Journal of Anxiety Disorders xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Journal of Anxiety Disorders

journal homepage: www.elsevier.com/locate/janxdis

Deterioration rates in Virtual Reality Therapy: An individual patient data level meta-analysis

Javier Fernández-Álvarez[a,*], Alexander Rozental[b,c], Per Carlbring[d], Desirée Colombo[e], Giuseppe Riva[a,f], Page L. Anderson[g], Rosa María Baños[h,i], Amanda A. Benbow[g], Stéphane Bouchard[j], Juana María Bretón-López[e,i], Georgina Cárdenas[k], JoAnn Difede[l], Paul Emmelkamp[m], Azucena García-Palacios[e,i], Verónica Guillén[h,i], Hunter Hoffman[n], Isabel Kampann[m], Ramona Moldovan[o], Andreas Mühlberger[p], Max North[q], Paul Pauli[r], Wenceslao Peñate Castro[s], Soledad Quero[e,i], Miquel Tortella-Feliu[t], Kataryzna Wyka[l], Cristina Botella[e,i]

[a] Università Cattolica del Sacro Cuore, Milan, Italy
[b] Centre for Psychiatry Research, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden
[c] Institute of Child Health, University College London, London, United Kingdom
[d] Division of Clinical Psychology, Department of Psychology, Stockholm University, Stockholm, Sweden
[e] Department of Basic and Clinical Psychology and Psychobiology, Jaume I University. Castellon, Spain
[f] Applied Technology for Neuro-Psychology Laboratory, Instituto Auxologico Italiano, Milan, Italy
[g] Department of Psychology, Georgia State University, Atlanta, United States
[h] Department of Personality, Assessment, and Psychological Treatments, Valencia University, Valencia, Spain
[i] CIBER of Physiopathology of Obesity and Nutrition CIBERobn, CB06/03 Instituto de Salud Carlos III, Spain
[j] Département de psychoéducation et de psychologie, Université du Québec en Outaouais, & Centre Intégré de Santé et de Services Sociaux de l'Outaouais, Gatineau, Québec, Canada
[k] School of Psychology, National Autonomous University of Mexico, Mexico City, Mexico
[l] Department of Psychiatry, Weill Medical College of Cornell University, New York, United States
[m] Department of Clinical Psychology, University of Amsterdam, Amsterdam, The Netherlands
[n] Virtual Reality Research Center at the Human Photonics Lab, Mechanical Engineering, University of Washington, Seattle, United States
[o] Department of Psychology, Babes-Bolyai University, Cluj-Napoca, Romania
[p] Department of Psychology, University of Regensburg, Regensburg, Germany
[q] Visualization and Simulation Research Cluster, Information Systems Department, Kennesaw State University, Georgia, United States
[r] Center of Mental Health, Department of Biological Psychology, Clinical Psychology, and Psychotherapy, Julius-Maximilians-Universität Würzburg, Würzburg, Germany
[s] Facultad de Ciencias de la Salud – Sección Psicología, Universidad de La Laguna, Tenerife, Spain
[t] University Research Institute on Health Sciences - IUNICS, University of the Balearic Islands, Palma (Mallorca), Spain

ARTICLE INFO

Keywords:
Virtual reality
Anxiety disorders
Deterioration rates
Individual patient data analysis

ABSTRACT

Ample evidence supports the use of Virtual Reality (VR) for anxiety disorders. Nonetheless, currently there is no evidence about moderators or potential negative effects of VR treatment strategies. An Individual Patient Data (IPD) approach was employed with 15 retrieved datasets. The current study sample was composed of 810 patients. Randomized control trials (RCTs) for each primary outcome measure were performed, in addition to moderator analyses of the socio-demographic variables. Deterioration rates were 14 patients (4.0%) in VR, 8 (2.8%) in active control conditions, and 27 (15%) in the WL condition. With regard to receiving treatment, patients in a waiting list control condition had greater odds of deteriorating than in the two active conditions, odds ratios (*ORs*) 4.87, 95% confidence interval (CI) [0.05, 0.67]. In the case of the socio-demographic variables, none of them were associated with higher or lower odds of deterioration, with the exception of marital status in the WL condition; married people presented a significantly lower probability of deterioration, *OR* 0.19, 95% CI [0.05, 0.67]. Finally, when comparing pooled effects of VR versus all control conditions, the OR was 0.61 (95% CI 0.31–1.23) in favor of VR, although this result was not statistically significant. This study provides evidence about the deterioration rates of a therapeutic VR approach, showing that the number of deteriorated patients coincides with other therapeutic approaches, and that deterioration is less likely to occur, compared to patients in WL control groups.

* Corresponding author.
E-mail address: javier.fernandezkirszman@unicatt.it (J. Fernández-Álvarez).

https://doi.org/10.1016/j.janxdis.2018.06.005
Received 20 December 2017; Received in revised form 23 May 2018; Accepted 18 June 2018

1. Introduction

It has been well established that psychological treatments are effective for different existing clinical conditions comprised under the umbrella of *anxiety disorders* (Emmelkamp & Ehring, 2014). They include social anxiety disorders (Hofmann & Di Bartolo, 2014), generalized anxiety disorder (Cuijpers et al., 2014), panic disorder and agoraphobia (Pompoli et al., 2016), specific phobias (Wolitzky-Taylor, Horowitz, Powers, & Telch, 2008), and stress-related disorders, such as post-traumatic stress disorder (Cusack et al., 2016).

However, the majority of this evidence is based on results of efficacy and face-to-face approaches. That is, conducting randomized control trials in non-naturalistic settings, which principally are tested in classical modalities in which one therapist attends one person. In this regard, the number of studies carried out in the past 20 years within the field of virtual reality (VR) has signified a great leap forward in incorporating technological advancements into psychological treatments. Indeed, VR is an illustrative example of how the potential of traditional face-to-face approaches can be enhanced, coinciding with what Kazdin and Blase (2011) described as *rebooting psychotherapy* (2011).

A large body of evidence has shown the wide range of advantages of the use of VR for the treatment of anxiety disorders, including more ecological, personalized, and controlled assessments and interventions, an increase in the acceptability of treatments, and, consequently, greater adherence (Botella et al., 1998; Lindner et al., 2017; Riva, 2005). Hence, VR has become a widely-used tool to provide patients with less invasive, and in some cases more powerful, interventions for a wide range of psychopathological conditions (Botella, Fernández-Álvarez, Guillén, García-Palacios, & Baños, 2017). Anxiety disorders, and particularly specific phobias, have become paradigmatic in VR implementation because exposure is undoubtedly the main specific element that must be addressed; VR provides the ability to increase exposure (Wiederhold & Bouchard, 2014; Morina, Ijntema, Meyerbröker, & Emmelkamp, 2015; Opriş et al., 2012; Powers & Emmelkamp, 2008).

1.1. The unwanted shall not be avoided

Although all the aforementioned evidence undoubtedly supports the efficacy of VR-based treatments for anxiety disorders, agreeing with the clinical scientific tradition, there has been a long-standing bias toward showing what is effective instead of recognizing possible negative effects (Barlow, 2010; Castonguay, Boswell, Constantino, Goldfried, & Hill, 2010). In fact, a systematic review investigating the extent to which harm was reported in randomized control trials in clinical psychology indicated that only 21% of the analyzed trials reported harm on a patient level, and only 3% explicitly described procedures to analyze harmful treatments (Jonsson, Alaie, Parling, & Arnberg, 2014).

This is a major challenge that has been undermined by the research, and, therefore, clinicians rightly perceive a lack of tools and mechanisms to measure, prevent, and deal with negative effects (Bystedt, Rozental, Andersson, Boettcher, & Carlbring, 2014; Peterson, Roache, Raj, & Young-McCaughan, 2013). Moreover, negative effects not only consist of iatrogenic effects and the vast array of worsening effects that a treatment can entail, but also the fact that a patient may not experience any benefits from a psychological treatment. For several reasons, an inert therapeutic process can be understood as a negative effect because it can make the patient more treatment resistant, create negative treatment expectations in the future, and increase the economic costs for the healthcare system and/or the patient (Dimidjian & Hollon, 2010).

1.2. New promising trend

Although there has been some sporadic research on negative effects from the 1960s onward (Bergin, 1963; Mohr, 1995; Strupp, Hadley, & Gomes-Schwartz, 1977), only in the past fifteen years has this topic sparked an interest within the fields of psychotherapy and clinical psychology. Illustrative examples are studies focused on operationalizing the concept of negative effects, which has led to creating a more accurate taxonomy of the intervening variables and the different subtypes that are assumed to occur during treatment (Linden & Schermuly-Haupt, 2014; Parry, Crawford, & Duggan, 2016).

While it may be true that no clear consensus exists over the classification of negative effects, Linden (2013) presented a thorough examination of the topic, in which seven types of side effects were described in relation to a checklist. First, unwanted events (UE) are defined as all kind of negative effect that occur at any moment of treatment; treatment-emergent reaction is conceived as an UE provoked by the treatment; adverse treatment reaction is any UE probably caused by a correctly administered treatment; malpractice reaction is understood as the consequence of an improperly applied treatment; treatment non-response is the lack of improvement throughout a therapeutic process; deterioration of illness, which is conceptualized as a worsening in symptoms at any moment during or after treatment; therapeutic risk, that is, the known adverse treatment reactions that a treatment can entail; and contraindications.

In addition, assessing each of the abovementioned aspects can be done in different ways. For instance, there are diverse areas to look for those negative effects, such as the appearance of novel symptoms, the deterioration of the existing symptoms, the decrease of social relationships and functioning, the dependency in the relationship with the therapist, among others.

Furthermore, several self-report measures have been developed over the years (e.g. Hatfield, McCullough, Frantz, & Krieger, 2010; Ladwig, Rief, & Nestoriuc, 2014). Among the developed instruments are, for instance, the Vanderbilt Negative Indicator Scale (VNIS; Suh, Strupp, & O'Malley, 1986), the Inventory for the Assessment of Negative Effects in Psychotherapy (INEP; Ladwig et al., 2014) the Unwanted Events and Adverse Treatment Reaction Checklist for Psychotherapy (UE-ATR; Linden, 2013), the Experience of Therapy Questionnaire (ETQ; Parker et al., 2013) and the Negative Effects Questionnaire (NEQ; Rozental, Kottorp, Boettcher, Andersson, & Carlbring, 2016). However, as explained in detail by Rozental et al. (2016), each scale has its own theoretical background as well as strengths and weaknesses. For example, the VNIS was based on a solid theory, but lacked practical utility given its comprehensive rating system. Similarly, the UE-ATR was proposed as a tool to help therapists detect negative effects, but was never intended to be used as a scale with psychometric properties. Moreover, the ETQ and the INEP both enjoys stronger empirical support, but the former includes items of both positive and negative nature, while the latter is comprised of items that are more related to malpractice than negative effects of properly implemented treatments.

Another important advancement with regard to negative effects has been the study of trajectories of change during treatment. In this vein, a key issue has been the increasing incorporation of routine outcome monitoring via algorithms to identify patterns of deterioration that therapists may not detect using clinical judgement alone. This technique has been used successfully in both research and clinical settings as a way of tracking patients throughout treatment in order to prevent trajectories of worsening (Lambert & Shimokawa, 2011; Lutz, De Jong, & Rubel, 2015; Whipple & Lambert, 2011). In this direction, a number of tools, such as the Outcome Questionnaire 45 (Lambert et al., 2004), CORE-OM (Barkham et al., 2001), Treatment Outcome Package (Kraus, Seligman, & Jordan, 2005) or the PCOMS (Miller, Duncan, Sorrell, & Brown, 2005) among others, have been developed (for a discussion on the topic see Boswell, Kraus, Miller, & Lambert, 2015). In addition, both quantitative (e.g. Nordberg, Castonguay, Mcaleavey, Locke, & Hayes, 2016) and qualitative research (Solstad, Castonguay, & Moltu, 2017) supports the use of routine outcome monitoring as a more objective way of determining how the patients are progressing through the therapeutic process.

Table 1
Types of negative effects and assessment tools.

Negative effects	Main target	Assessment tool
Unwanted events and adverse treatment reactions	General approaches to assess different facets of a treatment that directly or indirectly provoked a negative effect.	Unwanted Events and Adverse Treatment Reaction Checklist for Psychotherapy (Linden, 2013) Inventory for the Assessment of Negative Effects in Psychotherapy (Ladwig et al., 2014) Experience of Therapy Questionnaire (Parker et al., 2013) Vanderbilt Negative Indicator (Suh et al., 1986) Negative Effects Questionnaire (Rozental et al., 2016)
Sudden losses	Identifying significant worsening processes between two sessions	Any symptomatic tool
Routine Outcome Monitoring	Identification of deterioration	OQ-45 (Lambert et al., 2004); CORE-OM (Barkham et al., 2001); TOP (Kraus et al., 2005); PCOMS (Miller et al., 2005)
Alliance Ruptures	Detecting breakdown in the therapeutic relationship	Alliance Negotiation Scale (Waizmann et al., 2015)

A further development that has permitted to better determine the nature of negative trajectories has been the conceptualization of sudden losses within psychotherapy treatments. Sudden losses derived from the reverse concept of sudden gains that have been extensively researched (Tang & DeRubeis, 1999) and is therefore conceptualized as a significant worsening between two sessions. It has been identified as a marker of possible treatment failure (Lutz et al., 2013).

