Journal of Anxiety Disorders 28 (2014) 625–631

Contents lists available at ScienceDirect

Journal of Anxiety Disorders

Review

Virtual reality exposure therapy for the treatment of anxiety disorders: An evaluation of research quality

Russell A. McCann, Christina M. Armstrong, Nancy A. Skopp, Amanda Edwards-Stewart, Derek J. Smolenski, Jennifer D. June, Melinda Metzger-Abamukong, Greg M. Reger *

National Center for Telehealth and Technology, Madigan Army Medical Center, Attn: T2 (OMAMC, Bldg 9933), 9040 A Jackson Avenue, Tacoma, WA 98431-1100, United States

ARTICLE INFO

Article history:
Received 7 October 2013
Received in revised form 21 May 2014
Accepted 23 May 2014
Available online 7 June 2014

Keywords:
Randomized controlled trials
Effectiveness
Adherence

ABSTRACT

Randomized controlled trials (RCTs) support the effectiveness of virtual reality exposure therapy (VRET) for anxiety disorders; however, the overall quality of the VRET RCT literature base has yet to be evaluated. This study reviewed 27 VRET RCTs and the degree of adherence to 8 RCT research design criteria derived from existing standards. Adherence to the study quality criteria was generally low as the articles met an average 2.85 criteria (SD = 1.56). None of the studies met more than six quality criteria. Study quality did not predict effect size; however, a reduction in effect size magnitude was observed for studies with larger sample sizes when comparing VRET to non-active control groups. VRET may be an effective method of treatment but caution should be exercised in interpreting the existing body of literature supporting VRET relative to existing standards of care. The need for well-designed VRET research is discussed.

Contents

1. Introduction

With a lifetime prevalence rate for anxiety disorders at 28.8% (Kessler et al., 2005), research demonstrating efficacious therapeutic interventions for such disorders has the potential to benefit a large population. Cognitive behavioral therapy (CBT) is considered one of the most effective treatments for anxiety disorders (Arch & Craske, 2009; Norton & Price, 2007). Exposure therapy (ET) is an effective CBT component for the treatment of many anxiety disorders including posttraumatic stress disorder (PTSD; (Institute of Medicine of the National Academies, 2007; Rothbaum & Schwartz, 2002), panic disorder (Marks et al., 1993), generalized anxiety disorder (Stanley et al., 2009), obsessive compulsive disorder (Foa et al., 2005), and specific phobias (Davidson et al., 2004). The ET is accomplished through in vivo and imaginal exposure, which involves the confrontation of feared but objectively safe stimuli, situations, or memories. The use of multi-sensory virtual reality (VR) has been proposed as a cost-effective and logistically convenient clinical tool for ET, relative to traditional in vivo exposure procedures (Rothbaum et al., 2006). It has also been proposed as an exposure technique for those who

* Corresponding author at: VA Puget Sound, 9600 Veterans Drive, Tacoma, WA 98493-1100, United States. Tel.: +1 253 534 5579; fax: +1 253 968 4192.
E-mail address: Greg.reger2@gmail.com (G.M. Reger).

http://dx.doi.org/10.1016/j.janxdis.2014.05.010
0887-6185/

may fail to effectively activate fear networks (Difede & Hoffman, 2002), which is deemed necessary to achieve a therapeutic effect (Jaycox, Foa, & Morral, 1998).

The VR incorporates computer graphics, visual displays and sensory inputs to create an immersive virtual environment that facilitates the psychological sense of participating in the computer world. Given that VR permits the creation of customized virtual environments, this modality lends itself well to ET. Prior research has studied the use of VR to treat a range of anxiety disorders to include fear of flying, social phobia, panic disorder, and PTSD (Choi et al., 2005; Difede et al., 2007; Klinger et al., 2005; Maltby, Kirsch, Mayers, & Allen, 2002; Reger et al., 2011).

Three meta-analyses conducted on VRET for anxiety disorders have concluded that VRET is superior to waitlist control and no difference relative to active treatments (Opriş et al., 2012; Parsons & Rizzo, 2008; Powers & Emmelkamp, 2008). Similarly, Meyerbröker and Emmelkamp (2010) concluded in a narrative review that VRET is a promising treatment for anxiety disorders; however, the authors noted that the literature base for this treatment would benefit from studies with stronger methodologies. Additional concerns have been raised about the quality of the current VRET literature due to the use of small sample sizes, and a lack of breadth and uniformity in the reporting of data (Parsons & Rizzo, 2008). While concerns about the quality of VRET studies have been raised, the quality of this literature has yet to be assessed in a systematic way.

A previous study reviewed the literature on psychotherapy for depression (Cuijpers, van Straten, Bohlmeijer, Hollon, & Andersson, 2010) and found that the studies rated as being of high quality reported smaller treatment effect sizes compared to low quality studies. The authors concluded that while the effects of psychotherapy for depression remain significant, meta-analyses have over-estimated the effects of this intervention. Cuijpers et al. posited that this over-estimation is largely due to the "inadequately rigorous methods" found in the literature.

The primary goal of this study is to systemically evaluate the quality of the VRET literature, quantify the extent to which quality research design characteristics were present, and to examine whether or not study quality relates to treatment effect size. Toward this end, the eight criteria laid out by Cuijpers et al. (2010) were applied to randomized-controlled trials (RCTs) conducted to evaluate VRET for the treatment of anxiety disorders. An additional goal of this study was to assess for a change in VRET RCT study quality and treatment effect size over time. It is possible that as VRET has become a more established treatment over the last two decades, the quality and effect size values associated with the VRET RCTs have increased and decreased, respectively. Finally, an analysis was conducted to assess for a relationship between sample size and effect size. It was hypothesized that there would be a negative relationship between study quality and effect size, that study quality of VRET RCTs would improve over time, and that there would be a negative relationship between sample size and effect size.

