Outcomes associated with Virtual Reality in Psychological Interventions:

Where Are We Now?

Wesley A. Turner[a,b*]

Leanne M. Casey[a,b]

[a]School of Applied Psychology

[b]Griffith Health Institute

Griffith University, Brisbane, Queensland, Australia

*Correspondence should be addressed to Wesley A. Turner, School of Applied Psychology, Griffith University, Mt Gravatt Campus, 176 Messines Ridge Road, Mt Gravatt, QLD 4122, Australia. (email: wesley.turner@griffithuni.edu.au). Telephone: +61 7 37353343.

Abstract

The impending commercial release of affordable VR systems is likely to accelerate both the opportunity and demand for VR applications that specifically target psychological conditions. The aim of this study was to conduct a meta-analysis of outcomes associated with VR psychological interventions and to examine the methodological rigour used in these interventions. Literature search was conducted via Ovid, ProQuest Psychology Journals and ScienceDirect (Psychology) databases. Interventions were required to: be published between 1980 to 2013; use a randomised controlled trial design; be published in a scholarly journal; focused primarily on psychological/behavioural intervention; include validated measures; include reported means and standard deviations of outcome measures; and include one group with clinical/subclinical disorders, syndromes or distressing behaviours. Thirty-one eligible studies were identified. Random effects meta-analysis found an overall moderate effect size for VR interventions. Individual meta-analyses found an overall large effect size against non-intervention wait-lists and an overall moderate effect size against active interventions. No correlation was found between treatment outcomes and methodological rigour. Limitations include limited study numbers, small sample sizes, and a need for more in-depth analyses. The current review supports VR interventions as efficacious, promising forms of psychological treatment. Use of reporting guidelines such as the CONSORT and CONSORT-EHEALTH statements should promote greater emphasis on methodological rigour, providing a firm foundation for the further development of clinical VR applications.

Keywords: virtual reality; VR; technology, eHealth, review, methodology, meta-analysis

Outcomes Associated with Virtual Reality in Psychological Interventions: Where Are We Now?

Virtual Reality in Psychological Intervention

Despite several decades of research, use of virtual reality (VR) in psychological interventions has only grown more recently (Gorini & Riva, 2008; Repetto & Riva, 2011). Rise in the use of VR interventions is likely due to rapid advancements in underlying technologies. Substantial improvements have been made in several areas, including computer graphics, speed and processing power; head-mounted displays (HMD) and VR glasses/goggles quality; and motion tracking technology (Gregg & Tarrier, 2007). Costs associated with purchasing and maintaining VR systems have also dropped markedly, resulting in the impending commercial release of affordable VR systems such as the Oculus Rift and Sony HMZ-T2 Personal 3D Viewer (Gregg & Tarrier, 2007; Rougeau & Hawkins, 2013). Release of such systems into the general market is likely to accelerate both the opportunity and demand for VR applications that specifically target psychological conditions.

Existing VR interventions provide a range of interactive systems, environments and mechanisms by which psychological and behavioural change can be targeted in novel and engaging ways. Often making use of similar if not identical hypothesized mechanisms of action to traditional face to face interventions, VR interventions are now available to treat a variety of psychological disorders and behavioural issues (Fox et al., 2009) while providing greater flexibility in intervention timing, greater cost effectiveness, and an increased ability to tailor interventions to individual preferences (Carlbring & Andersson, 2006; Clough & Casey, 2011). Though the use of VR technology in the field is not yet widespread (Repetto & Riva, 2011), the substantial increase in use of VR within psychological research has resulted

in several recent meta-analyses (Opris et al., 2012; Parsons & Rizzo, 2008; Powers & Emmelkamp, 2008).

Results have been promising: VR interventions demonstrate strong pre-post effect sizes (Cohen's *d* = 0.95, Parsons & Rizzo, 2008) and strong overall effect sizes when compared to non-intervention wait lists (*d* = 1.12, Opris et al., 2012; *d* = 1.11 Powers & Emmelkamp, 2008), although low effect sizes were observed across VRET and in-vivo exposure/cognitive behavioural therapy comparisons (d = no effect, Opris et al., 2012; d = 0.35, Powers & Emmelkamp, 2008), However, these reviews have focused only on VR exposure interventions (virtual reality exposure therapy: VRET) aimed at treating anxiety disorders. It is unknown whether these findings can be generalised to VR interventions overall.

Additionally, there has been little assessment of the methodological rigour of research into VR interventions, despite the need for improvement in this area (Parsons & Rizzo, 2008). Sub-optimal methodology and methodological reporting can raise numerous concerns; inadequate randomisation and blinding can lead to bias (Strech, 2012); and participant and administration setting information can influence the generalisability of research findings (Knuppel et al., 2013). To position VR research in clinical psychology to take advantage of the potential expansion offered by current development in technology, it is timely to review the outcomes associated across the available range of VR based psychological interventions and to examine the methodological rigour used to substantiate these outcomes.

Defining VR

Definitions vary in what technological devices constitute VR systems. They may be regarded as being strictly comprised of HMDs or VR glasses/goggles and 3D virtual environments (Gregg & Tarrier, 2007), or may be used in a much broader sense, referring to any technological system that immerses a user in a virtual environment (VE). Discrepancy in

definitions may be due to the evolving nature of the technology itself, with rapid improvements resulting in greater options in interactive technologies (Adamovich, Fluet, Tunik, & Merians, 2009; Bohil, Alicea, & Biocca, 2011). Early images of bulky, cumbersome and heavily wired headwear and peripherals in the 1980s have given way to the promise of sleek, easily portable and relatively affordable VR systems that can be purchased for home use (Fox et al., 2009). Despite variations in definition and technology, VR is defined by its capacity to allow users to explore and engage with a VE, experiencing a sense of presence ('losing oneself) in a computer generated world (Baños et al., 2011; Bordnick, Traylor, Carter, & Graap, 2011; Botella et al., 2007; Fox, Arena, & Bailenson, 2009; Repetto & Riva, 2011; Rothbaum et al., 2006).

Additional devices are often used to aid in recreating real life scenarios, thus fostering a sense of presence, a factor seen by some as vital to successful immersion (Gorini & Riva, 2008; Gregg & Tarrier, 2007; Rothbaum et al., 2006). Though presence may be of greater importance in the treatment of clinically anxious individuals via VR based exposure (Ling et al., 2014). These devices may include delivery of tactile (haptic) and aural sensation (Bordnick et al., 2011; Krijn, Emmelkamp, Biemond, et al., 2004), as well as simulations of real life steering wheels, gears and pedals in VR driving simulators (Cox et al., 2010), and replica seats, windows and partial cabins in VR flight simulators (Muhlberger, Wiedemann, & Pauli, 2003; Rothbaum et al., 2006).

Methodological Rigour in VR Research and the CONSORT Statements

A vital question in assessing outcomes is whether the methodology used to produce these outcomes is of an appropriate standard. Transparent and detailed methodological reporting can aid in replication, clarify generalisability and validity, identify and reduce sources of potential bias, and promote confidence in research findings (Boutron, Moher, Altman, Schulz, & Ravaud, 2008; Knuppel et al., 2013; Moher, 1998; Moher, Schulz, &

Altman, 200; Strech, 2012). Accurate reporting of the conducting of randomised controlled trials (RCTs) is imperative to ensure that results are as free of bias as is possible (Boutron, Moher, Altman, Schulz, & Ravaud, 2008; Moher, 1998; Moher, Schulz, & Altman, 2001). To promote methodological rigour, a number of standardised reporting tools have been developed: for example, the Critical Appraisal Skills Program (CASP): Randomised Controlled Trial Appraisal Tool (CASP International Network, 2014) and the Physiotherapy Evidence Database (PEDro) Scale (Centre of Evidence-Based Physiotherapy, 2014).

One of the most widely used of these tools is the CONSORT Statement, which was designed as an essential items checklist and participant flow diagram to be included by RCT study authors (Moher, 1998; Moher et al., 2010). Use of CONSORT guidelines provides a concise method of systematically assessing methodological reliability and validity (Moher, 1998; Moher et al., 2010).

