How Effective Are Active Videogames Among the Young and the Old? Adding Meta-analyses to Two Recent Systematic Reviews

Jonathan van ’t Riet, PhD,¹ Rik Crutzen, PhD,² and Amy Shirong Lu, PhD³

Abstract

Objective: Two recent systematic reviews have surveyed the existing evidence for the effectiveness of active videogames in children/adolescents and in elderly people. In the present study, effect sizes were added to these systematic reviews, and meta-analyses were performed.

Materials and Methods: All reviewed studies were considered for inclusion in the meta-analyses, but only studies were included that investigated the effectiveness of active videogames, used an experimental design, and used actual health outcomes as the outcome measures (body mass index for children/adolescents [k = 5] and functional balance for the elderly [k = 6]).

Results: The average effect of active videogames in children and adolescents was small and nonsignificant: Hedges’ g = 0.20 (95 percent confidence interval, −0.08 to 0.48). Limited heterogeneity was observed, and no moderator analyses were performed. For the effect of active videogames on functional balance in the elderly, the analyses revealed a medium-sized and significant effect of g = 0.68 (95 percent confidence interval, 0.13–1.24). For the elderly studies, substantial heterogeneity was observed. Moderator analyses showed that there were no significant effects of using a no-treatment control group versus an alternative treatment control group or of using games that were especially created for health-promotion purposes versus off-the-shelf games. Also, intervention duration and frequency, sample size, study quality, and dropout did not significantly moderate the effect of active videogames.

Conclusions: The results of these meta-analyses provide preliminary evidence that active videogames can have positive effects on relevant outcome measures in children/adolescents and elderly individuals.

Introduction

Physical activity contributes to a healthy body weight in children and adolescents,¹ as well as the quality of life in the general adult population,² and is a major predictor of physical function in the elderly.³ Promoting physical activity is challenging, however, because behavior is influenced by many factors.⁴ Interventions have generally had small effects⁵,⁶ and have not been able to reverse an alarming increase in obesity rates.⁷ Undiminished efforts are needed, therefore, to identify new approaches to promoting physical activity.

One example of such a new approach is to incorporate active videogames in interventions. Recently, a new generation of digital gaming systems, which require physical exertion to play the game, has become commercially available. These games are denoted as “active videogames”⁸ or “exergames”⁹ (e.g., “Dance Dance Revolution™” [Konami Digital Entertainment, El Segundo, CA] and “Wii Sports Resort™” [Nintendo, Kyoto, Japan]). Several studies have shown that active videogames can have beneficial effects on physical activity and related health outcomes.¹⁰-¹⁵ On the other hand, several other studies have failed to find significant effects.¹⁴,¹⁵

So how effective are active videogames? Two recent systematic reviews investigated the effects of active videogames among the young and the old. Both provided a thorough overview of the available research in each domain: One focused on the effects of videogames targeting nutrition behavior and physical activity among children and adolescents (18 years of age and under),¹⁶ and the other focused on the effects of active videogames on physical function in the elderly (60 years of age and above).¹⁷ In both reviews conclusions regarding effectiveness were based on P values, which
is not surprising in light of behavioral science’s traditional reliance on P values as the key indicator. Nowadays, however, effect sizes are deemed increasingly important because effect sizes, unlike P values, are informative to determine whether an effect is meaningful and substantive. In this article, we therefore report the effect sizes of the included studies and use meta-analytic procedures to gain a more detailed insight into the meaningfulness and substance of the study results.

At present, several systematic reviews on active videogames have been conducted, but only one used meta-analytic procedures. That study synthesized the results of 18 studies on the effects of active videogames on acute energy expenditure and concluded that active videogames do indeed facilitate light- to moderate-intensity activities. No meta-analysis, however, has assessed whether active videogames can result in increased long-term physical activity, which is more important from a public health perspective than one-time physical exertion. Also, no published meta-analysis has focused on direct health outcomes, such as body mass index (BMI).