Finally, alliance ruptures, that is, a breakdown in the therapeutic relationship, not only is a relevant phenomenon to understand the trajectory of change but also as a key moment that may be taken to greatly improve the therapeutic outcomes by means of its repair (Safran, Muran, & Eubanks-Carter, 2011). Table 1 summarizes all the mentioned types of negative effects.

1.3. Negative effects in technology-mediated treatments

In a context where the implementation of technology is increasing exponentially in clinical research and practice (e.g. Botella et al., 2012; Imel, Caperton, Tanana, & Atkins, 2017; Mohr, Weingardt, Reddy, & Schueller, 2017; Riva, 2005), negative effects should be further explored. Internet-based interventions are a positive exception to the rule because, due to the novelty of the approach, a considerable number of studies have been conducted (Carlbring, Andersson, Cuijpers, Riper., & Hedman-Lagerlöf, 2018 In press). Diverse qualitative endeavors have focused on the patients' experiences of negative effects (Rozental, Boettcher, Andersson, Schmidt, & Carlbring, 2015) and related constructs, such as non-adherence (Johansson, Michel, Andersson, & Paxling, 2015) or dropout experiences (Fernández-Álvarez et al., 2017). Recently, three high quality individual patient data (IPD) meta-analyses were published that assessed deterioration rates in Internet interventions (Ebert et al., 2016; Karyotaki et al., 2018; Rozental, Magnusson, Boettcher, Andersson, & Carlbring, 2017).

However, other domains within the technology-mediated realm have not explored the negative effects of the therapeutic processes. VR is a paradigmatic example because it emerged more than 20 years ago as a treatment option for different psychological disorders; however, very few studies have focused on the possible negative effects of the treatments. In the first years of the use of VR as a therapeutic tool, there was concern about possible negative effects that the technology could introduce, such as an effect on reality judgment, especially for severe mental disorders like schizophrenia (Baños, Botella, & Perpiña, 1999; Rizzo, Wiederhold, & Buckwalter, 1998). One decade later, other theoretical studies pointed out the need to take this topic into account (Botella et al., 2007; Gregg & Tarrier, 2007); however, no empirical study has yet focused on the potential negative effects that might exist. The only exception is a study of a PTSD treatment that mentioned possible worsening due to the exposure. The assessed domains were in relation to risk of suicide and alcohol consumption, so it was not deterioration of the post-traumatic symptoms but exacerbation of other side effects (Beidel, Neer, Bowers, Frueh, & Rizzo, 2014). In addition, although not related to treatment outcomes, studies have explored more physiological side effects of using technological devices, which has been labeled cybersickness (Rebenitsch & Owen, 2016).

Given the scarcity of empirical studies on the negative effects of VR-based treatments, and the fact that the use of VR is expected to significantly increase due to emerging affordable models (e.g. Gear VR and Google Cardboard), it is time to explore the possible deleterious effects of VR-based treatments.

1.4. The current study

The main goal of the present study was to examine the deterioration rates in the existing randomized control trials for anxiety disorders conducted with VR. Among the different available options to assess negative effects, deterioration rates have been proposed as the most straightforward method to capture the phenomenon. While other ways of assessing negative effects can be of great clinical value, like the subjective experience of the participants or a decrease in interpersonal functioning, a through quantitative study is difficult given the lack of specific scales adapted to VR and the complex nature of the intervening variables (Rozental et al., 2017). Deterioration can be defined as a worsening in symptomatology, and given that the outcomes of treatment are always investigated in randomized control trials, the procedure for exploring the number of patients who are worse off is fairly easy to employ in clinical psychology.

2. Method

2.1. Eligibility criteria

Adults diagnosed with an anxiety disorder (including those categorized as anxiety disorders in the DSM-IV, e.g., PTSD and OCD) who received a manualized treatment that included VR and an active condition or a waiting list as control to the VR condition were eligible for inclusion. By validated protocol is meant a manualized treatment The VR condition had to have at least 10 patients, following the cut-off point identified by Meyerbröker and Emmelkamp (2010). Only papers published in peer-reviewed journals in English, Spanish, German, or Italian were considered for inclusion. Retrieved data also had to include values for the primary outcome measure for each individual patient because the assessment of deterioration is explored per patient.

2.2. Data items

P: adult population with anxiety disorders or stress related disorders
I: psychological intervention with VR[1]
C: non-VR active condition or waiting list

[1] One study was conducted with Augmented Reality (Botella et al., 2016), but because it lies on the mixed reality continuum, the decision was made to include it in the aggregated sample.

O: symptomatology
S: randomized control trials

2.3. Search strategy

First, through a systematic search of *Pubmed* and *Web of Science,* two reviewers independently searched potentially relevant articles. There was no restriction on the year of publication, and the last update of the search strings was carried out in March 2017. For every database consultation, specific keywords and mesh terms were used (see Appendix 1 for the string used for *Pubmed*). In all cases, our strings consisted of two components: (1) virtual reality and all synonyms; and (2) anxiety disorders and all related and specific terms.

Supplementary strategies were also applied. One involved manual searches of literature reviews and book chapters and their reference lists. The other involved examining references of all the included articles; relevant excluded articles were also manually examined.

2.4. Study selection process

All the articles were exported to Mendeley. The overall search yielded 1272 records, of which 646 were retrieved for further examination after eliminating duplicates. In the next step, the two reviewers independently read all the titles, removing all titles that were not relevant to the purpose of the systematic review (e.g. did not mention VR or related terms, or were not based on psychological interventions). There were no discrepancies between the two reviewers. After this step, 124 full-text articles were considered potentially relevant studies. Finally, 36 studies were found eligible according to the eligibility criteria.

2.5. Studies included in the IPD meta-analysis

Unlike standard meta-analyses, after systematically reviewing the existing literature on VR-based trials, all of the datasets were requested. From the obtained datasets, a new dataset was created with the aim of assessing the raw scores for each patient in order to achieve a more accurate examination, rather than one based on group means and standard deviations. In the case of deterioration rates, the individual patient level approach is the only way to determine their occurrence; these are the effects that were meta-analyzed.

The last author of this paper used email to contact all the corresponding authors of the 36 papers included. This e-mail contained the purpose of the study, and corresponding authors were invited to participate in the manuscript as co-authors, given the major contribution they were making by sharing the data.

Finally, 15 studies were included for the data analysis, as 21 datasets could not be retrieved from the authors for different reasons, including irretrievable data, impossibility of sharing the material due to legal constraints, or no answer from the corresponding authors (after 3 attempts). Fig. 1 presents a flow chart of the systematic review.

2.6. Quality Assessment of included studies

To assess the risk of bias, the *Cochrane Handbook for Systematic Review Interventions* (Higgins et al., 2011) was used. The following domains were assessed: (1) Random sequence generation and allocation concealment (selection bias); (2) Therapist and researcher allegiance, treatment fidelity (performance bias); (3) Blinding of outcome assessor (detection bias); (4) Incomplete outcome data reporting (attrition bias); and (5) Selective outcome reporting (reporting bias). We assessed and categorized the risk of bias following the *Cochrane* guidelines: (1) low risk of bias, plausible bias unlikely to seriously alter the results; (2) high risk of bias, plausible bias that seriously weakens confidence in the results; and (3) unclear risk of bias, plausible bias that raises some doubt about the results. Two independent reviewers screened the studies, with discrepancies being resolved by consensus. Overall, 92.22% agreement was reached in the initial independent review.

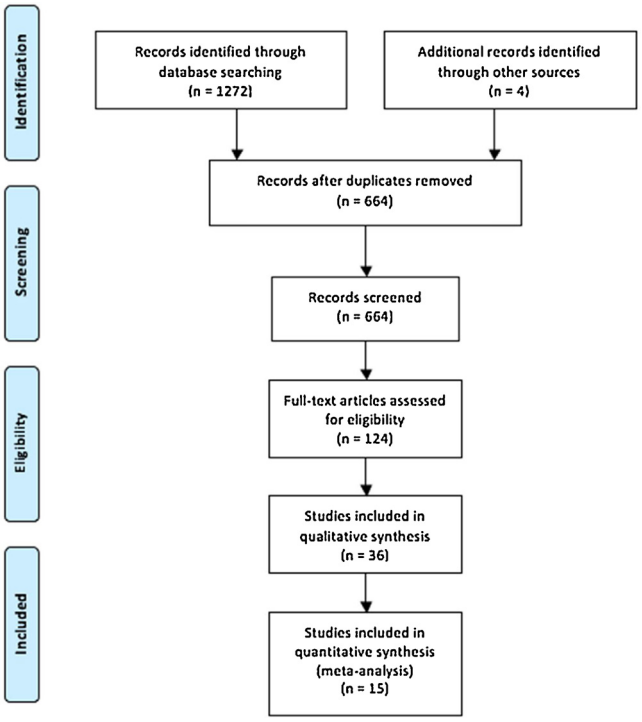

Fig. 1. Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) Diagram of selected studies.

2.7. Statistical analysis

Deterioration was determined using the Reliable Change Index (RCI), following the recommendations given by Jacobson, Follette, and Revenstorf, (1984). The RCI reflects whether a change score between two measurement points is reliable and not caused solely by measurement error. For the current study, it was used to assess a reliable deterioration. In the IPD meta-analysis, the RCI was calculated by subtracting the scores on the pre-treatment assessment from those of the post treatment assessment and dividing the difference by the standard error of difference between the two test scores (Evans, Margison, & Barkham, 1998). This was done separately for each clinical trial because they used different primary outcome measures, and the RCI is affected by the different standard deviations at pretreatment assessment. Furthermore, the standard error of difference was derived from the test-retest reliabilities previously obtained for each primary outcome measure (see Table 2). This strategy is preferred over implementing internal consistencies, e.g., Cronbach's α, as the test-retest reliability also introduces some variability between the two test scores (Speer, 1992). However, if there was more than one reliability estimate, as in the case of different subscales in the same self-report measure, the highest value was selected. In addition, for the Attitudes Towards Agoraphobia Questionnaire (North, North, & Coble, 1996), it was not possible to attain a reliability estimate; thus, instead it was derived from the Agoraphobia Inventory (Echeburúa, Corral, García, Páez, & Borda, 1992). Moreover, on one of the clinical trials, García-Palacios, Botella, Hoffman, and Fabregat, (2007), only the Behavioral Approach Test was administered as a measure of treatment outcome. Because this test cannot be used when assessing deterioration with the RCI, this study was removed from the analysis.

Usually, an RCI of 1.96 is considered a reliable change ($p = .05$). However, Wise (2004) defended the use of a less conservative threshold to examine moderate and mild deterioration and improvement. This

Table 2
Test-Retest Reliability Estimates Used to Calculate the Reliable Change Index.

Primary outcome	Test-retest reliability	Time period	Population	Reference
Fear of Negative Evaluation Scale	$r = 0.75$	4 weeks	Normal	Leary (1983)
Fear and Avoidance Scale	$r = 0.85$	1 weeks	Patient	Marks and Mathews (1979)
Liebowitz Social Anxiety Scale – Self-Report	$r = 0.93$	8 weeks	Normal	Heeren et al. (2012)
Attitudes Towards Agoraphobia Questionnaire	$r = 0.69$[a]	6 weeks	Patient	n.a.
Clinician Administered PTSD Scale	$r = 0.96$	–	Patient	Blake et al. (1995)
Panic Disorder Severity Scale	$r = 0.94$	2 days	Patient	Lee, Kim, & Yu (2009)
Agoraphobia Inventory	$r = 0.69$	6 weeks	Patient	Echeburúa, Corral, García, Páez, & Borda (1992)
Fear of Flying Scale	$r = 0.94$	12 weeks	Patient	Haug et al. (1987)
Acrophobia Questionnaire	$r = 0.86$	12 weeks	Patient	Baker, Cohen, and Saunders (1973)
Flight Anxiety Situation Questionnaire	$r = 0.92$	–	Patient	Van Gerwen, Spinhoven, Van Dyck, and Diekstra (1999)

n.a. = not available
[a]Obtained from the Agoraphobia Inventory due to non-existing reliability estimate

lowers the confidence levels, but could be important in identifying cases of worsening that might otherwise be overlooked. In addition, because the average rate of deterioration was quite low overall, with only about 5–10% faring worse during the treatment period (Lambert, 2013), capturing mild deterioration could increase the power for subsequent predictor analyses (Edwards, Yarvis, Mueller, Zingale, & Wagman, 1978). Therefore, in the IPD meta-analysis, an RCI of 0.84 was selected ($p = .20$). The same procedure was recently used by Rozental et al. (2017) in a similar IPD meta-analysis of deterioration in Internet-based cognitive behavior therapy.

