Can virtual reality exposure therapy gains be generalized to real-life? A meta-analysis of studies applying behavioral assessments

Article *in* Behaviour Research and Therapy · November 2015

DOI: 10.1016/j.brat.2015.08.010

CITATIONS
242

READS
3,021

4 authors:

Nexhmedin Morina
University of Münster
120 PUBLICATIONS **3,723** CITATIONS
SEE PROFILE

Hiske Ijntema
University of Amsterdam
2 PUBLICATIONS **242** CITATIONS
SEE PROFILE

Katharina Meyerbröker
Utrecht University
33 PUBLICATIONS **688** CITATIONS
SEE PROFILE

Paul Emmelkamp
University of Amsterdam
614 PUBLICATIONS **21,859** CITATIONS
SEE PROFILE

Post-print:
Morina, N., Ijntema, H., Meyerbröker, K., & Emmelkamp, P.M.G. (2015). Can virtual reality exposure therapy gains be generalized to real life? A meta-analysis of studies applying behavioral assessments. *Behaviour Research and Therapy* 74, 18 – 24. DOI: 10.1016/j.brat.2015.08.010.

Can virtual reality exposure therapy gains be generalized to real-life? A meta-analysis of studies applying behavioral assessments

Nexhmedin Morina [a, b, *], Hiske Ijntema, [a,], Katharina Meyerbröker, [c], Paul M. G. Emmelkamp [d, e]

[a] Department of Clinical Psychology, University of Amsterdam, Amsterdam, The Netherlands

[b] Amsterdam Brain and Cognition Center, University of Amsterdam, Amsterdam, The Netherlands

[c] Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands

[d] Netherlands Institute for Advanced Study, Wassenaar, The Netherlands

[e] Center for Social and Humanities Research, King AbdulAziz University, Jeddah, Saudi Arabia

[*] Nexhmedin Morina, University of Amsterdam, Department of Clinical Psychology; Weesperplein 4, 1018 XA Amsterdam, The Netherlands; Tel: +31(0) 20 525 8607; Fax: +31(0) 20 525 6971; e-mail: n.morina@uva.nl

Can virtual reality exposure therapy gains be generalized to real-life? A meta-analysis of studies applying behavioral assessments

ABSTRACT

In virtual reality exposure therapy (VRET), patients are exposed to virtual environments that resemble feared real-life situations. The aim of the current study was to assess the extent to which VRET gains can be observed in real-life situations. We conducted a meta-analysis of clinical trials applying VRET to specific phobias and measuring treatment outcome by means of behavioral laboratory tests or recordings of behavioral activities in real-life. Data sources were searches of databases (Medline, PsycInfo, and Cochrane). We included in total 14 clinical trials on specific phobias. Results revealed that patients undergoing VRET did significantly better on behavioral assessments following treatment than before treatment, with an aggregated uncontrolled effect size of g = 1.23. Furthermore, patients undergoing VRET performed better on behavioral assessments at post-treatment than patients on wait-list (g = 1.41). Additionally, results of behavioral assessment at post-treatment and at follow-up revealed no significant differences between VRET and exposure in vivo (g = - 0.09 and 0.53, respectively). Finally, behavioral measurement effect sizes were similar to those calculated from self-report measures. The findings demonstrate that VRET can produce significant behavior change in real-life situations and support its application in treating specific phobias.

Keywords: Virtual reality therapy, behavioral assessment, specific phobias, anxiety disorders, meta-analysis

Introduction

In the last two decades, virtual reality exposure therapy (VRET) has been increasingly applied in treating individuals with anxiety disorders, in particular specific phobias. The therapeutic goals in VRET are based on treatment strategies used in behavior therapy while making use of virtual worlds that resemble feared real-life situations. Accordingly, virtual worlds are used to enable systematic exposure to feared stimuli within a contextually relevant situation. The advantage of using VRET rather than exposure in vivo (i.e., carried out in real-life situations) or imaginal exposure (i.e., carried out through imagination) lies in the possibility of controlling the quality, intensity, duration and frequency of exposure (Emmelkamp, 2005). VRET integrates real-time computer graphics, body tracking devices, visual displays and other sensory inputs to immerse individuals in computer-generated virtual environments. As a result, the perception of an interactive, three-dimensional world is constructed. The control of exposure elements might be more manageable than in exposure in vivo or imaginal exposure as the stimuli eliciting anxiety can be more easily modified and manipulated by therapists.

Three published meta-analyses have reviewed the literature on the efficacy of VRET for anxiety disorders. First, in a meta-analysis on the efficacy of VRET as compared to control conditions, Powers and Emmelkamp (2008) included 13 studies (nine on specific phobias, two on social phobia, one on panic disorder, and one on post-traumatic stress disorder). The authors reported an overall controlled effect size (ES) of 1.11 for the efficacy of VRET as assessed with self-report measures. Furthermore, the authors concluded that in vivo interventions were not significantly more effective than VRET. Lastly, the authors reported on two studies on specific phobias that had assessed treatment efficacy on behavioral measures (Garcia-Palacios, Hoffman, Carlin, Furness, & Botella, 2002; Krijn et al., 2004) and those showed a mean effect size of 1.27 as compared to wait-list. Next, Parsons and Rizzo (2008) conducted a meta-analysis with clinical and non-clinical samples coming from 12 trials on specific phobias, four on social phobia, three on agoraphobia, and two on post-traumatic stress disorder (i.e., a total of 21 trials). The authors conducted uncontrolled