Although research on active videogames is still in its infancy and insights can be expected to evolve quickly, it is important to obtain a quantified estimate of active videogames’ effects on health-related outcomes in a “considered synthesis of multiple studies.” In this study, therefore, we added effect sizes to two existing systematic reviews and performed meta-analyses to provide more fine-tuned insight into the effectiveness of active videogames. The two systematic reviews were chosen because they were published recently and contain all the evidence available to date. Close inspection of three other recent reviews did not yield additional studies that would have met the inclusion criteria of Lu et al. or Larsen et al. 

Materials and Methods

Data sources

Active videogames for children and adolescents. In their systematic review, Lu et al. identified studies on the effects of health videogames on childhood obesity-related outcomes. The inclusion criteria were (1) focus on improving or maintaining health, (2) one or more videogames were used as the intervention, (3) use of quantitative outcome measures, (4) target population being the healthcare receiver population (e.g., overweight and healthy children) instead of healthcare providers (e.g., doctors), and (5) original study only. Additionally, included studies must (6) target participants 18 years or younger and (7) include one or more obesity-related health outcome measures such as BMI.

Because of this latter focus on actual biophysical outcomes, the included studies are very much comparable to other studies in terms of outcome measures. Indeed, all included studies used BMI as the primary outcome. In terms of intervention approaches and study designs, however, considerable variation was observed, limiting the extent to which the studies are comparable, as the authors noted. We therefore added two criteria for inclusion in the present meta-analysis. First, the intervention had to consist of an active videogame in which physical exertion was required for gameplay. This excluded two studies examining the effect of nutrition games and two others that targeted the determinants of physical activity (e.g., knowledge) but did not require physical exertion (see below). Second, we only included studies that compared the effects of the active videogame with an alternative treatment control group or a no-game control group (i.e., using an experimental design). Table 1 shows all studies included in the review of Lu et al. and indicates which studies were excluded and why. In total, five studies were included in the present meta-analysis.

Active videogames for elderly individuals. In their systematic review, Larsen et al. set out to determine the effects of active videogames on physical outcome measures in elderly individuals. They included research that (1) investigated the effects of active videogames, (2) used randomized controlled trials to compare active videogames with an alternative intervention or no intervention, (3) targeted healthy elderly individuals (> 60 years of age) as study participants, and (4) assessed quantitative physical variables, such as aerobic fitness, muscle strength, balance, or body composition, as the outcome measures, using validated assessment tools. It turned out that most studies were performed in the context of functional balance and fall prevention and assessed (aspects of) functional balance as the outcome measure. Therefore, we added the criterion that the studies had to use functional balance, or other strong predictors of falling incidents, such as speed of forward- and backward step, as outcome measures. This excluded one study from the analysis that assessed muscle strength (see Table 1).

Study characteristics

All included studies were coded for several intervention characteristics. Intervention duration (in weeks) and frequency (number of sessions per week) were coded from the original publications. Also, it was coded whether the active videogame was especially developed for health-promotion purposes or constituted a commercially available videogame. The latter variable was incorporated because previous research suggests that off-the-shelf (commercial) videogames may be a particular promising tool for health promotion as they tend to be more affordable, accessible, and technologically advanced than videogames developed by researchers for health-promotion purposes.

With regard to characteristics of the methodology used, sample size was recorded, and it was coded whether the control group received some alternative treatment, such as physical activity knowledge games, or no treatment. Additionally, in the case of the studies of children/adolescents, the methodological quality of the included studies was assessed, using the Cochrane Collaborations risk of bias tool. This tool assesses risk of bias in seven categories: Sequence generation, allocation concealment, blinding of participants, personnel and outcome assessors, incomplete outcome data, selective reporting, and other sources of bias. Risk of bias for each category was determined to be low risk (1 point), unclear risk (0 points), or high risk (~1 point), resulting in a total score for each study between 1 and 7. The first two authors rated the studies independently and then compared their assessments; any disagreements were discussed and resolved by consensus. In the case of the elderly studies, Larsen et al. provided an assessment of methodological quality in their systematic review. Therefore, these
scores did not have to be obtained anew but could be calculated from their original article by awarding 1 point to a low risk score, 0 points to an unclear risk score, and −1 point to a high risk score. The dropout rate was recorded as an additional indicator of methodological quality.