Patients exceeding an RCI of 0.84 in a negative direction were dummy coded as being reliably deteriorated (1 = yes, 0 = no). This information was later used to investigate possible predictors of worsening, using deterioration as the dependent variable in a logistic regression that implemented forced entry, i.e., all predictors were entered simultaneously. The independent variables, i.e., predictors, were chosen a priori: 1) clinical severity at pre-treatment assessment, 2) marital status, 3) educational level, 4) age, and 5) gender. This is a replication of the study by Rozental et al. (2017), except for prior psychological treatment, current use of psychotropic medication, and sick leave, as information about these aspects was not available in the current study. The results of the logistic regression are presented as odds ratios (OR), which refers to an increase or decrease in the odds of deterioration, compared to a predetermined reference category. For nominal variables, e.g., gender, the OR indicates an increase in the odds of worsening when going from female (0) to male (1) (see Table 6). For the continuous variables, i.e., clinical severity at pre-treatment assessment and age, the OR stands for an increase of 1 SD above the mean. Predictors with an OR where the 95% Confidence Interval (CI) does not contain 1 are considered significant, which implies that they might predict deterioration. However, as with all analyses of predictors, the results should be considered tentative, requiring further research before any firm conclusions can be drawn about their importance (Clarke, 2005). Lastly, because there is a risk of ceiling effects when determining deterioration during treatment using self-report measures, scoring near the instrument's maximum at pretreatment assessment was assessed by dummy coding of patients who were within the RCI's upper boundary on each respective clinical trial (1 = yes, 0 = no). All statistical analyses were conducted in SPSS 24.0.0.1.

Finally, to summarize the evidence obtained from the included studies, deterioration rates from VR group and a combination of deterioration rates from both control groups[2] were used to calculate Odds Ratios (ORs). The summary statistic is reported as an OR with a CI of 95%. Heterogeneity was calculated with the Q statistic and the I^2 index (Higgins, Thompson, Deeks, & Altman, 2003). The I^2 index is interpreted as the percentage of between-study variance that is not explained by random sampling error of the primary studies. In other words, this index makes it possible to establish whether the variability is due to differences in the effect sizes or not (Borenstein et al., 2009). The percentages are used to discriminate low (25%), middle (50%), and high heterogeneity (75%). However, it must be taken into account that the I2 statistic can be biased when it is calculated based on small sample sizes (Von Hippel, 2015). Given the ORs obtained, a forest plot was also prepared. This analysis was conducted using Comprehensive Meta-analysis software (Comprehensive Meta-Analysis, 2013). The odds ratio comparing the likelihood of deterioration between VR and waitlist was calculated using the MH method. Most commonly used meta-analytic approaches are based on assumptions that rely on large samples and may be inappropriate for assessment of rare events such as deterioration. The Mantel–Haenszel (MH) method has been shown to have minimal bias under conditions consistent with the present analyses (Bradburn, Deeks, Berlin, & Localio, 2007).

2.8. Ethical considerations

All the raw data are part of published studies approved by the respective ethical review boards. For more information, see the original studies. All the patients were coded so that they were unidentifiable within the new dataset created. In line with the description provided by Rozental et al. (2017), the present study utilized only the raw scores from already completed RCTs, making it impossible to intervene in those cases where deterioration was detected. However, the current study will hopefully encourage future RCTs to include a deterioration rate assessment, or other ways of determining potential negative effect domains during the trial.

3. Results

3.1. Study characteristics

The 36 studies that were eligible to meet all the inclusion criteria are presented in Table 3. After contacting corresponding authors of the published articles, datasets from 15 of these 36 studies were retrieved (for the whole procedure, see 2.5.). Raw scores from these studies were aggregated into a unified dataset totaling 810 patients. Of this total number, 348 (42.96%) received a VR treatment, 282 (34.81%) received other active treatment, and 180 (22.22%) were in a waiting list condition. Regarding the clinical conditions, 230 patients were diagnosed with social anxiety disorder (4 studies), 60 with agoraphobia (4 studies), 17 with panic disorder (1 study), 225 with specific phobias including small animal phobias (2 studies) and fear of flying (4 studies), and 80 with post-traumatic stress disorder (3 studies).[3]

[2] This has been done to obtain only one funnel plot and retrieve more data (there were groups with no deterioration rates in one of the two conditions, or there was no third condition in other cases).

[3] Note that the sum of these numbers is larger than fifteen because Moldovan and David (2014) included three conditions.

Table 3
Study Characteristics of all Clinical Trials that fulfilled the inclusion criteria.

Study	N (F/M)	Age	Clinical sample	Condition (N)	Sessions	Primary outcome measure	Post-assessment	Description of protocol utilized
(Anderson et al., 2013)*	97 (60/37)	19-60 M = 39	SAD	−1: VRET (n = 25) −2: EGT (n = 25) −3: WL (n = 25)	8	PRCS FNE-B	Post: (1 = 2) > 3 12m: (1 = 2) > 3	(Anderson et al., 2013) (Hofmann, 2004)
(Baños et al., 2011)*	39 (27/12)	18-50 M = 30.85	PTSD (n = 10) PG (n = 16) AD (n = 13)	−1: CBT (n = 20) −2: CBT + VR (n = 19)	9	FAS	Post: 1 = 2	(Baños et al., 2009)
(Botella et al., 2016)*	63 (59/4)	20-70 M = 31,73	Small animal phobia	−1: IVET (n = 31) −2: ARS (n = 32)	1	BAT	Post: 1 > 2 3 m: 1 = 2 6m: 1 = 2	"One-session treatment" (Öst) (Ost, 1989)
(Botella et al., 2007)*	37 (26/11)	18-72 M = 34,7	PD with agoraphobia	−1: VRET (n = 12) −2: iVET (n = 12) −3: WL (n = 13)	9	FAS PDSS	Post: (1 = 2) > 3 12m: (1 = 2) > 3	(Cristina Botella, Osma, Garcia-Palacios, Quero, & Banõs, 2004)
(Bouchard et al., 2017)*	59 (43/16)	M = 34,5	SAD	−1: CBT + VR: (n = 17) −2: CBT (n = 22) −3: WL (n = 20)	14	LSAS-SR	Post: 1 > 2 > 3 6m: 1 > 2 > 3	Clark & Wells (Czerniak et al., 2016)
(De La Rosa and Cárdenas, 2012)*	20 (12/8)	M = 35,8	PTSD	−1: VR: (n = 10) −2: CBT (n = 10)	12	CAPS PSS	Post: 1 > 2 > 3 6m: 1 > 2 > 3	(Rothbaum, Difede, & Rizzo, 2008)
(Choi et al., 2005)	40 (18/22)	M = 36,2	PD with agoraphobia	−1: CBT + VR (n = 20) −2: PCP (n = 20)	4	ACQ BSQ	Post: 1 = 2 6m: 1 = 2	(Vincelli, Choi, Molinari, Wiederhold, & Riva, 2000)
(Difede et al., 2007)*	21 (3/18)	M = 43,02	PTSD	−1: VRET (n = 13) −2: WL (n = 8)	6/13	CAPS	Post: 1 > 2 6m: 1 > 2	Unspecified
(Emmelkamp et al., 2002)	33 (15/18)	M = 49,97	Acrophobia	−1: VRET (n = 16) −2: iVET (n = 17)	3	BAT	Post:1 = 2 6m: 1 = 2	(Emmelkamp, Bruynzeel, Drost, & van der Mast, 2001)
(Garcia-Palacios, Hoffman, Carlin, Furness, & Botella, 2002)	23 (21/2)	18-58 M = 29,25	Spider phobia	−1: VRET (n = 13) −2: WL (n = 8)	4	FSQ BAT	Post: 1 > 2	Unspecified
(Lorenzo González et al., 2011)	64 (41/23)	28-61 M = 38,82	Agoraphobia	−1: CBT + Paroxetine (n = 11) −2: CBT + Venlafaxine (n = 11) −3: VRET + Paroxetine (n = 11) −4: VRET + Venlafaxine (n = 11) -WL (n = 20)	11	BSQ ACQ	Post: 1 = 2 = 3 = 4 = 5 6m: 1 = 2 = 3 = 4 = 5	(Penate, Pitti, Manuel Bethencourt, de la Fuente, & Gracia, 2008)
(Kampmann et al., 2016)*	60 (38/22)	18-65 M = 36,88	SAD	−1: VRET (n = 20) −2: iVET (n = 20) −3: WL (n = 20)	10	LSAS-SR FNE-B	Post: (1 = 2) > 3 3 m: 2 > 1 > 3	(Scholing & Emmelkamp, 1993) and Hofmann and Otto (2008)
(Klinger & Bouchard, 2005)	36 (19/17)	30,5	SAD	−1: VRET (n = 18) −2: CBT (n = 18)	12	LSAS-SR CGI HAD SCIA	Post: 1 = 2	(Roy et al., 2003)
(Krijn et al., 2004)	30 (12/18)	M = 50,6	Acrophobia	−1: VRET (n = 17) −2: WL (n = 11)	3	AQ ATHQ BAT	Post:1 > 2 6m: 1 > 2	(Emmelkamp et al., 2001)
(Krijn et al., 2007)	86	M = 38,58	FOF	−1: VRET (n = 43) −2: CBT (n = 18) −3: BIB (n = 25)	4	FAS FAM	Post: 2 > 1 > 3	Unspecified
(Maltby, Kirsch, Mayers, & Allen, 2002)	43 (34/9)	20-72 M = 45,34	Specific phobia (n = 28) Agoraphobia with PD (n = 5) Agoraphobia without PD (n = 10)	−1: VRET (n = 20) −2: APGT (n = 23)	5	FAS FAM	Post: 2 > 1 6m: 1 = 2	Unspecified
(McLay et al., 2011)	20 (1/19)	M = 28,4	PTSD	1: VRET (n = 10) 2: TAU (n = 10)	9	CAPS	Post: 1 > 2	(Spira, Pyne, & Wiederhold, 2007)
(Meyerbroeker, Morina, Kerkhof, & Emmelkamp, 2013)	55	18-65	Agoraphobia	1: VRET (n = 19) 2: iVET (n = 18) 3: WL (n = 18)	10	ACQ	PDSS: 2 > 1 > 3 ACQ, BSQ, MIA: Post: (1 = 2) > 3	(Barlow, 2007)

(*continued on next page*)

Table 3 (*continued*)

