A RE-AIM Evaluation of Theory-Based Physical Activity Interventions

Iina Antikainen and Rebecca Ellis
Georgia State University

Although physical activity interventions have been shown to effectively modify behavior, little research has examined the potential of these interventions for adoption in real-world settings. The purpose of this literature review was to evaluate the external validity of 57 theory-based physical activity interventions using the RE-AIM framework. The physical activity interventions included were more likely to report on issues of internal, rather than external validity and on individual, rather than organizational components of the RE-AIM framework, making the translation of many interventions into practice difficult. Furthermore, most studies included motivated, healthy participants, thus reducing the generalizability of the interventions to real-world settings that provide services to more diverse populations. To determine if a given intervention is feasible and effective in translational research, more information should be reported about the factors that affect external validity.

Keywords: exercise, external validity, translational research

Despite the numerous health benefits of physical activity and increased focus on physical activity interventions, activity levels in the United States remain low, with only about 15% of adults achieving the recommended amount of activity (United States Department of Health and Human Services, 2000). An estimated 70% of Americans are considered sedentary based on no leisure time activity or inadequate levels of activity (President’s Council on Physical Fitness and Sports, 2002) and the direct and indirect costs of sedentary living for 1987 incidences were estimated at over $150 billion (Pratt, Macera, & Wang, 2000). Lack of physical activity along with poor nutrition and associated weight gain also affect mortality rates and they are now considered the second leading cause of preventable death in the United States (Mokdad, Marks, Stroup, & Gerberding, 2004).

Programs that promote physical activity vary from simple knowledge-based programs and exercise prescription to theory-based behavior modification programs. Results of a meta-analytic review by Dishman and Buckworth (1996) showed that exercise adherence can be improved by intervention strategies. Some of the characteristics of an effective intervention included those based on behavior modification techniques, targeting groups rather than individuals, utilizing a mediated approach,
and emphasizing lower intensity, leisure activities. Since the review, more emphasis has been placed on the use of theory in guiding intervention design as opposed to health education, risk appraisal, and exercise prescription (Baranowski, Cullen, Nicklas, Thompson, & Baranowski, 2003; Biddle & Nigg, 2000).

According to a review by Biddle and Nigg (2000), the most supported theories in the physical activity domain include the transtheoretical model (TTM; Prochaska & DiClemente, 1984), the social cognitive theory (SCT; Bandura, 1986), and the theory of planned behavior (TPB; Ajzen, 1991). More recently, self-determination theory (SDT; Ryan & Deci, 2000) has gained popularity in exercise psychology research (Lox, Martin Ginis, & Petruzzello, 2010). All of these theories have been used successfully in physical activity interventions (Biddle and Nigg, 2000; Chatzisarantis & Hagger, 2009).

The transtheoretical model includes the following constructs: stages of change (SOC), decisional balance, processes of change, and self-efficacy (Prochaska & DiClemente, 1984). The main hypothesis of the model is that behavior change does not occur all at once, but rather it is a dynamic and gradual process with individuals progressing or relapsing through five SOC including precontemplation, contemplation, preparation, action, and maintenance.

The hypotheses of the SCT are that each of the three main constructs—including personal factors (internal thoughts and feelings about a behavior such as self-efficacy and outcome expectations), behavioral factors (knowledge and skills related to a health behavior), and environmental factors (perceptions of and the actual physical and social environment)—are reciprocally related (Bandura, 1986). The reciprocal nature of the relationship among the constructs indicates that changes in personal and environmental factors will influence changes in behavior, and in turn, changes in behavior will influence changes in personal and environmental factors.

The hypotheses of the TPB are that participation in a behavior is determined by an individual’s intention (i.e., motivation), as well as their perceived behavioral control over the activity (Ajzen, 1991). The person’s intention (INT), in turn, is determined by the person’s attitude (ATT; the overall evaluation of the behavior), subjective norm (SN; the perceived pressure to participate or not participate in the behavior), and perceived behavioral control (PBC; the person’s estimation about their ability to perform a given task and the amount of control they feel they have over participation; Ajzen).