Effect size measures

For the studies of children/adolescents, BMI scores (including z-scores, percentiles) were chosen as the outcome measure for the meta-analysis as all five included studies assessed BMI as an outcome measure. For the elderly studies, five out of six studies assessed functional balance as the primary outcome measure. To ensure optimal comparability between studies, we calculated the effect sizes for functional balance from these five studies to be used in the meta-analysis, thereby ignoring other outcome measures such as self-reported “perceived balance” and depression. When several measures of functional balance were available, for instance, scores on the Unipedal test and on the Tinetti test, we aimed to aggregate these effect sizes within the same studies as much as possible. To this aim, we calculated effect sizes for each measure of functional balance and averaged these to arrive at a total effect size. The remaining study assessed speed of forward and backward step as the outcome measure. Because speed of forward and backward step is an important predictor of falling and has been shown to be closely associated with functional balance, an effect size for this outcome measure was calculated and used in the meta-analysis alongside the five effect sizes for functional balance. As with functional balance, an average effect size was calculated in the case of several different measures of the same construct.

Each study provided a comparison between an active videogame and a control condition. This comparison was summarized using the standardized mean difference as the effect size index. Because Cohen’s $d$, an often-used estimate, is slightly biased toward overestimating the standardized mean difference in small samples, we used Hedges’ $g$, which corrects for sample size and yields an unbiased estimate. Conventionally, an effect is considered small when $g = 0.20$ and medium when $g = 0.50$, whereas an effect sizes of $g = 0.80$ is considered large. The data were coded such that higher scores indicate lower BMI and better physical function in the intervention condition versus the control condition. Thus, a positive $g$ indicates a positive intervention effect.

Effect sizes were computed from reported means, standard deviations, and $n$ values for the post-test comparisons. If these were not available, means, standard deviations, and $n$ values for the difference scores were used. Otherwise, means, standard deviations, and $n$ values were requested from the authors. If we did not hear from the authors or received incomplete answers, we calculated $g$ from reported test statistics, preferably from analyses without covariates, such that individual difference factors were put back into the error term.

Data analysis

We synthesized the individual effect sizes using both a fixed-effect model and a random-effects model. A fixed-effect model produces a straightforward summary of what was found in our five and six studies, respectively. Fixed-effect
models can be used when we want to compute the common effect size for the identified population. However, a disadvantage of fixed-effect models is that the result cannot be generalized to other populations. When we do want to generalize to a greater population of studies, random-effects models are generally recommended. However, the estimate of between-studies variance that is used in random-effects models lacks precision for meta-analyses using a small number of studies. In other words, when using random-effects models on small datasets, there is a risk that the model is not applied correctly. Also, a random-effects model makes the assumption that there is a population of studies and that the k studies included in the meta-analysis are a random selection from this hypothetical population. Thus, although generalizing the results to a larger number of populations and interventions is very useful, a random-effects model does require making additional assumptions, and these assumptions are essentially not verifiable. Because both approaches have advantages and disadvantages, we chose to report the results of both fixed-effect and random-effects models.

We calculated the within-class goodness-of-fit statistic $Q$ (which is approximately chi-squared distributed, with $df = k - 1$, where $k$ is the number of effect sizes), which tests for homogeneity in the true effect sizes across studies. A significant $Q$ statistic indicates that moderators can explain the variability in effect sizes across studies. Because the $Q$ statistic has been found to rely greatly on the number of included studies and as a result has limited power in the case of few studies, we also calculated the $I^2$ statistic, which can be interpreted as the proportion of total variability explained by heterogeneity. In the case of large heterogeneity, we performed moderator analyses with the coded intervention and study characteristics. We tested for categorical moderators with the categorical model test, which results in the between-class goodness-of-fit statistic $Q_B$, with $df = j - 1$ (where $j$ is the number of categories or groups). A significant between-groups effect indicates that the variance in effect sizes is at least partially explained by the moderator. We tested for continuous moderators using weighted least square regression of the effect sizes onto the continuous moderator. Significant prediction indicates that the effect sizes vary in a linear manner with the continuous moderator. It should be noted, however, that the statistical power of moderator analyses in meta-analysis is not always high and that a large number of studies is generally needed to detect effects. Given that research about active videogames is still in its infancy, so that not many published studies exist, the results of these analyses should be considered with caution. No moderator analyses were performed in case of limited heterogeneity. No formal test of publication bias (e.g., examining funnel plots) was performed because these tests are generally not recommended when less than 10 effect sizes are available for analysis. Data were analyzed using an Excel spreadsheet and with RevMan software. The spreadsheet is made available through https://osf.io/rjsg9/

