



UiT The Arctic University of Norway

Department of Psychology – The Faculty of Health Sciences

The Effects of Virtual Reality on Procedural Pain and Anxiety in Pediatrics

A systematic review and meta-analysis

Rikke Nordgård

Thesis, cand.psychol., PSY-2901, December 2020

The Effects of Virtual Reality on Procedural Pain and Anxiety in Pediatrics

Effekten av Virtuell Virkelighet (VR) på Barns Smerte og Angst i forbindelse med
Medisinske Prosedyrer

Rikke Nordgård

Department of Psychology

University of Tromsø – The Arctic University of Norway

PSY-2901 Hovedoppgave

cand.psychol.

Supervisor: Dr. Torstein Låg

December 2020

Preface

This thesis is submitted as partial fulfillment of the cand.psyhol. degree at the University of Tromsø – The Arctic University of Norway.

While I came up with the initial idea for the study, my supervisor and I designed the study together and I received continuous feedback on the review protocol. The supervisor also provided methodological literature and guidance. We both assessed studies for eligibility. I conducted the risk of bias judgements and data analyses independently but consulted with the supervisor when necessary. Finally, the supervisor provided invaluable input on the interpretation of results and feedback on the writing.

I would like to thank my supervisor, Torstein Låg, for introducing me to meta-analyses and systematic reviews, and for his encouragement during the process of writing this thesis.

Keywords: Virtual reality, pediatrics, pain, anxiety, distraction, procedural preparation, immersion, interactivity

Abstract

Distraction and procedural preparation techniques are frequently used to manage pain and anxiety in children undergoing medical procedures. An increasing number of studies have indicated that Virtual Reality (VR) can be used to deliver these interventions, but treatment effects vary greatly. The present study is a systematic review and meta-analysis of studies that have used VR to reduce procedural pain and anxiety in children. It is the first meta-analytic assessment of the potential influence of technical specifications (immersion) and degree of user-system interactivity on treatment effects. 65 studies were identified, of which 42 reported pain outcomes and 35 reported anxiety outcomes. Results indicate large effect sizes in favor of VR for both outcomes. Larger effects were observed in dental studies and studies that used non-interactive VR. No relationship was found between the degree of immersion or participant age and treatment effects. Most studies were found to have a high risk of bias and there are strong indications of publication bias. The results and their implications are discussed in context of these limitations, and modified effect sizes are suggested. Finally, recommendations for future investigations are provided.

Introduction

The management of pain and anxiety in children undergoing medical procedures remains sub-optimal (Birnie et al., 2014; Friedrichsdorf & Goubert, 2020; Stevens et al., 2011). As well as causing excessive and unnecessary suffering, undertreated procedural distress may have long-term negative effects on child health and development, as well as treatment outcomes (Young, 2005). Current best practice guidelines recommend that non-pharmacological interventions are routinely implemented in treatment plans (Wilson-Smith, 2011). Two common, non-pharmacological approaches are distraction and procedural preparation. Distraction involves the use of distractors like music and television to divert attention away from noxious stimuli, whereas preparation techniques usually entail information about the procedure or exposure to the procedural setting (e.g., a tour of the clinic). Over the last couple of decades, researchers have explored whether virtual reality (VR) can be used to deliver and possibly enhance distraction and preparation interventions in pediatrics.

Previous reviews have indicated the potential of VR in pediatrics (e.g., Eijlers, Utens, et al., 2019; Georgescu et al., 2020; Indovina et al., 2018; Lambert et al., 2020). Its immersive, interactive nature is thought to provide particularly captivating distraction, as well as a cost-effective and engaging medium for procedural preparation. However, previous meta-analyses have revealed great heterogeneity in treatment effects and little is known about the underlying mechanisms and factors that determine the effectiveness of VR interventions (Li et al., 2011).

The present study is a systematic review and meta-analysis of studies that have used VR to reduce procedural pain and anxiety in pediatrics. To address the variability of effect sizes that have been observed across studies, the potential influence of various VR,

procedural, and participant characteristics will be explored. The main focus will be on characteristics of VR systems, including the technical specifications and degree of user-system interaction. While some evidence suggest that VR characteristics influence treatment effects (e.g., Hoffman et al., 2006; Johnson & Coxon, 2016; Wender et al., 2009), this has not yet been assessed in a meta-analysis.

Virtual Reality in Healthcare

Virtual reality (VR) may be described as an interactive, immersive, computer-generated environment or experience (Gigante, 1993; Pan & Hamilton, 2018). Typically presented on a head-mounted display (HMD), the screens are positioned close to the users' eyes with full or partial occlusion of their physical surroundings. Images are often three-dimensional and continuously adjusted in accordance with the user's head movements (Slater & Sanchez-Vives, 2016, p. 3). Such features contribute to the sense of being surrounded by or present in the virtual environment that is unique to VR.

Various applications of VR in health have been explored, including in the assessment and treatment of patients. Reviews of the literature have reported significant methodological issues and a need for further research, but nevertheless indicate a considerable potential for VR in various clinical settings. For example, VR interventions have been applied in rehabilitation (Laver et al., 2017), habilitation (Snider et al., 2010), psychiatry (Freeman et al., 2017), geriatrics (Neri et al., 2017), and palliative care (Niki et al., 2019). An increasing number of studies have demonstrated its utility in the management of pain and anxiety caused by medical procedures in adult and pediatric populations (Chan et al., 2018; Eijlers, Utens, et al., 2019; Georgescu et al., 2020; Malloy & Milling, 2010).

Procedural Pain and Anxiety in Pediatrics

Children in developed countries undergo an increasing number of potentially painful and anxiety-inducing medical procedures (Curtis et al., 2012). Depending on their age and development, children may experience these procedures as more aversive than adults due to limitations in their ability to communicate their pain and need for pain management, to understand why the procedure is necessary, and to self-regulate (Cohen et al., 2008; McMurtry et al., 2015; Slifer, 2003). While conditions like cancer and burn injuries often require repeated or particularly distressing procedures (Gandhi et al., 2010; Twycross et al., 2015), routine procedures like venipuncture and immunizations are also known to induce considerable pain and anxiety in children (Reid et al., 2014). If poorly managed, procedural pain and anxiety could have detrimental effects on child health and development, as well as treatment outcomes (Mathews, 2011; Wilson-Smith, 2011). For example, painful and frightening medical procedures in childhood have been linked to alterations in pain responses later in life (Kennedy et al., 2008; Pate et al., 1996; Taddio et al., 1997), reduced effects of future pharmacological analgesia (Weisman et al., 1998), and development of needle phobia (McMurtry et al., 2015).

The International Association for the Study of Pain ([IASP], 2011) defines pain as “an unpleasant sensory and emotional experience associated with, or resembling that associated with, actual or potential tissue damage”. Procedural pain refers to pain associated with medical (or dental) procedures. Procedural anxiety may be described as a response to such procedures characterized by feelings of dread and apprehensiveness, accompanied by physical symptoms such as sweating and increased heart rate (Lavoie, 2013). The relationship between procedural pain and anxiety is intertwined and complex - for example, they frequently co-

occur and exacerbate each other (Cohen et al., 2004; Kao & Schwartz, 2019; McMurtry et al., 2015).

The experience of pain is modulated by multiple biological, psychological, and social processes (Bentley, 2014, p. 27). Some factors known to modulate pain top-down include attention towards painful stimuli, expectation of pain, anxiety, and previous experiences with pain (Bentley, 2014; Linton & Shaw, 2011). Knowledge of these and other pain-modulating mechanisms have informed the development of various non-pharmacological pain management approaches, including distraction and procedural preparation (Curtis et al., 2011). Current best practice guidelines recommend a combination of pharmacological and non-pharmacological interventions in the treatment of procedural pain and anxiety (e.g., Association of Paediatric Anaesthetists of Great Britain and Ireland, 2012). Over the last couple of decades, researchers have explored whether VR can be used to effectively deliver distraction and preparation interventions in pediatrics.

Distraction and Preparation Techniques

Distraction techniques are commonly used during painful or frightening procedures of shorter durations (DeMore & Cohen, 2005). They involve the use of stimuli such as videos, music, and conversation to divert attention away from noxious stimuli (Schechter et al., 2007). No single theory can fully account for the effects of distraction analgesia (DeMore & Cohen, 2005), but they are often understood in terms of attentional capacities. It is assumed that pain perception requires attention, and that by focusing on distractors, less attentional resources are available for pain perception (Gupta et al., 2017; McCaul & Malott, 1984). However, distraction may also work through other mechanisms. For example, pleasant distractors may have inherent positive effects on mood, arousal, and anxiety, all of which have the capacity to alter pain perception (Johnson, 2005). Attention, mood, arousal, and

anxiety can all be understood as processes inhibiting nociceptive signals as described in the gate control and neuromatrix theories of pain (Melzack, 1999; Melzack & Wall, 1965). Due to its immersive, interactive, and multisensory properties, VR is thought to be particularly captivating and thus provide superior distraction (Slifer, 2013, p. 93).