effect-sizes only and reported an overall pre- vs. post-treatment ES of 0.95. Parsons and Rizzo did not specifically report on the nature of the assessment instruments used in the included studies. Finally, a recent meta-analysis by Opris et al. (2012) included 23 studies that compared virtual reality interventions for anxiety disorders with an evidence-based intervention. Here too, the majority of the included trials were on specific phobias (n = 12) and the remaining trials were on social phobia (n=5), panic disorder with or without agoraphobia (n=5), and post-traumatic stress disorder (n=1). The authors reported an overall ES of 1.12 for eight studies comparing VRET with wait-list conditions as measured with self-reports. Additionally, they reported no overall difference between VRET and cognitive behavior interventions at post-treatment (k = 15) and at 3-6 months follow-up (k = 7) on self-report measures. Opris et al. (2012) also reviewed the results of studies comparing VRET with cognitive behavior interventions on behavioral assessments. They included eight studies at post-treatment and four studies at follow-up and reported an overall effect size of -0.03 and 0.24 respectively, indicating that there is no evidence for a difference between VRET and cognitive behavior interventions on behavioral assessments. However, it must be noted that the generalization of the findings of this meta-analysis regarding the extent to which VRET gains can be observed in real-life situations is rather limited. First, the authors included both studies with clinical as well as non-clinical samples. Furthermore, they focused on what they labeled “virtual reality exposure enhanced evidence-based interventions” and consequently they included clinical trials that had applied VRET within a larger package of behavioral interventions. For example, in one of the included studies, VRET sessions had consisted of less than half of total amount of the offered sessions (Penate, Pitti, Manuel Bethencourt, de la Fuente, & Gracia, 2008) and one can argue that any measured change in this study might be a result of other parts of treatment than VRET. Additionally, in line with the focus of the meta-analysis by Opris et al on how effective virtual reality enhanced interventions are compared to evidence-based interventions, the authors excluded VRET trials that had not compared VRET to evidence-based interventions, yet still had included observable

behavioral outcomes, such as the study by Garcia-Palacios et al. (2002) that was included in the meta-analysis by Powers and Emmelkamp (2008).

In summary, previous meta-analyses on the efficacy of VRET for anxiety disorders have primarily focused on self-reports of inner states rather than behavioral laboratory tests or behavioral activities in real-life. In general, the increasing reliance on self-reports of inner states and thus the decreasing use of behavioral assessments and observations in psychology has been criticized (Baumeister, Vohs, & Funder, 2007; Furr, 2009). In fact, our knowledge about the efficacy of psychotherapy in general can be enhanced if both self-reports of inner states as well as behavioral tasks and observations of daily life activities are applied to better assess the impact of interventions in changing behavior. This seems particularly relevant with regard to VRET given that this approach makes use of virtual environments to treat psychological complaints rather than real life situations. Accordingly, it is relevant to examine the extent in which VRET has the potential to produce behavioral change that can be observed in daily life. For example, with regard to fear of spiders, self-reports of inner states would rather focus on thoughts, emotions and attitudes associated with fear of spiders. Yet, as the ultimate aim of treatment is to help clients with fear of spiders to better cope when confronted with spiders, we also need to assess how clients are coping with the specific situation in question on the behavioral level.

In sum, previous meta-analyses have shown that VRET is effective in treating anxiety disorders as measured by self-reports. However, as clients undergoing VRET get exposed to feared stimuli within virtual environments, it seems essential that we have a better understanding of the generalization of treatment effects in real-life situations. The aim of this study was to provide an updated and comprehensive systematic review and meta-analysis of the extent in which VRET gains can be transferred to real-life. We aimed at including studies on the efficacy of VRET as measured in behavioral laboratory tests (such as approaching a real spider or speech length in minutes) or behavioral activities in real-life (such as having flown in the past six months). We hypothesized first that VRET would lead to significant uncontrolled changes as measured in behavioral laboratory tests

or behavioral activities in real-life. Second, we hypothesized that VRET would outperform inactive control conditions on these behavioral assessments. Finally, we hypothesized that there would be no significant difference between VRET and in vivo conditions on behavioral laboratory tests or behavioral activities in real-life.

Method

Identification and Selection of Studies

The criteria for including studies into the current meta-analysis were: 1) participants were diagnosed with an anxiety disorder according to DSM or ICD criteria; 2) a virtual reality-based intervention was applied to address core symptoms of the relevant psychological disorder; 3) at least 50% of treatment consisted of virtual reality-based interventions; 4) at least ten patients were treated in the virtual reality condition; and 5) efficacy of treatment was assessed with some sort of behavioral laboratory tests or behavioral activities in real-life. If a publication did not provide enough data to calculate effect-sizes, its authors were contacted by e-mail to retrieve the data. After two contact-attempts, the studies were excluded if the authors did not respond with sufficient data to perform the meta-analysis. No restrictions were made upon publication language.

We searched the following databases: PsycINFO, MEDLINE, and Cochrane. The last search was conducted on March 04, 2014 and included the following search terms: “virtual reality” alone and in combination with “treatment” or “intervention” or “therapy” or “psychotherapy” or “exposure” or “trial”. The search string yielded 335 hits. When duplicate publications were removed, 165 publications remained. An examination of the abstracts led to the exclusion of 105 studies that were evaluated as irrelevant. The evaluation of the full text of the remaining 60 publications led to the exclusion of 45 publications. Of the remaining 15 publications, 14 were conducted with patients with specific phobia and with patients with social anxiety disorder (P. L. Anderson et al., 2013). Given this outcome, we decided to focus on trials on specific phobias only. Fig. 1 presents a flow diagram of the study selection process. As shown in Fig. 1, publications were excluded for a variety

of reasons, including three publications that did not report enough data to be included in our meta-analysis (P. Anderson et al., 2006; Bouchard et al., 2011; Kahan, Tanzer, Darvin, & Borer, 2000).