### Results

**Active videogames for children and adolescents**

Effect sizes were available for five studies, with a total of 561 participants (Table 2). An analysis using a fixed-effect model revealed a small but significant composite effect size of $g = 0.20$ (95 percent confidence interval, 0.04–0.37). Using a random-effects model to estimate the composite effect size yielded a nonsignificant effect size of $g = 0.20$ (95 percent confidence interval, −0.08 to 0.48). Figure 1 shows a forest plot of the random-effects model. Note that the composite effect size is represented by a diamond, with the width of the diamond indicating the 95 percent confidence interval (i.e., we can be 95 percent certain that our mean effect size falls within this range) and the horizontal line indicating the prediction interval (i.e., we estimate that the true effect in 95 percent of future studies will fall within this range). The $Q$ statistic yielded a nonsignificant effect, $Q(4) = 7.46, P = 0.11$, indicating limited heterogeneity. Because the $Q$ statistic has been shown to have low power in the case of few studies, we also calculated the $I^2$ statistic. This revealed that 46 percent of the variance in effect sizes across studies was attributable to systematic differences between studies, $I^2 = 0.46$, a proportion that can be classified as low to moderate. This result indicates that a larger share of the variance in effect sizes across studies can be attributed to sampling error rather than to moderator variables. We therefore did not perform moderator analyses.

**Active videogames for elderly individuals**

Effect sizes were available for six studies, with a total of 142 participants (Table 3). Using a fixed-effect model, the meta-analysis revealed a medium to large and significant

### Table 2. Sample Sizes, Effect Sizes, and Study Characteristics for the Studies of Children and Adolescents

<table>
<thead>
<tr>
<th>Study</th>
<th>n</th>
<th>g</th>
<th>Duration (weeks)</th>
<th>Frequency (session/week)</th>
<th>Game commercially available</th>
<th>Control condition</th>
<th>Study qualitya</th>
<th>Dropout (percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gao et al.</td>
<td>126</td>
<td>−0.04</td>
<td>36</td>
<td>3</td>
<td>Yes</td>
<td>No treatment</td>
<td>−5</td>
<td>22</td>
</tr>
<tr>
<td>Maddison et al.</td>
<td>322</td>
<td>0.25</td>
<td>24</td>
<td>NA</td>
<td>Yes</td>
<td>No treatment</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>Maloney et al.</td>
<td>58</td>
<td>0.30</td>
<td>10</td>
<td>NA</td>
<td>Yes</td>
<td>No treatment</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Murphy et al.</td>
<td>35</td>
<td>0.88</td>
<td>12</td>
<td>NA</td>
<td>Yes</td>
<td>No treatment</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Ni Mhurchu et al.</td>
<td>20</td>
<td>−0.38</td>
<td>12</td>
<td>NA</td>
<td>Yes</td>
<td>No treatment</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>

aRange, −7 to 7.

NA, not available.
composite effect size of $g = 0.64$ (95 percent confidence interval, 0.29–0.99). A random-effects model showed a medium to large and significant effect size of $g = 0.68$ (95 percent confidence interval, 0.13–1.24). Figure 2 shows a forest plot of the random-effects model. The $Q$ statistic yielded a significant effect, $Q(5) = 12.37$, $P = 0.03$, indicating heterogeneity. Calculation of the $I^2$ statistic revealed that 68 percent of the variance in effect sizes across studies was attributable to systematic differences between studies, a proportion that can be classified as moderate to high.37 This result indicates that a larger share of the variance in effect sizes across studies can be attributed to moderator variables than to sampling error, which calls for an investigation of potential moderators.