Finally, SDT proposes that motivation that is based on intrinsic factors and is self-driven or an autonomous choice is more likely to translate into behavioral change rather than extrinsically motivated behavior based on rewards or coercion (Ryan & Deci, 2000). Each of these theories provides a framework of constructs for understanding behavior. Furthermore, many of the constructs can be targeted as mediators, which are variables that can be manipulated to produce behavior change and explain the mechanism by which the behavior change has occurred (Marcus & Forsyth, 2009). To design more effective interventions, Marcus and Forsyth suggest using several theories, building the intervention around several mediators, and measuring the change in theory-based mediators.

In addition to using behavioral theory to guide intervention design, others have recommended evaluating the public health impact of interventions with an increased focus on external validity (Estabrooks, Dzewaltowski, Russell, Glasgow, & Klesges, 2003). In the past, based on a medical research model, physical activity research
has focused on clinically significant results leading to highly controlled, short-term interventions with healthy, motivated participants. These types of programs are difficult and costly to maintain and virtually impossible to adopt in real-world settings where participants are more likely to have a variety of health issues and be less motivated to engage in physical activity. The RE-AIM framework was developed to evaluate the external validity of health promotion programs (Glasgow, Vogt, & Boles, 1999) and it allows for the evaluation of the potential of a program for translational research in real-world settings (Eakin et al., 2007, Estabrooks et al.).

Glasgow and colleagues (1999) define the five dimensions (reach, efficacy, adoption, implementation, and maintenance) of external validity that may be assessed in the RE-AIM framework, as well as associated metrics/measures that can be used to assess the degree to which a study addressed each dimension. They define reach as a measure of individual-level participation including the proportion of the population targeted that are affected by the intervention, as well as the representativeness of the participants to the target population. Efficacy measures should include the effectiveness of the intervention, positive or negative consequences of the intervention, as well as behavioral, quality of life, and participant satisfaction outcomes. Adoption is concerned with the proportion of existing or available settings that offer the intervention and how representative these settings are of the community as a whole. Implementation refers to the degree to which the intervention is delivered as intended, and it is evaluated based on the faithfulness of the program administrators to the design of the intervention through proper process evaluation. Finally, maintenance includes an individual and an institutional level component. On the individual level, maintenance refers to long-term (≥ 6 months) change in behavior, and on the institutional level it refers to the extent that the new intervention becomes an established program in the organization.

To date, few studies have used the RE-AIM framework to review the reporting of dimensions important to the external validity of physical activity interventions. Estabrooks et al. (2003) evaluated 32 school health promotion studies (these programs targeted physical activity, good nutrition, and/or smoking cessation or prevention) and found that efficacy was the only component of RE-AIM that all studies reported. Although most studies reported number of participants, few reported the number of all eligible participants or participant characteristics, and reporting in the areas of adoption and maintenance was equally poor. The authors concluded that health promotion studies rarely reported on the RE-AIM dimensions concerned with external validity, which limits the translation of such results into practice. To assist future researchers, the authors developed a template to standardize the reporting of the characteristics of internal and external validity and they made specific recommendations about how to clearly describe participation and implementation on the individual and organizational levels to design interventions with a potential for large-scale adoption (Glasgow et al., 1999). However, to date no published reviews have used the RE-AIM framework to examine the external validity of theory-based physical activity interventions in various populations and settings.

The physical activity literature includes several meta-analytic and narrative reviews that focus on the efficacy of physical activity interventions, but none have specifically addressed the public health impact and translation of these programs to real-world settings. Therefore, the overall purpose of this review was to evaluate the external validity of theory-based physical activity interventions designed...
according to the TTM, TPB, SCT, and SDT using the RE-AIM framework and criteria established by previous researchers (Glasgow et al., 1999). These behavior change theories were chosen for this review because they were identified as the most supported and most commonly studied in the physical activity literature (Biddle & Nigg, 2000; Lox et al., 2010).

Method

Search and Inclusion Criteria

Using the keywords physical activity, exercise, theory of planned behavior, social cognitive theory, transtheoretical model, self-determination theory, theory-based, and intervention, we searched the electronic databases of PsychInfo (EBSCO host) and PubMed (EBSCO host), as well as reference lists of retrieved articles and review articles to identify physical activity intervention studies published between 1996 and July of 2009. The year 1996 was selected as the starting point because of the increased emphasis on theory-based interventions since that time (Biddle and Nigg, 2000; Dishman & Buckworth, 1996).