However, not all CONSORT inclusion recommendations are applicable to VR intervention studies. Technologically based interventions include specific details and challenges not yet fully addressed by existing RCT reporting guidelines (Eysenbach, 2002). A need to address this rapid technological growth resulted in the development of an extension to the CONSORT Statement: the CONSORT-EHEALTH Statement (Eysenbach & CONSORT-EHEALTH Group, 2011). The CONSORT-EHEALTH Statement is a checklist of items specific to web-based and mobile health RCT interventions (Eysenbach & CONSORT-EHEALTH Group, 2011). Although the CONSORT-EHEALTH Statement focuses primarily on web-based and mobile health interventions, its inclusion of technologically related variables and information enables assessment of the methodological rigour of studies using delivery platforms such as VR applications (Eysenbach & CONSORT-EHEALTH Group, 2011).

Aim

The aim of this study was to conduct a meta-analysis of outcomes associated with VR psychological treatment across the range of mental health /psychological concerns. As effect sizes can be influenced by comparison conditions (Furukawa et al., 2014), this analysis looked at both wait and active treatment controls. Additionally, as outcomes can be influenced by methodological rigour, the CONSORT and CONSORT-EHEALTH checklists were used to establish methodological rigour and evaluate potential methodological bias. Achievement of these objectives would serve not only to present the current state of VR psychological treatment, but would also identify areas requiring further investigation and advancement.

Method

This study was conducted and reported according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA; Moher et al., 2009) guidelines.

Selection of Articles

An extensive literature search was conducted via Ovid, ProQuest Psychology Journals and ScienceDirect (Psychology) databases. Articles published from 1980 to 2013 considered for inclusion. An additional manual search was conducted of referenced articles. The search terms used for this literature search were ‘game’, ‘virtual’, ‘reality’, ‘VR’, ‘Wii’, ‘Xbox’ and ‘PlayStation’. To be included in this review, VR based psychological interventions were required to have used a randomised controlled trial design; to have been published in a scholarly journal; focused primarily on psychological or behavioural intervention (as opposed to assessment or education); include validated measures; and to have reported means and standard deviations of outcome measures.

Additionally, interventions were required to have included one group with clinical or subclinical psychological or neurological disorders, syndromes or distressing behaviours. Disorders, syndromes and/or behaviours were required to have been identified via validated,

standardised clinical measures or to have been made according to DSM IV-TR criteria. Sub-clinical samples were defined as those partially meeting DSM IV-TR diagnosis or criteria specified by standardised clinical measures. Single case studies, dissertations and preliminary reports were excluded. Figure 1 displays selection and exclusion of studies at each stage of review. Thirty-one eligible studies were identified (Appendix A), with descriptions of the included studies displayed in Table 1.

Data Synthesis

Means and sample sizes for VR interventions and control groups were used in the calculation of effect sizes. To control for upwards bias effects caused by small sample sizes, Hedge's adjusted *g* (*g**) effect sizes were calculated for measures of improvement in psychological/behavioural functioning. Effect sizes were calculated such that positive outcomes reflected greater improvements in VR based intervention groups and negative outcomes reflected greater improvements in control groups. Mean effect sizes were calculated for each group comparison so as to reduce potential biases of study overrepresentation.

Publication bias was assessed via funnel plot (Figure 2) and Rosenthal (1979) *Fail-Safe N* value. Control types were coded as either non-intervention (wait lists control groups) or active interventions. Active interventions were defined as comparative, established interventions involving active, psychologically therapeutic mechanisms of action (e.g., traditional exposure therapy, traditional cognitive behavioural therapy, and active non-VR technological interventions).

Meta-analyses were conducted for overall effect sizes, as well as for effect sizes specific to non-intervention and active intervention groups. Comparison between VR interventions and all forms of controls was not desirable due to the variability in controls

(e.g., nicotine replacement therapy, traditional face to face exposure therapy, and non VR audio-visual distraction).

Additional moderator analyses were conducted to explore variability in effect sizes. Potential moderators were grouped into relevant clusters: clinical/subclinical sample, form of VR based intervention, form of control. Weighted meta-analytic multiple regressions were used to evaluate the impact of the aforementioned categorical moderators, goodness of fit of moderator variable sets, and estimates for fixed and random effects meta-analysis models (Wilson, 2005).

In order to assess methodological rigour, included studies were evaluated according to the CONSORT 2010 Statement and CONSORT-EHEALTH 2011 Statement (Moher et al., 2010; Eysenbach & CONSORT-EHEALTH Group, 2011). Due to considerable overlap in Statement items, CONSORT-EHEALTH 2011 items were only included when unique to those presented in the CONSORT 2010 Statement. The combination of these two CONSORT statements resulted in a total of 68 items for analysis.

Associations between treatment efficacy and methodological rigour were assessed via Pearson product-moment correlation coefficients, with total met CONSORT and CONSORT-EHEALTH checklist items used as measures of methodological rigour. Three correlation analyses were conducted: total checklist scores and mean effect sizes; total checklist scores and waitlist comparison mean effect sizes; and total checklist scores and active intervention comparison mean effect sizes.

All analyses were completed via IBM SPSS Statistics for Windows, Version 22.0 (IBM Corp, 2013). Meta-analysis and meta-analytic weighted multiple regression macros were used in accordance with Lipsey and Wilson (2000; Wilson 2005).

Results

Targets of Intervention

Reported targets of intervention were treatment of specific phobias such as driving, flying, heights, spiders, etc. (*n* = 15), pain management (*n* = 3), intellectual and developmental disabilities (*n* = 2), cerebral palsy (*n* = 1), multiple sclerosis (*n* = 1), panic disorder (*n* = 1), social anxiety (*n* = 1), traumatic brain injury (*n* = 1), post-traumatic stress disorder, adjustment disorder, pathological guilt (*n* = 3), schizophrenia (*n* = 1), addiction (*n* = 1) and dementia (*n* = 1).

Type of Intervention

Interventions employed purely varied greatly, including purely behavioural VR based exposure (*n* = 18), VR skills training (*n* = 6), VR cognitive behavioural therapy (*n* = 2), VR occupational therapy/physical rehabilitation (*n* = 3), and comparably more immersive forms of VR exposure (*n* = 2).

Comparison Conditions

Comparisons were made against non-intervention wait-lists (*n* = 11), in-vivo exposure (*n* = 8), imaginal exposure (*n* = 1), cognitive therapy/cognitive behavioural therapy (*n* = 4), therapist led interventions (*n* = 2), and active technologically based interventions (e.g., non VR audio-visual distraction and computer aided therapist run modules; *n* = 4). Comparisons were additionally made against 'treatment as usual' (*n* = 8), consisting of interventions such as person centred therapy (Ready et al., 2010) and nicotine replacement (Bordnick et al., 2012).

Analysis of Effect Sizes

Table 1 displays the number of effect sizes (*k*) and mean effect sizes calculated for analysis. Homogeneity analysis revealed significant results ($Q = 134.66$, $p = < .001$), indicating heterogeneity in effect sizes. A random effects model meta-analysis was employed to address heterogeneity, whilst inverse variance weighting was employed to afford larger studies more analytical weight (Lipsey & Wilson, 2000).

The mean effect size was 0.75, 95% CI [0.54, 0.96] and possessed a highly significant z score ($z = 7.06$, $p < .001$). According to Cohen's (1988) criterion, this figure indicates an overall moderate effect size for VR interventions. However, as the stem-and-leaf plot of all calculated effect sizes displayed in Table 2 indicates, there was considerable variability in outcomes. Effect sizes were observed to cluster around 0 to 1.3, with some effect sizes clustering around 2.0 to 2.8.

To understand the role that different comparisons may play in outcome levels, separate stem-and-leaf plots of effect sizes were calculated for non-intervention and active intervention comparison types (Table 3 and Table 4). Visual inspection of these effect sizes suggested greater variability in intervention effect sizes, as well as larger effects for non-intervention wait-list control comparisons. Individual meta-analyses were conducted in order to identify differences in effect sizes, with VR based interventions obtaining significant effect sizes against both non-intervention wait-list and active intervention controls (Table 5). VR interventions were observed to possess an overall large effect size when compared to non-intervention wait-lists, indicating greater outcomes via VR intervention than those observed naturally over time. Comparisons between VR interventions and active interventions resulted in an overall moderate effect size, suggesting that whilst VR interventions remained advantageous, less difference was observed in comparison to outcomes from active interventions.

A random effects meta-analytic multiple regression analyses was conducted for three variables: clinical/subclinical sample, form of VR based intervention, and form of control. Only form of VR based intervention was observed to significantly moderate observed effect sizes ($\beta = .33$, 95%, CI [.000, .323], $p = .049$).