Quality Assessment

All studies were rated with a methodology rating form for psychotherapy outcome studies developed by Öst (2008). This rating form consists of 22 items that are rated as 0 (poor), 1 (fair), or 2 (good). Examples of the scale include *'representativeness of the sample*', *'reliability of the diagnosis', 'number of therapists', or 'statistical analysis and presentation of results*'. Two raters (the first and third author) independently rated all studies. The Intraclass Correlation Coefficient (ICC) of the total score for all studies combined was 0.90, 95% CI [0.88, 0.92], indicating good inter-rater reliability.

Coding of treatment characteristics and effect size calculation

First, all conditions were coded as either virtual reality-based or control conditions. Then, control conditions were further specified as active or inactive. Active treatments were subdivided into cognitive behavior therapy or other interventions. Inactive conditions included wait-list (n = 8) and attention-placebo conditions (n = 1). Based on the aim of the study, behavioral assessment outcomes were collected for all available measurement points.

We first computed uncontrolled effect sizes (e.g. change from pre- to post-treatment; change from pre-treatment to follow-up) for virtual reality-based conditions. Second, controlled effect sizes were computed for all trials that compared virtual reality-based conditions with active or inactive treatment groups. Finally, we directly compared behavioral measurement effect sizes with self-report measurement effect sizes as assessed in the same clinical trials. Within- and between-group effect sizes were computed using Hedge's *g* that provides a better estimate of sample sizes based on small samples as compared to Cohen's *d* (Field & Gillett, 2010). This was obtained by first subtracting the post-treatment mean or follow-up mean from the pre-treatment mean (uncontrolled effect size) or the control group mean from the treatment group mean at post-treatment (controlled effect size)

respectively and dividing the outcome by the pooled standard deviation. The outcome was then multiplied by a sample size correction factor $J = 1-(3/(4df - 1))$ to obtain the effect size Hedges's *g* (Lipsey & Wilson, 2001). Effect size *g* can conservatively be interpreted using suggestions by Cohen (1988), with 0.2 indicating a small, 0.5 a medium, and 0.8 a large effect, respectively. Furthermore, we used a random effects model to calculate effect sizes given the heterogeneity of the studies (Field & Gillett, 2010). Only one of the 16 publications reported intent-to-treat analyses related to behavioral assessment scores (P. L. Anderson et al., 2013). Accordingly, all analyses were calculated on data based on treatment completers. Finally, 15 trials were conducted with adult participants and one with children and adolescents (age 8 – 15 years, with a mean age of 10 years) (St-Jacques, Bouchard, & Belanger, 2010). The mean age in the 13 trials with adults ranged from 29 to 51 years (with a mean age of 39 years). The publication by St-Jacques et al. (2010) was included in both uncontrolled (pre- vs. post-treatment only) and controlled effect (VRET vs. active conditions at post-treatment only) sizes. Given the age difference, these analyses were first conducted with all studies included and then repeated without the study by St-Jacques et al. All analyses were completed with comprehensive meta-analysis (CMA; version 3) (Borenstein, Hedges, Higgins, & Rothstein, 2009).

Results

Description of Studies

Table 1 presents a summary of the 14 included studies on specific phobias. The control conditions utilized in the included studies were cognitive-behavior therapy (including exposure-only, eight comparisons), a waitlist condition (six comparisons), and a psychological placebo control intervention (one comparison). Finally, three studies applied an uncontrolled design and data from these studies were used for computing uncontrolled effect sizes only. See Table 1 for information on behavioral assessment applied in the single studies as well as number of participants and sessions.

Table 2 provides an overview of the methodological characteristics across the included studies as measured by the methodology rating form for psychotherapy outcome studies developed

by Öst (2008). As stated above, items are rated as 0 (poor), 1 (fair), or 2 (good). The ratings yielded a mean score of 0.95 (SD = 0.49), with 50% of the evaluated characteristics of the included studies having a score between poor and fair. A meta-regression revealed no significant relation between study quality ratings and pre-post effect sizes for behavioral assessments (Q = 1.11, df = 1, p = .292).

Uncontrolled effect sizes

We first computed effect sizes for the impact of virtual reality-based interventions on behavioral assessment scores pre- to post-treatment. Out of 14 publications, 11 virtual reality-based interventions provided pre- and post-treatment behavioral assessment data. Across all virtual reality-based treatments, a large pre-post effect size was found, g = 1.23 95% CI = [1.00;1.46] (see also Fig. 2 for a forest plot including all 11 studies). The exclusion of the one trial with children and adolescents (St-Jacques et al., 2010) led to similar results, g = 1.28; 95% CI = [1.04; 1.53]. Four studies reported behavioral changes from pre-treatment to follow-up in virtual reality-based interventions. Results showed large pre-follow-up-effect sizes, g = 1.63, 95% CI = [0.68; 2.84].

Controlled effect sizes

Trials comparing a virtual reality-based condition to at least one control group were used to compute controlled (i.e. between-group) effect sizes at post-treatment. Four studies compared a virtual reality-based intervention to a wait-list (i.e., inactive control) group at post-treatment. The results of this comparison produced a large mean overall between-group effect size in favor of VRET, g = 1.41; 95% CI = [0.82; 1.99]. Eight studies compared a virtual reality-based intervention to an active control group at post-treatment. There was no significant difference between the two groups, g = -0.13; 95% CI = [-0.43; 0.17] (see also Fig. 3 for a forest plot). The exclusion of the trial with children and adolescents (St-Jacques et al., 2010) produced similar results, g = -0.04; 95% CI = [-0.31; 0.25]. Five studies compared a virtual reality-based intervention to an active control group at

follow-up. Results revealed no significant difference between the two groups, g = 0.44; 95% CI = [-0.12; 1.00].