The inclusion criteria for articles were (a) articles written in English; (b) publication in peer-reviewed journals; (c) a theory-based physical activity intervention, as defined by the authors, designed according to the TTM, SCT, TPB, SDT or a combination of these four theories; (d) at least one intervention arm that included a behavioral component using behavior modification, cognitive behavior modification, health education, or exercise prescription strategies; and (e) the intervention had to compare physical activity behavior from pre- to postintervention (i.e., assessment of physical activity behavior change). Initially, we excluded articles by screening titles and abstracts of the hundreds of articles that met the keywords, but did not clearly meet the inclusion criteria. As a result of this initial screening process, we were able to identify 80 physical activity intervention studies using the selected behavioral theories reported in 88 journal articles (results of some studies were reported in several articles). Of the 88 articles that remained following our initial screening process, 24 were eliminated because following a more thorough review they did not use a behavioral theory for the physical activity component of the study (n = 10), they used a theory not included in this review (n = 8), physical activity behavior change was not assessed or results were not reported (n = 5), or the study compared two interventions based on a different theory (n = 1). The final review included 57 interventions that were reported in 66 journal articles (some interventions were reported in more than one article and one article covered two different intervention programs). When the results from one intervention were reported in multiple publications, the RE-AIM evaluation was based on all available data on the intervention.

Coding Protocol

The publications were coded based on whether they reported on the dimensions of RE-AIM as outlined in the definitions for each dimension below. Any information that was reported was coded for further analysis. Reach was coded on the following levels: (a) participation rate was the percentage of eligible people from the targeted
population who completed the intervention; (b) inclusion criteria, exclusion criteria, and percentage of people excluded; and (c) representativeness of the participants as compared with the targeted population. \textit{Efficacy} was evaluated based on the following criteria: (a) assessment of physical activity rate, (b) attrition rate at the completion of the intervention, (c) use of intent-to-treat strategies, (d) inclusion of a quality of life measure, and (e) reporting of any negative consequences of the intervention. \textit{Adoption} was assessed on the following levels: (a) percentage of eligible organizations or sites that offered the intervention, (b) eligibility criteria and percentage of eligible sites participating, (c) reporting of the exclusion criteria and rate at the organizational level, and (d) representativeness of the participating sites as compared with sites in the region not offering the intervention. \textit{Implementation} was assessed based on whether the researchers reported data on the faithfulness of intervention delivery assessed by process evaluation. \textit{Maintenance} was reported on the following levels: (a) whether the study included at least a 6-month follow-up on individual behavior following last contact in the intervention, (b) if change in physical activity was maintained at follow-up, and (c) if the intervention continued on the organizational level after study completion. The frequencies and percentages were recorded and calculated using Microsoft Excel 2007.

\textbf{Results}

Of the 57 interventions reviewed, 27 were based on the TTM, 5 were based on the TPB, 9 were based on the SCT, 2 were based on the SDT, and 14 used a combination of 2 or more of these theories. The most commonly targeted mediator across the four theories was the SOC, which was included in 52.6\% ($n = 30$) of all studies. Fourteen TTM-based interventions also reported targeting processes of change, 13 focused on self-efficacy, and 10 targeted decisional balance.

Interventions based on the TPB most often targeted INT ($n = 4$) or the underlying beliefs about physical activity ($n = 4$). Some studies also reported targeting ATT ($n = 3$), PBC ($n = 3$), and SN ($n = 2$). Two of the five interventions based on the TPB reportedly targeted all of the TPB constructs and their underlying beliefs (i.e., Hardeman, Kinmonth, Michie, & Sutton, 2009; Vallance, Courneya, Plotnikoff, & Mackey, 2008).

Among interventions based on the SCT, the most commonly targeted mediators were self-efficacy and outcome expectations with seven out of the nine SCT-based interventions targeting self-efficacy and five interventions targeting outcome expectations. Other mediators targeted within SCT-based interventions included physical activity barriers ($n = 5$), benefits of physical activity ($n = 3$), social support ($n = 2$), and SOC ($n = 1$).