Moderator analyses revealed that active videogames that were especially developed for health-promotion purposes did not result in significantly different intervention effects than commercially available videogames: $Q_B(1) = 0.03$, $P = 0.87$. Also, studies in which the control group received some alternative treatment, such as physical activity knowledge games, did not result in significantly different intervention effects than studies in which the control group received no treatment: $Q_B(1) = 0.20$, $P = 0.65$. Nonsignificant effects were also found for sample size [$b = -0.12$, $t(5) = -0.22$, $P = 0.84$], intervention frequency (number of sessions per week) [$b = -0.49$, $t(5) = -0.88$, $P = 0.42$], intervention duration (in weeks) [$b = 0.77$, $t(5) = 1.38$, $P = 0.23$], study quality as assessed by Larsen et al.9 [$b = 0.50$, $t(5) = 0.88$, $P = 0.42$], and dropout [$b = -0.76$, $t(3) = -1.33$, $P = 0.24$].

Discussion

The present research aimed to add to two recent systematic reviews9,16 by calculating effect sizes for the reviewed studies and synthesizing those using meta-analytic procedures. The results showed a small positive effect of active videogames on children’s BMI ($g = 0.20$) and a medium to large positive effect on functional balance in the elderly ($g = 0.68$). This synthesis suggests that active videogames can be helpful in improving BMI among children and functional balance among the elderly.

As active videogames research is still in its infancy, only a limited number of studies could be included in the meta-analyses ($k$ for children/adolescents $= 5$; $k$ for elderly $= 6$). As such, the present research offers a first, preliminary, estimate of active videogames’ effects. When more effectiveness studies become available, this estimate will likely be revised.

As a first estimate of the effects of active videogames, however, we argue that the present meta-analyses are both necessary and informative.

Meta-analyses are necessary to obtain a meaningful estimate of the substance of active videogame effects. Previous empirical studies and systematic reviews have primarily relied on $P$ values.8,9,16 Effect sizes, however, are more appropriate for intervention evaluation.17,18 Although the flaws of null-hypothesis significance testing have been discussed

---

### Table 3. Sample Sizes, Effect Sizes, and Study Characteristics for Studies of the Elderly

<table>
<thead>
<tr>
<th>Study</th>
<th>n</th>
<th>$g$</th>
<th>Duration (weeks)</th>
<th>Frequency (sessions/week)</th>
<th>Game commercially available</th>
<th>Control condition</th>
<th>Study quality$^a$ (percent)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Franco et al.46</td>
<td>21</td>
<td>0.04</td>
<td>3</td>
<td>2</td>
<td>Yes</td>
<td>No treatment</td>
<td>-1</td>
</tr>
<tr>
<td>Pichierri et al.30</td>
<td>15</td>
<td>0.42</td>
<td>12</td>
<td>2</td>
<td>No</td>
<td>Usual care</td>
<td>-3</td>
</tr>
<tr>
<td>Pluchino et al.29</td>
<td>27</td>
<td>-0.05</td>
<td>8</td>
<td>2</td>
<td>Yes</td>
<td>Alternative training program</td>
<td>-1</td>
</tr>
<tr>
<td>Rendon et al.12</td>
<td>34</td>
<td>0.86</td>
<td>6</td>
<td>3</td>
<td>Yes</td>
<td>No treatment</td>
<td>4</td>
</tr>
<tr>
<td>Szturm et al.13</td>
<td>27</td>
<td>1.07</td>
<td>8</td>
<td>2</td>
<td>No</td>
<td>Usual care</td>
<td>3</td>
</tr>
<tr>
<td>Toullette et al.60</td>
<td>18</td>
<td>2.11</td>
<td>20</td>
<td>1</td>
<td>Yes</td>
<td>No treatment</td>
<td>1</td>
</tr>
</tbody>
</table>

$^a$Range, -7 to 7.
for some time,\textsuperscript{44} this discussion has recently increased in intensity. Some authors go so far as to proposing wholly abolishing null-hypothesis significance testing in favor of “the new statistics,” which rely on effect sizes, confidence intervals, and meta-analyses.\textsuperscript{45} In keeping with this trend, this article obtained effect sizes for relevant studies to obtain an estimate of the substance of active videogame effects.