As six of the comparison conditions were behavioral interventions, we also compared virtual reality-based treatment to behavioral interventions. This comparison also produced non-significant results at post-treatment, g = -0.09; 95% CI = [-0.40; 0.23]. Finally, a comparison between virtual reality-based interventions and behavioral interventions at follow-up (k = 4) resulted also in a non-significant difference between the two, g = 0.53; 95% CI = [-0.24; 1.35].

Heterogeneity

Heterogeneity was not significant neither for within-group effect sizes (I^2 = 23.92; Q = 13.14, df = 10, p = .22) nor between-group effect sizes (I^2 = 20.17; Q = 8.77, df = 7, p = .27).

Publication Bias

To account for the possibility that our meta-analysis may be missing unpublished non-significant results and therefore could overestimate the overall effect size, a fail-safe N was computed. This is a conservative method to address this potential problem, which assumes that the effect sizes of unpublished studies are equal to zero and then computes the number of studies that would be required to reduce the overall effect size of the analysis to a non-significant level (Rosenthal & Rubin, 1988). We computed the "fail-safe N" (Rosenthal, 1991) for addressing the publication bias. The following equation is suggested for computing a fail-safe N:

$$X = \frac{K(K\overline{Z}^2 - 2.706)}{2.706}$$

where K is the number of studies in the analysis and $\overline{Z}$ is the mean Z obtained from the K studies. A common practice for determining whether the findings may be considered robust is to determine whether the required number of studies to reduce the analysis to non-significance is greater than 5K + 10 (Rosenthal, 1991). In this study, the required number of studies would be 80. An analysis of

publication bias revealed a fail-safe N of 420, indicating that it would require more than 420 current or future unpublished studies with an effect size of zero to reduce the effect size of the current analysis to non-significant. This suggests that the findings of the current study are robust.

Additional Analyses

Mean age of participants in the included trials was further analyzed as a potential moderator of treatment efficacy regarding within-group effect sizes (for between-group effect sizes, the number of trials was considered too small to be subdivided into two age categories). See Table 1 for information on mean age in the single trials. Among the trials with adult participants, a comparison of trials with a mean age of less than 38 years (g= 1.50, 95% CI= [1.18; 1.82], k = 5) and those with a mean age of higher than 38 years (g = 1.10, 95% CI = [0.74; 1.46], k = 5) did not produce any significant result at post-treatment, Qm (1) = 2.71, p = 0.10.

Additionally, we compared behavioral measurement effect sizes with self-report measurement effect sizes as assessed in the same clinical trials. Table 1 shows the self-report measurements applied in each study. As can be seen in Table 3, the effect sizes calculated from self-report scores were similar to effect sizes calculated from behavioral measurements. The only difference was found with regards to the uncontrolled effect size from pre- to post-treatment, with self-report scores producing a somewhat larger effect size than behavioral assessment scores (see Table 3).

Discussion

In this meta-analysis, we evaluated the efficacy of VRET for specific phobias as assessed by means of behavioral laboratory tests or behavioral activities in real-life. Results of the included publications indicated that patients undergoing VRET score significantly better on behavioral assessments at post-treatment and at follow-up as compared to pre-treatment. Furthermore, patients treated by VRET reported significantly better behavioral assessment scores than patients in inactive

conditions. Additionally, results yielded no significant difference on behavioral assessment scores between VRET and behavior therapy. Finally, effect sizes resulting from behavioral assessment scores were similar to those calculated from self-report measures. Taken together, our results suggest that VRET can effectively enable behavior change in real-life situations.

Self-report measurements have been criticized as they might introduce systematic measurement bias into data (Baumeister et al., 2007). Yet, our results yielded that effect sizes calculated from assessment of behavioral laboratory tests or behavioral activities in real-life are similar to effect sizes based on self-report measures. Furthermore, our findings are in line with previous meta-analyses on the efficacy of VRET for anxiety disorders (Opris et al., 2012; Parsons & Rizzo, 2008; Powers & Emmelkamp, 2008). As previous meta-analyses relied mainly on non-behavioral assessments, our meta-analysis with a specific focus on behavior changes as assessed in behavioral tasks or activities in real-life adds further support for the efficacy of VRET in treating specific phobias. With regard to other disorders than specific phobias, future research needs to apply behavioral assessments in real-life as a means of measuring the efficacy of VRET.

The heterogeneity test revealed that the variation in effect sizes (i.e., VRET compared to control conditions) was compatible with chance alone as heterogeneity was not significant. This indicates that the evidence is consistent across the included trials. Additional analyses revealed that mean age of participants was not associated with effect-sizes. However, this needs to further be examined in future research.

The finding that VRET can produce similar effect sizes as other behavioral interventions in the treatment of specific phobias indicates that VRET should be applied if it is justified by potential advantages as compared to evidence-based treatments. Such advantages might be related to the content of the intervention in question, such as intensity, quality, duration or frequency of exposure. Potential advantages might also be related to relevant issues surrounding the interventions, e.g., if VRET is perceived as more tolerable by patients than for example exposure in vivo or its use might be cheaper than forms of psychotherapy.