Two of the interventions based on the SDT targeted intrinsic motivation (i.e., Chatzisarantis & Hagger, 2009; Wilson, Evans, Williams, Mixon, Sirard, & Pate, 2005) and autonomy (i.e., Chatzisarantis & Hagger; Levy & Cardinal, 2004. Other mediators targeted by the SDT-based interventions included intention ($n = 1$), self-efficacy ($n = 1$), self-concept ($n = 1$), competence ($n = 1$), and relatedness ($n = 1$).

The interventions were designed for a variety of populations including sedentary adults ($n = 14$), employees ($n = 8$), older adults ($n = 9$), people with diabetes ($n = 7$), children or adolescents ($n = 7$), primary care patients ($n = 5$), cancer patients ($n = 4$), and other ($n = 3$). The majority of the studies assessed physical activity solely
by self-report questionnaires \((n = 40, 70.2\%)\), followed by using a combination of self-report and an objective measure \((n = 14)\), and using objective measures such as pedometers or accelerometers only \((n = 3)\). The length of the interventions varied greatly with most of the studies being single contact interventions \((n = 13)\). The second most common intervention lengths were interventions lasting for 3 months \((n = 9)\), 6 months \((n = 7)\), and 1 year \((n = 7)\). The majority of the interventions lasted for 6 months or less \((n = 46, 80.7\%)\).

The interventions that recommended home-based activities were coded as home-based and those that offered physical activity classes or programs were coded as center based. All but one intervention targeted home-based physical activity or home and center-based activity, and one study targeted active commuting behavior. The interventions included mail or e-mail interventions \((n = 17, 29.8\%)\); interventions delivered in person at a center, school, or clinic \((n = 17, 29.8\%)\); phone-delivered interventions \((n = 2, 3.5\%)\); mass media \((n = 1, 1.8\%)\); and a combination of these strategies \((n = 20, 35.1\%)\).

Reach

Only 19.3\% \((n = 11)\) of the interventions reviewed reported participation rate as a percentage of those eligible to participate (see Table 1). Many studies did not include information about the specific target population or researchers were using convenience samples with various recruitment methods within a community. When participation rate was reported, it was between 0.86\% and 66.7\% of the population with a median reach of 18.7\% with only two interventions being able to attract more than 50\% of the target population (i.e., Dishman et al., 2004; Jones, Courneya, Fairey, & Mackey, 2004).

Exclusion criteria were reported in 78.9\% \((n = 45)\) of the studies; however, only 26.3\% \((n = 15)\) of the studies that specified exclusion criteria reported the proportion of people excluded from the study. In addition, 86.0\% \((n = 49)\) of the studies reported inclusion criteria. Finally, only one study reported on the representativeness of the participants to the total target population (i.e., Hallam & Petosa, 2004).

The low reach (0.87\%) in the intervention by Hallam and Petosa (2004) was caused by a limitation on the number of participants that could be accommodated by this worksite program. Specifically, their study was limited to 60 participants with a total eligible work force of 7000. With so few studies reporting participation rate, it is difficult to make any conclusions about trends, but some of the more successful programs in reference to reach were one single contact intervention (66.7\%; i.e., Jones et al., 2004), a school physical education intervention (51.6\%; i.e., Dishman et al., 2004), and an intensive face-to-face intervention (49.8\%; i.e., Proper, Hildebrand, van der Beek, Twist, & van Mechelen, 2003).

Efficacy

All articles included in this review had physical activity as one of the outcome measures; therefore, all of the studies selected reported efficacy as a difference in physical activity participation from pre- to postintervention with 75.4\% \((n = 43)\) of the studies reporting an increase in physical activity levels (Table 1). Attrition rates were reported in 93.0\% \((n = 53)\) of the studies and ranged from 0\% to 62.5\% with a median of 22.7\%. The highest attrition rates occurred in a 9-month
face-to-face intervention among people who were recruited through a general practice (62.5%; i.e., Proper et al., 2003). Some of the lowest attrition rates occurred when the participants were not asked for a large commitment to a program. For example, a single contact intervention targeting school children with a 5-week follow-up assessment was able to retain all participants (i.e., Chatzisarantis & Hagger, 2005) and a brief telephone based intervention with breast cancer survivors lost only 4.7% of the participants (i.e., Pinto, Frierson, Rabin, Trunzo, & Marcus, 2005).