2. Method and materials

Inclusion criteria for reviewed articles were: a randomized total sample size equal to or greater than ten, at least two different comparison groups with an active or inactive control condition and at least one VR condition, report of interval or ratio data, use of an anxiety outcome measure, and written in English. Studies were excluded if a non-clinical population was employed. Databases searched included PsycINFO, PubMed, MEDLINE, Academic Complete, Cochrane, and EMBASE. Keywords used to search were: "virtual reality" and "treatment"; "specific phobia"; "generalized anxiety disorder"; "obsessive compulsive disorder"; "anxiety"; "posttraumatic stress disorder"; "claustrophobia"; "driving"; "flying"; "aviophobia"; "panic"; "acrophobia"; "agoraphobia"; "social phobia"; "spiders"; "arachnophobia"; "public speaking"; "heights"; and "insects." "Virtual reality" was individually paired with each of the above terms to encompass as many articles as possible. Articles were also identified for inclusion by way of review of VR article reference sections.

2.1. Procedure

Twenty-seven articles met inclusion criteria and were coded independently by two psychologists. The two coders were blind to each other's ratings. Articles were coded for the presence of Cuijpers et al. (2010) eight quality criteria, which were based on Chambless and Hollon's (1998) review of empirically supported psychotherapies and the Cochrane Collaboration's (Higgins & Green, 2011) criteria on study methodology. The eight quality criteria required: 1) that participants met diagnostic criteria for an anxiety disorder according to a personal diagnostic interview; 2) use of a treatment manual; 3) providers received treatment specific training; 4) treatment fidelity was evaluated throughout the study; 5) intent-to-treat analyses were used; 6) the comparison of treatment and control included $\geq$50 participants; 7) a third party independent to assessment and treatment conducted randomization, and; 8) assessors were blind to condition. Each criterion was evaluated for each article and assigned either 1 point (if the study fit the criterion) or 0 points (if the study did not fit the criterion). Quality criterion adherence was not assumed or inferred. Items were only coded as 1 if the information was explicitly stated in the article. After the coders rated each article, they met with two additional investigators to resolve any rating discrepancies and reach consensus.

2.2. Analyses

To determine which variable from each study would be used to calculate an effect size, articles were first organized into groups by diagnosis treated. Primary outcome measures within diagnostic criteria were then identified and these values were used when available in articles. If an article did not use the modal measure for its diagnostic category, behaviorally anchored outcome measures of avoidance were used, after which the first measure reported in Section 3 germane to the diagnosis became the variable of focus. If data were only available for one outcome measure, that variable was used. Finally, measures were only included in the effect size calculation and the related analyses when sufficient data were reported to calculate an effect size. Four of the 27 articles met inclusion/exclusion criteria but did not report sufficient information to calculate an effect size. Accordingly, these studies were included in the review of quality criteria but excluded from the effect size analyses.

Given the small sample sizes of the included studies, Hedge's g effect sizes were estimated to correct for small-sample bias (Deeks, Altman, & Bradburn, 2001). The primary goal of the analysis was to examine the magnitude, not the direction, of effect sizes as a function of study quality. To that end, the absolute value of the effect size estimates was used as the outcome in the analysis. Weights for each study were calculated using the inverse variance. We used meta-regression to compare the difference in the average effect size magnitude between groups based on quality score and total sample size. We used a joint effects exposure definition to separate the effects of study quality and total sample size. This allowed us to examine the association between study quality and average effect size within strata of total sample size, and vice versa. To account for nonindependence of studies that reported comparisons with both an active and a non-active comparison

group, all analyses were conducted separately by type of comparison group so that individual estimates within strata were independent. Stata version 12.1 (2013), was used to conduct all analyses; the metaan package (Kontopantelis & Reeves, 2013) was used to estimate average effect sizes, overall and for each stratum, and the metareg package (Harbord & Higgins, 2008) to estimate the meta-regression models.

Quality scores and study sample sizes were dichotomized at the sample medians for stratification to provide large enough groups for comparison.

3. Results

The final sample of 27 articles had publication dates ranging from 1995 to 2012 and included outcome data for 1080 participants (see Table 2 for description of studies). The mean sample size across studies was 40.00 (SD = 23.50). Twenty-two of the studies described treatment protocols with an established number of sessions (M = 7.95 sessions, SD = 3.57, range = 1–16 sessions). Five studies described protocols with a variable number of sessions (e.g., "up to 6 sessions"). The clinical diagnoses studied included fear of flying (n = 7), public speaking/social anxiety (n = 4), acrophobia (n = 3), arachnophobia (n = 3), panic disorder with agoraphobia (n = 3), PTSD (n = 3), generalized anxiety disorder (n = 1), mixed stress-related disorders (n = 1), panic disorder with or without agoraphobia (n = 1), and school phobia (n = 1). Most studies compared the experimental treatment to at least one active control group (n = 19), while a sizable minority of studies compared treatment to only a wait-list control or active placebo (n = 8).

Results of the final rating for all eight criteria can be found in Table 1. The articles met an average of 2.85 quality items (SD = 1.56, minimum = 0, maximum = 6). Six studies met more than half of the quality items and the modal number of quality items met was three. The correlation between number of study quality criteria present and publication year trended positive; however, the correlation coefficient was not statistically significant (ρ = .24, p = .23).