Several limitations of the current meta-analysis must be acknowledged. First, the number of studies to meet the inclusion and exclusion criteria was rather limited. Second, the used rating scale of the methodology of the included clinical trials (Ost, 2008) revealed that several relevant characteristics of the trials can be rated as methodologically relatively weak (e.g., assessor and therapist training or checks for therapist competence). It must be noted, however, that the methodology of the included studies was rated based on the information provided in the article, which may or may not be consistent with the way the study was conducted. For example, in a particular study an assessor training might have taken place, yet, if the authors had not reported that information in the publication then that particular item is rated as poor. A further limitation of this meta-analysis is that the number of participants in most studies was rather low and the controlled studies were not adequately powered as non-inferiority trials with regard to the comparison between VRET and other active conditions (i.e, behavior therapy).

In line with these limitations, our review offers several recommendations for future research. To draw any strong conclusions about the efficacy of VRET on psychological disorders in general, we need more methodologically strong randomized controlled trials with adequate power. Additionally, clinical trials on other disorders than specific phobias are needed. Furthermore, there is lack of research on assessing potential mechanisms underlying therapeutic change within VRET. One suggested factor in this regard has been sense of presence (i.e., the extent to which virtual reality worlds feel realistic to participants) and its relation to perceived anxiety during VRET. However, a recent meta-analysis revealed that the association between sense of presence and perceived anxiety depends on the disorder. Whereas large correlations were found in virtual reality trials involving fear of animals, there was no significant association between sense of presence and perceived anxiety in individuals with social anxiety (Ling, Nefs, Morina, Heynderickx, & Brinkman, 2014). The authors further reported that there is lack of research on the validity of existing presence scales when it comes to accurately measuring essential aspects of sense of presence in VRET (Ling et al., 2014). Additionally, only one clinical trial conducted with children and adolescents could be included in our

meta-analysis. Consequently, more research is needed to assess the extent to which VRET gains can be generalized to real-life situations among children and adolescents. Finally, research on effectiveness or clinical utility of virtual reality interventions is needed to assess feasibility and cost-effectiveness of virtual reality interventions being delivered in a local setting.

In conclusion, the current findings support the efficacy of VRET for specific phobias. Results gained by VRET seem to significantly affect positive change in real-life. More research is needed to make reliable statements about other disorders than specific phobias and to address mechanisms underlying behavior change in VRET.

References

* = studies included in the meta-analysis are indicated by an asterisk.

Anderson, P., Jacobs, C. H., Lindner, G. K., Edwards, S., Zimand, E., Hodges, L., & Rothbaum, B. O. (2006). Cognitive behavior therapy for fear of flying: Sustainability of treatment gains after september 11. *Behavior Therapy, 37*(1), 91-97.

Anderson, P. L., Price, M., Edwards, S. M., Obasaju, M. A., Schmertz, S. K., Zimand, E., & Calamaras, M. R. (2013). Virtual reality exposure therapy for social anxiety disorder: A randomized controlled trial. *Journal of Consulting and Clinical Psychology, 81*(5), 751-760. doi:10.1037/a0033559

Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science, 2*(4), 396-403. doi:10.1111/j.1745-6916.2007.00051.x

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Software, in introduction to meta-analysis*. Chichester, UK: John Wiley & Sons. doi:10.1002/9780470743386.ch44

* Bouchard, S., Cote, S., St-Jacques, J., Robillard, G., & Renaud, P. (2006). Effectiveness of virtual reality exposure in the treatment of arachnophobia using 3D games. *Technology & Health Care, 14*(1), 19-27.

Bouchard, S., Dumoulin, S., Robillard, G., Guitard, T., Klinger, É, Forget, H., & Roucaut, F. (2011). A randomized control trial for the use of in virtuo exposure in the treatment of social phobia: Final results. *Journal of Cybertherapy and Rehabilitation, 4*(2), 197-199.

* Coelho, C. M., Silva, C. F., Santos, J. A., Tichon, J., & Wallis, G. (2008). Contrasting the effectiveness and efficiency of virtual reality and real environments in the treatment of acrophobia. *PsychNology Journal, 6*(2), 203-216.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Erlbaum: Hillsdale.

* Cote, S., & Bouchard, S. (2005). Documenting the efficacy of virtual reality exposure with psychophysiological and information processing measures. *Applied Psychophysiology & Biofeedback, 30*(3), 217-232.

Emmelkamp, P. M. (2005). Technological innovations in clinical assessment and psychotherapy. *Psychotherapy and Psychosomatics, 74*(6), 336-43. doi:10.1159/000087780

* Emmelkamp, P. M. G., Krijn, M., Hulsbosch, A. M., de Vries, S., Schuemie, M. J., & van der Mast, C. A. P. G. (2002). Virtual reality treatment versus exposure in vivo: A comparative evaluation in acrophobia. *Behaviour Research and Therapy, 40*(5), 509-516.

Field, A. P., & Gillett, R. (2010). How to do a meta-analysis. *British Journal of Mathematical & Statistical Psychology, 63*(3), 665-694. doi:10.1348/000711010X502733

Furr, R. M. (2009). Personality psychology as a truly behavioural science. *European Journal of Personality, 23*(5), 369-401. doi:10.1002/per.724

* Garcia-Palacios, A., Hoffman, H., Carlin, A., Furness, T. A. I. I. I., & Botella, C. (2002). Virtual reality in the treatment of spider phobia: A controlled study. *Behaviour Research and Therapy, 40*(9), 983-993.

Kahan, M., Tanzer, J., Darvin, D., & Borer, F. (2000). Virtual reality-assisted cognitive-behavioral treatment for fear of flying: Acute treatment and follow-up. *CyberPsychology & Behavior, 3*, 387-392.