Table 1  Percentage and Number of Articles Reporting on the RE-AIM Components

<table>
<thead>
<tr>
<th>Component</th>
<th>Number Reporting</th>
<th>Percent Reporting</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Reach</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participation rate</td>
<td>11</td>
<td>19.3</td>
</tr>
<tr>
<td>Exclusion criteria</td>
<td>45</td>
<td>78.9</td>
</tr>
<tr>
<td>Percentage of people excluded</td>
<td>15</td>
<td>26.3</td>
</tr>
<tr>
<td>Inclusion criteria</td>
<td>49</td>
<td>86.0</td>
</tr>
<tr>
<td>Representativeness of participants</td>
<td>1</td>
<td>1.8</td>
</tr>
<tr>
<td><strong>Efficacy</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Behavioral measure</td>
<td>57</td>
<td>100.0</td>
</tr>
<tr>
<td>Change in physical activity reported</td>
<td>43</td>
<td>75.4</td>
</tr>
<tr>
<td>Attrition</td>
<td>53</td>
<td>93.0</td>
</tr>
<tr>
<td>Intent to treat / Imputation</td>
<td>16</td>
<td>28.0</td>
</tr>
<tr>
<td>Quality of life measure</td>
<td>13</td>
<td>22.8</td>
</tr>
<tr>
<td>Negative consequences</td>
<td>2</td>
<td>3.5</td>
</tr>
<tr>
<td><strong>Adoption</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of sites participating</td>
<td>32</td>
<td>56.1</td>
</tr>
<tr>
<td>Number of sites eligible</td>
<td>2</td>
<td>3.5</td>
</tr>
<tr>
<td>Percentage of eligible sites participating</td>
<td>2</td>
<td>3.5</td>
</tr>
<tr>
<td>Exclusion criteria</td>
<td>6</td>
<td>10.5</td>
</tr>
<tr>
<td>Exclusion rate</td>
<td>1</td>
<td>1.8</td>
</tr>
<tr>
<td>Representativeness of sites</td>
<td>1</td>
<td>1.8</td>
</tr>
<tr>
<td><strong>Implementation</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Process evaluation data reported</td>
<td>19</td>
<td>33.3</td>
</tr>
<tr>
<td>Changes made to intervention</td>
<td>6</td>
<td>10.5</td>
</tr>
<tr>
<td><strong>Maintenance</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Individual level—6 months after contact</td>
<td>14</td>
<td>24.6</td>
</tr>
<tr>
<td>Change in physical activity at follow-up</td>
<td>8</td>
<td>14.0</td>
</tr>
<tr>
<td>Institutional level—program continued</td>
<td>3</td>
<td>5.3</td>
</tr>
</tbody>
</table>

*Note.* The percentages calculated are based on a total of 57 interventions.
Intent-to-treat or imputation procedures were used to account for missing values by 28.0% \((n = 16)\) of the studies. Only 22.8% \((n = 13)\) of the studies measured changes in quality of life and any unintended negative consequences of the intervention were reported by only two studies. In addition to reporting on program effectiveness, quality of life, and attrition, Mutrie et al. (2002) reported that their active commuting program had no adverse effects such as increased traffic accidents. Hardeman et al. (2009) reported that 32 of their 321 participants reported having to visit a physician because of exercise related pain, but that none experienced severe complications.

**Adoption**

Reporting of information necessary to evaluate adoption such as specifying sites or recruitment strategies was not included in 14 of the 57 studies included in this review. Of the remaining 43 studies, 32 (56.1%) included the number of sites participating in recruitment or implementation of the program, but only two studies reported the number of eligible sites for the study and the percentage of sites participating (Table 1; i.e., Dishman et al., 2004; Smith, Bauman, Bull, Booth, & Harris, 2000). Exclusion criteria at the organizational level was reported by 10.5% \((n = 6)\) of the studies (i.e., Chatzisarantis & Hagger, 2009; Dishman et al., 2004; Elbel, Aldana, Bloswick, & Lyon, 2003; Hooker et al., 2005; Reger et al., 2002; Resnick, Luisi, & Vogel, 2008), and the percentage of sites excluded and representativeness of the sites participating as compared with those that elected not to participate was reported by only one study (i.e., Reger et al.).