Another common way of reducing pain and anxiety is procedural preparation, often in the form of a verbal briefing, written materials, or a tour of the clinic (Curtis et al., 2011). Such techniques are meant to reduce anxiety (and possibly also pain) by promoting a sense of control and adaptive behaviors, as well as desensitizing the child to the medical procedure and the setting in which it takes place (Edward et al., 2015; Jaaniste et al., 2007). Research on virtual reality exposure therapy (VRET) has established that VR can be used to expose users effectively and ecologically to feared stimuli (Boeldt et al., 2019; Botella et al., 2017). Based on these findings, researchers have recently begun exploring whether VR can be used for procedural preparation (Eijlers, Utens, et al., 2019). In addition to exposure to the medical procedure and the environment in which it takes place, VR preparation may involve modelling, instructions, and rehearsal of the procedure (e.g., Han et al., 2019; Liszio et al., 2020; Ryu et al., 2018).

The Influence of Virtual Reality Characteristics

VR systems offer varying degrees of interaction with the user. Less interactive forms of VR include videos converted to a 360/180-degree format for viewing on a VR headset. While the user may effect changes in perception (i.e., looking around the virtual environment in 360/180 degrees through tracking of head movements), he or she is nevertheless a passive spectator of the virtual environment. On the other hand, VR games or simulations may offer interactivity *beyond* head tracking, such as navigation in the virtual environment, social

interaction with avatars, or manipulation of virtual objects. In the present study, head tracking will be considered an aspect of immersion, and not interactivity.

A potential impact of VR interactivity on procedural pain and anxiety seems plausible. It is generally assumed that active distraction poses greater attentional demands on patients than passive distraction, thus providing superior analgesia (Slifer, 2003, p. 91). Some studies have reported this pattern for VR specifically (e.g., Dahlquist et al., 2007; Gutiérrez-Maldonado et al., 2011; Gutiérrez-Martínez et al., 2011; Wender et al., 2009). In addition, VR interactivity may augment learning and memory (e.g., James et al., 2002; Tuena et al., 2019), which could be beneficial when used for procedural preparation.

VR systems also vary in terms of technological sophistication, which may be conceptualized as varying degrees of immersion (Agrawal et al., 2019; Nilsson et al., 2016). According to Slater and Wilbur (1997), a highly immersive system should minimize signals from the physical world (e.g., fully occlude the user's physical surroundings), stimulate multiple senses (e.g., visual, auditive, and tactile), visually surround the user (e.g., a wide field of view), provide a vivid representation of the virtual environment (e.g., high screen resolution) and match the actions of the participant with the sensory output of the system (e.g., low latency between head rotation and subsequent change in images displayed). This concept of immersion provides a useful framework for comparison of VR systems, as it can be operationalized and objectively measured (Cummings & Bailenson, 2016; Slater et al., 2009).

The degree of immersion may have an impact on the effectiveness of VR interventions. According to Slater (2018), higher levels of immersion facilitate the perceptive illusion that the virtual environment is real, which he referred to as presence. Presence is commonly thought to increase the effectiveness of various forms of VR interventions

(Cummings & Bailenson, 2016, p. 273). VR studies have indicated a possible relationship between immersion/presence and the effectiveness of VR distraction analgesia (e.g., Hoffman et al., 2006; Hoffman et al., 2004).

Objectives

Previous reviews have indicated the potential of VR in pediatrics (e.g., Eijlers, Utens, et al., 2019; Georgescu et al., 2020; Iannicelli et al., 2019). However, nearly half of the studies included in the present review were published in 2019 and 2020. As the literature search of the most recent review (Georgescu et al., 2020) was conducted in 2018, an updated review is necessary. Another motivation for the present study is that previous reviews have not quantitatively assessed the differences between VR interventions. Considering the potential impact immersion and interactivity may have on treatment effects, such assessments could have important clinical implications.

Some previous reviews have employed somewhat vague definitions of VR in their inclusion criteria. For example, some authors have specified that they would only include ‘immersive VR’ (Chan et al., 2018) or ‘fully immersive VR’ (Eijlers, Utens, et al., 2019), but did not explicitly state their definition of these terms. It is crucial that these terms are clearly defined and consistently applied to avoid confusion. For example, it can be argued that some of the technologies (e.g., the eMagin 3DVisor) included in Eijlers, Utens, et al. (2019) are not fully immersive because their users can still see some of their physical surroundings (see Slater & Wilbur, 1997). Perhaps more importantly, unclear definitions of VR and immersion have resulted in an inconsistent inclusion of less advanced technologies that are often referred to as ‘audiovisual glasses’ (AV-glasses), rather than ‘VR’. These often lack features such as stereoscopy and head tracking, and often have a narrower field of view (Wismeijer & Vingerhoets, 2005). However, as review authors do not include ‘audiovisual glasses’ in their

search strategies, many studies using comparable technologies have previously been overlooked. The present review will therefore employ an inclusive definition of VR and a wider search strategy that also includes AV-glasses. The impact of VR characteristics (immersion and interactivity) will then be assessed to explore whether some technologies are more effective than others. The term ‘VR’ will mostly be used in the current study.

Previous reviews have reported great heterogeneity in effect sizes, which may reflect VR characteristics, but also differences between medical procedures and patients (e.g., age). The increased number of studies gained from also including AV-glasses will provide greater statistical power to explore these variables as potential sources of the heterogeneity. Identifying any such moderators of treatment effects may help inform the process of designing and implementing VR interventions for clinical use. Moreover, the increased number of studies may also provide more accurate estimates of the true effects of using VR during medical procedures.

The present study consists of a systematic literature review and meta-analysis of studies that have used VR to reduce procedural pain and anxiety in pediatrics. It also provides a meta-analytic assessment of the role of VR hardware specifications (i.e., immersion) and the degree of interaction between the patient and the VR system. The different groups of medical procedures and the age of participants will also be explored as potential moderators of treatment effects.

The research questions were as follows:

1. Do VR interventions reduce pain and anxiety in pediatric patients undergoing medical/dental procedures more than standard procedures?
2. Does effectiveness of VR interventions vary depending on the type of medical procedure, VR characteristics, and the age of patients?

Methods

The effects of VR interventions on procedural pain and anxiety in children will be evaluated through a systematic literature review and meta-analysis. Reporting will follow the Preferred Reporting Items of Systematic Review and Meta-Analysis (PRISMA) guidelines (Moher et al., 2009).

Protocol and Registration

A study protocol (CRD42020155056) was submitted to the Prospective Register of Systematic Reviews (PROSPERO) in May 2019. Some deviations from the protocol were deemed necessary. Firstly, as the differentiation between ‘VR’ and ‘audiovisual glasses’ was somewhat inconsistent in the literature, the search strategies were changed to also include ‘audiovisual glasses’ and variants of this term. Due to the resulting increase in search results, it was necessary to limit the volume of retrieved studies by also adding the terms ‘preparation’, ‘distraction’, ‘pain’, and ‘anxiety’. Secondly, it was discovered that the reporting of technical specifications of VR systems was poor and inconsistent, particularly in older studies. Selective reporting of technical specifications by authors and VR manufacturers hindered calculations that are required for accurate quantitative comparison in terms of screen resolution and field of view (see subsections ‘screen resolution’ and ‘field of view’). The screen refresh rate was also rarely disclosed in older studies. Screen resolution, field of view and refresh rate were thus omitted from quantitative analyses.

Eligibility Criteria

Study and Publication Characteristics

Studies were considered eligible if a VR intervention was compared experimentally or quasi-experimentally with any non-VR interventions or a control group. Studies with single-case studies and pretest-posttest designs without control groups were excluded.

Unpublished studies were eligible for inclusion. Only publications in English or one of the Scandinavian languages were considered eligible. No time restraints were applied.

Participant Characteristics

Only pediatric samples were eligible for inclusion. Pediatric patients were defined as 0-21 years of age, in accordance with recommendations issued by the American Academy of Pediatrics (Hardin & Hackell, 2017).

Intervention Characteristics

Studies were considered eligible if the VR intervention was used to reduce pain and/or anxiety in pediatric patients associated with medical or dental procedures through distraction or procedural preparation. VR was defined as a computer-generated virtual environment presented on a head-mounted device or other VR system that perceptually surrounds the user (i.e., cover all or most of the field of view). VR presented on conventional screens (with or without 3D-effects) were thus not eligible for inclusion. So-called audiovisual glasses were eligible for inclusion. Augmented reality (AR) technologies render images on a transparent screen that reveals the user's physical surroundings and were thus excluded.

Outcomes

Questionnaire and observational measures of pain and (state) anxiety were considered eligible. Stress and fear measures were accepted as anxiety measures, as these were thought to have a high degree of conceptual overlap with state anxiety (Öhman, 2008). Studies that used measures of procedural distress were excluded, as this concept includes dimensions of both pain and anxiety (McMurtry et al., 2015). Physiological measures and measures of maladaptive behavior were not considered valid pain or anxiety measures for the same reason.

Comparison Groups

Studies were eligible for inclusion if they compared VR interventions with no intervention, standard of care (SOC), or other non-VR interventions, such as television, videogames, and conventional verbal/written information about the procedure. The inclusion of both no intervention, SOC and other non-VR conditions was deemed necessary as Eijlers, Utens, et al. (2019) found that standard of care was often poorly defined, and often involved a variety of both pharmacological and non-pharmacological interventions.