Methodological Rigour

Table 6 lists the degree to which included VR studies met both CONSORT and CONSORT-EHEALTH Statement checklist items. Mean items meeting CONSORT guidelines were observed to be 33.84 (SD = 8.42), or approximately 49% of checklist items. Twenty-nine studies included a scientific background and explanation of their research, thus justifying the purpose and need for their study (item 2a). Twenty-eight of the 31 studies also described the eligibility criteria used to select participants, allowing readers to assess any potential issues in target population validity (item 4a).

Encouragingly, 30 of the 31 studies described the interventions employed in sufficient detail, allowing accurate replication and modification (item 5). Twenty-nine of the 31 studies addressed potential trial limitations (item 20), as well as the generalisability of their findings and accurately interpreting their results (item 21). However, few of the reviewed studies detailed information regarding randomisation method (*n* = 5; item 8a), randomisation type (*n* = 5; item 8b), who generated the random allocation sequence (*n* = 3; item 10), or the mechanism used to implement random allocation (*n* = 5; item 9).

Twenty-four of the studies described how participants accessed the VR system, in what context they accessed the system and whether they had to be members of specific groups in order to access the intervention (item 5 vii). Twenty-seven of the 31 studies described the mode of delivery employed and unique features of the utilised VR interventions (item 5 viii). However, only four studies reported demographics associated with digital divide issues including age, education, gender, socio-economic status, computer literacy, etc. (item 15 i), with 21 of the 31 studies also failing to include screenshots or sufficiently describe technical details of graphical engines and VEs employed (item 5 v). Information pertaining to conflict of interests (item 5 i), the display of institutional affiliations (item 4b ii) and the relation of the study team towards evaluated systems (item 27) was similarly under-reported, with 8, 0 and 10 studies (respectively) including this information.

Treatment Efficacy and Methodological Rigour

Correlation analyses were conducted to test potential associations between reported treatment outcomes and methodological rigour. Pearson product-moment correlation coefficients were calculated between total checklist scores and mean effect sizes; total checklist scores and waitlist comparison mean effect sizes; and total checklist scores and active intervention comparison mean effect sizes. Overall, no significant correlation was found between treatment outcomes and methodological rigour ($r[36] = -.21$, $p = .213$). VR intervention and active control group comparison correlations were non-significant ($r[25] = -.03$, $p = .899$), as were VR intervention and wait-list group comparison correlations ($r[9] = -.49$, $p = .130$). Methodological reporting appeared independent of mean effect sizes, suggesting outcomes reported in the studies included in the meta-analysis provide a robust gauge of VR treatment efficacy.

The impact of potential publication bias was examined via funnel plot and Rosenthal's (1979) *Fail-Safe N* value. Rosenthal (1979) suggests that findings can be deemed robust if the *Fail-Safe N* value exceeds 5 times the number of studies included plus 10. Subsequently, the *Fail-Safe N* number required to support the current study's findings was 205. A *Fail-Safe N* analysis demonstrated that 2417 additional unpublished studies on VR psychological interventions with effect sizes of zero would be required to lower the overall effect size to that of non-significance. These results suggest that the current study's findings are indeed robust.

Discussion

The objectives of this review were three-fold: firstly, to determine the efficacy of randomised controlled trials of VR based interventions via random effects meta-analysis; secondly, to assess the methodological rigour of these VR based interventions according to

CONSORT 2010 and CONSORT-EHEALTH 2011 guidelines; and thirdly, to explore whether methodological rigour was associated with treatment efficacy.

VR interventions demonstrated large and moderate effect sizes when compared to non-intervention and active intervention control groups, respectively. Importantly, VR interventions were more effective in ameliorating psychological disorders, syndromes or behaviours than active interventions, a finding that remained unaffected via publication bias analysis. As such, the current review clearly supports VR interventions as efficacious forms of psychological treatment and as a promising addition to existing treatment options.

VR Based Intervention Effectiveness

Thirty-one studies were included in the meta-analysis. Despite a significant meta-analytic multiple regression analysis, only form of VR based intervention was observed as an individual moderating variables. Though this finding suggests that effect sizes may differ depending on the form of VR intervention being employed, alternative explanations for these effect sizes (e.g., weak active intervention responses, expectancy biases, allegiance effects and outcome measure bias; Powers and Emmelkamp, 2008) cannot yet be ruled out. However, the present findings nonetheless indicate that the relative effectiveness of VR interventions are not simply limited to exposure based applications.

Methodological Rigour

No associations were found between treatment efficacy and methodological rigour, with methodological reporting observed to be independent of effect size. This would suggest that VR intervention outcomes remain significant and substantial in the face of sub-optimal methodological reporting, and lend further weight to their use. However, continued development in this field may be assisted by adoption of more stringent standards of methodological rigour.

Considerable variation existed in the level of detail reported across the reviewed studies. Some variation may be due to differences in journal reporting criteria (Kane, Wang, & Garrard, 2007; Ladd, McCrady, Manuel, & Campbell, 2010; Plint et al., 2006) and to methodological developments made between 1995 and 2013 (Ladd et al., 2010). Nevertheless, examination of the methodologies used to test and report VR interventions clearly demonstrates that numerous relevant factors are still not being reported adequately.

Individual differences in computer and VR familiarity are often seen to act as mediators or moderators of intervention efficacy and completion (Hamilton et al., 2011; Schulz et al., 2010), yet went largely unreported. Subsequently, it is strongly recommended that they be reported in subsequent VR intervention studies.

The replicability of several of the reviewed studies was unnecessarily hampered, with studies failing to include screenshots or sufficiently describe technical details of graphical engines and VEs employed. As differences in visual presentation may result in subsequent study replications and extensions obtaining erroneously dissimilar results, this information should be reported in future studies.

As numerous VR intervention studies are conducted in order to assess commercial VR systems in which researchers themselves possess financial interests, omissions of information regarding conflicts of interest should be avoided. If VR is to move into mainstream clinical practice, it is important that clinicians are confident that VR interventions has demonstrable benefit for clients. Reporting of potential bias by researchers will not only improve transparency in research, but may serve to allay concerns regarding potential bias.

Use of CONSORT Statement and CONSORT-EHEALTH Statement items included in our analyses checklists will aid subsequent replications and permit detailed evaluations of reliability and validity, fostering a greater understanding, acceptance and knowledge of the variety, effectiveness and efficacy of VR interventions.

Associations between Treatment Efficacy and Methodological Rigour

Our finding that treatment efficacy was not associated with methodological rigour was unexpected, and may raise questions concerning the importance and influence that methodological rigour has on interpreting treatment outcomes.

However, there may be possible explanations for this finding. The potential bias of researcher allegiance, a researcher's belief in the superiority of their treatment (Munder et al., 2013), may pose a substantial threat to internal validity. Researcher allegiance has been regarded as a determinant of treatment outcome (Munder et al., 2013). This is concerning, not least because several VR intervention studies are conducted in order to assess commercial VR systems. Adoption of researcher allegiance minimisation strategies such as mixed allegiance therapists would aid in negating potential bias (Leykin & DeRubeis, 2009; Luborsky et al., 1999; and Munder et al., 2013), as would the use of trial registries, declarations of conflicts of interest and affiliations (Moher et al., 2001).

Limitations

There are several limitations to this study. Firstly, these analyses were based upon the limited number of studies that met the inclusion criteria of RCTs. More studies that meet the gold standard of RCTS are required to further substantiate the effectiveness of VR based psychological interventions. However, this may be in part due to a second limitation. Though Ovid, ScienceDirect (Psychology) and Proquest Psychology provide substantial coverage of literature in the field, databases such as Web of Science and Cochrane Central Register of Controlled Trials were not utilised. Review of these and similar databases may result in a more comprehensive analysis. Thirdly, assessment of methodological rigour was conducted by only a single coder. As the use of multiple coders is necessary to ensure inter-rater agreement over methodological rigour, the current study's findings may not be free from bias.

Fourthly, variations in both the forms and number of VR interventions reviewed, disorders targeted and active intervention control groups suggests a need for greater analyses of these specific differences, as variations in mechanisms of action, treatment resistance, delivery, cohort, and other factors may mediate outcomes (Kazdin, 2007). Fifthly, this review purposely targeted a broad range of VR interventions in order to present a comprehensive view of current research. As VR continues to develop, more specific areas within VR intervention will need to be reviewed (e.g. specific clinical/subclinical samples, proposed mechanisms of action, target populations).