Study	N (F/M)	Age	Clinical sample	Condition (N)	Sessions	Primary outcome measure	Post-assessment	Description of protocol utilized
(Michaliszyn, Marchand, Bouchard, Martel, & Poirier-Bisson, 2010)	43 (42/1)	18-51 M = 29,1	Spider phobia	−1: VRET (n = 16) −2: iVET (n = 16) −3: WL (n = 11)	8	FSQ-F SBQ-F BAT	Post: (1 = 2) > 3 (SBQ-F: 2 > 1) 3 m: 1 = 2	Unspecified
(Miyahira et al., 2012)	22	Unspecified	PTSD	−1: CBT + VRET (n = 12) −2: MA (n = 10)	10	CAPS	Post: 1 > 2 (CAPS criterion C, TRGI)	Unspecified
(Moldovan & David, 2014)*	32 (15/17)	Over 18	FOF (n = 9); Social Anxiety Disorder (n = 15); Acrophobia (n = 8)	−1: VRET + CBT (n = 16) −2: WL (n = 16)	1	LSAS FAS FAM STAI FNE-B SSPS	Post: 1 = 2 Follow up (unspecified when): 1 = 2	"One-session treatment" (Öst) (Ost, 1989) CBT: REBT theory
(Mühlberger et al., 2001)*	30 (26/4)	M = 43,05	FOF	−1: VRET (n = 15) −2: Relaxation (n = 15)		FFS AFA DES AES ASI	Post: 1 > 2	Unspecified
(Mühlberger, Wiedemann, Pauli, & Wiedemann, 2003)*	47 (34/13)	M = 39,8	FOF	−1: VRET + motion simulation (n = 13) −2: VRET without motor simulation (n = 13) −3: WL (n = 11)	1	FFS	Post: (1 = 2) > 3 6m: (1 = 2) > 3	(Mühlberger et al., 2001)
(North et al., 1996)*	60	Unspecified	Agoraphobia	−1: VRET (n = 30) −2: WL (n = 30)	8	ATAQ SUDS	Post: 1 > 2	Unspecified
(Pelissolo et al., 2012)	92 (62/30)	24-72 M = 37,1	PD with agoraphobia	−1: VRET (n = 29) −2: CBT (n = 31) −3: WL (n = 32)	12	FQ PDSS	Post: 1 = 2 = 3 3 m: 1 = 2 = 3 6m: 1 = 2 = 3 12m: 1 = 2 = 3	(Cottraux, Bouvard, & Légeron, 1985 amd Landon & Barlow, 2004)
(Peñate Castro, Pitti, Manuel Bethencourt, de la Fuente, & Gracia, 2008)	37 (27/10)	17/60 M = 36,75	Agorapobia	−1: CBT + VRET (n = 21) −2: CBT (n = 16)	11	ACQ	Post: 1 = 2 3 m: 1 = 2 Effect size: 1 > 2 (Post + 3 m)	Unspecified
(Peñate Castro et al., 2014)*	50	24-60	Agoraphobia	1: VRET (n = 30) 2: CBT (n = 30) 3: Medication (n = 20)	11	ACQ	Post: 1 > (2 = 3) 6m: 1 > (2 = 3)	Unspecified
(Pitti et al., 2015)*	128	M = 39	Agoraphobia	1: PX-CBT (n = 27) 2: PX-CBT-VRET (n = 27) 3: PX (n = 32)	11	ACQ	Post:(1 = 2) > 3 6m: 1 = 2	Unspecified
(Reger et al., 2016)	162 (6/156)	M = 30,27	PTSD	−1: VRET (n = 54) −2: PE (n = 54) −3: WL (n = 54)	10	CAPS PCL-C	Post: (1 = 2) > 3 3 m: 2 > (1, 3) 6m: 2 > (1,3)	(Foa, Chrestman, & Gilboa-Schechtman, 2009)
(Rothbaum et al., 2006)	83 (67/16)	M = 40,1	FOF	−1: VRET (n = 29) −2: iVET (n = 29) −3: WL (n = 25)	8	FFI QAF	Post:(1 = 2) > 3 6m: (1 = 2) > 3 12m: (1 = 2) > 3	(Rothbaum, Hodges, Smith, Lee, & Price, 2000)
(Rothbaum et al., 1995a)	20 (8/12)	M = 20	Acrophobia	−1: VRET (n = 12) −2: WL (n = 8)	7	AQ ATHQ FQ	Post: 1 > 2	(Rothbaum et al., 1995b)
(Rothbaum et al., 2000)	49 (32/17)	24-69 (M = 405)	FOF	−1: VRET (n = 18) −2: iVET (n = 16) −3: WL (n = 15)	8	QAF FFI	Post:(1 = 2) > 3 6m: (1 = 2) > 3	(Rothbaum, Hodges, & Smith, 1999)
(Tortella-Feliú et al., 2011)*	60 (35/25)	M = 37,04	FOF	−1: VRET (n = 19) −2: CAE-T (n = 20) −3: CAE-SA (n = 21)	6	FFS FFQ	Post: 1 = 2 = 3 12m: 1 = 2 = 3	(Botella-Arbona, Osma, Garcia-Palacios, Quero, & Banõs, 2004; Botella et al., 2008)
(Triscari, Faraci, Catalisano, D'Angelo, & Urso, 2015)	65	24-70 M = 43,52	FOF	1: CBT-SD (systematic desensitization) (n = 22) 2: CBT – EMRD (n = 22) 3: CBT – VRET (n = 21)	10	FAS FAM	Post: 1 = 2 = 3 12m: 1 = 2 = 3	Unspecified

(*continued on next page*)

Table 3 (*continued*)

Study	N (F/M)	Age	Clinical sample	Condition (N)	Sessions	Primary outcome measure	Post-assessment	Description of protocol utilized
(Wallach, Safir, & Bar-Zvi, 2009)	88 (68/20)	M = 27,35	Public speaking anxiety	−1: CBT + VRET (n = 28) −2: CBT (n = 30) −3: WL (n = 30)	12	FNE LSAS SSPS	Post: (1 = 2) > 3	CBT guidelines for social phobias (Heimberg et al., 1998)
(Wiederhold et al., 2002)	30 (18/12)	24-55 M = 39,8	FOF	−1: VRET (n = 20) −2: IET (n = 10)	8	SUDS	Post: 1 > 2	Unspecified

Note: F = Feminine; M = Masculine; EGT = Exposure Group Therapy; WL = Waiting List; FOF = Fear Of Flying; PRCS = Self-Report Of Public Speaking Fears; FNE-B = Self-Report Of Social Anxiety Disorder Symptoms; AGS = Augmented Reality System; BAT = Behavioural Avoidance Test ; SAD = Social Anxiety Disorder; PTSD = Post Traumatic Stress Disorder; PG = Pathological Grief; AD = Adjustment Disorders; PD = Panic Disorder; CBT + VR = Cognitive Behavioural Therapy Plus Virtual Reality; LSAS-SR = Liebowitz Social Anxiety Scale-Self Report; VRET = Virtual Reality Exposure Therapy; iVET = In Vivo Exposure Therapy; PANAS = Positive And Negative Affect Schedule; DASS = Depression Anxiety Stress Scale; ASI = Anxiety Sensitivity Index; CGI = Clinician Global Impression; MS = Maladjustment Scale; PBQ = Panic Belief Questionnaire; CAPS = Virtual Reality Exposure Therapy For The Treatment Of Posttraumatic Stress Disorder; BSI = Brief Symptom Inventory; MIA = Mobility Inventory For Agoraphobia; ACQ = Agoraphobic Cognitions Questionnaire; BSQ = Body Sensations Questionnaire; PDSS = Panic Disorder Severity Scale; FAS = Flight Anxiety Situations; AQ = Acrophobia Questionnaire; FAM = Flight Anxiety Modality; STAI = State And Trait Anxiety Questionnaire; SPSS = Self Statements During Public Speaking Scale; FQ = Fear Questionnaire; REBT Theory = Rational Emotive Behaviour Therapy; CAS = Chambless Agoraphobic Cognitions; PPGAS = Panic, Phobia And Generalized Anxiety Scale; HAR S = Hamilton Anxiety Rating Scale; BDI = Beck Depression Inventory; WSA = Work And Social Adjustment Scale ;SDS = Sheehan Disability Scale; BAI = Beck Anxiety Inventory; SUA = Subjective Units Of Anxiety; PX = Paroxetine; SD = Systematic Desensitization; EMRD = Eye Movement Desensitization And Reprocessing; PCP = Panic Control Program; BIB = Bibliotherapy; CAE-T = Computer-Aided Exposure With A Therapist's; CAE-SA = Self-Administered Computer-Aided Exposure; IET = Imaginal Exposure Therapy; ATHQ = Attitude Towards Heights Questionnaire; FSQ = Fear of Spiders Questionnaire; HAD = Hospital Anxiety And Depression; SCIA = Social Contexts Inducing Anxiety; APGT = Attention Placebo Group Treatment; TAU = Treatment As Usual; SBQ-F = Spider Beliefs Questionnaire; SSPS = Self Statements During Public Speaking Scale; FFS = Fear Of Flying Scale; AFA = General Fear Of Flying Questionnaire (Allgemeiner Flugangstfragebogen); DES = Danger Expectancy Scale; AES = Anxiety Expectancy Scale; ATAQ = Attitude Towards Agoraphobia Questionnaire; SUDS = Subjective Units Of Discomfort Scale; AGPH = Agoraphobia Questionnaire; FFI = Fear Of Flying Inventory; QAF = Questionnaire On Attitudes Toward Flying, FFQ = Fear Of Flying Questionnaire; ADIS-IV = Anxiety Disorders Interview Schedule For DSM-IV; MA = Minimal Attention control condition; PDS = PTSD Diagnostic Scale; QOLI = Quality of Life Inventory; TRGI = Trauma-Related Guilt Inventory; PCL-C = PTSD Checklist; SSRPH = Stigma Scale for Receiving Psychological Help; IASMHS = Inventory of Attitudes toward Seeking Mental Health Services; PE = Prolonged Exposure.
* Articles included in the IPD analysis.

Analyzing the 36 papers, the only two studies that included deterioration rate analysis were Botella et al. (2016) and Reger et al. (2016). From a total sample of 162 patients, only 1 was deteriorated (in the VR condition). Other studies indicate the reliable change (e.g. Kampmann et al., 2016), but they do not report the deterioration rates.

3.2. Representativeness of the included studies in relation to the 36 papers that met the criteria in the systematic review

A total of 1956 patients with anxiety disorders treated with VR met the inclusion criteria specified for this study. This means that the recruited sample for the IPD analysis represents 41.10% of the overall sample. In terms of the clinical conditions, 22.43% of the PTSD patients, 23.66% of PD patients, 73.83% of AG patients, 21.09% of SP patients, and 65.07% of SAD patients treated with VR were included in the analyses.

3.3. Risk of bias in the included studies

As depicted in Fig. 2, the quality of the included papers varied. Whereas the studies presented low methodological quality overall on randomization and blinding aspects, their high quality was observed in terms of attrition and reporting biases. The fact that only 46.66% of the studies included a low risk of bias with regard to the random sequence generation and the blinding of participants and personnel, and only 20% of them followed an adequate allocation concealment, reveals the low methodological quality of the VR studies. Nevertheless, more troubling is the random sequence generation, which is often impossible to conduct in psychotherapeutic interventions. However, because the selection bias is mainly unclear, it is possible that proper randomizations were conducted and that a reporting bias is present instead. In any case, improvements within the field must be made in future studies. Finally, the fact that 21 datasets were irretrievable constitutes a possible source of bias.

3.4. Missing values and ceiling effects

In all, 76 patients (9.4%) did not complete either the pre or post treatment assessments in the clinical trials that were included in the IPD meta-analysis, which is necessary to determine whether deterioration has occurred. However, examining these patients more closely did not reveal a relationship with any of the sociodemographic variables, $\chi^2(1) = 0.15–3.63$, $p = .06-.70$. As for the predictors, 14 patients (1.7%) did not complete the pre-treatment assessment, i.e., clinical severity. In addition, values were missing for 165 patients (20.4%) in terms of marital status, 93 (11.5%) for educational level, and 19 (2.4%) for age, but there were no missing values for gender (see Table 5). With regard to ceiling effects, i.e., being within the RCI's upper boundary in each respective clinical trial, 58 patients (7.2%) were identified as being close to reaching the ceiling of the self-report measures at the pre-treatment assessment, but these were still kept in the analysis of deterioration rates. However, as another way of determining their progress during treatment, an independent-samples *t*-test was performed post hoc on the available data for the ceiling-cases. Of the 51 patients who had both pre- and post-treatment data, there was a significant average decrease of 6.43 points (SD = 4.43), t(50) = 10.37, 95% CI [5.19, 7.68], suggesting that they tended to improve during treatment.

3.5. Deterioration rates

After dropping the 76 participants who did not complete either the pre or post treatment assessments, the remaining 734 (90.6%) were included in the investigation of deterioration in the IPD meta-analysis. From those 734 participants, 49 (6.0%) were identified as reliably deteriorated, i.e., with an RCI of at least 0.84. These numbers, however, varied between the conditions, ranging from 14 (4.0%) for patients receiving VR, 8 (2.8%) when assigned to another form of treatment, and 27 (15.0%) in wait-list control (see Table 4). Given the large difference in rates, the two treatment conditions were combined and compared only to the waiting list, resulting in an OR = 4.87, 95% CI [2.69, 8.80], indicating that the odds for deterioration were significantly higher for patients in wait-list control.

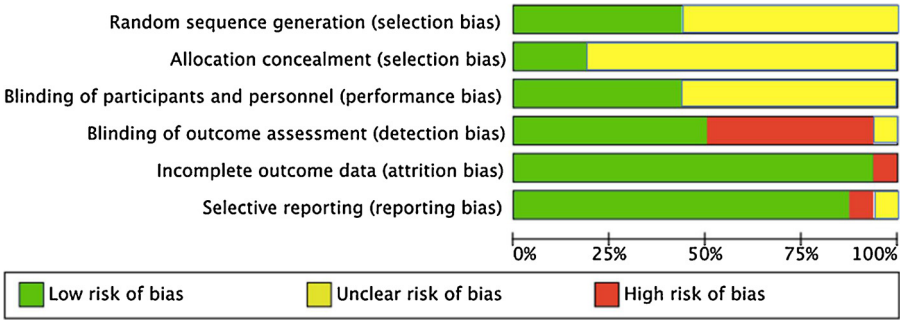

Fig. 2. Risk of bias assessment.