Information Sources

The following databases were searched for research articles: PsycINFO, Web of Science, MEDLINE, EMBASE, SveMed+, Scopus, Google Scholar, Cochrane Central Register of Controlled Trials (CENTRAL), Bielefield Academic Search Engine (BASE), ClinicalTrials.gov, and International Clinical Trials Registry Platform (ICTRP). The latter three databases were included to also identify any ‘grey literature’, such as unpublished studies and theses. Only the first 150 publications were extracted from Google Scholar. Unpublished studies were collected by contacting researchers identified in bibliographies, search results or elsewhere. Article reference lists of included studies were also searched manually.

Search

Databases were searched using the following terms and their synonyms: Virtual reality/audiovisual glasses + pediatrics/child + anxiety/pain/preparation/distraction. Search strategies were adapted for each database. The complete search strategy for PsycINFO is presented in Table 1. The last search was conducted October 1, 2020, but manuscripts were received from contacted authors until November 25, 2020.

Table 1*Search strategy for the PsycINFO database*

1	exp Pediatrics/
2	child*.mp.
3	adolescen*.mp.
4	boy*.mp.
5	girl*.mp.
6	1 OR 2 OR 3 OR 4 OR 5
7	exp Virtual Reality/
8	audiovisual*.mp.
9	7 OR 8
10	exp. Distraction/
11	prepar*
12	exp Exposure/
13	exp Pain/
14	exp Analgesia/
15	exp Anxiety
16	anx*.mp.
17	10 OR 11 OR 12 OR 13 OR 14 OR 15 OR 16
18	6 AND 9 AND 17

Note. mp. = field code for title, abstract, heading word, table of contents, key concepts, original title, tests and measures, mesh.

Study Selection

Upon completion of the literature search and after removal of duplicates, each publication was screened for potential eligibility by the author. Researchers identified in trial registries and conference abstracts were contacted if any corresponding, published research

articles were not identified in the search results. The resulting list of studies were considered for eligibility by both the author and supervisor. Reasons for exclusions were recorded at this point. Any disagreements were resolved through discussion.

Data Collection Process

Data extraction was conducted using an Excel spreadsheet. The spreadsheet was piloted with five randomly selected studies that were coded independently by both the student and supervisor. As coding agreement was deemed satisfactory, the remaining data was collected independently by the student. Numerical study results were coded by the student and double checked for accuracy by the supervisor. Any disagreements were resolved through discussion. If sufficient information was not available in the articles, information was requested from corresponding authors on multiple occasions between May and November 2020. Co-authors were contacted if corresponding authors could not be reached. Efforts were made to locate updated contact information for researchers that did not respond. VR hardware or software specifications were also sourced from direct communication with manufacturers, technical manuals published online or vendors. Specifications sourced directly from articles were preferred, as authors may have reconfigured HMD settings.

Data Items

All data items were extracted as specified in the review protocol. If more than one measure of pain or anxiety were available, retrospectively, self-reported measures were prioritized. Self-reported measures were preferred as pain and anxiety are subjective and private experiences, and because observers' ability to accurately describe the patient's distress may be compromised as the VR headsets cover parts of the patient's face. For pain specifically, measures of sensory pain were preferred over measures of the affective or cognitive aspects of pain. Final values were preferred over change scores.

The following information was extracted from each primary study: (a) publication and study details (author(s), year published, study design, sample sizes, description of comparison groups); (b) participant characteristics (average age and a measure of dispersion, gender distribution, other health-related characteristics); (c) details regarding the pain and anxiety measures that were used (name of measures, timing of administration, informant); (d) the procedural setting (clinical context in which the procedure took place, the kind of medical procedure, timing of VR intervention); (e) results (key findings, summary statistics for VR and non-VR groups); (f) VR characteristics (technical specifications, degree and form of interactivity, and descriptions of media displayed). The VR characteristics (immersion and interactivity) are described in further detail below.

Immersion

The variables describing technical specifications are primarily based on Cummings and Bailenson (2016), who compiled a list of VR features that increase the level of immersion and thus the sense of being present in the virtual environment. The list of VR characteristics included in the present study is not exhaustive, but rather focused on the objective, purely technical properties that were deemed realistic to code. For example, the overall level of detail and realism in virtual environments were not included. In addition to hardware specifications, information was extracted regarding the number of senses stimulated, the level of user-system interactivity, and the media displayed to participants.

Screen Resolution. The screen resolution refers to the number of pixels the screen displays per frame (Kourtesis et al., 2019). A screen with a high resolution will be perceived to have greater fidelity, or ‘crispness’, of images displayed. Resolution is typically reported as horizontal x vertical pixels (e.g. 1280 x 1800), or pixels per inch (ppi). However, as pointed out by Hugues (2019), the pixel per degree (ppd) format more truly reflects the fidelity of the

display, as it is independent of the field of view. Calculating the ppd requires knowledge of the horizontal field of view, which is rarely disclosed. The screen resolutions were therefore not compared quantitatively.

Field of View. The field of view (FoV) refers to the degrees of the VR user's visual field that is occupied by the virtual environment (Cummings & Bailenson, 2016). FoV may be reported as diagonal, horizontal or vertical. Manufacturers oftentimes reveal only one measure (diagonal) of the FoV, whereas others withhold this information completely. The FoV may also be artificially increased by reducing the stereo overlap, i.e., the area of the screen in which the user can perceive depth (Hugues, 2019). It was thus decided that the field of view of devices could not be quantitatively, fairly compared and this variable was omitted from quantitative synthesis.

Screen Refresh Rate. The screen refresh rate refers to the rate at which the screens update the images displayed on the screen, based on input generated by the computer (Kourtesis et al., 2019). A low screen refresh rate would be perceived as a lack of fluency in images, or a lag between the user's actions and visual input. The screen refresh rate is either reported in cycles per second (Hz) or frames per second (FPS). As this information was frequently missing, particularly in older studies, the screen refresh rate was not used to compare VR interventions.

Stereoscopy/Three-Dimensional Graphics. Stereoscopy is achieved by presenting separate images to each eye with slight differences in perspective that reflects the interpupillary distance. It provides an illusion of depth in the virtual environment and may increase immersion (Yang et al., 2012).

Head Tracking. Some VR systems track user movements and use this information to adjust images (and sometimes sound) accordingly. All parts of the body can be tracked, but

tracking of head movements is the most common. According to Slater (2009), tracking strengthens the illusion of being present in the virtual environment as the participant can perceive through natural sensorimotor contingencies (O'Regan & Noë, 2001). For example, a participant may tilt his or her head to inspect a virtual object from several angles, which is not possible on conventional screens.

Visual Occlusion. This variable refers to whether the VR system fully covered the participant's physical surroundings. HMDs that are not fully occlusive may have a gap between the device and the participant's face that lets light through and allows the participant to see parts of the procedural setting. Minimizing input from the physical reality may strengthen the illusion of being present in the virtual environment (Slater & Wilbur, 1997).

Non-Visual Sensory Stimulation. This variable described whether the VR intervention involved any non-visual, sensory stimulation. This would typically be in the form of auditory stimuli (e.g., music or sound effects from games), but also tactile stimuli (e.g., force feedback or vibration from controllers). Researchers may choose not to include audio to avoid disruption in communication between patients and personnel delivering the medical procedures. However, it is commonly assumed that multisensory stimuli provide greater immersion and sense of presence (Cummings & Bailenson, 2016).

Interactivity

This variable was used to declare whether the VR system offered any user-system interaction beyond control of the field of view (i.e., tracking of head movements).

Interactivity may for example include navigation in the virtual environment or manipulation of virtual objects.

Risk of Bias Assessment in Individual Studies

Assessment of study risk of bias was conducted in accordance with the Cochrane Handbook of Systematic Reviews (Higgins et al., 2020). The effect of interest was the Intention-To-Treat (ITT) effect, i.e., the effect of allocation to intervention. Risk of bias was assessed at outcome level independently by the student. The ROB 2.0 (Sterne et al., 2019) and ROBINS-I (Sterne et al., 2016) tools were used for RCTs and non-randomized studies, respectively. The RCT characteristics assessed were (a) bias arising from the randomization process, (b) bias due to deviations from intended interventions, (c) bias due to missing outcome data, (d) bias in measurement of the outcome, and (e) bias in selection of the reported result. Additional considerations for cross-over trials were applied (Sterne et al., 2019). However, they were evaluated with the parallel design tool if only data from the first study period was analyzed. Non-randomized studies were evaluated in terms of the following domains: (a) confounding, (b) selection bias, (c) bias in classification of interventions, (d) bias due to deviations from intended interventions, (e) bias due to missing data, (f) bias in measurement of outcomes, and (g) bias in selection of the reported result. The risk of bias judgements for each domain are illustrated in separate figures for randomized and non-randomized studies. Additional bar plots illustrate the overall judgement for each domain across studies, with each study's contribution weighted by their standard error. The figures were constructed using the robvis web application (McGuinness & Higgins, 2020). A separate, additional analysis excluding studies deemed to have a high risk of bias in more than two domains was conducted.