Finally, the effect of 'presence', the extent to which VR feels realistic (Alsina-Jurnet, et al., 2011; Hoffman et al., 2004), was not assessed. Although evidence supporting the role of presence in improving outcomes is mixed (Powers & Emmelkamp, 2008; Riva et al., 2011), its importance to VR intervention can only be determined if future studies routinely incorporate measures of presence.

Conclusion

It is clear that VR psychological interventions show considerable promise, allowing clients to engage in novel, highly interactive and effective adaptations of active face to face interventions. However, larger sample sizes, more control intervention comparisons, replication studies, and measures of potentially important variables such as presence are required to develop the field of VR based interventions.

Reporting guidelines such as those outlined in the CONSORT and CONSORT-EHEALTH statements will promote a greater emphasis on methodological rigour in the reporting of RCT studies, adding further support to the development of VR applications in clinical settings. With the imminent release of affordable VR systems, it is likely that the future will see a new wave of VR intervention development. For this reason, it is timely to consolidate and build on these promising results to guide future research.

Declaration of Conflicting Interests

The authors declare no potential conflict of interest with respect to funding, authorship and/or publication of this article.

References

Adamovich, S. V., Fluet, G. G., Tunik, E., & Merians, A. S. (2009). Sensorimotor training in virtual reality: a review. *NeuroRehabilitation, 25*(1), 29-44. doi: 10.3233/NRE-2009-0497

Alsina-Jurnet, I., Gutiérrez-Maldonado, J., & Rangel-Gómez, M.-V. (2011). The role of presence in the level of anxiety experienced in clinical virtual environments. *Computers in Human Behavior, 27*(1), 504-512. doi: 10.1016/j.chb.2010.09.018

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. doi: 10.1016/j.ijhcs.2011.06.002

Bohil, C. J., Alicea, B., & Biocca, f. A. (2011). Virtual reality in neuroscience research and therapy. *Nature Reviews Neuroscience, 12*, 752-762. doi: 10.1038/nrn3122

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

Boutron, I., Moher, D., Altman, D. G., Schulz, K., & Ravaud, P. C. G. (2008). Extending the CONSORT Statement to randomized trials of nonpharmacologic treatment: explanation and elaboration. *Annals of Internal Medicine*, *148*(4), 295-309. Retrieved from http://annals.org/

Carlbring, P., & Andersson, G. (2006). Internet and psychological treatment. How well can they be combined? *Computers in Human Behavior, 22*(3), 545-553. doi: 10.1016/j.chb.2004.10.009

CASP International Network (2014, March 3). CASP Appraisal Tools. Retrieved from http://www.caspinternational.org/?o=1012

Centre of Evidence-Based Physiotherapy (2014, March 3). Pedro: Physiotherapy Evidence Database. Retrieved from http://www.predo.org.au

Clough, B. A., & Casey, L. M. (2011). Technological adjuncts to increase adherence to therapy: a review. *Clinical Psychology Review, 31*(5), 697-710. doi: 10.1016/j.cpr.2011.03.006

Cohen, J. (1988). *Statistical Power for the Behavioral Sciences (2nd Edition)*. Hillsdale, New Jersey: Lawrence Eribaum Associates, Inc.

Cox, D. J., Davis, M., Sing, H., Barbour, B., Nidiffer, F. D., Trudel, T., Moncrief, R. & Moncrief, R. (2010). Driving rehabilitation for military personnel recovering from traumatic brain injury using virtual reality driving simulation. *Military Medicine, 175*(6), 411-416. Retrived from http://www.amsus.org/index.php/journal

Eysenbach, G. (2002). Issues in evaluating health websites in an Internet-based randomized controlled trial. *Journal of Medical Internet Research, 4*(3), e17. doi: 10.2196/jmir.4.3.e17

Eysenbach, G., & CONSORT-EHEALTH Group. (2011). CONSORT-EHEALTH: improving and standardizing evaluation reports of Web-based and mobile health interventions. *Journal of Medical Internet Research, 13*(4), 126. doi: 10.2196/jmir.1923

Furukawa, T. A., Noma, H., Caldwell, D. M., Honyashiki, M., Shinohara, K., Imai, H., Chen, P., Hunot, V. & Churchill, R. (2014). Waiting list may be a nocebo condition in

psychotherapy trials: a contribution from network meta-analysis. *Acta Psychiatrica Scandinavica*, n/a-n/a. doi: 10.1111/acps.12275

Gorini, A., & Riva, G. (2008). Virtual reality in anxiety disorders: the past and the future. *Expert Review of Neurotherapeutics, 8*(2), 215-233. doi: 10.1586/14737175.8.2.215

Gregg, L., & Tarrier, N. (2007). Virtual reality in mental health : a review of the literature. *Social Psychiatry and Psychiatric Epidemiology, 42*(5), 343-354. doi: 10.1007/s00127-007-0173-4

Hamilton, S., Moore, A. M., Crane, D. R., & Payne, S. H. (2011). Psychotherapy dropouts: differences by modality, license, and DSM-IV diagnosis. *Journal of Marital and Family Therapy, 37*(3), 333-343. doi: 10.1111/j.1752-0606.2010.00204.x

Hoffman, H. G., Sharar, S. R., Coda, B., Everett, J. J., Ciol, M., Richards, T., & Patterson, D. R. (2004). Manipulating presence influences the magnitude of virtual reality analgesia. *Pain, 111*(1–2), 162-168. doi: 10.1016/j.pain.2004.06.013

International Centre for Allied Health Evidence. (n.d). *Critical Appraisal Tools*. Retrieved from http://www.unisa.edu.au/Research/Sansom-Institute-for-Health-Research/Research-at-the-Sansom/Research-Concentrations/Allied-Health-Evidence/Resources/CAT/

Kane, R. L., Wang, J., & Garrard, J. (2007). Reporting in randomized clinical trials improved after adoption of the CONSORT statement. *Journal of Clinical Epidemiology, 60*(3), 241-249. doi: 10.1016/j.jclinepi.2006.06.016

Kazdin, A. E. (2007). Mediators and mechanisms of change in psychotherapy research. *Annual Review of Clinical Psychology, 3*, 1-27. doi: 10.1146/annurev.clinpsy.3.022806.091432

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. doi: 10.1016/s0005-7967(03)00139-6

Knüppel, H., Metz, C., Meerpohl, J. J., & Strech, D. (2013). How psychiatry journals support the unbiased translation of clinical research. A cross-sectional study of editorial policies. *PLoS One, 8*(10) doi: http://dx.doi.org/10.1371/journal.pone.0075995

Ladd, B. O., McCrady, B. S., Manuel, J. K., & Campbell, W. (2010). Improving the quality of reporting alcohol outcome studies: Effects of the CONSORT statement. *Addictive Behaviors, 35*(7), 660. doi: 10.1016/j.addbeh.2010.02.009

Leykin, Y. & DeRubeis, R.J. (2009). Allegiance in psychotherapy outcome research: Separating association from bias. *Clinical Psychology: Science and Practice, 16(1),* 54-65. doi: 10.1111/j.1468-2850.2009.01143.x

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) doi:http://dx.doi.org/10.1371/journal.pone.0096144

Lipsey, M. W. & Wilson, D. (2000). . Practical Meta-Analysis. Thousand oaks, California: Sage Publications.