The effect size estimates for studies with an active comparison group were small in magnitude (Hedge's g = .19; $CI_{.95}$ = .03, .35), and variation across strata was not observed. In contrast, effect sizes were larger for studies that used a non-active comparison group (Hedge's g = .75; $CI_{.95}$ = .56, .94). Fig. 1 displays scatterplots of the effect sizes, overall, for the active and non-active comparison

Table 1
Description of articles evaluated in the current study.

Authors, year	Clinical sample	Groups (n per group)	Total sample size (N)	No. of sessions
Baños et al., 2011	Stress-related disorders	EMMA's World (CBT with VR; n = 19); Traditional CBT (n = 20)	39	4–9
Botella et al., 2007	Panic disorder with or without agoraphobia	VRET (n = 12); in vivo exposure (n = 12); wait list control (n = 13)	37	9
Choi et al., 2005	Panic disorder with agoraphobia	Experiential cognitive therapy (VR condition; n = 20; panic control program (CBT); n = 20)	40	12
Emmelkamp et al., 2002	Acrophobia	VRET (n = 17); in vivo exposure (n = 16)	33	3
Gamito et al., 2010	PTSD	VRET (n = 4); imaginal exposure (n = 2); wait list control (n = 3)	9	12
Garcia-Palacios, Hoffman, Carlin, Furness, and Botella, 2002	Arachnophobia	VRET (n = 12); wait list control (n = 11)	23	3–10
Gorini et al., 2010	Generalized anxiety disorder	VRET (n = 12); wait list control (n = 8)	20	8
Gutiérrez-Maldonado et al., 2009	School phobia	VRET (n = 18); Wait list control (n = 18)	36	8
Harris, Kemmerling, and North, 2002	Public speaking anxiety	VRET (n = 8); wait list control (n = 6)	14	4
Krijn et al., 2004	Acrophobia	VRET (n = 17); wait list control (n = 11)	28	3
Krijn et al., 2007	Fear of flying	VRET + bibliotherapy (n = 30); CBT + bibliotherapy (n = 23); bibliotherapy (n = 19)	72	2–4
Maltby et al., 2002	Fear of flying	VRET (n = 20); attention–placebo group (n = 23)	43	5
McLay et al., 2011	PTSD	VR-graded exposure therapy (n = 10); treatment as usual (n = 9)	19	4–20
Michaliszyn, Marchand, Bouchard, Martel, and Poirier-Bisson, 2010	Arachnophobia	VRET (n = 16); in vivo exposure (n = 16); wait list control (n = 11)	43	8
Mühlberger, Wiedemann, and Pauli, 2003	Fear of flying	Cognitive therapy + VRET (n = 26); cognitive therapy (n = 11); wait list control (n = 10)	47	1
Pelissolo, Zaoui, Aguayo, Yao, Roche, Ecochard, Gueyffier, Pull, Berthoz, Jouvent, and Cottraux, 2012	Panic disorder with agoraphobia	VRET (n = 33); CBT (n = 34)	67	12
Price and Anderson, 2011	Social anxiety	VRET (n = 32); exposure group therapy (n = 33); wait list control (n = 25)	90	8
Ready, Gerardi, Backscheider, Mascaro, and Rothbaum, 2010	PTSD	VRET (n = 5); present centered therapy (n = 4)	9	10
Robillard, Bouchard, Dumoulin, Guitard, and Klinger, 2010	Social anxiety	CBT with VR (n = 14); CBT with in vivo exposure (n = 16); wait list control (n = 15)	45	16
Rothbaum et al., 2006	Fear of flying	VRET (n = 29); in vivo (n = 29); wait list control (n = 25)	83	8
Rothbaum et al., 1995	Acrophobia	VRET (n = 10); wait list control (n = 7)	17	7
Rothbaum, Hodges, Smith, Lee, and Price, 2000	Fear of flying	VRET (n = 15); standard exposure (n = 15); wait list control (n = 15)	45	8
St-Jacques, Bouchard, and Belanger, 2010	Arachnophobia	VRET (n = 17); in vivo exposure (n = 14)	31	5
Tortella-Feliu, Botella, Llabrés, Bretón-López, del Amo, Baños, and Gelabert, 2011	Fear of flying	VRET (n = 19); computer aided exposure with therapist (n = 20); computer aided exposure (self-administered) (n = 21)	60	Up to 6
Vincelli et al., 2003	Panic disorder with agoraphobia	Experiential-cognitive therapy with VR (n = 4); CBT (n = 4); wait list control (n = 4)	12	8
Wallach, Safir, and Bar-Zvi, 2009	Public speaking anxiety	VR CBT (n = 28); CBT (n = 30); wait list control (n = 30)	88	12
Wiederhold et al., 2002	Fear of flying	VR-graded exposure therapy (n = 20); systematic desensitization with imaginal exposure (n = 10)	30	8

Notes: VRET: Virtual reality exposure therapy; VR: virtual reality; and CBT: cognitive behavioral therapy. Sample sizes were calculated based on the number of participants in each comparison group for the analysis. Total sample size was the sum of these groups.

Table 2
Number and percentage of trials meeting each quality criterion and quality score rating per item.