Summary Measures

The differences in mean pain and anxiety scores for the VR and control groups were calculated as Hedges' g (Hedges, 1981). While similar to d , the Hedges' g includes a

correction term that yields a less biased estimate, particularly when sample sizes are small (Borenstein et al., 2009). If a study had multiple VR or non-VR arms, their summary statistics were combined. Means and standard deviations were ideally extracted directly from articles or obtained from study authors. If necessary, they were estimated. Sample means were estimated from the median by the method of Shi et al. (2020). Estimation of variance based on the median, interquartile range and sample sizes were based on the method of Wan et al. (2014). For studies that also reported the minimum and maximum values, the formula proposed by Luo et al. (2018) was used for additional precision. These estimations were performed using an online calculator by Shi et al. (2020). The Campbell Collaboration effect size calculator (Wilson, n.d.) was further used to estimate effect sizes from *t*-statistics.

Cross-over trials were only included for quantitative synthesis if data from the first study period only was available, or if effect sizes could be calculated from paired analyses that account for the correlation between each participant's responses (Elbourne et al., 2002).

Several studies reported multiple measures of pain and anxiety. As specified in the review protocol, only one measure for each outcome was used for quantitative synthesis. The selection was based on the following pre-specified criteria: (a) Self-reported measures were preferred over observational measures; (b) measures of sensory pain were preferred over measures of the cognitive or affective aspects of pain. If two or more measures fit the abovementioned criteria, the most frequently used measure was selected.

Syntheses of Results

The methodology was guided by Borenstein et al. (2009) and the Cochrane handbook of systematic reviews (Higgins et al., 2020). All statistical analyses were conducted using Stata 16 (StataCorp, 2019). Standardized mean differences in pain and anxiety were combined using a random-effects model. The random-effects model assumes that the study effect sizes

are drawn from different populations of study effect sizes, i.e., that observed variance consists of both sampling error and differences in true effect sizes (Borenstein et al., 2009). This model was selected as the studies were expected to be diverse in terms of study designs, participant characteristics, medical procedures, and VR characteristics, to name a few. The restricted maximum likelihood estimator of between-studies variance (τ^2) was selected based on recommendations by Veroniki et al. (2016). The results of the two meta-analyses are presented in separate forest plots, with primary studies stratified by the type of medical procedure. The standardized mean effect will also be expressed as absolute mean differences on the Wong-Baker Faces scale and the Child Fear Scale. These scales were selected as they were the most frequently used one-item scales among the outcomes included in the meta-analysis. The absolute mean difference will be calculated by multiplying the standardized mean difference with the combined standard deviations from every study in which these measures were used in the meta-analysis (Schünemann et al., 2020).

Heterogeneity among all included studies was assessed by consulting the Cochran's Q test. A significant result indicates that the observed variation in effect sizes reflects true heterogeneity (Borenstein et al., 2009). The I^2 statistic was then used to quantify the magnitude of heterogeneity. It describes the percentage of total variation that is due to heterogeneity (Higgins et al., 2003), with higher values indicating greater heterogeneity.

Risk of Bias Across Studies

Publication bias compromises the validity of the results of meta-analyses and systematic reviews. The term is typically used to refer to the greater likelihood of studies with statistically significant results being published. This tendency leads to an over-estimation of the summary effect sizes. However, the availability of all relevant studies may also be compromised for other reasons, such as language and the cost of accessing articles (Rothstein

et al., 2006). Regardless of the reason for publication bias, studies retrieved from literature searches may not be representative of all studies conducted on a certain topic (Rothstein et al., 2005).

Publication bias was assessed visually with a funnel plot in which study effect sizes (horizontal axis) were plotted against their inverse standard error (vertical axis). Areas representing three intervals of p -values (contours) were added to facilitate interpretation (Peters et al., 2008). As the standard error is directly related to the number of participants, plot asymmetry may be indicative of small-study effects (Sterne et al., 2006). Funnel plot asymmetry was also assessed formally with Egger's tests, which involves regression analyses of the relationship between effect sizes and their standard error (Sterne & Egger, 2005). If the regression intercept differs from zero, publication bias is likely to be present.

The trim-and-fill algorithm was used to estimate an effect size adjusted for publication bias. This procedure is conducted in two steps. During the first step, studies that cause funnel plot asymmetry are removed from the mean effect size estimate until symmetry is achieved (iteration step) (Borenstein et al., 2009, p. 286). An adjusted mean effect size is then estimated. The removed studies are finally re-applied, along with the studies that are assumed to be missing from either side of the funnel plot (pooling step). This final step estimates the variance of the new mean effect size. The trim-and-fill method is widely used, but its performance may vary depending on the presence of substantial heterogeneity or outlying studies, as well as which combination of models, methods, and estimators that is used. Researchers are thus encouraged to use various versions of the trim-and-fill method (Shi & Lin, 2019). In the present study, all possible combinations of the fixed- and random-effects (restricted maximum likelihood method) models with the linear (L_0), run (R_0) and quadratic (Q_0) estimators will be used.

Additional Analyses

Subgroup and Meta-regression Analyses

Moderator analyses were conducted to explore potential sources of heterogeneity in effect sizes. The differences between subsets of the studies were initially explored with subgroup analyses. Categorical and continuous variables were then used as predictors in a random-effects meta-regression analysis. It is generally recommended that there are approximately ten studies per predictor (Borenstein et al., 2009). As the present study was focused on the differences between VR interventions, these variables were prioritized in the meta-regression analysis rather than the kind of medical procedure.

As previously discussed, the screen refresh rate, resolution and field of view were omitted from quantitative analysis due to insufficient information. After coding the remaining immersion variables, it was discovered that only one study included any non-visual stimuli. This variable was thus also omitted from the composite immersion variable. As information regarding the four remaining immersion variables was lacking for several studies, it was decided to code VR interventions as either highly immersive (included auditory stimuli, head tracking, stereoscopy/three-dimensional images, and full visual occlusion) or less immersive/insufficient information. The VR interventions were also coded as either interactive or passive (i.e., no interactivity beyond head tracking). Medical procedures were categorized as either 'dental', 'needle-related procedures', 'pre-operative', or 'wound care'. The mean study-level age was included as a continuous variable. All potential moderators were pre-specified in the review protocol.

Sensitivity Analyses

Sensitivity analyses were conducted to ensure that the summary effect estimates were robust to the removal of the following studies: (a) under-powered studies, (b) non-randomized

studies, and (c) studies deemed to have a high risk of bias in two or more domains. Assuming a one-tailed alpha of .05 and an 80 % power to detect an effect size of 0.50, studies were considered under-powered if they had less than 50 participants in each group (Cohen, 1998).

Results

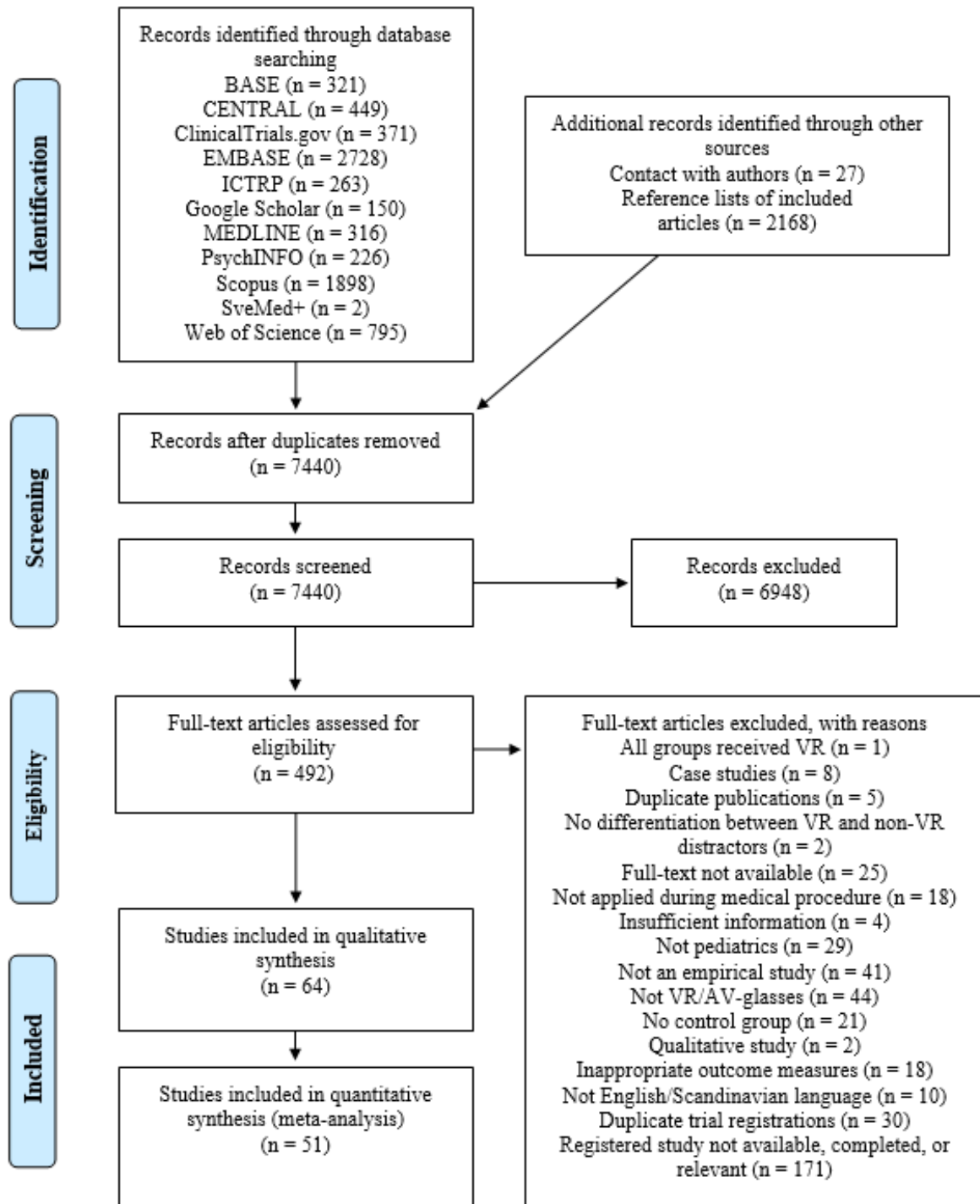
65 primary studies derived from 64 articles published between 2004-2020 were included in qualitative synthesis. 13 studies were not included in the meta-analyses due to missing numerical results (Gershon et al., 2004; Khan et al., 2019), only change from baseline scores being reported (Kipping et al., 2012), or insufficient data to include cross-over trials (Attar & Baghdadi, 2015; Atzori, Grotto, et al., 2018; Atzori, Hoffman, et al., 2018; Chan et al., 2007; Das et al., 2005; El-Sharkawi et al., 2012; Garrocho-Rangel et al., 2018; Hoffman et al., 2019; Koticha et al., 2019; Sullivan et al., 2000). Two data sets were obtained from contact with authors to calculate the effect size for the first study period only (Schmitt et al., 2011) and summary statistics (Jeffs et al., 2014). Two unpublished studies were acquired by contacting authors identified in the trial registries (Gerceker et al., in press; Osmanlliu et al., in press). Another two published manuscripts were received from contacted authors after the final database search was conducted (Buldur & Candan, 2020; Litwin et al., 2020). The process of study selection is illustrated in Figure 1. Study and VR characteristics are presented in separate tables (Appendix A and B, respectively), while the data used in the meta-analysis is presented in Appendix C. Stata output is listed in Appendix F. A narrative synthesis of study and VR characteristics is presented in the following paragraphs.