Furthermore, the MH ORs were conducted for VR versus all controls on the one hand, all active treatments versus waitlist on the other hand, and finally VR versus active treatments to confirm no differences in odds of deterioration between these two conditions. With regard to the comparison of VR versus all controls, twelve of the 15 studies were included in the comparison because 3 studies (Baños et al., 2011; De La Rosa & Cárdenas, 2012; Mühlberger, Herrmann, Wiedemann, Ellgring, & Pauli, 2001) did not present deteriorated patients in any of the conditions, and, thus, it was not possible to establish the MH OR. As Fig. 3 shows, the pooled OR was 0.61 (95% CI 0.31–1.23) in favor of VR, establishing that, on average, the VR-based treatments were 39% less likely to produce a deteriorated effect than all the control conditions. Nevertheless, the 95% confidence interval did not reach statistical significance ($z = 1.17$, p = 0.17). Heterogeneity between studies is low ($Q(11) = 11.402$, $p = 0.410$, $I^2 = 3.525$). While not significant, the odds ratio indicates that the odds of deterioration in VR are lower than in waitlist. Finally, Duval and Tweedie (2000) trim and fill method was used to identify any potentially missing studies and their impact on the summary MH odds ratio. Four potentially missing studies were identified, with an adjusted MH odds ratio of 0.98 [0.46–2.11].

Likewise, as Fig. 4 shows, the comparison of all active conditions (VR and other treatments) versus waiting list yielded an odds ratio of 0.165 [−3.879 – 0.411], $p < 0.001$ reflecting that the odds of deterioration are significantly higher in waitlist compared to treatment (VR or other treatment). Although a higher variability between studies is noticeable in comparison to VR versus all controls, following the standards suggested in the literature (Higgins et al., 2003) the detected is also low ($Q(9) = 11.343$, $p = 0.253$, $I^2 = 20.658$). One potentially missing study was identified and imputed based on the Duval and Tweedie trim and fill method. The adjusted OR was 0.129 [0.049 – 0.346].

Finally, a comparison of VR and active conditions was performed with the aim of confirming that no significant difference was present between these two conditions. The MH odds ratio comparing the odds of deterioration from VR to other treatment was 1.68 [0.65–4.33], $p = 0.287$. Variability between studies was low ($Q(6) = 2.64$, $p = 0.853$, $I^2 = 0.00$). By doing this, the previous comparison of VR and other treatments versus WL is consistent.

As for the specific clinical trials, Anderson et al. (2013) had the highest number of deteriorated patients in VR, 5 (13.9%), Tortella-Feliú et al. (2011) in another form of treatment, 4 (9.8%), and Difede et al. (2007) for wait-list control, 4 (50.0%). One of the worrisome aspects of utilizing RCI is the possibility of not identifying patients who deteriorated, due to the ceiling effects produced by the limitations of self-report measures (Martinovich, Saunders, & Howard, 1996). However, within this sample, only 58 (7.2%) were at risk of ceiling effects.

3.6. Predictors of deterioration

Given the large difference in deterioration rates, the analyses of potential predictors using logistic regressions were carried out separately for the three conditions. None of the predictors chosen a priori were related to deterioration in any of the cases; 1) clinical severity at pre-treatment assessment, 2) marital status, 3) educational level, 4) age, and 5) gender. The only exception was marital status in patients waiting for treatment, OR = 0.19, 95% CI [0.05, 0.67], suggesting that being in a relationship lowered the odds of deterioration somewhat. See Table 6 for an overview of the predictors.

4. Discussion

The current study examined the deterioration rates of VR-based treatments for anxiety disorders. The dataset consisted of 15 studies, which, overall, were representative of the whole sample of 36 RCTs conducted within the field of VR for anxiety disorders and stress related disorders; each study had a sample of at least 10 patients in the VR condition. Raw scores from 810 patients were entered into an aggregated dataset, and the RCI was utilized to detect the deteriorated cases.

First, it must be mentioned that, based on the results presented, overall, VR seems to be a non-deleterious treatment for patients with anxiety disorders. To the best of our knowledge, three previous studies have examined deterioration rates in technology-mediated treatments conducted with patients who experienced an Internet intervention (Ebert et al., 2016; Karyotaki et al., 2018; Rozental et al., 2017), all of which did not show a larger deterioration in Internet conditions compared to active control groups. Hence, the present results constitute a novel finding in the same line as the cited Internet intervention studies. Taken together, all these studies permit to establish preliminary though solid evidence about the non-harmful effect of technology-mediated treatments. In this regard, the initial concerns about the potential negative impact of VR have been dissipated, at least with regard to exposure for anxiety disorders.

Regarding the predictors, none of the included variables were related to higher or lower odds of deterioration, with the exception of marital status in the wait-list control. This result agrees with the existing IPD meta-analysis on deterioration rates for ICBT (Rozental et al., 2017). However, because the total amount of deterioration was low, it was not possible to examine this issue further. Only 49 (6.0%) patients deteriorated in all, and dividing them between the three conditions made the predictor analysis lack sufficient power, even though the RCI of 0.84 was used, instead of 1.96, to determine deterioration.

In the same vein, when ORs are compared, aggregated groups should be created due to the low frequencies. First, all the active conditions (VR and other active conditions) were compared to waiting list groups. This difference was statistically significant, showing that exposure, regardless of the type of treatment, is less likely to induce a deteriorated effect compared to patients assigned to waiting list controls. Nevertheless, when comparing VR versus all controls, no significant differences were found, although the tendency was in favor of VR. In all cases, the low frequencies of deterioration rates made it difficult to arrive at more conclusive statements.

Table 4
Characteristics and Deterioration Rates for the Acquired Clinical Trials.

Study	Recruitment	Screening interview	Primary diagnosis	Virtual reality *n* (%)	Other treatment *n* (%)	Wait-list control *n* (%)	Primary outcome	Deterioration virtual reality *n* (%)	Deterioration other treatment *n* (%)	Deterioration waiting-list *n* (%)	OR (VR vs. All controls)
Anderson et al. (2013)	General population	SCID	SAD	36 (37.5)	49 (51.0)	11 (11.5)	FNE	5 (13.9)	2 (4.1)	1 (9.1)	2.77
Botella et al. (2016)	General population	ADIS	SP	32 (50.8)	31 (49.2)	–	BAT FAS[1]	1 (3.1)	1 (3.2)	–	0.97
Bouchard et al. (2017)	General population	SCID	SAD	17 (28.8)	22 (37.7)	20 (33.9)	LSAS	0 (0.0)	0 (0.0)	5 (25.0)	–
Baños et al. (2011)	General population	ADIS	TEPT PG AD	20 (51.3)	19 (48.7)	–	FAS	0 (0.0)	0 (0.0)	–	–
Kampmann et al. (2016)	General population	SCID	SAD	20 (33.3)	20 (33.3)	20 (33.3)	LSAS FNE	1 (5.0)	0 (0.0)	2 (10.)	1.00
North et al. (1996)	University students	DSM-IV items	AG	30 (50.0)	–	30 (50.0)	ATAQ	0 (0.0)	–	2 (6.7)	–
De la Rosa et al. (2012)	General population	CAPS	TEPT	10 (50.0)	10 (50.0)	7 (35.0)	CAPS	0 (0.0)	0 (0.0)	–	–
Difede et al. (2007)	General population	CAPS	TEPT	13 (61.9)	–	8 (38.1)	CAPS	0 (0.0)	–	4 (50.0)	–
Botella et al. (2007)	General population	ADIS	PD	12 (32.4)	12 (32.4)	13 (35.1)	PDSS	0 (0.0)	0 (0.0)	1 (7.7)	–
Peñate Castro et al. (2014)	Primary care	CIDI	AG	23 (26.0)	14 (48.0)	13 (26.0)	AI	1 (4.3)	0 (0.0)	2 (15.4)	0.82
Pitti et al. (2015)	Primary care	CIDI	AG	49 (38.3)	38 (29.7)	41 (32.0)	AI	1 (2.0)	0 (0.0)	7 (17.1)	0.16
Mühlberger et al. (2001)	General population	DSM-IV items	SP	15 (50.0)	12 (50.0)	–	FFS	0 (0.0)	0 (0.0)	–	–
Mühlberger et al. (2003)	General population	DSM-IV items	SP	26 (55.3)	11 (23.4)	10 (21.3)	FFS	2 (7.7)	1 (9.1)	3 (30.0)	0.4
Moldovan and David, (2014)	Unspecified	SCID	SP SAD AG	26 (65.0)	–	14 (35.0)	LSAS AQ FASQ	2 (7.7)	0 (0.0)	0 (0.0)	–
Tortella-Feliú et al., (2011)	General population	DSM-IV items	SP	19 (31.7)	41 (68.3)	–	FFS	1 (5.3)	4 (9.8)	0 (0.0)	0.54
								14 (4.0%)	**8 (2.8%)**	**27 (15.0%)**	

Table 5
Socio-demographic Characteristics of the Patients.

Baseline characteristic	Virtual reality (n = 348)	Other treatment (n = 282)	Wait-list control (n = 180)	Full sample (n = 810)	Missing data
Gender: *n* (% female)	248 (71.3)	203 (72.0)	99 (55.0)	550 (67.9)	0 (0.0)
Age (years): *M* (*SD*)	35.6 (11.2)	36.6 (11.7)	36.9 (11.2)	36.2 (11.4)	19 (2.4)
Marital status: *n* (%)					165 (20.4)
Single	135 (38.8)	125 (44.3)	65 (36.1)	325 (40.1)	
Relationship	138 (39.7)	115 (40.8)	67 (37.2)	320 (39.5)	
Highest educational level: *n* (%)					93 (11.5)
Primary school	20 (5.7)	8 (2.8)	7 (3.9)	35 (4.3)	
Secondary school	142 (40.8)	108 (38.3)	58 (32.2)	308 (38.0)	
University	149 (42.8)	134 (47.5)	91 (50.6)	374 (46.1)	
Employment: *n* (%)					359 (44.3)
Unemployed	18 (5.2)	28 (9.9)	15 (8.3)	61 (7.5)	
Student	35 (10.1)	28 (9.9)	13 (7.2)	76 (9.4)	
Employed	134 (38.5)	104 (36.9)	70 (38.9)	308 (38.0)	
Retired	2 (0.6)	3 (1.1)	1 (0.6)	6 (0.7)	
Primary diagnosis: *n* (%)					2 (0.2)
Social anxiety disorder	98 (28.2)	90 (31.9)	42 (23.3)	230 (28.4)	
Specific phobia	87 (25.0)	79 (28.0)	59 (32.8)	225 (27.8)	
Agoraphobia	33 (9.5)	23 (8.2)	4 (2.2)	60 (7.4)	
Posttraumatic stress disorder	33 (9.5)	25 (8.9)	22 (12.2)	80 (9.9)	
Panic disorder	3 (0.9)	1 (0.4)	2 (1.1)	6 (0.7)	
Panic disorder with agoraphobia	92 (26.4)	64 (22.7)	51 (28.3)	207 (25.6)	

Table 6
Odds Ratios for Each Predictor Variable Using the Full Imputed Model and Divided by Virtual Reality, Other Treatments, and Wait-list Control.

	Virtual Reality				Other treatments				Wait-list control			
Predictor (*reference*)	OR	Lower CI	Upper CI	*p*	OR	Lower CI	Upper CI	*p*	OR	Lower CI	Upper CI	*p*
Clinical severity at pre treatment assessment (*lower severity*)	1.00	0.99	1.02	0.90	0.92	0.84	1.01	0.09	1.00	0.98	1.02	0.74
Marital status, single/relationship (*single*)	1.19	0.37	3.86	0.77	n.a.	n.a.	n.a.	n.a.	0.19	0.05	0.67	0.01*
Educational level, less than/at least university (*less than university*)	0.64	0.18	2.26	0.49	n.a.	n.a.	n.a.	n.a.	1.11	0.31	4.02	0.87
Age (*lower age*)	1.01	0.96	1.07	0.63	1.25	0.98	1.60	0.08	1.04	0.98	1.10	0.16
Gender (*female*)	2.13	0.63	7.17	0.22	n.a.	n.a.	n.a.	n.a.	0.87	0.21	3.56	0.84

Note. *OR* odds ratio; CI 95% confidence interval. *p .05. n.a. = not applicable, i.e., too few cases for the logistic regression to converge.

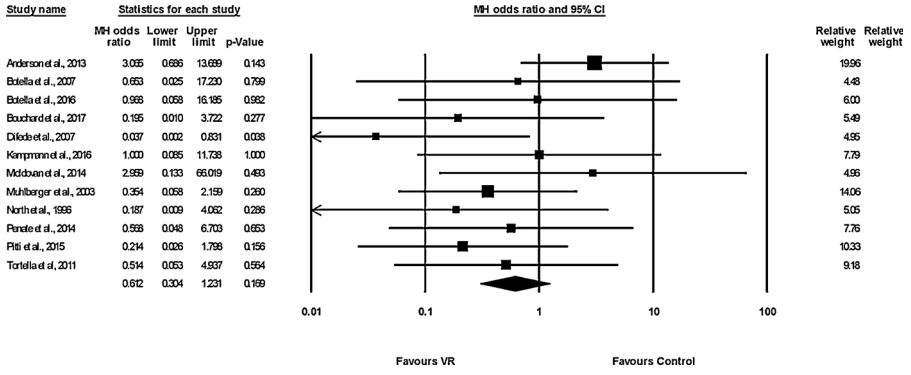

Fig. 3. Forest plot of included studies for the comparison VR versus all controls.