Luborsky, L., Diguer, L., Singer, B., Dickter, D. & Schmidt, K. (1999). The researcher's own therapy allegiances: A 'wild card' in comparisons of treatment efficacy. *Clinical Psychology: Science and Practice, 6(1),* 95-100. doi: 10.1093/clipsy.6.1.95

Moher, D. (1998). Consort: An evolving tool to help improve the quality of reports of randomized controlled trials. *JAMA: The Journal of the American Medical Association, 279*(18), 1489-1491. doi: 10.1001/jama.279.18.1489

Moher, D., Liberati, A., Tetzlaff , J. & Altman, D.G. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *BMJ, 339,* 332-336. doi: 10.1136/bmj.b2535

Moher, D., Hopewell, S., Schulz, K. F., Montori, V., Gøtzsche, P. C., Devereaux, P. J., Elbourne, E., Egger, M. & Altman, D. G. (2010). CONSORT 2010 Explanation and Elaboration: updated guidelines for reporting parallel group randomised trials. *British Medical Journal, 340*(c869). doi: 10.1136/bmj.c869

Moher, D., Schulz, K. F., & Altman, D. G. (2001). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials. *The Lancet, 357*(9263), 1191-1194. doi: 10.1016/s0140-6736(00)04337-3

Moher, D. C., Ho, J., Duffecy, J., Baron, K. G., Lehman, K. A., Jin, L., & Reifler, D. (2010). Perceived barriers to psychological treatments and their relationship to depression. *Journal of Clinical Psychology, 66*(4), 394-409. doi: 10.1002/jclp.20659

Muhlberger, A., Wiedemann, G., & Pauli, P. (2003). Efficacy of a one-session virtual reality exposure treatment for fear of flying. *Psychother Research, 13*(3), 323-336. doi: 10.1093/ptr/kpg030

Munder, T., Brutsch, O., Leonhart, R., gerger, H. & Barth, J. (2013). Researcher allegiance in psychotherapy outcome research: An overview of reviews. *Clinical Psychology Review, 33(4),* 501-511. doi: 10.1016/j.cpr.2013.02.002

Opris, D., Pintea, S., Garcia-Palacios, A., Botella, C., Szamoskozi, 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

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

Plint, A. C., Moher, D., Morrison, A., Schulz, K., Altman, D. G., Hill, C., & Gaboury, I. (2006). Does the CONSORT checklist improve the quality of reports of randomised controlled trials? A systematic review. *Medical Journal of Australia, 185*(5), 263-267. Retrieved from https://www.mja.com.au/

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

Ready, D. J., Gerardi, R. ZJ., 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. doi: 10.1089=cyber.2009.0239

Repetto, C., & Riva, G. (2011). From virtual reality to interreality in the treatment of anxiety disorders. *Neuropsychiatry, 1*(1), 31-43. doi: 10.2217/npy.11.5

Riva, G., Waterworth, J. A., Waterworth, E. L., & Mantovani, F. (2011). From intention to action: The role of presence. *New Ideas in Psychology, 29*(1), 24-37. doi: 10.1016/j.newideapsych.2009.11.002

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. doi: 10.1016/j.beth.2005.04.004

Rougeau, M. & Hawkins, M. (2013, June 22). Hands on: Oculus Rift review. [Review of the Virtual Reality hardware Oculus Rift by Oculus VR™]. Retrieved from http://www.techradar.com/au/reviews/gaming/gaming-accessories/oculus-rift-1123963/review

Schulz, K. F., Altman, D. G., & Moher, D. (2010). CONSORT 2010 Statement: Updated guidelines for reporting parallel group randomised trials. *Journal of Clinical Epidemiology, 63*(8), 834-840. doi: 10.1016/j.jclinepi.2010.02.005

Strech, D. (2012). Normative arguments and new solutions for the unbiased registration and publication of clinical trials.*Journal of Clinical Epidemiology, 65*(3), 276-81. doi: http://dx.doi.org/10.1016/j.jclinepi.2011.07.002

Wilson, D.B. (2005). Meta-analysis macros for SAS, SPSS, and Sata. Retrieved June, 30th, 2014, from http://mason.gmu.edu/~dwilsonb/ma.html

Appendix A - Meta-Analysis References

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. *J Consult Clin Psychol, 81*(5), 751.

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. doi: 10.1016/j.ijhcs.2011.06.002

Bordnick, P. S., Traylor, A. C., Carter, B. L., & Graap, K. M. (2012). A feasibility study of virtual reality-based coping skills training for nicotine dependence. *Research on Social Work Practice, 22*(3), 293-300. doi: 10.1177/1049731511426880

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

Cox, D. J., Davis, M., Sing, H., Barbour, B., Nidiffer, F. D., Trudel, T., Moncrief, R. & Moncrief, R. (2010). Driving rehabilitation for military personnel recovering from traumatic brain injury using virtual reality driving simulation. *Military Medicine, 175*(6), 411-416. Retrived from http://www.amsus.org/index.php/journal

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. doi: 10.1016/s0005-7967(01)00023-7

Garcia-Palacios, A., Hoffman, H., Carlin, A., Furness, T. A., & Botella, C. (2002). Virtual reality in the treatment of spider phobia: a controlled study. *Behaviour Research and Therapy, 40*(9), 983-993. doi: 10.1016/s0005-7967(01)00068-7

Gutiérrez, R. O., Galán del Río, F., Cano de la Cuerda, R., Alguacil Diego, I. M., González, R. A., & Page, J. C. M. (2013). A telerehabilitation program by virtual reality-video games improves balance and postural control in multiple sclerosis patients. *NeuroRehabilitation, 33*(4), 545-554. doi: 10.3233/nre-130995

Hoffman, H. G., Sharar, S. R., Coda, B., Everett, J. J., Ciol, M., Richards, T., & Patterson, D. R. (2004). Manipulating presence influences the magnitude of virtual reality analgesia. *Pain, 111*(1–2), 162-168. doi: 10.1016/j.pain.2004.06.013

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. doi: 10.1016/s0005-7967(03)00139-6

Krijn, M., Emmelkamp P. M. G., O´Lafsson R. P., Bouwman M., van Gerwen L. J., SpinhovenP.p, Schuemie M. J. & van der Mast C. A. P. G. (2007). Fear of flying treatment methods: Virtual reality exposure vs. cognitive behavioral therapy. *Aviation, Space, and Environmental Medicine, 78*(2), 121-128.Retrieved from http://www.ingentaconnect.com/content/asma/asem

Leibovici, V., Magora, F., Cohen, S., & Ingber, A. (2009). Effects of virtual reality immersion and audiovisual distraction techniques for patients with pruritus. *Pain Research & Management: The Journal of the Canadian Pain Society, 14*(4), 283-286. Retrieved from http://www.pulsus.com/pain/home.htm

Lotan, M., Yalon-Chamovitz, S., & Weiss, P. L. (2010). Virtual reality as means to improve physical fitness of individuals at a severe level of intellectual and developmental

disability. *Research in Developmental Disabilities, 31*(4), 869-874. doi: 10.1016/j.ridd.2010.01.010

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. doi: 10.1037//0022-006x.70.5.1112

Man, D. W., Chung, J. C., & Lee, G. Y. (2012). Evaluation of a virtual reality-based memory training programme for Hong Kong Chinese older adults with questionable dementia: a pilot study. *International Journal of Geriatric Psychiatry, 27*(5), 513-520. doi: 10.1002/gps.2746

McLay, R., Wood, D. P., Webb-Murphy, J. A., Spira, J. L., Wiederhold, M. D., Pyne, J M. & Wiederhold, B. K. (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, Behaviour, and Social Networking, 14*(4), 223-229. doi: 10.1089/cyber.2011.0003

Meyerbroeker, K., Morina, N., Kerkhof, G. A., & Emmelkamp, P. M. G. (2013). Virtual Reality Exposure Therapy Does Not Provide Any Additional Value in Agoraphobic Patients: A Randomized Controlled Trial. *Psychotherapy and Psychosomatics, 82*(3), 170-176. doi: http://dx.doi.org/10.1159/000342715

Park, K.-M., Ku, J., Choi, S.-H., Jang, H.-J., Park, J.-Y., Kim, S. I., & Kim, J.-J. (2011). A virtual reality application in role-plays of social skills training for schizophrenia: A randomized, controlled trial. *Psychiatry Research, 189*(2), 166-172. doi: 10.1016/j.psychres.2011.04.003

Price, M. & Anderson, P. L. (2012). Outcome expectancy as a predictor of treatment response in cognitive behavioral therapy for public speaking fears within social anxiety disorder. Psychotherapy (Chicago, III.), 49(2), 173-179. doi: 10.1037/a0024734

Ready, D. J., Gerardi, R. ZJ., 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. doi: 10.1089=cyber.2009.0239

Reid, D., & Campbell, K. (2006). The use of virtual reality with children with cerebral palsy: A pilot randomized trial. *Therapeutic Recreation Journal, 40*(4), 255-268. Retrieved from http://js.sagamorepub.com/trj

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. doi: 10.1016/j.beth.2005.04.004

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. doi: 10.1037//0022-006x.70.2.428

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. doi: 10.1037//0022-006x.68.6.1020

Rothbaum, B. O., Hodges, L. F., Kooper, R., Opdyke, D., & et al. (1995). Effectiveness of computer-generated (Virtual Reality) graded exposure in the treatment of acrophobia. *The American Journal of Psychiatry, 152*(4), 626-628. Retrieved from http://ajp.psychiatryonline.org/