Item	Description	*n*	$n/27 \times 100$ (%)
1	Participants met diagnostic criteria for anxiety disorder as determined by a personal diagnostic interview	14	51.85
2	A treatment manual was used by the providers	10	37.04
3	The providers received treatment specific training	9	33.33
4	Treatment fidelity was evaluated throughout the study	8	29.63
5	An intent-to-treat analysis was conducted	11	40.74
6	Comparison of treatment with controls included ≥50 participants	4	14.81
7	Randomization was conducted by a third party, independent to assessment and treatment	5	18.52
8	Assessors were blind to condition	16	59.26

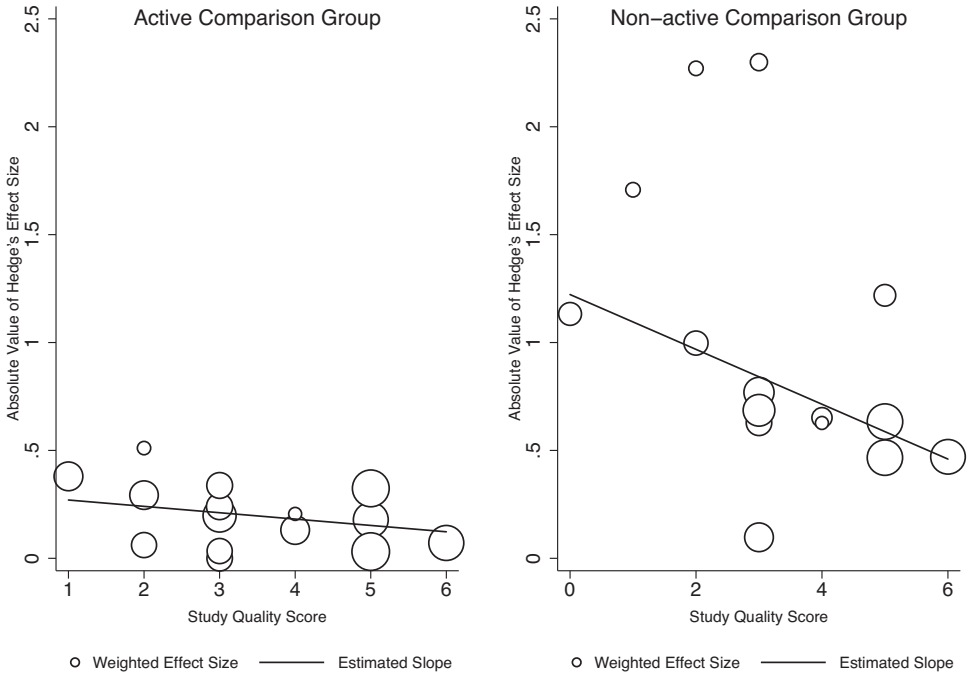

Fig. 1. Scatterplot of the absolute value of study effect sizes and study quality with an overlaid best-fit line, stratified by control type (active, not active). Larger plot points indicate studies with greater weight (inverse variance).

groups with overlaid slopes. The overall slope estimates were −.03 ($CI_{.95}$ = −.15, .09) among studies with an active comparison group and −.13 ($CI_{.95}$ = −.30, .04) among studies with a non-active comparison group, indicating a stronger association between study quality and effect size for studies with a non-active comparison group. Division of the quality scores into two groups (Table 3) yielded a similar inference. For studies with an active comparison group, there was no observed difference in effect size as a function of total sample size. In contrast, among studies with a non-active comparison group, those with a larger total sample size

Table 3
Average absolute value effect size, overall and within strata of quality score, sample size, and their joint effects.

Variable	Active comparison group			Non-active comparison group		
	No. of studies	ES (95% CI)[a]	b (95% CI)[b]	No. Studies	ES (95% CI)[a]	b (95% CI)[b]
All studies	16	0.19 (0.03, 0.35)		15	0.75 (0.56, 0.94)	
Quality score (QS)						
<4	9	0.21 (−0.02, 0.44)	Ref.	9	0.88 (0.62, 1.14)	
≥4	7	0.17 (−0.06, 0.39)	-0.05 (-0.39, 0.30)	6	0.61 (0.34, 0.88)	−0.31 (−0.91, 0.29)
Total sample size (TSS)						
<33	7	0.17 (−0.14, 0.48)	Ref.	9	1.15 (0.83, 1.47)	Ref.
≥33	9	0.19 (0.01, 0.38)	0.02 (−0.37, 0.42)	6	0.53 (0.30, 0.76)	−0.62 (−1.07, −0.16)
Joint effects						
QS < 4, TSS < 33	5	0.14 (−0.21, 0.48)	Ref.	6	1.25 (0.87, 1.63)	Ref.
QS ≥ 4, TSS < 33	2	0.32 (−0.38, 1.02)	0.18 (−0.68, 1.05)	3	0.90 (0.32, 1.48)	−0.35 (−1.20, 0.51)
QS < 4, TSS ≥ 33	4	0.27 (−0.04, 0.58)	0.14 (−0.38, 0.65)	3	0.54 (0.18, 0.90)	−0.72 (−1.36, −0.07)
QS ≥ 4, TSS ≥ 33	5	0.15 (−0.08, 0.38)	0.01 (−0.45, 0.47)	3	0.52 (0.22, 0.83)	−0.73 (−1.33, −0.13)

Notes: ES: Effect size, CI: confidence interval.
[a] Effect sizes are the absolute value of the effect size estimate to evaluate magnitude.
[b] Difference from meta-regression.

had a substantial reduction in average effect size compared to those with smaller total sample sizes. In the joint effects model, there were no associations between either study quality or total sample size among studies with an active comparison group. Finally, among studies with a non-active comparison group, there were no differences between studies of greater and lesser quality within strata of total sample size. Irrespective of study quality, those that had larger sample sizes had smaller average effect sizes as compared to studies with smaller sample sizes.