Exclusion criteria were reported by one choice-based, telephone-assisted program for older adults offered through 13 diverse agencies. They required that participating sites were experienced in providing wellness-related services to older adults, experienced in recruiting and managing volunteers, had adequate infrastructure, and the ability to provide $10,000 in contribution to the program (i.e., Hooker et al., 2005). Although they failed to specifically describe the representativeness of the 13 sites that participated in this study as compared with those that did not, the researchers did describe each site and the populations served by the site. Another intervention that focused on increasing activity levels on the community level chose the target city because of its adequate and affordable media, co-operation of the local health agencies, and its proximity to the university conducting the study (i.e., Reger et al., 2002). Researchers also briefly described the city’s representativeness with characteristics such as annual average income, as compared with the rest of the nation and the state.

**Implementation**

Implementation was evaluated based on whether the authors reported any data on the faithfulness of intervention delivery. Reporting on implementation and process evaluation varied greatly between the programs. In one case a complete article (i.e., Stewart et al., 2006) was dedicated to describing implementation while some provided no information. Of the articles reviewed, 33.3% \((n = 19)\) addressed process evaluation on some level and 6 (10.5%) of these studies reported changing the intervention in some way based on the results of a process evaluation or feedback
from participants (see Table 1). However, of the 19 studies that reported conducting a process evaluation, 6 did not provide information about who conducted the process evaluation (i.e., Hardeman et al., 2009; Hooker et al., 2005; Patrick et al., 2001; Pinto et al., 2002; Stewart et al., 2006; Wilcox et al., 2008) and 2 studies reported that the evaluation was completed by the same person conducting the intervention (i.e., Basler, Bertalanffy, Quint, Wilke, & Wolf, 2007; Jones et al., 2004). Four interventions based the evaluations on information provided by participants (i.e., Levy & Cardinal, 2004; Marshall et al., 2003; Napolitano et al., 2003; Rovniak, Howell, Wojcik, Winett, & Martinez-Donate, 2005) and six articles specified that process evaluation was conducted by an independent evaluator or a researcher not delivering the intervention (i.e., Calfas et al., 2000; Chatzisarantis & Hagger, 2009; Dishman et al., 2004; Marcus et al., 1998; Pinto et al., 2005; Wilson et al., 2005).

The actual reporting on process evaluation ranged from a few simple statements to a complete analysis. One study by Patrick et al. (2001) reported that staff received frequent supervision and mailings were sent as planned and they reported the average number of phone calls completed as a part of the intervention. In a physical therapy setting, Basler et al. (2007) mentioned that the therapists recorded their actual behavior during the treatment and marked any deviation from the plan. Their recordings were then reviewed by two of the authors and encouragement to adhere to the program was provided to the therapists.

A multisite, school-based program evaluated implementation by categorizing each school as high or low implementers based on a comprehensive evaluation of an independent process evaluator’s records and the program staff rating of all implementation components and adherence to the physical education class criteria (i.e., Dishman et al., 2004).

**Maintenance**

Individual level maintenance for at least 6 months after the intervention was reported by 24.6% \((n = 14)\) and 14.0% \((n = 8)\) reported that significant increases in physical activity levels were maintained (Table 1). The follow-up period in these studies ranged from 1 to 24 months. Institutional maintenance past the study period was reported by 5.3% \((n = 3)\) interventions (i.e., Mutrie et al., 2002; Stewart et al., 2006; Wilcox et al., 2008) and because such maintenance is rare, it can be assumed that most of the programs were not continued.

A study by Mutrie et al. (2002) reported on both individual and institutional level maintenance. This worksite program was designed to encourage active commuting and it was effective in maintaining change in active commuting at a 12-month postintervention assessment. Based on the positive results of this study, the program was printed for national distribution in Scotland free of charge. Another program that was specifically designed for community diffusion targeting diverse older adults at three community centers reported that each site continued to provide physical activity programming at the site after the study conclusion by applying for grants or using volunteer workers (i.e., Stewart et al., 2006).
Discussion

This review focused on examining the extent to which current theory-based physical activity interventions report on the issues of external validity by using the RE-AIM framework. Similar to a school health promotion RE-AIM evaluation by Estabrooks et al. (2003), most of the studies reviewed focused on internal validity and statistically significant findings rather than on issues of external validity. Few studies included a thorough description of the intervention protocol and materials.