Rus-Calafell, M., Gutierrez-Maldonado, J., Botella, C., & Banos, R. M. (2013). Virtual Reality Exposure and Imaginal Exposure in the Treatment of Fear of Flying: A Pilot

Study. *Behavior Modification, 37*(4), 568-590. doi: http://dx.doi.org/10.1177/0145445513482969

Safir, M. P., Wallach, H. S., & Bar-Zvi, M. (2012). Virtual reality cognitive-behavior therapy for public speaking anxiety: one-year follow-up. *Behavior Modification, 36*(2), 235-246. doi: 10.1177/0145445511429999

Tam, S.-F., Man, D. W.-K., Chan, Y.-P., Sze, P.-C., & Wong, C.-M. (2005). Evaluation of a computer-assisted, 2-d virtual reality system for training people with intellectual disabilities on how to shop. *Rehabilitation Psychology, 50*(3), 285-291. doi: 10.1037/0090-5550.50.3.285

Tortella-Feliu, M., Botella, C., Llabres, J., Breton-Lopez, J. M., del Amo, A. R., Banos, R. M., & Gelabert, J. M. (2011). Virtual reality versus computer-aided exposure treatments for fear of flying. *Behavior Modification, 35*(1), 3-30. doi: 10.1177/0145445510390801

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

Wuang, Y.-P., Chiang, C.-S., Su, C.-Y., & Wang, C.-C. (2011). Effectiveness of virtual reality using wii gaming technology in children with down syndrome. *Research in Devolpmental Disabilies, 32*(1), 312-321. doi: 10.1016/j.ridd.2010.10.002

Table 1
Reviewed Studies Descriptive Information and Calculated Effect Sizes

Author(s)	Year	Disorder, Syndrome or Behaviour	Intervention	*(n)*	Control	(*n)*	*k*	Provided/Calculated Effect Sizes (Adjusted Hedge's *g*)	Mean (Adjusted Hedge's *g)*
Rothbaum, Hodges, Kooper, Opdyke, Williford & North	1995	Specific Phobia	VR Exposure	10	Non-Intervention	7	*3*	2.69, 2.39, 1.01	*2.03*
Rothbaum, Hodges, Smith, Lee & Price	2000	Specific Phobia	VR Exposure	15	Non-Intervention	15	*3*	0.65, 1.07, 1.80	*1.17*
Rothbaum, Hodges, Smith, Lee & Price	2000	Specific Phobia	VR Exposure	15	In-vivo Exposure	15	*3*	0.44, 0.09, 0.13	*0.22*
Emmelkamp, Krijn, Hulsbosch, de Vries, Schuemie & van der Mast	2002	Specific Phobia	VR Exposure	16	In-vivo Exposure	13	*3*	0.11, 0.04, 0.26	*0.14*
Garcia-Palacios, Hoffman, Carlin, Furness & Botella	2002	Specific Phobia	VR Exposure	12	Non-Intervention	11	*5*	2.33, 0.87, 2.85, 2.39, 2.53	*2.19*
Maltby, Kirsch, Mayers & Allen	2002	Specific Phobia	VR Exposure	20	Non-Intervention	23	6	0.39, 0.47, 0.37, 0.82, 0.61, 0.69	0.56
Rothbaum, Hodges, Anderson, Price & Smith	2002	Specific Phobia	VR Exposure	15	In-vivo Exposure	15	3	0.14, 0.05, 0.06	0.08
Hoffman, Sharar, Coda, Everett, Coil, Richards & Patterson	2004	Pain	More Immersive VR Exposure	20	Alternative Technology	19	3	1.13, 1.47, 1.48	1.36
Krijn, Emmelkamp, Biemond, de Wilde de Ligny, Scheumie & van der Mast	2004	Specific Phobia	VR Exposure	17	Non-Intervention	11	4	1.17, 0.82, 0.79, 0.53	0.83
Tam, Man, Chan, Sze & Wong	2005	Intellectual Impairment	VR Skills Training	8	Treatment as Usual	8	1	0.63	0.63
Reid & Campbell	2006	Cerebral Palsy	VR Skills Training	19	Treatment as Usual	12	1	0.44	0.44
Rothbaum, Anderson, Zimand, Hodges, Lang &	2006	Specific Phobia	VR Exposure	29	Non-Intervention	25	3	0.48, 0.83, 1.00	0.77

Wilson									
Rothbaum, Anderson, Zimand, Hodges, Lang & Wilson	2006	Specific Phobia	VR Exposure	42	In-vivo Exposure	40	3	0.17, 0.24, 0.22	0.21
Botella, Garcia-Palacios, Villa, Banos, Quero, Alcaniz & Riva	2007	Panic Disorder	VR Exposure	12	Non-Intervention	13	9	2.00, 1.64, 1.93, 2.04, 1.83, 1.26, 1.44, 1.42, 1.64	1.69
Botella, Garcia-Palacios, Villa, Banos, Quero, Alcaniz & Riva	2007	Panic Disorder	VR Exposure	12	In-vivo Exposure	12	9	0.29, 0.19, 0.34, 0.24, 0.15, 0.13, 0.34, 0.25, 0.23	0.24
Krijn, Emmelkamp, Olafsson, Bouwman, van Gerwen, Spinhoven, Schuemie & van der Mast	2007	Specific Phobia	VR Exposure	30	Cognitive Therapy/Cognitive Behavioural Therapy	23	2	0.50, 0.35	0.43
Leibovici, Magora, Cohen & Ingber	2009	Pain	VR Exposure	12	Alternative Technology	12	2	0.63, 0.62	0.63
Wallach, Safir & Bar-Zvi	2009	Specific Phobia	VR CBT	28	Cognitive Therapy/Cognitive Behavioural Therapy	30	5	0.01, 0.14, 0.19, 0.46, 0.07	0.17
Wallach, Safir & Bar-Zvi	2009	Specific Phobia	VR CBT	28	Non-Intervention	30	5	0.29, 0.90, 0.95, 0.67, 0.43	0.65
Cox, Davis, Singh, Barbour, Nidiffer, Trudel, Mourant & Moncrief	2010	TBI	VR Skills Training	6	Non-Intervention	5	3	1.55, 0.55, 1.36	1.15
Lotan, Yalon-Chamovitz & Weiss	2010	Pain	VR OT-PT	20	Alternative Technology	24	1	2.21	2.21
Ready, Gerardi, Backscheider, Mascaro & Rothbaum	2010	PTSD-AD	VR Exposure	5	Treatment as Usual	4	5	0.85, 1.83, 0.88, 0.61, 2.04	1.24
Banos, Guillen, Quero, Garcia-Palacios, Alcaniz & Botella	2011	PTSD-AD	More Immersive VR Exposure	19	Cognitive Therapy/Cognitive Behavioural Therapy	20	3	0.09, 0.04, 0.35	0.16

McLay, Wood, Webb-Murphy, Spira, Wiederhold, Pyne & Wiederhold	2011	PTSD-AD	VR Exposure	10	Treatment as Usual	10	1	1.01	1.01
Park, Ku, Choi, Jang, Park, Kim & Kim	2011	Schizophrenia	VR Skills Training	33	Treatment as Usual	31	6	0.64, 0.62, 0.70, 0.13, 0.28, 0.37	0.46
Tortella-Feliu, Botella, Llabres, Breton-Lopez, Riera del Amo, Banos & Gelabert	2011	Specific Phobia	VR Exposure	12	Therapist Led Intervention	14	6	0.12, 1.88, 0.02, 0.09, 0.05, 0.53	0.45
Tortella-Feliu, Botella, Llabres, Breton-Lopez, Riera del Amo, Banos & Gelabert	2011	Specific Phobia	VR Exposure	12	Alternative Technology	12	6	0.89, 2.56, 0.11, 0.06, 0.22, 0.40	0.71
Wuang, Chiang, Su & Wang	2011	Intellectual Impairment	VR OT-PT	52	Non-Intervention	50	17	1.58, 1.36, 1.23, 1.86, 2.40, 3.57, 4.57, 3.16, 3.34, 1.20, 3.38, 4.00, 4.93, 5.88, 2.20, 1.80, 0.42	2.76
Wuang, Chiang, Su & Wang	2011	Intellectual Impairment	VR OT-PT	52	Treatment as Usual	53	17	0.44, 0.53, 0.07, 0.76, 0.42, 0.52, 0.56, 0.70, 1.22, 1.67, 1.04, 1.11, 0.07, .099, 1.48, 2.41, 1.73	0.91
Bordnick, Traylor, Carter & Graap	2012	Addictions	VR Skills Training	11	Treatment as Usual	11	2	1.35, 0.79	1.07
Man, Chung & Lee	2012	Dementia	VR Skills Training	20	Therapist Led Intervention	24	7	0.97, 2.26, 0.96, 1.35, 0.14, 1.34, 0.51	1.08
Price & Anderson	2012	Specific Phobia	VR Exposure	33	In-vivo Exposure	34	3	0.63, 0.61, 0.24	0.49
Safir, Wallach & Bar-Zvi	2012	Specific Phobia	VR CBT	25	Cognitive Therapy/Cognitive Behavioural	24	5	0.10, 0.05, 0.08, 0.33, 0.02	0.12