4. Discussion

This study evaluated the quality of research design and study methodology as reported in 27 RCTs of VRET for the treatment of anxiety disorders. This was accomplished by systematically applying eight criteria designed to evaluate study quality (Cuijpers et al., 2010), which stemmed from the review of empirically supported psychotherapies conducted by Chambless and Hollon (1998) and the criteria on study methodology as proposed by the Cochrane Collaboration (Higgins & Green, 2011).

The VRET RCTs, on average, met approximately three of the eight quality items. The number of study quality criteria observed in the VRET RCTs is considerably lower than the number of quality criteria observed when these criteria were applied to 16 RCTs evaluating psychotherapy for adult depression (Cuijpers et al., 2010). Specifically, none of the VRET articles included in this analysis met all eight study quality criteria, whereas 11 of the 16 articles in Cuijpers et al.'s sample met all items. Although previous reviews of the literature of VRET in the treatment of anxiety disorders by Parsons and Rizzo (2008) and Meyerbröker and Emmelkamp (2010) commented on the limitations of research design quality, the current study is the first to systematically examine the quality of VRET RCTs.

It was hypothesized that the quality of VRET RCTs would improve over time as VR transitioned from being a novel application of technology to a more established intervention. It was found, however, that study quality did not improve consistently since the first VRET RCT nearly 20 years ago. Perhaps one of the most compelling explanations for the quality of VRET RCTs is that the research questions examined in these RCTs has changed frequently due to the many different VR hardware types (e.g., head-mounted display (HMD), cave automatic virtual environment (CAVE)), software programs (e.g., EMMA's World, Virtual Iraq), and diagnostic applications (e.g., fear of flying, PTSD). The lack of a single technological approach to VRET and diagnostic application of this intervention may have then led to additional studies focused on providing initial support for many potential types and uses for VRET, as opposed to a growing research base where studies may tend to become more methodologically rigorous over time.

Effect sizes were calculated for the 23 VRET RCTs that reported adequate data to do so. The effect sizes obtained are consistent with VRET meta-analyses (Opriş et al., 2012; Parsons & Rizzo, 2008; Powers & Emmelkamp, 2008). VRET was found to have a small effect size advantage over active comparison groups and a large effect size advantage over non-active comparison groups, regardless of study quality.

The study quality was also compared to corresponding effect sizes. Contrary to what was hypothesized, lower quality studies were not found to have larger effect sizes; however, a non-significant negative trend between these variables was observed. This result differs from Cuijpers et al.'s (2010) finding that study quality and effect size were negatively related when examining psychotherapy for adult depression RCTs. It should be noted that the Cuijpers et al.'s analysis only included studies with extreme high and low quality scores, whereas this study included all studies regardless of quality. Cuijpers et al.'s analytic strategy could not be replicated with this sample because none of the VRET RCTs met all quality items, whereas studies that met all quality items made up the entirety of the high-quality comparison group in their analysis. The methodical difference between these two analyses makes comparison of these results somewhat difficult.

While this study did not find a significant relationship between study quality and effect size, a relationship was observed between sample size and effect size. It was found that effect sizes for studies comparing VRET to active control groups were relatively consistent regardless of study quality or sample size, whereas effect sizes tended to be larger when VRET was compared to non-active control groups with small sample sizes, regardless of study quality. This makes intuitive sense in that the larger the sample size, the more likely an observed effect size represents an accurate estimate of how a given treatment may work with the population relative to another intervention. Thus, as sample sizes increase, the possibility of an erroneously large effect size is reduced. Furthermore, the file drawer effect, or the tendency for lower publication rates of studies supporting the null hypothesis, may contribute to the increase in effect sizes for studies with small samples. Published VRET studies with small samples may be more likely to demonstrate results that are particularly positive for VRET. It is not clear why this relationship between sample size and effect size was observed only when VRET was compared to non-active control groups, and not when compared to active control groups.

Study quality has an important impact on internal and external validity. For example, internal validity might be compromised if a treatment manual was not used to ensure the administration of the same intervention by study clinicians. Such a situation might also obfuscate the external validity by making it difficult to replicate the treatment intervention, since no treatment was firmly established to begin with. Therefore, low study quality has the potential to render results somewhat difficult to interpret.

The call for more clarity and comprehensiveness with regard to reporting is not new. The Consolidated Standards of Reporting Trials (CONSORT) was developed to improve the quality of reporting in RCTs (Begg et al., 1996; Moher, Schulz, & Altman, 2001; Schulz, Altman, & Moher, 2010). The goal of the CONSORT statement is to promote the "clear, transparent, and accurate" reporting of RCTs. The 2010 CONSORT statement contains a 37-item checklist (which could be expanded to a total of 58 items (Davidson et al., 2003; Hopewell et al., 2008; Moher et al., 2010) and a flow diagram for researchers to use when reporting the results of RCTs. The CONSORT statement provides a standard method for researchers to report trial findings, facilitating their complete and transparent reporting, and aiding critical appraisal and interpretation. Toward this end, it is recommended that future VRET RCTs adhere to the CONSORT guidelines, with appropriate consideration for reporting in behavioral health (Davidson et al., 2003).

4.1. Limitations

One limitation of this study is that the validity of the count of study quality criteria obtained in this study is dependent on whether the articles are reported in a comprehensive manner. It is possible, if not likely, that some studies included in this analysis are of higher quality than the observed quality score suggests because the articles may not have reported information needed to determine that a given quality item was met. The quality scores calculated in this study, as well as others, might be best characterized as the study quality as reported in articles, not necessarily the actual quality of the studies. The extent to which this may be true is unclear; however, it should be noted Cuijpers et al. (2010) observed study quality levels higher than those

observed in our VRET RCT sample despite the aforementioned limitation. It is possible that because the treatment literature on depression is better established relative to the VRET treatment literature, standardized methodologies are more common among depression treatment outcome studies.