Study Characteristics

Most of the studies ($k = 61$) were RCTs, of which 43 employed a parallel-groups design and 18 studies employed a cross-over design. Four non-randomized studies were included.

Figure 1

PRISMA flowchart showing the study selection process



Participant Characteristics

The total number of participants was 4654, with sample sizes ranging from 5 to 220, and averaging at 72 participants. Included participants were between 6 months-21 years of age, and the mean study-level age was 9.23 years.

Measures

Self-reported measures of pain were available in all but two studies (Khadra et al., 2020; Wolitzky et al., 2005), whereas observational measures had to be used for 11 of the anxiety studies. The Wong-Baker Faces Scale (Wong-Baker FACES Foundation, 2018) and the (revised) Faces Pain Scale (Hicks et al., 2001) were the most widely used pain measures, followed by visual analogue scales ([VAS], Bailey et al., 2012). VAS scales were also frequently used to measure anxiety. The most used observational measure of anxiety was the modified Yale Preoperative Anxiety Scale (Kain et al., 1997).

Settings and Medical procedures

Studies were mostly conducted in pediatric hospitals or dental clinics. Most of the procedures were classified as needle-related procedures ($k = 25$), followed by dental ($k = 24$), pre-operative ($k = 8$), and wound care ($k = 8$).

Intervention Characteristics

Most of the distraction studies ($k = 61$) used VR as a distraction during the medical procedures. Only Al-Nerabieah et al. (2020) used VR as a distraction *before* the procedure (i.e., in the waiting room before dental procedures). In one cross-over trial, the effect of receiving VR distraction during the first treatment on pre-operative anxiety before the second treatment could be extracted (Fakhrudin et al., 2015).

Four studies (Eijlers, Dierckx, et al., 2019; Ryu et al., 2017, 2018, 2019) were categorized as preparation studies. These VR interventions involved virtual tours of the pre-

operative settings, in which children were exposed to the procedural environment and medical personnel, as well as information about the procedures. Ryu and colleagues incorporated popular cartoon figures that explained and modelled the procedures. Participants in Eijlers, Dierckx, et al. (2019) and Ryu et al. (2018) were also able to interact with virtual medical devices and receive further information about them.

VR Characteristics

Head-mounted devices (HMDs) were used in all but three studies ($k = 62$). In Khadra (2020), patients were placed in front of a wide, curved screen that images were displayed on with a projector. This study was included as the screen covered the majority of the patient's field of view and resembled a surrounding, dome-based VR system. Jeffs (2014) and Hoffman (2019) used HMDs that were mounted on either a tripod or a robotic arm to facilitate participation by patients with burn injuries in the head and neck region, or to facilitate use during hydrotherapy. In 28 studies, so-called smartphone-based systems were used in which a smartphone or other device is inserted into the HMD to serve as the screen and tracking device (Fuchs, 2019). The most common combination was the Samsung Gear headset coupled with various Samsung smartphones.

As previously mentioned, information regarding at least some technical specifications were lacking for many studies, particularly in older studies and in studies that used less advanced VR systems. However, it was clear that the quality of the VR equipment varied considerably between studies. 37 of the VR systems offered stereoscopy/three-dimensional graphics, whereas seven did not. Unfortunately, this information was not available for 21 studies. Nearly half of the VR interventions ($k = 32$) involved head tracking, 17 VR interventions did not, and information was lacking for the remaining 16 studies. Most of the VR devices fully covered the patient's field of view ($k = 41$), whereas 13 did not. For 11 of

the studies, this information was not available. Nearly all of the VR interventions involved auditive stimuli ($k = 60$), and one study also included tactile feedback in the form of tactile feedback from controllers (Gold et al., 2006). Two studies did not include any audio (Aydin & Ozyazicioglu, 2019; Dumoulin et al., 2015), whereas this information could not be confirmed for three studies (Attar & Baghdadi, 2015; Das et al., 2005; Isong et al., 2014).

27 VR systems were classified as interactive, meaning that the system afforded interactivity beyond head tracking. Four studies (Chaudhary et al., 2020; Gerceker et al., 2020, in press; Piskorz et al., 2020), included both interactive and non-interactive subgroups. The interactive group of VR interventions was diverse; while some merely involved visual effects as the patient focused his or her gaze on a virtual object (e.g., Aydin & Ozyazicioglu, 2019), others involved more interactivity with virtual objects (e.g., Eijlers, Dierckx, et al., 2019) or more demanding tasks and games (e.g., Piskorz & Czub, 2018).

In most of the studies, patients viewed videos ($k = 29$), followed by games ($k = 19$), and finally simulations ($k = 11$). Two studies included both video and game conditions (Chaudhary et al., 2020; Piskorz et al., 2020). Information regarding the VR software was not available for Attar and Baghdadi (2015).

Comparison Groups

Comparison groups were diverse and not always clearly described. They included a range of non-VR distractions (e.g., other electronic devices or conversation) or procedural preparation (e.g., informative videos or verbal briefings), behavior management techniques (e.g., positive reinforcements, tell-show-do technique), or standard of care procedures (SOC). The SOC conditions were also diverse, with some involving no intervention at all and others a combination of several interventions. Three dental studies used sunglasses or protective

eyeglasses, either as part of standard care (Hoge et al., 2012), as a behavior management technique (Bagattoni et al., 2018) or as a form of placebo (Buldur & Candan, 2020).