					Therapy				
Anderson, Price, Edwards, Obasaju, Schmertiz, Zimand & Calamaras	2013	Social Anxiety	VR Exposure	32	Non-intervention	21	4	1.45, 0.58, 0.65, 0.85	0.88
Anderson, Price, Edwards, Obasaju, Schmertiz, Zimand & Calamaras	2013	Social Anxiety	VR Exposure	26	In-vivo Exposure	24	2	0.69, 0.00	0.35
Gutiérrez, Galán del Río, Cano de la Cuerda, Alguacil Diego, González & Page	2013	Multiple Sclerosis	VR OT-PT	24	Treatment as Usual	23	6	1.36, 0.58, 2.38, 2.29, 0.56, 1.04	1.37
Meyerbroeker, Morina, Kerkhof & Emmelkamp	2013	Specific Phobia	VR exposure	23-24	In-vivo Exposure	21-22	4	0.27, 0.22, 0.10, 0.01	0.15
Rus-Calafell, Gutierrez-Maldonado, Botella & Banos	2013	Specific Phobia	VR Exposure	7	Imaginal Exposure	8	3	0.42, 0.09, 0.78	0.43

k = number of effect sizes calculated per study.

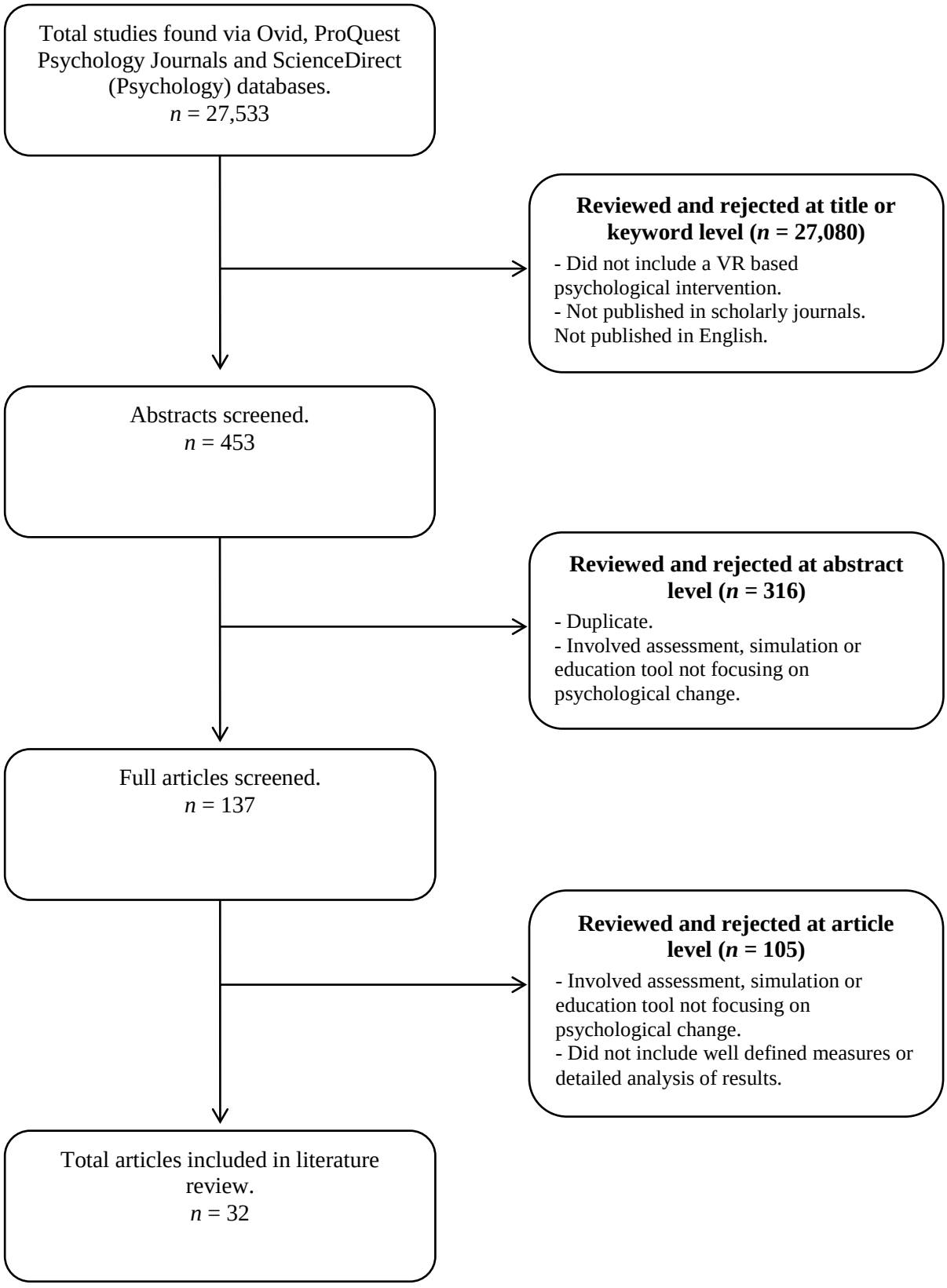

Figure 1. *Virtual reality intervention literature search and study selection.*

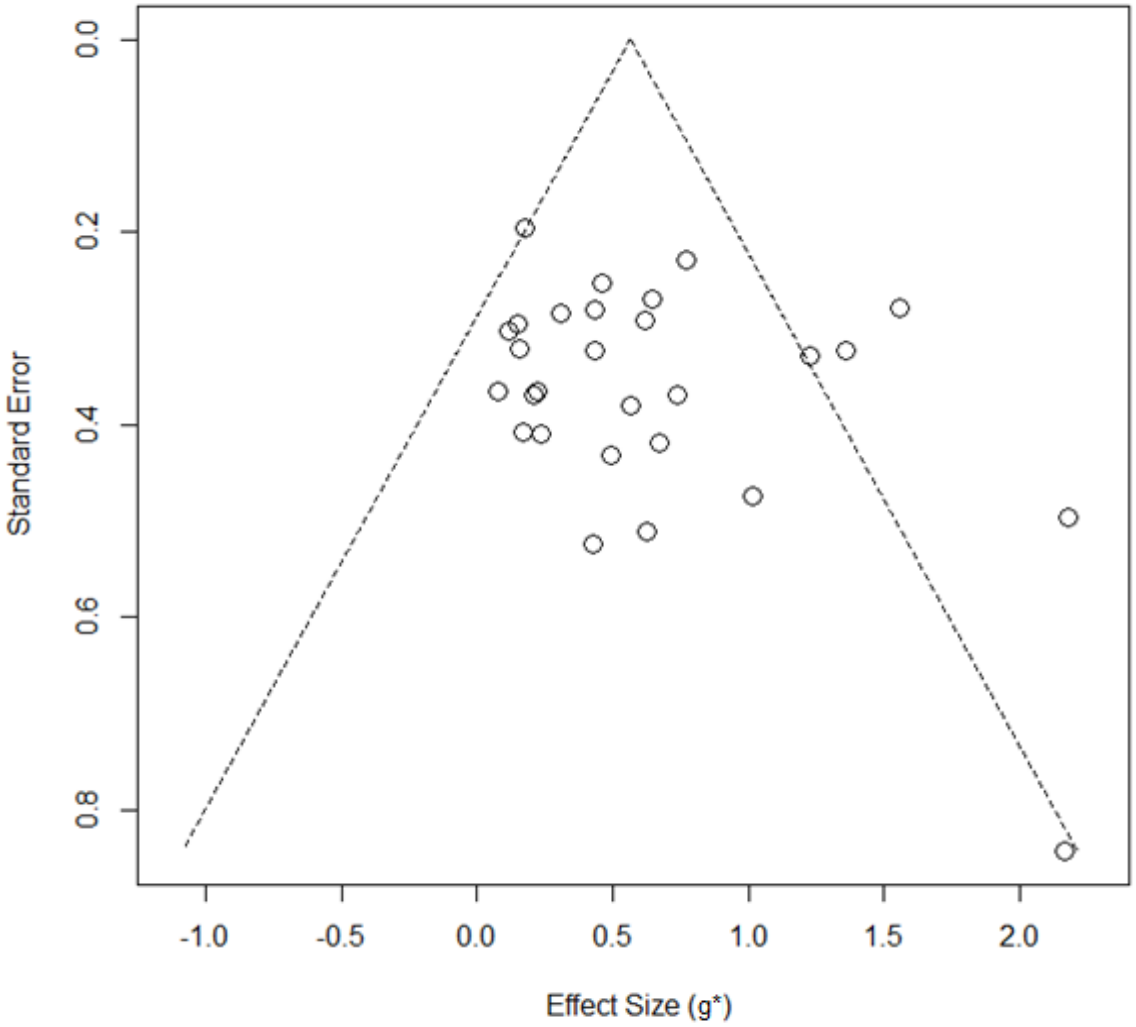

Figure 2. *Potential publication bias funnel plot for reviewed studies.*

Table 2
Stem-and-Leaf Plot of All Effect Sizes

Stem	Leaf
0.	0, 1, 1, 1, 1, 1, 1, 2, 2, 2, 3, 4, 4, 4, 4, 4, 4
0.	5, 6, 6, 6, 7, 7, 8, 8, 9
1.	0, 0, 1, 1, 1, 2, 3, 3
2.	0, 1, 2
2.	8

Table 3
Stem-and-Leaf Plot of Non-Intervention Wait-List Comparison Effect Sizes

Stem	Leaf
0.	1
0.	5, 6, 7, 8, 8, 9
1.	1, 1
2.	0, 1

Table 4
Stem-and-Leaf Plot of Alternative Intervention Comparison Types' Effect Sizes

Alternative Intervention Comparison Type	Stem	Leaf
Alternative Technology	0.	6, 7
	1.	3
	2.	2
Cognitive Therapy	0.	1, 1, 1
/Cognitive Behavioural Therapy	0.	4
Imaginal Exposure	0.	4
In-Vivo Exposure	0.	0, 1, 1, 2, 2, 2
	0.	3, 4
Therapist Led Intervention	0.	4
	1.	0
Treatment As Usual	0.	4, 4
	0.	6
	0.	0, 1, 2, 3
	2.	8

Table 5
Individual Random Effects Meta-Analyses by Non-Intervention and Comparative Intervention Forms of Control

	k	*Q*	*g**	95% *CI*	*z*
Non-Intervention	11	6.70	0.95	0.55, 1.35	4.71***
Comparative-Intervention	27	27.52	0.67	0.43, 0.92	5.34***

*** $p < .001$, one-tailed.

Table 5. *Reviewed Virtual Reality Studies Scored Against the CONSORT 2010 and CONSORT-EHEALTH 2011 Checklists*

CONSORT Item	Item Description	Studies including item (*N* = 31)
Item 1a	Identification as a randomised trial in title	6
Item 1b	Structured summary of trial design, methods, results, and conclusions	28
Item 2a	Scientific background and explanation of rationale	29
Item 2b	Specific objectives or hypotheses	24
Item 3a	Description of trial design (such as parallel, factorial) including allocation ratio	25
Item 3b	Important changes to methods after trial commencement (such as eligibility criteria), with reasons	2
Item 4a	Eligibility criteria for participants	28
Item 4a i[a]	Computer / Internet literacy	3
Item 4a ii[a]	Open vs. closed, web-based vs. face-to-face assessments	21
Item 4a iii[a]	Information given during recruitment	14
Item 4b	Settings and locations where the data were collected	13
Item 4b i[a]	Clearly report if outcomes were (self-)assessed through online questionnaires (as common in web-based trials) or otherwise	24
Item 4b ii[a]	Report how institutional affiliations are displayed to potential participants (describe only if this may bias results)	0
Item 5	The interventions for each group with sufficient details to allow replication, including how and when they were actually administered	30
Item 5 i[a]	If authors/evaluators are owners or developer of the software, this needs to be declared in a “conflict of interest” section or mentioned elsewhere in the manuscript	8
Item 5 ii[a]	Describe the history/development process of the application	1
Item 5 iii[a]	Revisions and updating	0
Item 5 iv[a]	Provide information on quality assurance methods	4
Item 5 v[a]	Ensure replicability (sourcecode, screenshots, flowchart, etc.)	10
Item 5 vi[a]	Digital preservation (url, digital archive, etc.)	2
Item 5 vii[a]	Describe how participants accessed the application, in what setting/context, if they had to pay (or were paid) or not, whether they had to be a member of specific group	23