Although few would argue that the rating system used for this research assesses core aspects of study quality, there may be additional variables that would be important to include. It is also likely that not all study quality items are of equal weight and importance. For example, a large sample size may or may not, be a better indicator of study quality than the use of treatment fidelity monitoring throughout the course of treatment. Still another limitation of this study is that the authors chose to exclude studies with very small sample sizes (<10), which in turn, may have led to a positive study quality bias. Finally, four studies were excluded from the effect size analysis because insufficient data were available for analysis. The average study quality score for the studies excluded from this analysis is lower than the score for those included.

5. Conclusion

There may be great promise for VRET as an effective therapy for anxiety disorders. Indeed, recent well designed research investigating VRET combined with medications (Rothbaum et al., 2014 in press) supports its effectiveness. However, the results of this review suggest that the VRET body of literature would be strengthened by additional high quality, well-designed RCTs that compare this intervention control conditions, particularly other standards of care. RCTs represent the highest methodological standard in research, thus future studies should aim to achieve the highest standard of quality. Without such quality, advancement and further adoption of this intervention will be limited. It is only with improved standards that the literature in this area will be elevated to the standard of RCTs of more well-established interventions.

References

Arch, J. J., & Craske, M. G. (2009). First-line treatment: A critical appraisal of cognitive behavioral therapy developments and alternatives. *Psychiatric Clinics of North America*, *32*(3), 525–547. http://dx.doi.org/10.1016/j.psc.2009.05.001.

Baños, R. M., Guillen, V., Quero, S., García-Palacios, A., Alcaniz, M., & Botella, C. (2011). A virtual reality system for the treatment of stress-related disorders: A preliminary analysis of efficacy compared to a standard cognitive behavioral program. *International Journal of Human-Computer Studies*, *69*(9), 602–613. http://dx.doi.org/10.1016/j.ijhcs.2011.06.002.

Begg, C., Cho, M., Eastwood, S., Horton, R., Moher, D., & Olkin, I. (1996). Improving the quality of reporting of randomized controlled trials. The CONSORT statement. *JAMA: the Journal of the American Medical Association*, *276*(8), 637–639.

Botella, C., García-Palacios, A., Villa, H., Baños, R. M., Quero, S., & Alcañiz, M. (2007). Virtual reality exposure in the treatment of panic disorder and agoraphobia: A controlled study. *Clinical Psychology & Psychotherapy*, *14*(3), 164–175. http://dx.doi.org/10.1002/cpp.524.

Chambless, D. L., & Hollon, S. D. (1998). Defining empirically supported therapies. *Journal of Consulting and Clinical Psychology*, *66*(1), 7–18. http://dx.doi.org/10.1037/0022-006x.66.1.7.

Choi, Y.-H., Vincelli, F., Riva, G., Wiederhold, B. K., Lee, J.-H., & Park, K.-H. (2005). Effects of group experiential cognitive therapy for the treatment of panic disorder with agoraphobia. *Cyberpsychology & Behavior: the Impact of the Internet Multimedia and Virtual Reality on Behavior and Society*, *8*(4), 387–393.

Cuijpers, P., van Straten, A., Bohlmeijer, E., Hollon, S. D., & Andersson, G. (2010). The effects of psychotherapy for adult depression are overestimated: A meta-analysis of study quality and effect size. *Psychological Medicine*, *40*(2), 211–223. http://dx.doi.org/10.1017/s0033291709006114.

Davidson, J. R. T., Foa, E. B., Huppert, J. D., Keefe, F. J., Franklin, M. E., & Compton, J. S. (2004). Fluoxetine, comprehensive cognitive behavioral therapy, and placebo in generalized social phobia. *Archives of General Psychiatry*, *61*(10), 1005–1013. http://dx.doi.org/10.1001/archpsyc.61.10.1005.

Davidson, K. W., Goldstein, M., Kaplan, R. M., Kaufmann, P. G., Knatterud, G. L., & Orleans, C. T. (2003). Evidence-based behavioral medicine: What is it and how do we achieve it? *Annals of Behavioral Medicine: a Publication of the Society of Behavioral Medicine*, *26*(3), 161–171.

In: J.J. Deeks, D.G. Altman, & J. Bradburn (Eds.), (2001). *Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis*. 2nd ed. London: BMJ Publishing Group.

Difede, J., Cukor, J., Jayasinghe, N., Patt, I., Jedel, S., & Spielman, L. ([152_TD$DIFF] 2007). Virtual reality exposure therapy for the treatment of posttraumatic stress disorder following September 11. *Journal of Clinical Psychiatry*, *68*(11), 1639–1647. http://dx.doi.org/10.4088/JCP.v68n1102.

Difede, J., & Hoffman, H. G. (2002). Virtual reality exposure therapy for World Trade Center post-traumatic stress disorder: A case report. *CyberPsychology & Behavior*, *5*(6), 529–535. http://dx.doi.org/10.1089/109493102321018169.

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.

Foa, E. B., Liebowitz, M. R., Kozak, M. J., Davies, S., Campeas, R., & Franklin, M. E. (2005). Randomized, placebo-controlled trial of exposure and ritual prevention, clomipramine, and their combination in the treatment of obsessive-compulsive disorder. *The American Journal of Psychiatry*, *162*(1), 151–161.

Gamito, P., Oliveira, J., Rosa, P., Morais, D., Duarte, N., & Oliveira, S. (2010). PTSD elderly war veterans: A clinical controlled pilot study. *Cyberpsychology Behavior and Social Networking*, *13*(1), 43–48.