Reach

When evaluating reach, few studies reported the intended target population. Compared with the RE-AIM evaluation of school-based health promotion programs, the percentage of interventions reporting the total population reached in our review is much lower (59.3% vs. 20.4%, respectively). However, specifying and determining the total target population in school-based programs where the population of a school is documented is easier than determining the target population when recruiting from various community-based programs open to the public. Future researchers should identify a specific target population and report the number of people targeted to provide other researchers or practitioners the ability to evaluate the potential reach of an intervention.

Reporting on other aspects of reach is important for determining the feasibility of translating the intervention to a real-world setting. Several of the reviewed studies excluded anyone with a chronic health condition and some studies targeting minorities excluded participants who did not speak adequate English (i.e., Albright et al., 2005). In addition, several studies excluded a large portion of participants who would have been interested in the intervention. For example, Pinto et al. (2005) excluded participants because of chronic health conditions including high blood pressure and diabetes, medications, and prior history of cancer or for being considered physically active. Being that the target population of this intervention was female breast cancer patients, these exclusion criteria lead to a 71% exclusion rate. The practical importance of reporting exclusion criteria and exclusion rate for translational research is twofold. First, researchers need to know who was excluded to determine if the results can be generalized to a potential population and setting, and second, researchers need to know how many people were excluded based on these standards to determine the feasibility of the intervention for a given target population. Future researchers should consider the impact a strict focus on internal validity has on the external validity of their study as they provide limited information for practitioners looking to implement the intervention in real-world settings where such factors cannot be controlled.

The last facet of reach that is important for determining public health impact and appropriateness of a program for translational research is the representativeness of the study participants to the target population. Unfortunately, only one of the physical activity interventions reviewed reported on the representativeness of the participants to the target population (i.e., Hallam & Petosa, 2004). Future researchers should report the characteristics of their participants as compared with the whole target population and determine if demographic differences exist in intervention efficacy.
Efficacy

Although all of the reviewed studies reported on the effectiveness of the intervention for changing physical activity behavior, the reporting was often vague. Some studies that first reported that levels of activity changed further stated that no statistical differences in activity levels were found (i.e., Jimmy & Martin, 2005). In other studies only percentages and raw data were reported with no statistical test to determine if the changes were significant. Some researchers reported that their program changed behavior in both the intervention and the control group without accounting for the fact that such changes could be caused by testing, seasonal effects, or other programs available to the participants (i.e., DuVall, Dinger, Taylor & Bemben, 2004). This type of reporting on efficacy may reflect researchers’ sense of urgency for finding statistically significant findings, but the results are not helpful for health promotion professionals and researchers trying to discern the clinical significance of the interventions.

Difficulties in recruiting adequate numbers of participants or having limited resources to include large samples are additional factors that impact the results and efficacy of intervention studies. In the studies reviewed, sample sizes ranged from 20 to 2087 with a median sample size of 126. In addition, 35.2% of the studies had less than 100 participants and several studies divided a relatively small sample into two or more intervention groups (i.e., Dinger, Heesch, Cipriani, & Qualls, 2007). However, when considering the public health impact of physical activity interventions, the focus should not necessarily be on recruiting large numbers of people for a single intervention because in real-world settings few individual programs or organizations have the funding to reach thousands of people. Thus, in future studies, it may be better to focus on proper reporting of the intervention protocol, results, and effects of interventions rather than focus only on the efficacy of a single program. This would provide the information needed for future reviews and meta-analyses to determine the actual public health impact and efficacy of interventions. Shifting focus in this manner would also encourage the publication of studies that are well designed and described and can be translated to real-world settings, but that may lack statistically significant findings because of a small sample size.