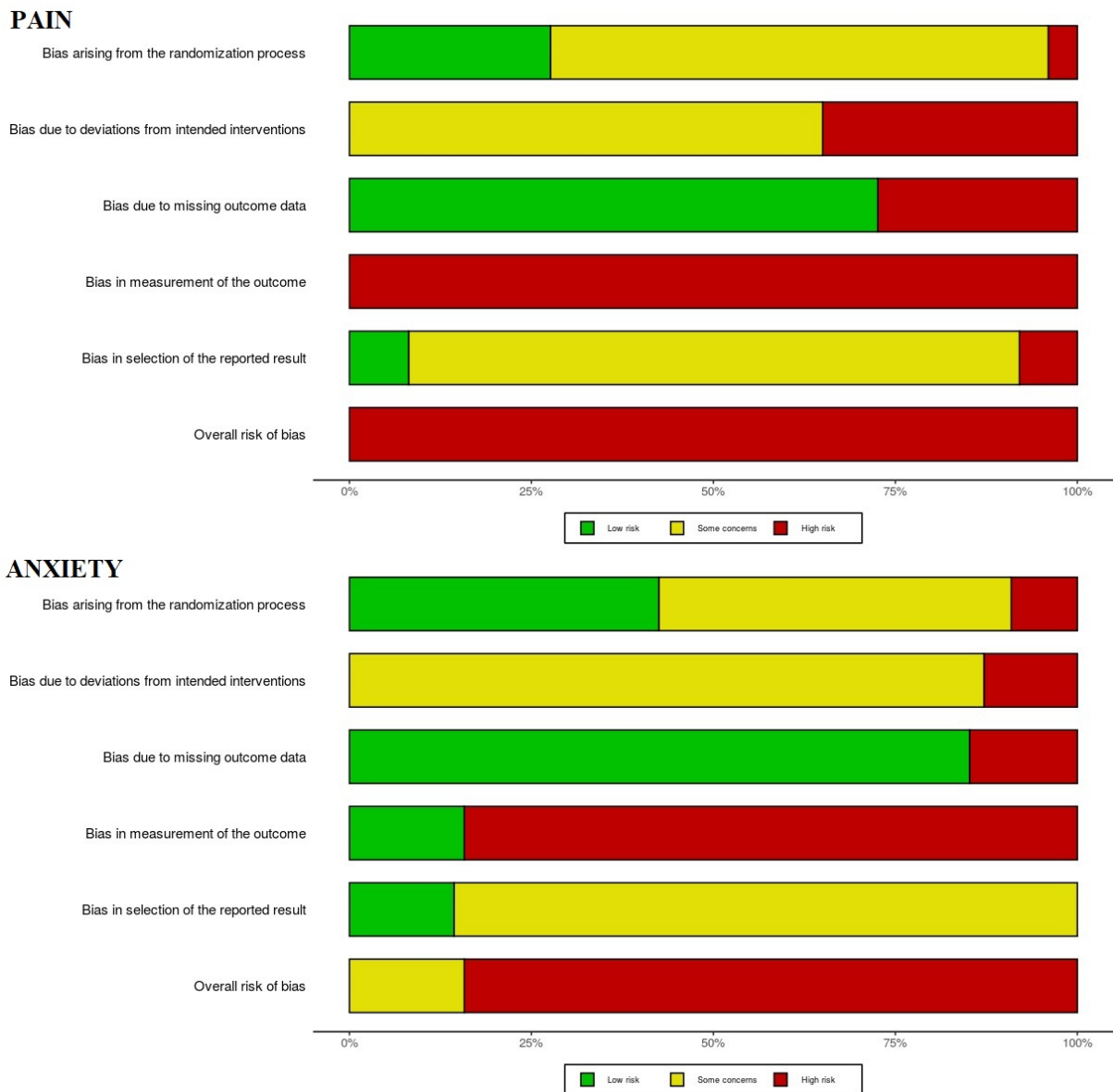
Risk of Bias Within Studies

Risk of bias was assessed per outcome for all included studies. The risk of bias judgements of each domain combined are illustrated in Figure 2 (randomized studies) and Figure 3 (non-randomized studies). Contributions from each study towards the combined risk of bias judgements are weighted by standard error of their effect sizes. Separate figures for pain and anxiety outcomes were not constructed for the non-randomized trials as these were identical. Risk of bias judgements for each study per domain are available in Appendix D (pain) and E (anxiety).

None of the included studies received an overall low risk of bias judgement, and the vast majority were deemed to have an overall high risk of bias. This was partially because it is not possible to blind patients, parents and personnel delivering the VR interventions. Reports of pain and anxiety are highly subjective and may be influenced by beliefs regarding the efficacy of distraction methods. As self-reported measures were prioritized, most of the studies thus received a high risk of bias judgement in domain 4 (bias in measurement of the outcome). Blinding of outcome assessors and personnel conducting the medical procedures was only feasible in studies that applied VR before the medical procedure and only reported observational measures of either pain or anxiety (Al-Nerabieah et al., 2020; Eijlers, Dierckx, et al., 2019; Ryu et al., 2017, 2018, 2019). The lack of blinding may also have affected the behavior of patients, parents, carers, and others. Most studies therefore received at least an intermediate risk of bias judgement in domain 3 (bias due to deviations from the intended interventions), and high if data was not analyzed in accordance with intention-to-treat principles.

Figure 2

Combined risk of bias judgements of pain and anxiety outcomes reported in randomized trials



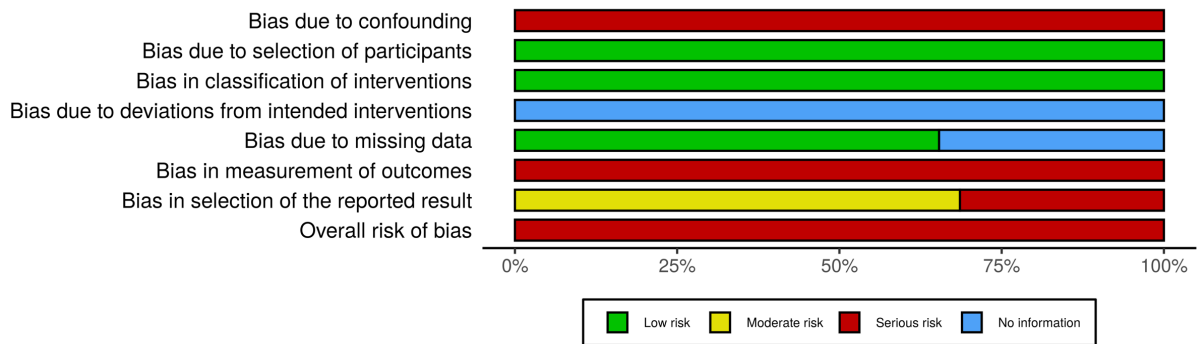
In addition to issues related to blinding, prospective trial registrations and/or pre-specified data analysis plans were identified for only a few studies. Many studies were thus

deemed to have at least an intermediate risk of bias due to selective reporting. Potential issues related to the randomization process were also observed in roughly half of the included studies. Frequently, the methods of randomization and concealment of allocation sequence were frequently not described in sufficient detail or at all. Some studies also performed block-randomizations with small, evenly sized blocks or used other methods that might enable prediction of the forthcoming allocation for at least some participants.

All the non-randomized trials (del Castillo et al., 2019; Piskorz & Czub, 2018; Piskorz et al., 2020; Sullivan et al., 2000) were deemed to have a serious risk of bias. Some of the issues observed in randomized trials were also seen in non-randomized trials, such as lacking pre-specified analysis intentions. Perhaps more importantly, the studies were considered to have a serious risk of bias due to confounding. For example, in Sullivan et al. (2000), children that were too anxious to receive VR on the first study day received VR on the second study day instead. In the remaining three studies, allocation was determined by either the timing of admission to the hospital in children that were regularly hospitalized for chronic disease (Piskorz & Czub, 2018; Piskorz et al., 2020), or whether the medical procedure was performed during the day or evening/night shifts (del Castillo et al., 2019). Although it is difficult to ascertain exactly how the timing of hospitalization or the medical procedure may have influenced study results, participants in the VR and non-VR groups may differ systematically in clinically relevant ways.

Figure 3

Combined risk of bias judgements of pain and anxiety outcomes reported in non-randomized trials



Results of Individual Studies and Syntheses of Results

Numerical results of each study and results of the meta-analyses are illustrated in forest plots for pain (Figure 4) and anxiety (Figure 5). Positive values (towards the right) indicate that results are in favor of VR. Qualitative results are presented in the study characteristics and results table (Appendix A). The results from studies that were not included in the meta-analyses were mixed; six studies reported results in favor of VR, two reported no difference between the groups, two studies did not find any difference in child and parent reported outcomes, and one study found that pain levels were higher in the VR group. The two anxiety studies both reported no difference between the VR and comparison groups.

Pain

42 studies reporting pain outcomes were synthesized. The overall mean effect (Hedges' g) for pain was estimated to 0.79 (95 % CI [0.48, 1.10], $z = 5.01$, $p < .001$). This effect size may be considered large, compared to effect sizes that have previously been obtained for educational or counselling interventions for medical patients (Lipsey & Wilson, 1993). Expressed in units of the 6-point Wong-Baker Faces scale, this would correspond to a

mean difference of 1.94 points. As will be discussed in sub-section ‘Risk of Bias Across Studies’, the true effect is likely considerably lower than the estimate that was obtained here.

The Q -statistic indicated statistically significant heterogeneity in effect sizes ($Q(41) = 441.46, p < .001$). A large proportion of the observed variation ($I^2 = 94.61\%$) was found to reflect differences in true effect sizes. Six studies reported results in favour of the control/non-VR group (Bagattoni et al., 2018; Eijlers, Dierckx, et al., 2019; Hoge et al., 2012; Jeffs et al., 2014; Mitrakul et al., 2015; Walther-Larsen et al., 2019). Potential sources of heterogeneity are assessed in the ‘Additional Analyses’ section.

Anxiety

35 studies reporting anxiety outcomes were synthesized. The mean effect size (Hedges’ g) for anxiety was estimated to 0.90 (95 % CI [0.55, 1.26], $z = 4.98, p < .001$), which too may be considered a large effect size compared to the effect sizes compiled in Lipsey & Wilson (1993). On the five-point Child Fear Scale (CFS), this would amount to a mean difference of 1.22 points. However, the true effect is likely to be smaller than this estimate (see ‘Risk of Bias Across Studies’). As for pain, the Q -statistic indicated statistically significant heterogeneity in effect sizes ($Q(34) = 437.69, p < .001$), with a large proportion ($I^2 = 95.43\%$) of variation attributable to differences in true effect sizes. Four studies reported results in favor of the control/non-VR treatment (Eijlers, Dierckx, et al., 2019; Litwin et al., 2020; Ryu et al., 2019; Shah & Bhatia, 2018). Potential sources of heterogeneity are further explored in the ‘Additional Analyses’ sub-section.