Garcia-Palacios, A., Hoffman, H., Carlin, A., Furness, T. A.3rd, & Botella, C. (2002). Virtual reality in the treatment of spider phobia: A controlled study. *Behaviour Research and Therapy*, *40*(9), 983–993.

Gorini, A., Pallavincini, F., Algeri, D., Repetto, C., Gaggioli, A., & Riva, G. (2010). Virtual reality in the treatment of generalized anxiety disorders. *Annual Review of CyberTherapy and Telemedicine*, *8*, 31–35.

Gutiérrez-Maldonado, J., Magallón-Neri, E., Rus-Calafell, M., & Peñaloza-Salazar, C. (2009). Virtual reality exposure therapy for school phobia. *Anuario de Psicologéa*, *40*(2), 223–236.

Harbord, R. M., & Higgins, J. P. (2008). Meta-regression in stata. *The Stata Journal*, *8*(4), 493–519.

Harris, S. R., Kemmerling, R. L., & North, M. M. (2002). Brief virtual reality therapy for public speaking anxiety. *CyberPsychology & Behavior*, *5*(6), 543–550. http://dx.doi.org/10.1089/109493102321018187.

Higgins, J. P., & Green, S. ([162_TD$DIFF]2011). *Cochrane Handbook for Systematic Reviews of Interventions. 5.1.0.*. . . accessed 30.05.13 http://handbook.cochrane.org/.

Hopewell, S., Clarke, M., Moher, D., Wager, E., Middleton, P., & Altman, D. G. (2008). CONSORT for reporting randomized controlled trials in journal and conference abstracts: Explanation and elaboration. *PLoS Medicine*, *5*(1), e20. http://dx.doi.org/10.1371/journal.pmed.0050020.

Institute of Medicine of the National Academies (2007). Treatment of PTSD: An assessment of the evidence. *Institute of Medicine of the National Academies* .

Jaycox, L. H., Foa, E. B., & Morral, A. R. (1998). Influence of emotional engagement and habituation on exposure therapy for PTSD. *Journal of Consulting and Clinical Psychology*, *66*(1), 185–192.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, *62*(6), 593–602.

Klinger, E., Bouchard, S., Légeron, P., Roy, S., Lauer, F., & Chemin, I. (2005). Virtual reality therapy versus cognitive behavior therapy for social phobia: A preliminary controlled study. *Cyberpsychology & Behavior: the Impact of the Internet Multimedia and Virtual Reality on Behavior and Society*, *8*(1), 76–88.

Kontopantelis, E., & Reeves, D. (2013). *METAAN: Stata module to perform fixed- or random-effects meta-analyses*. . . accessed 30.05.13 http://econpapers.repec.org/software/bocbocode/s457071.htm.

Krijn, M., Emmelkamp, P. M. G., Biemond, R., de Wilde de Ligny, C., Schuemie, M. J., & van der Mast, C. A. P. G. (2004). Treatment of acrophobia in virtual reality: The role of immersion and presence. *Behaviour Research and Therapy*, *42*(2), 229–239.

Krijn, M., Emmelkamp, P. M. G., Olafsson, R. P., Bouwman, M., van Gerwen, L. J., & Spinhoven, P. (2007). Fear of flying treatment methods: Virtual reality exposure vs. cognitive behavioral therapy. *Aviation, Space and Environmental Medicine*, *78*(2), 121–128.

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. http://dx.doi.org/10.1037/0022-006x.70.5.1112.

Marks, I. M., Swinson, R. P., Başoğlu, M., Kuch, K., Noshirvani, H., & O'Sullivan, G. (1993). Alprazolam and exposure alone and combined in panic disorder with agoraphobia. A controlled study in London and Toronto. *The British Journal of Psychiatry: the Journal of Mental Sciences*, *162*, 776–787.

McLay, R. N., Wood, D. P., Webb-Murphy, J. A., Spira, J. L., Wiederhold, M. D., & Pyne, J. M. (2011). A randomized, controlled trial of virtual reality-graded exposure therapy for post-traumatic stress disorder in active duty service members with combat-related post-traumatic stress disorder. *Cyberpsychology Behavior and Social Networking*, *14*(4), 223–229. http://dx.doi.org/10.1089/cyber.2011.0003.

Meyerbröker, K., & Emmelkamp, P. M. G. (2010). Virtual reality exposure therapy in anxiety disorders: A systematic review of process-and-outcome studies. *Depression and Anxiety*, *27*(10), 933–944. http://dx.doi.org/10.1002/da.20734.

Michaliszyn, D., Marchand, A., Bouchard, S., Martel, M.-O., & Poirier-Bisson, 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. http://dx.doi.org/10.1089/cyber.2009.0277.

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., & Devereaux, P. J. (2010). CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. *BMJ (Clinical Research Edition)*, *340*, c869. http://dx.doi.org/10.1136/bmj.c869.

Moher, D., Schulz, K. F., & Altman, D. (2001). The CONSORT statement: Revised recommendations for improving the quality of reports of parallel-group randomized trials. *JAMA: the Journal of the American Medical Association*, *285*(15), 1987–1991.

Mühlberger, A., Wiedemann, G., & Pauli, P. (2003). Efficacy of a one-session virtual reality exposure treatment for fear of flying. *Psychology Research: journal of the Society for Psychotherapy Research*, *13*(3), 323–336. http://dx.doi.org/10.1093/ptr/kpg030.