* Krijn, M., Emmelkamp, P., Biemond, R., de Ligny, C., Schuemie, M., & van der Mast, C. (2004). Treatment of acrophobia in virtual reality: The role of immersion and presence. *Behaviour Research and Therapy, 42*(2), 229-239. doi:10.1016/S0005-7967(03)00139-6

Ling, Y., Nefs, H. T., Morina, N., Heynderickx, I., & Brinkman, W. (2014). A meta-analysis on the relationship between self-reported presence and anxiety in virtual reality exposure therapy for anxiety disorders. *Plos One, 9*(5), e96144. doi:10.1371/journal.pone.0096144

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

* Maltby, N., Kirsch, I., Mayers, M., & Allen, G. J. (2002). Virtual reality exposure therapy for the treatment of fear of flying: A controlled investigation. *Journal of Consulting and Clinical Psychology, 70*(5), 1112-1118.

* Michaliszyn, D., Marchand, A., Bouchard, S., Martel, M., & PoirierBisson, J. (2010). A randomized, controlled clinical trial of in virtuo and in vivo exposure for spider phobia. *Cyberpsychology, Behavior, and Social Networking, 13*(6), 689-695.

* Muhlberger, A., Weik, A., Pauli, P., & Wiedemann, G. (2006). One-session virtual reality exposure treatment for fear of flying: 1-year follow-up and graduation flight accompaniment effects. *Psychotherapy Research, 16*(1), 26-40.

* Muhlberger, A., Wiedemann, G., & Pauli, P. (2003). Efficacy of a one-session virtual reality exposure treatment for fear of flying. *Psychotherapy Research, 13*(3), 323-336.

Opris, D., Pintea, S., Garcia-Palacios, A., Botella, C., Szamoskoezi, S., & David, D. (2012). Virtual reality exposure therapy in anxiety disorders: A quantitative meta-analysis. *Depression and Anxiety, 29*(2), 85-93. doi:10.1002/da.20910

Ost, L. (2008). Efficacy of the third wave of behavioral therapies: A systematic review and meta-analysis. *Behaviour Research and Therapy, 46*(3), 296-321. doi:10.1016/j.brat.2007.12.005

Parsons, T. D., & Rizzo, A. A. (2008). Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry, 39*(3), 250-261. doi:10.1016/j.jbtep.2007.07.007

Penate, W., Pitti, C. T., Manuel Bethencourt, J., de la Fuente, J., & Gracia, R. (2008). The effects of a treatment based on the use of virtual reality exposure and cognitive-behavioral therapy applied to patients with agoraphobia. *International Journal of Clinical and Health Psychology, 8*(1), 5-22.

Powers, M. B., & Emmelkamp, P. M. G. (2008). Virtual reality exposure therapy for anxiety disorders: A meta-analysis. *Journal of Anxiety Disorders, 22*(3), 561-569. doi:10.1016/j.janxdis.2007.04.006

Rosenthal, R. (1991). *Meta-analytic procedures for social research*. London: Sage Publications.

Rosenthal, R., & Rubin, D. B. (1988). Comment: Assumptions and procedures in the file drawer problem. *Statistical Science, 3*(1), 120-125.

* Rothbaum, B. O., Anderson, P., Zimand, E., Hodges, L., Lang, D., & Wilson, J. (2006). Virtual reality exposure therapy and standard (in vivo) exposure therapy in the treatment of fear of flying. *Behavior Therapy, 37*(1), 80-90.

* Rothbaum, B. O., Hodges, L., Anderson, P. L., Price, L., & Smith, S. (2002). Twelve-month follow-up of virtual reality and standard exposure therapies for the fear of flying. *Journal of Consulting and Clinical Psychology, 70*(2), 428-432.

* St-Jacques, J., Bouchard, S., & Belanger, C. (2010). Is virtual reality effective to motivate and raise interest in phobic children toward therapy? A clinical trial study of in vivo with in virtuo versus in vivo only treatment exposure. *Journal of Clinical Psychiatry, 71*(7), 924-931. doi:10.4088/JCP.08m04822blu

* Wiederhold, B. K., & Wiederhold, M. D. (2003). Three-year follow-up for virtual reality exposure for fear of flying. *CyberPsychology & Behavior, 6*(4), 441-445.