Item 5 viii[a]	Describe mode of delivery, features/functionalities/components of the intervention and comparator, and the theoretical framework	27
Item 5 ix[a]	Clarify what instructions or recommendations were given to the user, e.g., regarding timing, frequency, if any	20
Item 5 x[a]	Clarify the level of human involvement	28
Item 5 xi[a]	Report any prompts/reminders used	1
Item 5 xii[a]	Describe any co-interventions (incl. training/support)	18
Item 6a	Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed	28

Item 6a ii[a]	Describe whether and how “use” (including intensity of use/dosage) was defined/measured/monitored	27
Item 6a iii[a]	Describe whether, how and when qualitative feedback was obtained from participants	14
Item 6b	Any changes to trial outcomes after the trial commenced, with reasons	3
Item 7a	How sample size was determined	11
Item 7 a i[a]	Describe whether and how expected attrition was taken into account when calculating the sample size	8
Item 7b	When applicable, explanation of any interim analyses and stopping guidelines	5
Item 8a	Method used to generate the random allocation sequence	5
Item 8b	Type of randomisation; details of any restriction (such as blocking and block size)	5
Item 9	Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned	5
Item 10	Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions	3
Item 11a	If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how	9
Item 11a i[a]	Specify who was blinded, and who wasn't	10
Item 11a ii[a]	Discuss whether participants knew which intervention was the “intervention of interest” and which one was the “comparator”	3
Item 11b	If relevant, description of the similarity of interventions	16
Item 12a	Statistical methods used to compare groups for primary and secondary outcomes	29
Item 12 a[a]	Specify how participants who did not use the application or dropped out from the trial were treated in the statistical analysis	17
Item 12b	Methods for additional analyses, such as subgroup analyses and adjusted analyses	24
Item 13a	For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome	27
Item 13b	For each group, losses and exclusions after randomisation, together with reasons	28
Item 14a	Dates defining the periods of recruitment and follow-up	3
Item 14b	Why the trial ended or was stopped	21
Item 15	A table showing baseline demographic and clinical characteristics for each group	13
Item 15 i[a]	Report demographics associated with digital divide issues, such as age, education, gender, social-economic status, computer/Internet/eHealth literacy of the participants, if known	4
Item 16	For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups	27
Item 16 i[a]	Report N’s (and effect sizes) “across a range of study participation [and use] thresholds”	24

Item 16 ii[a]	Primary analysis should be intent-to-treat; secondary analyses could include comparing only “users”	30
Item 17a	For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence intervals)	29
Item 17b	For binary outcomes, presentation of both absolute and relative effect sizes is recommended	4
Item 18	Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory	24
Item 19	All important harms or unintended effects in each group	3
Item 19 i[a]	Include privacy breaches, technical problems	1
Item 19 ii[a]	Include qualitative feedback from participants or observations from staff/researchers	12
Item 20	Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses	29
Item 21	Generalisability (external validity, applicability) of the trial findings	29
Item 22	Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence	30
Item 22 i[a]	Restate study questions and summarize the answers suggested by the data, starting with primary outcomes and process outcomes (use)	28
Item 22 ii[a]	Highlight unanswered new questions, suggest future research	24
Item 23	Registration number and name of trial registry	3
item 24	Where the full trial protocol can be accessed, if available	2
Item 25	Sources of funding and other support (such as supply of drugs), role of funders	20
Item 27[a]	State the “relation of the study team towards the system being evaluated”	10

[a] = CONSORT-EHEALTH 2011 checklist specific item