Norton, P. J., & Price, E. C. (2007). A meta-analytic review of adult cognitive-behavioral treatment outcome across the anxiety disorders. *The Journal of Nervous And Mental Disease*, *195*(6), 521–531.

Opriş, D., Pintea, S., García-Palacios, A., Botella, C., Szamosközi Ş, & David, D. (2012). Virtual reality exposure therapy in anxiety disorders: A quantitative meta-analysis. *Depression and Anxiety*, *29*(2), 85–93. http://dx.doi.org/10.1002/da.20910.

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. http://dx.doi.org/10.1016/j.jbtep.2007.07.007.

Pelissolo, A., Zaoui, M., Aguayo, G., Yao, S. N., Roche, S., & Ecochard, R. (2012). Virtual reality exposure therapy versus cognitive behavior therapy for panic disorder with agoraphobia: A randomized comparison study. *Journal of CyberTherapy and Rehabilitation*, *5*(1), 35–43.

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. http://dx.doi.org/10.1016/j.janxdis.2007.04.006.

Price, M., & Anderson, P. L. (2011). The impact of cognitive behavioral therapy on post event processing among those with social anxiety disorder. *Behaviour Research and Therapy*, *49*(2), 132–137. http://dx.doi.org/10.1016/j.brat.2010.11.006.

Ready, D. J., Gerardi, R. J., Backscheider, A. G., Mascaro, N., & Rothbaum, B. O. (2010). Comparing virtual reality exposure therapy to present-centered therapy with 11 U.S. Vietnam veterans with PTSD. *Cyberpsychology Behavior and Social Networking*, *13*(1), 49–54.

Reger, G. M., Holloway, K. M., Candy, C., Rothbaum, B. O., Difede, J., & Rizzo, A. A. (2011). Effectiveness of virtual reality exposure therapy for active duty soldiers in a military mental health clinic. *Journal of Traumatic Stress*, *24*(1), 93–96. http://dx.doi.org/10.1002/jts.20574.

Robillard, G., Bouchard, S., Dumoulin, S., Guitard, T., & Klinger, E. (2010). Using virtual humans to alleviate social anxiety: Preliminary report from a comparative outcome study. *Studies in Health Technology and Informatics*, *154*, 57–60.

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. http://dx.doi.org/10.1016/j.beth.2005.04.004.

Rothbaum, B. O., Hodges, L., Smith, S., Lee, J. H., & Price, L. (2000). A controlled study of virtual reality exposure therapy for the fear of flying. *Journal of Consulting and Clinical Psychology*, *68*(6), 1020–1026.

Rothbaum, B. O., Hodges, L. F., Kooper, R., Opdyke, D., Williford, J. S., & North, M. (1995). Effectiveness of computer-generated (virtual reality) graded exposure in the treatment of acrophobia. *The American Journal of Psychiatry*, *152*(4), 626–628.

Rothbaum, B. O., Price, M., Jovanovic, T., Norrholm, S. D., Gerardi, M., & Dunlop, B., et al., (2014). A randomized, double-blind evaluation of D-cycloserine or alprazolam combined with virtual reality exposure therapy for posttraumatic stress disorder (PTSD) in Iraq and Afghanistan war veterans. *American Journal of Psychiatry* (in press).

Rothbaum, B. O., & Schwartz, A. C. (2002). Exposure therapy for posttraumatic stress disorder. *American Journal of Psychotherapy*, *56*(1), 59–75.

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT statement: Updated guidelines for reporting parallel group randomised trials. *Journal of Pharmacology & Pharmacotherapeutics*, *1*(2), 100–107. http://dx.doi.org/10.4103/0976-500x.72352.

Stanley, M. A., Wilson, N. L., Novy, D. M., Rhoades, H. M., Wagener, P. D., & Greisinger, A. J. (2009). Cognitive behavior therapy for generalized anxiety disorder among older adults in primary care: A randomized clinical trial. *JAMA: The Journal of the American Medical Association*, *301*(14), 1460–1467. http://dx.doi.org/10.1001/jama.2009.458.

STATA/MP, 12.1 ed. 2013; http://www.stata.com/products/

St-Jacques, J., Bouchard, S., & Bélanger, 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. *The Journal of Clinical Psychiatry*, *71*(7), 924–931. http://dx.doi.org/10.4088/JCP.08m04822blu.

Tortella-Feliu, M., Botella, C., Llabrés, J., Bretón-López, J. M., del Amo, A. R., & Baños, R. M. (2011). Virtual reality versus computer-aided exposure treatments for fear of flying. *Behavior Modification*, *35*(1), 3–30. http://dx.doi.org/10.1177/0145445510390801.

Vincelli, F., Anolli, L., Bouchard, S., Wiederhold, B. K., Zurloni, V., & Riva, G. (2003). Experiential cognitive therapy in the treatment of panic disorders with agoraphobia: A controlled study. *Cyberpsychology & Behavior: the Impact of the Internet Multimedia and Virtual Reality on Behavior and Society*, *6*(3), 321–328.

Wallach, H. S., Safir, M. P., & Bar-Zvi, M. (2009). Virtual reality cognitive behavior therapy for public speaking anxiety: A randomized clinical trial. *Behavior Modification*, *33*(3), 314–338. http://dx.doi.org/10.1177/0145445509331926.

Wiederhold, B. K., Jang, D. P., Gevirtz, R. G., Kim, S. I., Kim, I. Y., & Wiederhold, M. D. (2002). The treatment of fear of flying: a controlled study of imaginal and virtual reality graded exposure therapy. *IEEE Transactions on Information Technology in Biomedicine: a Publication of the IEEE Engineering in Medicine and Biology Society*, *6*(3), 218–223.