Adoption

Based on the articles reviewed and prior reports by other researchers (Estabrooks, et al., 2003), adoption appears to be the least reported dimension of external validity of interventions. The percentage of physical activity interventions that reported the number of eligible sites was lower than in the previous RE-AIM analysis of school-based programs (3.7% vs. 14.7%, respectively). Among the studies included in the review, several investigators relied on a convenience sample from the community while focusing solely on the efficacy of the program. Such studies cannot be translated to real-world settings and have limited external validity. Future researchers who recruit from community centers, physician’s offices, schools, or workplaces should report the number of sites that were screened or invited to participate in the study, as well as any institutional-level exclusion criteria and representativeness or characteristics of the sites choosing to adopt the program.
Implementation

Only about a third of the interventions reviewed reported any information on process evaluation. Considering the complexity of current behavior change programs this result is not surprising. The space limitations of most journals have led to brief descriptions of programs and a lack of reporting on process evaluation. Including process evaluation is particularly important with multisite programs where several individuals are independently responsible for delivering the interventions. If the intervention was not delivered as originally intended, the reader should be made aware of the changes that took place so they can evaluate the effectiveness of the program based on the actual process rather than a description of an ideal scenario. Process evaluation data can also provide practitioners with information about how flexible a given program is to changes while still maintaining efficacy. For example, Stewart et al. (2006) examined the CHAMPS physical activity program that was adopted for implementation at three sites and provided a detailed description of the differences in adoption and both organizational and individual level results at each site. Future researchers should consider including process evaluation data or submitting separate articles on implementation and process evaluation.

Maintenance

Individual-level maintenance levels reported in this review were similar to the review of school-based health promotion programs (25.9%). Although, the majority of studies in this review did not report maintenance data, it is promising that more than 50% of the studies that did assess individual level maintenance reported that increases in physical activity levels had been maintained. The three studies that reported on institutional level maintenance were community based interventions that focused on translating an intervention into a real-world setting (i.e., Mutrie et al., 2002; Stewart et al., 2006; Wilcox et al., 2008).

Study Limitations

One of the limitations of assessing the external validity of interventions is that the researchers of the reviewed studies may have collected some of the information required to complete a RE-AIM evaluation, but did not report it in the articles and their intention may be to publish this information in the future. In addition, while searching for theory-based interventions, it is possible that not all articles related to the studies reviewed were recovered. This may be the case particularly with articles related to process evaluation.

To evaluate the external validity of theory-based studies, all studies defined by the authors as theory-based were selected for this review. This selection criterion may be considered a limitation because some of the selected studies may have lacked fidelity to theory. Fidelity to theory refers to the extent to which the intervention adhered to the particular theory. Although reviewing the extent to which the intervention actually adhered to the author identified theory was beyond the scope of this study, future reviews on theoretical fidelity of intervention programs is recommended. In addition, future intervention programs should be designed with high fidelity to theory with specific descriptions of the theory-based strategies that were used and the identification of the theory-based constructs that were targeted
as the mediators for promoting change in physical activity behavior. These recommendations will help to improve the quality of future interventions.

Also, although each RE-AIM dimension is technically scored from 0 to 100, specific norms for the RE-AIM components have not been determined. Thus, the meaningfulness of the information provided for each dimension is left to be decided on a case-by-case basis. More studies evaluating various physical activity interventions are needed to determine standards for the RE-AIM dimensions.

**Conclusion**

To our knowledge this is the first study to review the external validity and translatable ability of theory-based physical activity interventions targeting various populations. The physical activity interventions included in this review were much more likely to report on issues of internal, rather than external validity and on individual, rather than organizational level components of RE-AIM. The practical implication of this is that the translation of many of the interventions into practice would be difficult or impossible. In addition, most studies included motivated, healthy participants reducing the generalizability of the interventions to real-world settings that provide services to diverse populations. The current trends of basing interventions on theory, writing separate articles on intervention design, efficacy, and process evaluation, and focusing on long-term maintenance of behavior change are encouraged to provide more meaningful information necessary for translational research.

To determine if a given intervention is feasible and effective in translational research and not only effective under highly controlled conditions, more information must be reported in future studies about the factors that affect external validity. Based on the findings of this review, future researchers are encouraged to focus on issues of external validity as outlined by the RE-AIM framework if significant changes in physical activity participation on the population level are desired.

**References**

*Note.* In the references list, the asterisk signifies a physical activity intervention study that was reviewed in the present article.


Manuscript submitted: June 3, 2010

Revision accepted: November 11, 2010