Figure 4

Forest plot with individual and combined results for pain

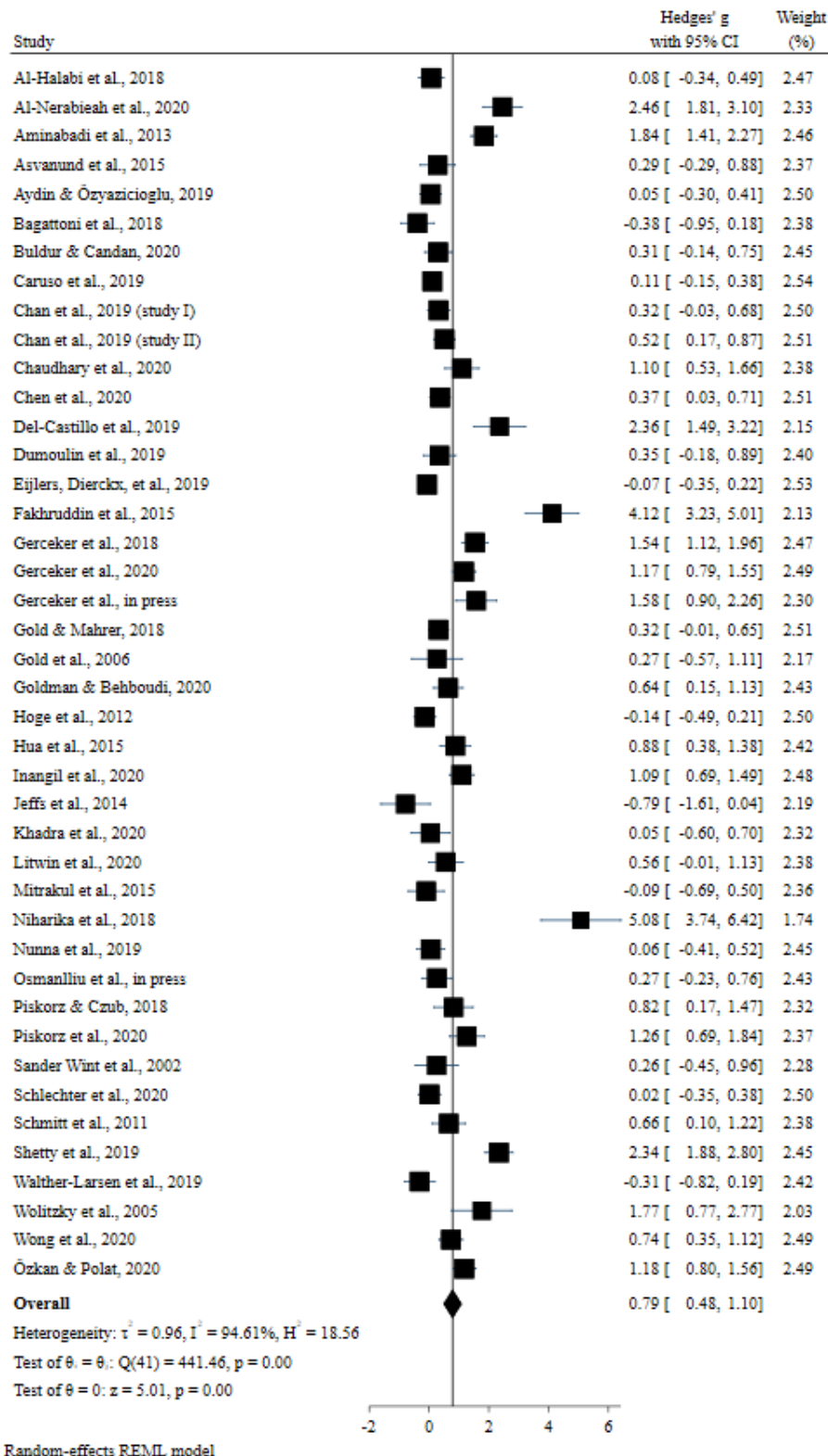
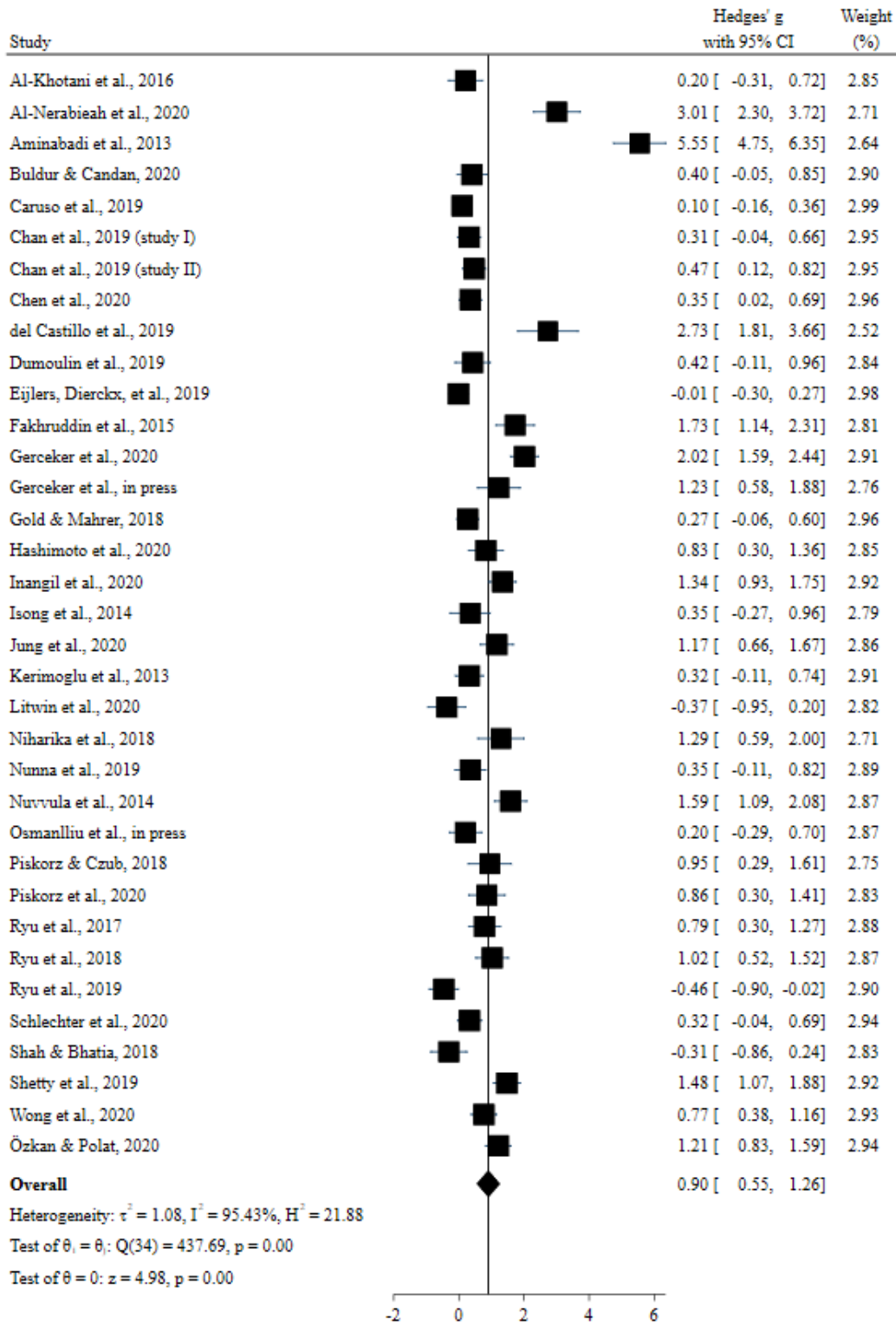