The present meta-analyses are informative in that they are generally supportive of claims that active videogames can be effective health-promoting interventions. Previous systematic reviews have provided comprehensive overviews of published empirical studies, but because these have shown mixed results, they have been unable to draw firm conclusions about active videogames’ effects. Lu et al.,\textsuperscript{16} for instance, do not comment specifically on the overall effectiveness of active videogames. The conclusion of Larsen et al.\textsuperscript{9} is that more research is necessary to study the effects of active videogames for the elderly. Another recent systematic review\textsuperscript{9} concluded that active videogames result in acute energy expenditure, but that it is not clear if active videogames result in longer-term health benefits. The preliminary estimates provided by the present meta-analytic procedures suggest that active videogames do indeed lead to improvements in children’s BMI and in functional balance in the elderly.

It should be noted that the two sets of studies were vastly different in terms of populations and outcome measures. Thus, we can observe that the composite effect for the studies in the elderly was larger than the composite effect for the studies in children, but such a comparison would hardly be meaningful considering the vast differences between the studies. At the same time, active videogames can be used for many other populations and in many other contexts, limiting the extent to which the present findings can be generalized to a “general” effectiveness of active videogames. As such, the present meta-analyses offer preliminary estimations of the effects of active videogames, but only in these two specific health domains.

It should also be noted that the relatively small number of included studies in both meta-analyses induced severe limits on the conclusions that we can derive. For one, when the number of included studies is small, the estimate of the between-studies variance is imprecise, therefore compromising the precision of random-effects models.\textsuperscript{32} In addition, moderator analyses have only limited power with a small set of studies.\textsuperscript{40,41} This latter limitation is particularly noteworthy given the substantial heterogeneity that was found in the elderly studies. Our results revealed that 68 percent of the variance in effect sizes across the elderly studies was attributable to systematic differences. Thus, despite the fact that inclusion criteria resulted in similar populations across studies (healthy over-60 year olds) and similar interventions (active videogames), the estimate of heterogeneity suggested that the extent to which studies could be compared was limited. Perhaps this was due to the use of different assessment procedures for functional balance, for instance, the Berg Balance Scale,\textsuperscript{13} the Tinetti test,\textsuperscript{46} and the speed of backward and forward step.\textsuperscript{30} When selecting studies for a meta-analysis, identical assessment procedures are preferable. However, there is often no general consensus on what is the best way to assess important constructs, and studies tend to differ in the procedures that they use. As a result, it is quite common in meta-analyses to include different assessment procedures (e.g., Conn et al.\textsuperscript{47}). In our case, we argue that it is justified to include studies with different assessment procedures for functional balance. However, it should be noted that the heterogeneity that may have resulted from this limits the extent to which studies could be compared.

In sum, the present results should be seen in light of the limited number of studies in both meta-analyses and the resultant limited possibilities to investigate the encountered heterogeneity. Caution is therefore warranted when interpreting the present results. Nevertheless, we argue that the composite effect sizes found in the present meta-analyses allow for some cautious optimism concerning the potential effects of active videogames for obesity prevention in the young and improvement of functional balance in the old. In future, meta-analytic procedures remain necessary to obtain valid estimates of active videogame effects. With additional empirical studies coming forth, effect size estimates are likely to become more reliable, moderator analyses will have increased power,\textsuperscript{30,42} and it will be possible to test for publication bias.\textsuperscript{42}

\textbf{Author Disclosure Statement}

No competing financial interests exist.

\textbf{References}


53. Maloney AE, Bethea TC, Kelsey KS, et al. A pilot of a video game (DDR) to promote physical activity and...
60. Toulotte C, Toursel C, Olivier N. Wii Fit® training vs. adapted physical activities: Which one is the most appropriate to improve the balance of independent senior subjects? A randomized controlled study. Clin Rehabil 2012; 26:827–835.

Address correspondence to:
Jonathan van ’t Riet, PhD
Department of Communication Science
Radboud University Nijmegen
P.O. Box 9104
6500 HE Nijmegen, The Netherlands
E-mail: j.vantriet@maw.ru.nl