In fact, the pitfall of small samples within VR research has been previously described in the literature (Page & Coxon, 2016), and it is a major issue to overcome in attempting to increase the quality of research in the field and, thus, better determine the extent to which VR can be useful in clinical practice. A recent example acknowledging this issue is the trial conducted by Reger et al. (2016), which presented a large sample. Besides, due to reductions in VR equipment costs, an increase in the number of participants in studies might be expected (Lindner et al., 2017), even when this does not constitute the main cost of a treatment. Examples of low-cost devices implemented in RCT trials are now being conducted, which are expected to be a powerful alternative to expensive devices (Carlbring, Andersson, Cuijpers, Riper, & Hedman-Lagerlöf, 2018).

4.1. Implications for research and practice

First and foremost, there is an important implication with regard to the control groups. Of the 36 studies that met the eligibility criteria to be included in the systematic review, 20 used a waiting list group. Given the present results, due to ethical considerations, waiting lists should be carefully considered for use in future VR trials.

Furthermore, new trials should assess and report negative effects, and patients suffering from deterioration or any other negative effect should be informed as an indispensable ethical requisite. Hence, researchers should address the rising concern about the topic by conducting more studies on the topic, and clinicians implementing the guidelines in their routine practice.

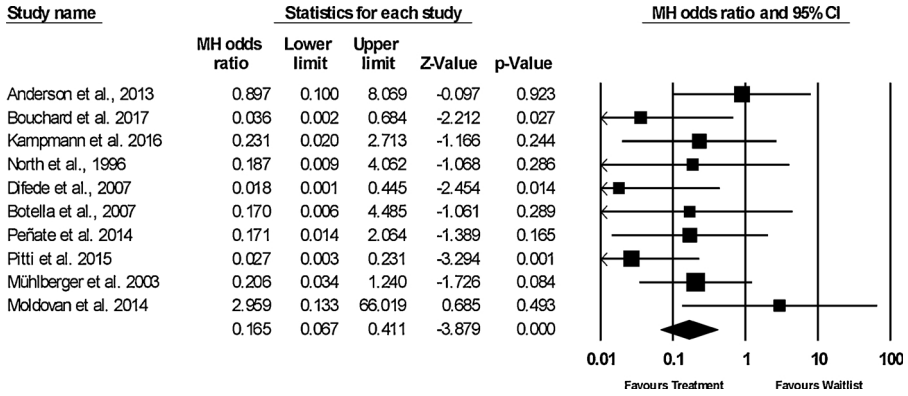

Fig. 4. Forest plot of studies for the comparison all active conditions versus WL.

4.2. Limitations and future research

If the total number of missing values was low, no imputation analyses were performed for missing values. However, this could have led to some biased results, and it would have been better to conduct these analyses. In addition, the lack of power due to the small samples in the VR literature hindered the ability to conduct regression analyses to determine predictors of deterioration. Thus, an important future study should build a more complete dataset in order to replicate the present study with a larger sample and apply more complex statistical analyses.

Another important aspect is that establishing deterioration rates is a straightforward way to obtain a negative effect, but adverse effects may be present in many other forms that should also be examined. Furthermore, despite the results obtained, it cannot be concluded that patients undergoing VR-based treatments will not experience any negative effects.

Declaration of conflict of interests

Stéphane Bouchard is president and part owner of In Virtuo, a company that distributes virtual environments, and conflict of interest are managed under UQO's conflict of interests policy.

Acknowledgments

This work was supported by the Marie Skłodowska-Curie Innovative Training Network AffecTech (project ID:722022) funded by the European Commission H2020.

Appendix 1

Pubmed string strategy

((virtual reality) OR (virtual reality exposure therapy) OR ((virtual reality) AND (exposure))) AND ((anxiety disorder*) OR (obsessive compulsive disorder) OR (OCD) OR (general anxiety disorder) OR (GAD) OR (fear of falling) OR (arachnophobia) OR (post traumatic stress disorder) OR (aviophobia) OR (flight phobia) OR (fear of flying) OR (phob*) OR (social anxiety*) OR (SAD))

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. *Journal of Consulting and Clinical Psychology, 81*(5), 751–760. https://doi.org/10.1037/a0033559.

Baker, B. L., Cohen, D. C., & Saunders, J. T. (1973). Self-directed desensitization for acrophobia. *Behaviour Research and Therapy, 11*(1), 79–89.

Baños, R. M., Botella, C., Guillen, V., García-Palacios, A., Quero, S., Bretón-López, J., ... Alcañiz, M. (2009). An adaptive display to treat stress-related disorders: EMMA's world. *British Journal of Guidance & Counselling, 37*(3), 347–356. https://doi.org/10.1080/03069880902957064.

Baños, R. M., Botella, C., & Perpiña, C. (1999). Virtual reality and psychopathology. *Cyberpsychology & Behavior, 2*(4).

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. https://doi.org/10.1016/j.ijhcs.2011.06.002.

Barkham, M., Margison, F., Leach, C., Lucock, M., Mellor-Clark, J., Evans, C., ... McGrath, G. (2001). Service profiling and outcomes benchmarking using the CORE-OM: Toward practice-based evidence in the psychological therapies. *Journal of Consulting and Clinical Psychology, 69*(2), 184–196. https://doi.org/10.1037/0022-006X.69.2.184.

Barlow, D. H. (2007). *Clinical handbook of psychological disorders.* New York: The Guilford Press.

Barlow, D. H. (2010). Negative effects from psychological treatments. *The American Psychologist, 65*(1), 13–20. https://doi.org/10.1037/a0015643.

Beidel, D. C., Neer, S. M., Bowers, C. A., Frueh, B. C., & Rizzo, A. (2014). *Using virtual reality as part of an intensive treatment program for PTSD. Proceedings of the Interservice/Industry Training, Simulation and Education Conference*.

Bergin, A. E. (1963). The effects of psychotherapy: Negative results revisited. *Journal of Counseling Psychology, 10*, 244–250.

Blake, D. D., Weathers, F. W., Nagy, L. M., Kaloupek, D. G., Gusman, F. D., Charney, D. S., ... Keane, T. M. (1995). The development of a clinician-administered PTSD scale. *Journal of Traumatic Stress, 8*(1), 75–90.

Boswell, J., Kraus, D., Miller, S., & Lambert, M. (2015). Implementing routine outcome monitoring in clinical practice: Benefits, challenges, and solutions. *Psychotherapy Research, 25*(1), 6–19. https://doi.org/10.1080/10503307.2013.817696.

Botella, C., Baños, R. M., Perpiñá, C., Villa, H., Alcaniz, M., & Rey, A. (1998). Virtual reality treatment of claustrophobia: A case report. *Behaviour Research and Therapy, 36*(2), 239–246.

Botella, C., Osma, J., Garcia-Palacios, A., Quero, S., & Banõs, R. M. (2004). Treatment of flying phobia using virtual reality: Data from a 1-year follow-up using a multiple baseline design. *Clinical Psychology & Psychotherapy, 11*(5), 311–323. https://doi.org/10.1002/cpp.404.

Botella, C., 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*, 164–175.

Botella, C., Quero, S., Banos, R. M., Garcia-Palacios, A., Breton-Lopez, J., Alcaniz, M., ... Fabregat, S. (2008). Telepsychology and self-help: The treatment of phobias using the internet. *CyberPsychology & Behavior, 11*(6), 659–664. https://doi.org/10.1089/cpb.2008.0012.

Botella, C., Riva, G., Gaggioli, A., Wiederhold, B. K., Alcaniz, M., & Banos, R. M. (2012). The present and future of positive technologies. *Cyberpsychology, Behavior, and Social Networking, 15*(2), 78–84.

Botella, C., Pérez-Ara, M.Á., Bretón-López, J., Quero, S., García-Palacios, A., & Baños, R. M. (2016). In Vivo versus augmented reality exposure in the treatment of small animal phobia: A randomized controlled trial. *PloS One, 11*(2), 1–22. https://doi.org/10.1371/journal.pone.0148237.

Botella, C., Fernández-Álvarez, J., Guillén, V., García-Palacios, A., & Baños, R. (2017). Recent progress in virtual reality exposure therapy for phobias: A systematic review. *Current Psychiatry Reports, 19*(7), https://doi.org/10.1007/s11920-017-0788-4.

Bouchard, S., Dumoulin, S., Robillard, G., Guitard, T., Klinger, E., Forget, H., ... Roucaut, F. X. (2017). Virtual reality compared with in vivo exposure in the treatment of social anxiety disorder: A three-arm randomised controlled trial. *The British Journal of Psychiatry,* 1–9. https://doi.org/10.1192/bjp.bp.116.184234.

Bradburn, M. J., Deeks, J. J., Berlin, J. A., & Localio, A. R. (2007). Much ado about nothing: A comparison of the performance of meta-analytical methods with rare events. *Statistics in Medicine, 26*(1), 53–77. https://doi.org/10.1002/sim.2528.

Bystedt, S., Rozental, A., Andersson, G., Boettcher, J., & Carlbring, P. (2014). Clinicians' perspectives on negative effects of psychological treatments. *Cognitive Behaviour Therapy, 43*(4), 319–331. https://doi.org/10.1080/16506073.2014.939593.

Carlbring, P., Andersson, G., Cuijpers, P., Riper, H., & Hedman-Lagerlöf, E. (2018). Internet-based vs. face-to-face cognitive behavior therapy for psychiatric and somatic disorders: An updated systematic review and meta-analysis (In press) *Cognitive*

