
</tr>
<tr>
<td>Examined long-term effects (follow-up ≥2 months)</td>
<td>28.6</td>
<td>36.4</td>
<td>45.0</td>
<td>14.3</td>
</tr>
<tr>
<td>Attrition rate</td>
<td>84.2</td>
<td>81.8</td>
<td>85.0</td>
<td>85.7</td>
</tr>
<tr>
<td>Differential attrition by condition tested</td>
<td>16.9</td>
<td>9.1</td>
<td>15.0</td>
<td>22.9</td>
</tr>
<tr>
<td>Dropout representativeness</td>
<td>27.3</td>
<td>13.6</td>
<td>20.0</td>
<td>40.0</td>
</tr>
</tbody>
</table>

\(^a\) External validity is defined according to Leviton.\(^112\)

\(^b\) See Green and Glasgow\(^6\) for a detailed description of coding dimensions.