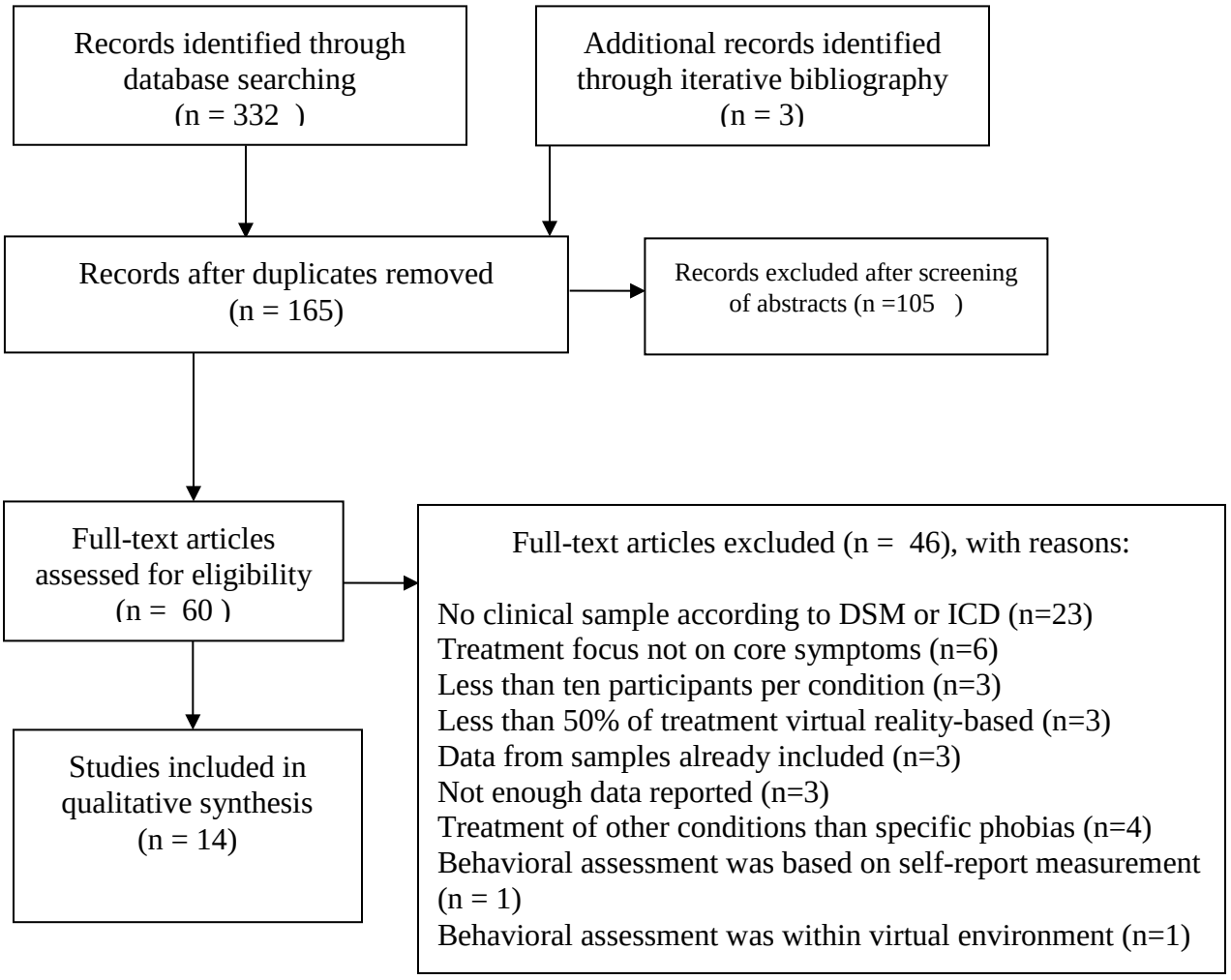

Fig. 1. Flow diagram of study selection process

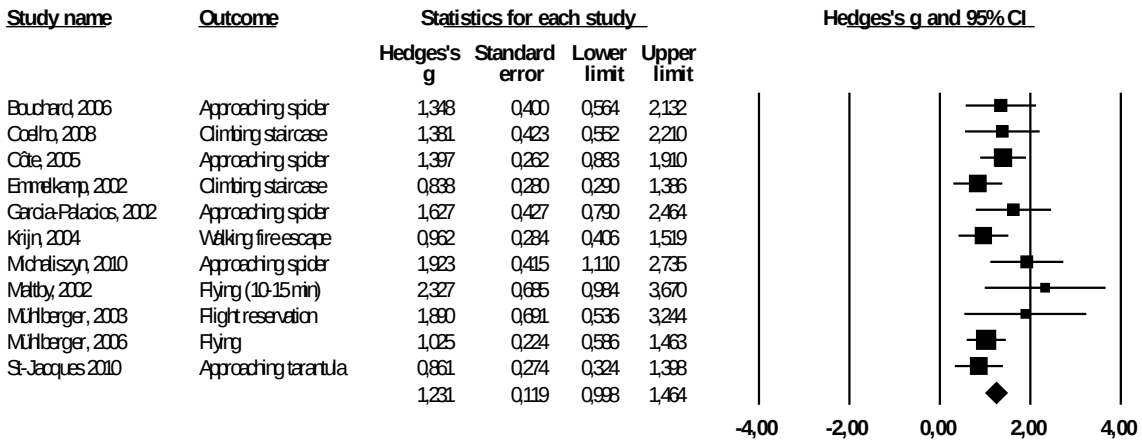

Fig. 2. Uncontrolled effect size estimates (pre- vs. post-treatment) for the efficacy of VRET on behavioral assessments

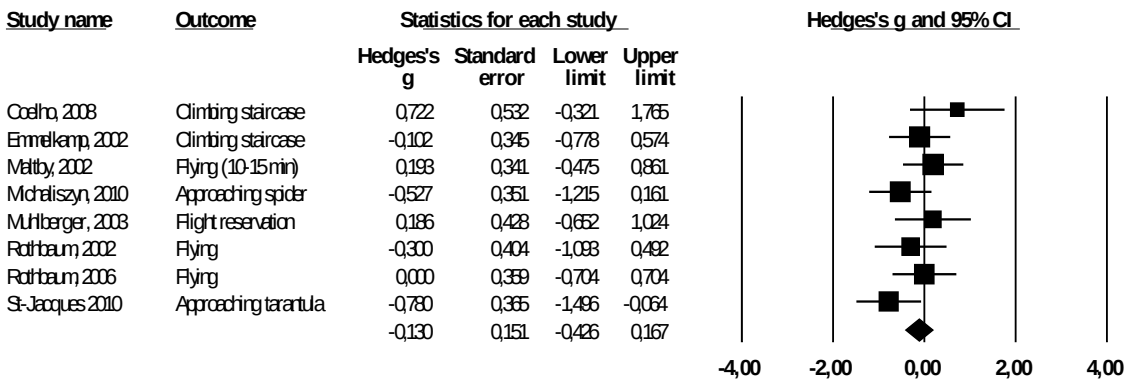

Fig. 3. Controlled effect size estimates for the efficacy of VRET compared to active control conditions on behavioral assessments