Behaviour Therapy. https://doi.org/10.1080/16506073.2017.1401115.
Castonguay, L. G., Boswell, J. F., Constantino, M. J., Goldfried, M. R., & Hill, C. E. (2010). Training Implications of harmful effects of psychological treatments. *The American Psychologist, 65*, 34–49.
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. https://doi.org/10.1089/cpb.2005.8.387.
Clarke, M. J. (2005). Individual patient data meta-analyses. *Best Practice & Research Clinical Obstetrics & Gynaecology, 19*, 47–55. https://doi.org/10.1016/j.bpobgyn.2004.10.011.
Comprehensive Meta-Analysis (2013). *Comprehensive meta-analysis.* Englewood, NJ: Biostat (Version 2).
Cottraux, J., Bouvard, M., & Légeron, P. (1985). Méthodes et échelles d'évaluation des comportements (Editions d). *Issy les Moulineaux.*. https://doi.org/10.1590/1516-4446-2014-1560.
Cuijpers, P., Sijbrandij, M., Koole, S., Huibers, M., Berking, M., & Andersson, G. (2014). Psychological treatment of generalized anxiety disorder: A meta-analysis. *Clinical Psycholy Review, 34*(2), 130–140. https://doi.org/10.1016/j.cpr.2014.01.002.
Cusack, K., Jonas, D. E., Forneris, C. A., Wines, C., Sonis, J., Middleton, J. C., ... Gaynes, B. N. (2016). Psychological treatments for adults with posttraumatic stress disorder: A systematic review and meta-analysis. *Clinical Psychology Review, 43*, 128–141. https://doi.org/10.1016/j.cpr.2015.10.003.
Czerniak, E., Caspi, A., Litvin, M., Amiaz, R., Bahat, Y., Baransi, H., ... Plotnik, M. (2016). A novel treatment of fear of flying using a large virtual reality system. *Aerospace Medicine and Human Performance, 87*(4), 411–416. https://doi.org/10.3357/AMHP.4485.2016.
De La Rosa, A., & Cárdenas, G. (2012). Trastorno por estrés postraumático: Eficacia de un programa de tratamiento mediante realidad virtual para víctimas de violencia criminal en población mexicana. *Anuario de PsicologÃa, 42*(3), 377–391.
Difede, J., Cukor, J., Jayasinghe, N., Patt, I., Jedel, S., Spielman, L., ... Hoffman, H. G. (2007). Virtual reality exposure therapy for the treatment of posttraumatic stress disorder following September 11, 2001. *The Journal of Clinical Psychiatry, 68*(11), 1639–1647. https://doi.org/10.4088/JCP.v68n1102.
Dimidjian, S., & Hollon, S. D. (2010). How would we know if psychotherapy were harmful? *The American Psychologist, 65*(1), 21–33. https://doi.org/10.1037/a0017299.
Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*, 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x.
Ebert, D. D., Donkin, L., Andersson, G., Andrews, G., Berger, T., Carlbring, P., ... Cuijpers, P. (2016). Does Internet-based guided-self-help for depression cause harm? An individual participant data meta-analysis on deterioration rates and its moderators in randomized controlled trials. *Psychological Medicine, 46*(13), 2679–2693. https://doi.org/10.1017/S0033291716001562.
Echeburúa, E., Corral, P., García, E., Páez, D., & Borda, M. (1992a). Un nuevo inventario de agorafobia (IA). *Análisis y Modificación de Conducta, 18*, 101–112.
Echeburúa, E., Corral, P., García, E., Páez, D., & Borda, M. (1992b). Un nuevo inventario de agorafobia (IA). *Análisis y Modificación de Conducta, 18*, 101–112.
Edwards, D. W., Yarvis, R. M., Mueller, D. P., Zingale, H. C., & Wagman, W. J. (1978). Test taking and the stability of adjustment scales: Can we assess patient deterioration? *Evaluation Quarterly, 2*, 275–291. https://doi.org/10.1177/0193841X7800200206.
Emmelkamp, P. M. G., & Ehring (2014). *The wiley handbook of anxiety disorders.* Chichester: Wiley Blackwell.
Emmelkamp, P. M. G., Bruynzeel, M., Drost, L., & van der Mast, C. A. (2001). Virtual reality treatment in acrophobia: A comparison with exposure in vivo. *Cyberpsychology & Behavior: The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society, 4*(3), 335–339. https://doi.org/10.1089/109493101300210222.
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. https://doi.org/10.1016/S0005-7967(01)00023-7.
Evans, C., Margison, F., & Barkham, M. (1998). The contribution of reliable and clinically significant change methods to evidence-based mental health. *Evidence-based Mental Health, 1*, 70–72. https://doi.org/10.1136/ebmh.1.3.70.
Fernández-Álvarez, J., Díaz-García, A., González-Robles, A., Baños, R., García-Palacios, A., & Botella, C. (2017). Dropping out of a transdiagnostic online intervention: A qualitative analysis of client's experiences. *Internet Interventions, 10*, 29–38. https://doi.org/10.1016/j.invent.2017.09.001.
Foa, E. B., Chrestman, K. R., & Gilboa-Schechtman, E. (2009). *Prolonged exposure therapy for adolescents with PTSD: Emotional processing of traumatic experiences: Therapist guide.* Programs that workhttps://doi.org/10.1093/med:psych/9780195308501.001.0001.
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. https://doi.org/10.1016/S0005-7967(01)00068-7.
García-Palacios, A., Botella, C., Hoffman, H., & Fabregat, B. A. (2007). Comparing acceptance and refusal rates of virtual reality exposure vs. in vivo exposure by participants with specific phobias. *CyberPsychology & Behavior, 10*(5), 722–724.
Gregg, L., & Tarrier, N. (2007). Virtual reality in mental health: A review of the literature. *Social Psychiatry and Psychiatric Epidemiology, 42*(5), 343–354.
Hatfield, D., McCullough, L., Frantz, S. H. B., & Krieger, K. (2010). Do we know when our clients get worse? An investigation of therapists' ability to detect negative client change. *Clinical Psychology & Psychotherapy, 17*(1), 25–32. https://doi.org/10.1002/cpp.656.
Haug, T., Brenne, L., Johnsen, B. H., Berntzen, D., Gotestam, K. G., & Hugdahl, K. (1987). A tree-systems analysis of fear of flying: A comparison of a consonant vs. A non-consonant treatment method. *Behaviour Research and Therapy, 25*, 187–194.
Heeren, A., Maurange, P., Rossignol, M., Vanhaelen, M., Peschard, V., Eeckhout, C., ... Phillippot, P. (2012). Self-report version of the liebowitz social anxiety scale: Psychometric properties of the french version. *Canadian Journal of Behavioural Science, 44*(2), 99–107.
Heimberg, R. G., Leibowitz, M. R., Hope, D. A., Schneier, F. R., Holt, C. S., Welkowitz, L. A., ... Klein, D. F. (1998). Cognitive behavioral group therapy vs. Phenelzine therapy for social phobia. *Archives of General Psychiatry, 55*(12), 1133–1141. https://doi.org/10.1001/archpsyc.55.12.1133.
Higgins, J. T., Altman, D. G., Gotzsche, P. C., Juni, P., Moher, D., Oxman, A. D., ... Sterne, J. A. (2011). The Cochrane Collaboration's tool for assessing risk of bias in randomised trials. *BMJ (Clinical Research Ed), 343* Article d5928.
Higgins, J. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ (Clinical Research Ed), 327*(7414), https://doi.org/10.1136/bmj.327.7414.557.
Hofmann, S. G. (2004). Cognitive mediation of treatment change in social phobia. *Journal of Consulting and Clinical Psychology, 72*(3), 393–399. https://doi.org/10.1037/0022-006X.72.3.393.
Hofmann, S. G., & Di Bartolo, P. M. (2014). *Social anxiety: Clinical, developmental, and social perspectives* (3rd ed.). New York, NY: Elsevier.
Hofmann, S. G., & Otto, M. W. (2008). *Cognitive-behavior therapy of social anxiety disorder: Evidence-based and disorder specific treatment techniques.* New York: Routledgehttps://doi.org/10.1016/j.bbr.2016.03.028.
Imel, Z. E., Caperton, D. D., Tanana, M., & Atkins, D. C. (2017). Technology-enhanced human interaction in psychotherapy. *Journal of Counseling Psychology, 64*(March), 385–393. https://doi.org/10.1037/cou0000213.
Jacobson, N. S., Follette, W. C., & Revenstorf, D. (1984). Psychotherapy outcome research: Methods for reporting variability and evaluating clinical significance. *Behavior Therapy, 15*, 336–352. https://doi.org/10.1016/S0005-7894(84)80002-7.
Johansson, O., Michel, T., Andersson, G., & Paxling, B. (2015). Experiences of non-adherence to Internet-delivered cognitive behavior therapy: A qualitative study. *Internet Interventions, 2*(2), 137–142. https://doi.org/10.1016/j.invent.2015.02.006.
Jonsson, U., Alaie, I., Parling, T., & Arnberg, F. K. (2014). Reporting of harms in randomized controlled trials of psychological interventions for mental and behavioral disorders: A review of current practice. *Contemporary Clinical Trials, 38*(1), 1–8. https://doi.org/10.1016/j.cct.2014.02.005.
Kampmann, I. L., Emmelkamp, G. P. M., Hartanto, D., Brinkman, W. P., Zijlstra, H. B. J., & Morina, N. (2016). Exposure to virtual social interactions in the treatment of social anxiety disorder: A randomized controlled trial. *Behaviour Research and Therapy, 77*, 147–156. https://doi.org/10.1016/j.brat.2015.12.016.
Karyotaki, E., et al. (2018). Is self-guided internet-based cognitive behavioural therapy (iCBT) harmful? An individual participant data meta-analysis. *Psychological Medicine,* 1–11. https://doi.org/10.1017/S0033291718000648.
Kazdin, A. E., & Blase, S. L. (2011). Rebooting psychotherapy research and practice to reduce the burden of mental illness. *Perspectives on Psychological Science, 6*(1), https://doi.org/10.1177/1745691610393527.
Klinger, E., & Bouchard, S. (2005). Virtual reality therapy versus cognitive behavior therapy for social phobia: A preliminary controlled study. *Cyberpsycholgy and Behavior, 8*(1), 76–89. https://doi.org/10.1089/cpb.2005.8.76.
Kraus, D. R., Seligman, D. A., & Jordan, J. R. (2005). Validation of a behavioral health treatment outcome and assessment tool designed for naturalistic settings: The Treatment Outcome Package. *Journal of Clinical Psychology, 61*(3), 285–314.
Krijn, M., Emmelkamp, P. M. G., Olafsson, R. P., Bouwman, M., Van Gerwen, L. J., Spinhoven, P., ... 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.
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. https://doi.org/10.1016/S0005-7967(03)00139-6.
Ladwig, I., Rief, W., & Nestoriuc, Y. (2014). Welche Risiken und Nebenwirkungen hat Psychotherapie? - Entwicklung des Inventars zur Erfassung Negativer Effekte von Psychotherapie (INEP). *Verhaltenstherapie, 24*(4), 252–263. https://doi.org/10.1159/000367928.
Lambert, M. J. (2013). Outcome in psychotherapy: The past and important advances. *Psychotherapy, 50*, 42–51. https://doi.org/10.1037/a0030682.
Lambert, M. J., & Shimokawa, K. (2011). Collecting client feedback. *Psychotherapy, 48*(1), 72–79. https://doi.org/10.1037/a0022238.
Lambert, M. J., Morton, J. J., Hatfield, D. R., Harmon, C., Hamilton, S., Shimokawa, K., et al. (2004). *Administration and scoring manual for the Outcome Questionnaire (OQ 45.2).* (3 ed.). Wilmington, DE: American Professional Credentialling Services LLC.
Landon, T. M., & Barlow, D. H. (2004). Cognitive-behavioral treatment for panic disorder: Current status. *Journal of Psychiatric Practice, 10*(4), 211–226.
Leary, M. R. (1983). A brief version of the fear of negative evaluation scale. *Personality & Social Psychology Bulletin, 9*, 371–375.
Lee, E. H., Kim, J. H., & Yu, B. H. (2009). Reliability and validity of the self-report version of the panic disorder severity scale in Korea. *Depression and Anxiety, 26*(8), E120–E123.
Linden, M. (2013). How to define, find and classify side effects in psychotherapy: From unwanted events to adverse treatment reactions. *Clinical Psychology & Psychotherapy, 20*(4), 286–296. https://doi.org/10.1002/cpp.1765.
Linden, M., & Schermuly-Haupt, M. L. (2014). Definition, assessment and rate of psychotherapy side effects. *World Psychiatry: Official Journal of the World Psychiatric Association (WPA), 13*(3), 306–309. https://doi.org/10.1002/wps.20153.

Lindner, P., Miloff, A., Hamilton, W., Reuterskiöld, L., Andersson, G., Powers, M., ... Carlbring, P. (2017). Creating state of the art, next-generationVirtual reality exposure therapies for anxiety disorders using con- sumer hardware platforms: Design considerations and future direc- tion. *Cognitive Behaviour Therapy, 46*(5), 404–420. https://doi.org/10.1080/16506073.2017.1280843.

Lorenzo González, M., Castro, W. P., Pitti González, C. T., Bethencourt Pérez, J. M., de la Fuente Portero, J. A., & Marco, R. G. (2011). Efficacy of virtual reality exposure therapy combined with two pharmacotherapies in the treatment of agoraphobia. *International Journal of Clinical and Health Psychology, 11*(2), 189–203.

Lutz, W., De Jong, K., & Rubel, J. (2015). Patient-focused and feedback research in psychotherapy: Where are we and where do we want to go? *Psychotherapy Research, 25*(6), 625–632. https://doi.org/10.1080/10503307.2015.1079661.

Lutz, W., Ehrlich, T., Rubel, J., Hallwachs, N., Röttger, M., Jorasz, C., ... Tschitsaz-Stucki, A. (2013). The ups and downs of psychotherapy: Sudden gains and sudden losses identified with session reports. *Psychotherapy Research, 23*(1), 14–24. https://doi.org/10.1080/10503307.2012.693837.

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. https://doi.org/10.1037//0022-006X.70.5.1112.

Marks, I. M., & Mathews, A. M. (1979). Brief standard self-rating for phobic participants. *Behaviour Research and Therapy, 17*(3), 263–267.

Martinovich, Z., Saunders, S., & Howard, K. (1996). Some comments on “assessing clinical significance.”. *Psychotherapy Research, 6*, 124–132. https://doi.org/10.1080/10503309612331331648.

McLay, R. N., 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, Behavior and Social Networking, 14*(4), 223–229. https://doi.org/10.1089/cyber.2011.0003.

Meyerbröker, K., & Emmelkamp, P. M. (2010). Virtual reality exposure therapy in anxiety disorders: A systematic review of processöandöoutcome studies. *Depression and Anxiety, 27*(10), 933–944.

Meyerbroeker, K., Morina, N., Kerkhof, G. A., & Emmelkamp, P. M. G. (2013). Virtual reality exposure therapy does not provide any additional value in agoraphobic participants: A randomized controlled trial. *Psychotherapy and Psychosomatics, 82*(3), 170–176. https://doi.org/10.1159/000342715.

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. https://doi.org/10.1089/cyber.2009.0277.

Miller, S. D., Duncan, B. L., Sorrell, R., & Brown, G. S. (2005). The partners for change outcome management system. *Journal of Clinical Psychology: In-Session, 61*, 199–208. https://doi.org/10.1002/jclp.20111.