Table 1

Study characteristics

Study	Disorder	*n* VRT	Number sessions	Measures: 1: Behavioral 2: Self-report	Mean age	Control	*n* control
Bouchard et al., 2006	Arachnophobia	11	4	1: Approaching spider 2: Spider Beliefs Questionnaire	31	None	n.a.
Coelho et al., 2008	Acrophobia	10	3	1: Climbing staircase 2: Acrophobia Questionnaire	37	Exposure in vivo	5
Côte & Bouchard, 2005	Arachnophobia	28	5	1: Approaching spider 2: Spider Beliefs Questionnaire	34	None	n.a.
Emmelkamp et al., 2002	Acrophobia	17	3	1: Climbing staircase 2: Acrophobia Questionnaire	43	Exposure in vivo	16
Garcia-Palacios et al., 2002	Arachnophobia	12	4	1: Approaching spider 2: Fear of Spiders Questionnaire	29	Waitlist	11
Krijn et al., 2004	Acrophobia	17	3	1: Walking fire escape 2: Acrophobia Questionnaire	51	Waitlist	11
Maltby et al, 2002	Acrophobia	20	5	1: Flying 2: Flight Anxiety Situations Questionnaire	45	Attention-placebo group therapy	23
Michaliszyn et al., 2010	Arachnophobia	16	8	1: Approaching spider 2: Fear of Spiders Questionnaire	29	Exposure in vivo	16
Muhlberger et al., 2003	Acrophobia	26	1	1: Flight reservation 2: Fear of Flying Scale	41	Cognitive Therapy# Waitlist*	11 10
Muhlberger et al., 2006	Acrophobia	30	1	1: Flying 2: Fear of Flying Scale	45	None	n.a.
Rothbaum et al., 2002	Acrophobia	15	8	1: Flying 2: Fear of Flying Inventory	41	CBT Waitlist	15 15
Rothbaum et al., 2006	Acrophobia	29	8	1: Flying 2: Fear of Flying Inventory	39	CBT Waitlist	29 25
St-Jacques et al., 2010	Specific Phobia (Arachnophobia)	14	4	1: Approaching tarantula 2: Spider Phobia Beliefs Questionnaire	10	Exposure in vivo	17
Wiederhold& Wiederhold, 2003	Acrophobia	20	8	1: Flying without alcohol/medication 2: Fear of Flying Inventory	40	Imaginal exposure *	10

Note. * = assessed at follow-up only; # = Cognitive Therapy without exposure elements; CBT= Cognitive-Behavior Therapy

Table 2

Means (with standard deviations in parentheses) for the methodology rating scale

Variable	Mean (SD)
1. Clarity of sample description	1.57 (0.51)
2. Severity/chronicity of the disorder	1.57 (0.65)
3. Sample representativeness	1.36 (0.50)
4. Reliability of the diagnosis	1.00 (0.78)
5. Specificity of outcome measures	2.00 (0.00)
6. Reliability and validity of outcome measures	2.00 (0.00)
7. Use of blind evaluators	0.36 (0.74)
8. Assessor training	0.21 (0.58)
9. Assignment to treatment (only controlled trials)	0.79 (0.43)*
10. Design	1.29 (0.73)
11. Power analysis	0.00 (0.00)
12. Assessment points	0.93 (0.73)
13. Manualized treatment programs	1.50 (0.52)
14. Number of therapists	0.36 (0.50)
15. Therapist training	0.50 (0.52)
16. Checks for treatment adherence	0.14 (0.53)
17. Checks for therapist competence	0.14 (0.36)
18. Control of concomitant treatments	0.64 (0.84)
19. Handling of attrition	1.07 (0.92)
20. Statistical analyses and presentation of results	1.79 (0.43)
21. Clinical significance	0.29 (0.47)
22. Equality of therapist hours (only active control)	2.00 (0.00)**

Note. * Based on 15 studies; ** Based on 12 studies.

Table 3

Effect sizes resulting from behavioral measures vs. those from self-report measures

Uncontrolled Effect Sizes					***Subgroup Analysis***	
	k	***g***	***SE***	***95% CI***	***Q (df = 1)***	***p***
Pre vs. Post					2.26	.13
Behavioral measures	11	1.23	0.12	[1.00; 1.46]		
Self-report measures	11	1.60	0.22	[1.18; 2.03]		
Pre vs. Follow-up					0.23	.63
Behavioral measures	4	1.63	0.62	[0.43; 2.84]		
Self-report measures	4	1.32	0.22	[0.89; 1.75]		

Controlled Effect Sizes					***Subgroup Analysis***	
	k	***g***	***SE***	***95% CI***	***Q (df = 1)***	***p***
VRET vs wait-list at post-treatment					1.35	.25
Behavioral measures	4	1.41	0.30	[0.82; 1.99]		
Self-report measures	4	0.88	0.35	[0.20; 1.55]		
VRET vs active control at post-treatment					0.97	.33
Behavioral measures	8	-0.13	0.15	[-0.43; 0.17]		
Self-report measures	8	0.15	0.25	[-0.33; 0.63]		
VRET vs active control at follow-up*					0.00	.99
Behavioral measures	4	0.20	0.21	[-0.21; 0.62]		
Self-report measures	4	0.20	0.24	[-0.27; 0.67]		

Note: Wiederhold & Wiederhold (2003) was not included in the analysis due to lack of self-report scores at follow-up.