Miyahira, S. D., Folen, R. A., Hoffman, H. G., Garcia-Palacios, A., Spira, J. L., & Kawasaki, M. (2012). The effectiveness of VR exposure therapy for PTSD in returning warfighters. *Annual Review of Cyber Therapy and Telemedicine, 10*, 128–132. https://doi.org/10.3233/978-1-61499-121-2-128.

Mohr, D. C. (1995). Negative outcome in psychotherapy: A critical review. *Clinical Psychology Science and Practice, 2*, 1–27.

Mohr, D. C., Weingardt, K. R., Reddy, M., & Schueller, S. M. (2017). Three problems with current digital mental health research…and three things we can do about them. *Psychiatric Services, 68*(5), 427–429. https://doi.org/10.1176/appi.ps.201600541.

Moldovan, R., & David, D. (2014). One session treatment of cognitive and behavioral therapy and virtual reality for social and specific phobias. Preliminary results from a randomized clinical trial. *Journal of Evidence-Based Psychotherapies, 14*(1), 67–83.

Morina, N., Ijntema, H., Meyerbröker, K., & Emmelkamp, P. M. G. (2015). Can virtual reality exposure therapy gains be generalized to real-life? A meta-analysis of studies applying behavioral assessments. *Behaviour Research and Therapy, 74*(9), 18–24. https://doi.org/10.1016/j.brat.2015.08.010.

Mühlberger, A., Herrmann, M. J., Wiedemann, G., Ellgring, H., & Pauli, P. (2001). Repeated exposure of flight phobics to flights in virtual reality. *Behaviour Research and Therapy, 39*, 1033–1050.

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

Nordberg, S. S., Castonguay, L. G., Mcaleavey, A. A., Locke, B. D., & Hayes, J. A. (2016). Enhancing feedback for clinical use: Creating and evaluating profiles of clients seeking counseling. *Journal of Counseling Psychology, 63*(3), 278–293. https://doi.org/10.1037/cou0000147.

North, M. N., North, S. M., & Coble, J. R. (1996). Effectiveness of virtual environment desensitization in the treatment of agoraphobia. *Presence Teleoperators & Virtual Environments, 5*(3), 346–352.

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. https://doi.org/10.1002/da.20910.

Ost, L. G. (1989). One-session treatment for specific phobias. *Behaviour Research and Therapy, 27*(1), 1–7.

Page, S., & Coxon, M. (2016). Virtual reality exposure therapy for anxiety disorders: Small samples and no controls? *Frontiers in Psychology, 7*, 326.

Parker, G., Fletcher, K., Berk, M., et al. (2013). Development of a measure quantifying adverse psychotherapeutic ingredients: The Experience of Therapy Questionnaire (ETQ). *Psychiatry Research, 206*(2–3), 293–301. https://doi.org/10.1016/j.psychres.2012.11.026.

Parry, G. D., Crawford, M. J., & Duggan, C. (2016). Iatrogenic harm from psychological therapies-time to move on. *The British Journal of Psychiatry, 208*(3), 210–212. https://doi.org/10.1192/bjp.bp.115.163618.

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

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

Peñate Castro, W., Roca Sánchez, M. J., Pitti González, C. T., Bethencourt, J. M., De la Fuente Portero, J. A., & Marco, R. G. (2014). Cognitive-behavioral treatment and antidepressants combined with virtual reality exposure for participants with chronic agoraphobia. *International Journal of Clinical and Health Psychology, 14*(1), 9–17. https://doi.org/10.1016/S1697-2600(14)70032-8.

Peterson, A. L., Roache, J. D., Raj, J., & Young-McCaughan, S. (2013). The need for expanded monitoring of adverse events in behavioral health clinical trials. *Contemporary Clinical Trials, 34*(1), 152–154. https://doi.org/10.1016/j.cct.2012.10.009.

Pitti, C. T., Peñate, W., De La Fuente, J., Bethencourt, J. M., Roca-Sánchez, M. J., Acosta, L., ... Gracia, R. (2015). The combined use of virtual reality exposure in the treatment of agoraphobia. *Actas Espanolas de Psiquiatria, 4343*(44), 133–41133.

Pompoli, A., Furukawa, T. A., Imai, H., Tajika, A., Efthimiou, O., & Salanti, G. (2016). Psychological therapies for panic disorder with or without agoraphobia in adults: A network meta-analysis. *Cochrane Database Systematic Review, 13*(4), CD011004.

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

Rebenitsch, L., & Owen, C. (2016). Review on cybersickness in applications and visual displays. *Virtual Reality, 20*(2), 101–125. https://doi.org/10.1007/s10055-016-0285-9.

Reger, G. M., Koenen-Woods, P., Zetocha, K., Smolenski, D. J., Holloway, K. M., Rothbaum, B. O., ... Gahm, G. A. (2016). Randomized controlled trial of prolonged exposure using imaginal exposure vs. Virtual reality exposure in active duty soldiers with deployment-related posttraumatic stress disorder (PTSD). *Journal of Consulting and Clinical Psychology, 84*(11), 946–959. https://doi.org/10.1037/ccp0000134.

Riva, G. (2005). Virtual reality in psychotherapy: Review. *CyberPsychology & Behavior, 8*(3), https://doi.org/10.1089/cpb.2005.8.220.

Rizzo, A. A., Wiederhold, M. D., & Buckwalter, J. G. (1998). Basic issues in the use of virtual environments for mental health applications. *Studies in Health Technology and Informatics*, 21–42.

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. https://doi.org/10.1016/j.beth.2005.04.004.

Rothbaum, B. O., Hodges, L., & Smith, S. (1999). Virtual reality exposure therapy abbreviated treatment manual: Fear of flying application. *Cognitive and Behavioral Practice, 6*(3), 234–244. https://doi.org/10.1016/S1077-7229(99)80081-9.

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

Rothbaum, B. O., Difede, J. Y., & Rizzo, A. (2008). *Therapist treatment manual for virtual reality exposure therapy*. USA: Virtually Better.

Rothbaum, B. O., Hodges, L. F., Kooper, R., Opdyke, D., Williford, J. S., & North, M. (1995a). Effectiveness of computer-generated (virtual reality) graded exposure in the treatment of acrophobia. *The American Journal of Psychiatry, 152*(4), 626–628. https://doi.org/10.1176/ajp.152.4.626.

Rothbaum, B. O., Hodges, L. F., Kooper, R., Opdyke, D., Williford, J. S., & North, M. (1995b). Virtual reality graded exposure in the treatment of acrophobia: A case report. *Behavior Therapy, 26*(3), 547–554. https://doi.org/10.1016/S0005-7894(05)80100-5.

Roy, S., Klinger, E., Légeron, P., Lauer, F., Chemin, I., & Nugues, P. (2003). Definition of a VR-Based protocol to treat social phobia. *CyberPsychology & Behavior, 6*(4), 411–420. https://doi.org/10.1089/109493103322278808.

Rozental, A., Boettcher, J., Andersson, G., Schmidt, B., & Carlbring, P. (2015). Negative effects of internet interventions: A qualitative content analysis of patients' experiences with treatments delivered online. *Cognitive Behaviour Therapy, 44*(3), 223–236.

Rozental, A., Kottorp, A., Boettcher, J., Andersson, G., & Carlbring, P. (2016). Negative effects of psychological treatments: An exploratory factor analysis of the negative effects questionnaire for monitoring and reporting adverse and unwanted events. *PloS One, 11*(6), 1–22. https://doi.org/10.1371/journal.pone.0157503.

Rozental, A., Magnusson, K., Boettcher, J., Andersson, G., & Carlbring, P. (2017). For better or worse: An individual patient data meta-analysis of deterioration among participants receiving Internet-based cognitive behavior therapy. *Journal of Consulting and Clinical Psychology, 85*(2), 160–177. https://doi.org/10.1037/ccp0000158.

Safran, J., Muran, C., & Eubanks-Carter, C. (2011). Repairing alliance ruptures. *Psychotherapy, 48*(1), 80–87. https://doi.org/10.1037/a0022140.

Scholing, A., & Emmelkamp, P. M. G. (1993). Exposure with and without cognitive therapy for generalized social phobia: Effects of individual and group treatment. *Behaviour Research and Therapy, 31*(7), 667–681. https://doi.org/10.1016/0005-7967(93)90120-J.

Solstad, S. M., Castonguay, L. G., & Moltu, C. (2017). Patients' experiences with routine outcome monitoring and clinical feedback systems: A systematic review and synthesis of qualitative empirical literature. *Psychotherapy Research*, 1–14.

Speer, D. C. (1992). Clinically significant change: Jacobson and Truax (1991) revisited.

Journal of Consulting and Clinical Psychology, 60, 402–408. https://doi.org/10.1037/0022-006X.60.3.402.

Spira, J. L., Pyne, J. M., & Wiederhold, B. K. (2007). Experiential methods in the treatment of combat PTSD. 2007 In Charles R Figley, & William P Nash (Eds.). *Combat stress injury: Theory, research, and management* (pp. 205–218). New York, NY, US: Routledge/Taylor US. https://doi.org/10.4324/9780203943694 (Pp 205-218) Xxi, 341 Pp.

Strupp, H. H., Hadley, S. W., & Gomes-Schwartz, B. (1977). *New York: Jason aronson. Psychotherapy for better or worse: The problem of negative effects*.

Suh, C. S., Strupp, H. H., & O'Malley, S. S. (1986). The vanderbilt process measures: The psychotherapy process scale (VPPS) and the negative indicators scale (VNIS). In L. S. Greenberg, & W. M. Pinsof (Eds.). *Guilford clinical psychology and psychotherapy series. The psychotherapeutic process: A research handbook* (pp. 285–323). New York, NY, US: Guilford Press.

Tang, T. Z., & DeRubeis, R. J. (1999). Sudden gains and critical sessions in cognitive-behavioral therapy for depression. *Journal of Consulting and Clinical Psychology, 67*(6), 894–904. https://doi.org/10.1037/0022-006X.67.6.894.

Tortella-Feliú, M., Botella, C., Llabrés, J., Bretón-lópez, J. M., Riera, A., Baños, R. M., ... Gelabert, J. M. (2011). Virtual reality versus computer-aided exposure treatments for fear of flying. *Behavior Modification, 35*(1), 3–30. https://doi.org/10.1177/0145445510390801.

Triscari, M. T., Faraci, P., Catalisano, D., D'Angelo, V., & Urso, V. (2015). Effectiveness of cognitive behavioral therapy integrated with systematic desensitization, cognitive behavioral therapy combined with eye movement desensitization and reprocessing therapy, and cognitive behavioral therapy combined with virtual reality expo. *Neuropsychiatric Disease and Treatment, 11*, 2591–2598. https://doi.org/10.2147/NDT.S93401.

Van Gerwen, L. J., Spinhoven, P., Van Dyck, R., & Diekstra, R. F. (1999). Construction and psychometric characteristics of two self-report questionnaires for the assessment of fear of flying. *Psychological Assessment, 11*(2), 146.

Vincelli, F., Choi, Y. H., Molinari, E., Wiederhold, B. K., & Riva, G. (2000). Experiential cognitive therapy for the treatment of panic disorder with agoraphobia: Definition of a clinical protocol. *CyberPsychology & Behavior, 3*, 375–385. https://doi.org/10.1089/10949310050078823.

Von Hippel, P. (2015). The heterogeneity statistic I 2 can be biased in small meta-analyses. *BMC Medical Research Methodology, 15*, 35. https://doi.org/10.1186/s12874-015-0024-z.

Waizmann, V., Doran, J., Bolger, K., Gómez Penedo, J. M., Safran, J., & Roussos, A. (2015). Alliance Negotiation Scale (Ans-A): Evidence on validity and reliability. *Revista Argentina de Clínica Psicológica, 24*(3), 243–254.

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. https://doi.org/10.1177/0145445509331926.

Whipple, J. L., & Lambert, M. J. (2011). Outcome measures for practice. *Annual Review of Clinical Psychology, 7*(1), 87–111. https://doi.org/10.1146/annurev-clinpsy-040510-143938.

Wiederhold, B., & Bouchard, S. (2014). *Advances in virtual reality and anxiety disorders.* Boston, MA: Springer.

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, 6*(3 SPEC), 218–223. https://doi.org/10.1109/TITB.2002.802378.

Wise, E. A. (2004). Methods for analyzing psychotherapy outcomes: A review of clinical significance, reliable change, and recommendations for future directions. *Journal of Personality Assessment, 82*, 50–59. https://doi.org/10.1207/s15327752jpa8201_10.

Wolitzky-Taylor, K. B., Horowitz, J. D., Powers, M. B., & Telch, M. J. (2008). Psychological approaches in the treatment of specific phobias: A meta-analysis. *Clinical Psychology Review, 28*(6), 1021–1037. https://doi.org/10.1016/j.cpr.2008.02.007.