Figure 5

Forest plot with individual and combined results for anxiety



Risk of Bias Across Studies

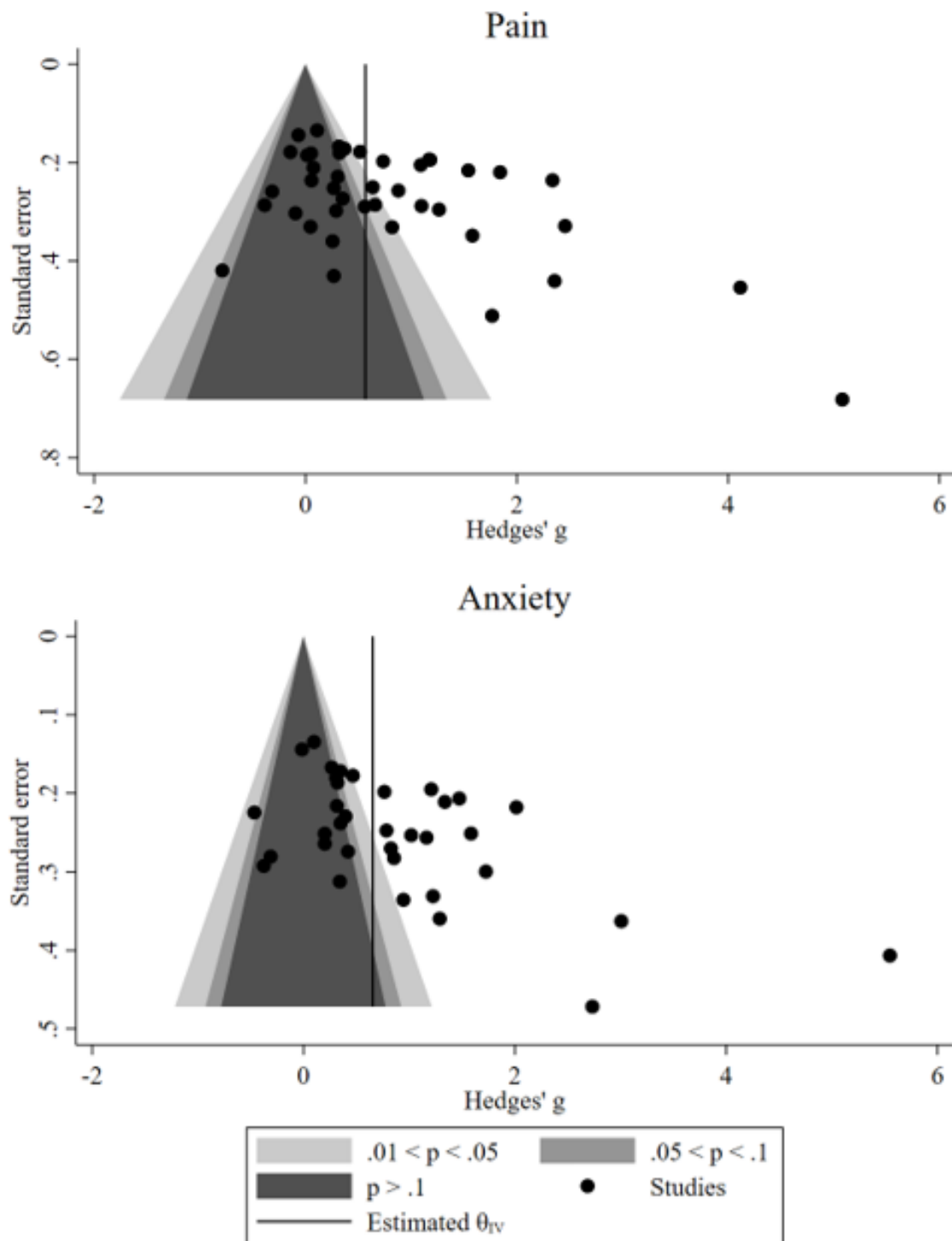
Publication bias was assessed by visually inspecting contour-enhanced funnel plots, and by performing Eggers' tests and the trim-and-fill-method. There is a clear lack of smaller studies reporting statistically non-significant results (i.e., towards the lower left part of the plot) among both the pain and anxiety studies (Figure 6). The plot asymmetries are further confirmed by a significant Eggers' regression tests ($p < .001$).

The trim-and-fill procedure was conducted with various settings as previously described. For the pain studies, four and six studies were imputed, with adjusted mean effect sizes ranging from 0.41 (95 % CI [0.34, 0.48]) (fixed-fixed with the R_0 estimator) to 0.55 (95 % [0.16, 0.94]) (random-random with the R_0 estimator). Based on these adjusted estimates, the true mean difference would be closer to 0.83 - 1.35 points on the Wong-Baker Faces scale. The results thus indicate that the true mean effect lies substantially below the observed mean effect.

For anxiety, 12 and 15 studies were imputed with the R_0 estimator, which yielded adjusted estimates of 0.48 (95 % CI [0.40, 0.55]) (fixed-fixed; fixed-random) and 0.58 (95 % CI [0.50, 0.65]) (random-random; random-fixed). The other estimators did not suggest any missing studies. This suggests that the true mean difference is closer to 0.65 – 0.79 points on the Child Fear Scale. These estimates are thus considerable moderations of the original effect size.

Figure 8

Contour-enhanced funnel plots for studies reporting pain and anxiety outcomes



Additional Analyses

Subgroup and Meta-Regression Analysis of the Effects of VR on Pain

The subgroup analyses (Table 2) revealed statistically significant differences in mean effects across the groups of medical procedures, most notably between the dental subgroup (Hedges' $g = 1.26$, 95 % CI [0.35, 2.17]) and the pre-operative (Hedges' $g = -0.13$, 95 % CI [-0.37, 0.12]) subgroups. In the pre-operative and wound care subgroups, the confidence intervals included zero, indicating the possibility of no or minimal differences between the VR and non-VR conditions. These subgroups were also quite small.

Table 2

Subgroup analyses of the effects of VR on pain

Subgroup	k	Hedges' g [95 % CI]	$Q(df)$	p
Baseline	42	0.79 [0.55, 1.26]		<.001
Medical procedure			27.83(3)	<.001
Dental	13	1.26 [0.35, 2.17]		
Needle-related	23	0.72 [0.49, 0.95]		
Pre-operative	2	-0.13 [-0.37, 0.12]		
Wound care	4	0.25 [-.45, 0.95]		
High immersion			0.18(1)	.67
Yes	19	0.72 [0.45, 1.00]		
No	23	0.86 [0.30, 1.41]		
Interactivity			7.69(1)	.006
Yes	18	0.28 [0.10, 0.45]		
No	20	1.16 [0.56, 1.75]		

Note. k = number of studies; CI = 95 % confidence interval. Q = test of homogeneity of effect sizes.

The mean effects were similar between the immersion subgroups. However, studies using less interactive VR systems reported significantly lower pain levels (Hedges' $g = 1.16$, 95 % CI [0.56, 1.75]) than those using interactive VR systems (Hedges' $g = 0.28$, 95 % CI [0.10, 0.45]). Four studies were not included in the subgroup analysis of interactivity, as they contained both interactive and non-interactive VR interventions (Chaudhary et al., 2020; Gerceker et al., 2020, in press; Piskorz et al., 2020).

Participants' age and the level of immersion and interactivity were applied as predictors in a meta-regression analysis (Table 3). Again, the four studies with both interactive and non-interactive interventions were not included. After controlling for the level of immersion and mean age of participants, the difference between interactive and non-interactive VR did not reach statistical significance. No relationship was found between the participants' age or level of immersion and mean pain scores.

Table 3

Results of meta-regression analysis on the effects of VR on pain

	Coefficient [95 % CI]	z	p
Intercept	1.77 [0.58, 2.97]	2.92	.004
Age	−.09 [−0.22, 0.04]	−1.32	.19
High immersion	0.17 [−0.51, 0.84]	0.48	.63
Interactivity	−0.65 [−1.35, .04]	−1.84	.07

Note. Residual heterogeneity: $I^2 = 94.35$ %, $R^2 = 12.54$ %.

Subgroup and Meta-regression Analysis of the Effects of VR on Anxiety

Subgroup analyses of studies reporting anxiety outcomes (Table 4) indicate similar patterns as those observed for pain outcomes, with the largest effect sizes reported in the

dental subgroup (Hedges' $g = 1.41$, 95 % CI [0.44, 2.37]). However, the difference between the groups of medical procedures was not statistically significant. The difference between the interactivity subgroups was statistically significant, with lower pain scores reported in the non-interactive group (Hedges' $g = 1.15$, 95 % CI [0.57, 1.73] than the interactive group (Hedges' $g = 0.38$, 95 % CI [0.15, 0.61]. Three studies were not included in the interactivity analysis as they included both interactive and non-interactive interventions (Gerceker et al., 2020, in press; Piskorz et al., 2020). The mean pain scores were markedly lower in the non-immersive subgroup, but this difference was not statistically significant.

Table 4

Subgroup analyses of the effects of VR on anxiety

Subgroup	k	Hedges' g [95 % CI]	$Q(df)$	p
Baseline	35	0.90 [0.55, 1.26]		<.001
Medical procedure			2.87(2)	.24
Dental	11	1.41 [0.44, 2.37]		
Needle-related	17	0.74 [0.41, 1.07]		
Pre-operative	7	0.50 [0.06, 0.95]		
High immersion			1.27(1)	.26
Yes	20	0.71 [0.37, 1.06]		
No	15	1.16 [0.46, 1.87]		
Interactivity			5.95(1)	.02
Yes	13	0.38 [0.15, 0.61]		
No	20	1.15 [0.57, 1.73]		

k = number of studies; CI = 95 % confidence interval. Q = test of homogeneity of effect sizes.

A meta-regression analysis with participants' age, the level of immersion and interactivity as predictors revealed no statistically significant relationships with anxiety scores (Table 5).

Table 5

Results of meta-regression analysis on the effects of VR on anxiety

	Coefficient [95 % CI]	<i>z</i>	<i>p</i>
Intercept	2.15 [0.72, 3.59]	2.95	.003
Age	-0.11 [-0.27, 0.06]	-1.23	.22
High immersion	-0.30 [-1.19, 0.58]	-0.67	.50
Interactivity	-0.36 [-1.36, 0.64]	-0.70	.48

Note. Residual heterogeneity: $I^2 = 94.96\%$, $R^2 = 10.75\%$.

Sensitivity Analyses

Sensitivity analyses were conducted to evaluate the robustness in results when removing studies that were not adequately powered (<100 participants), non-randomized studies, and studies with two or more individual domains considered at a high risk of bias. As previously discussed, most studies received an overall high risk of bias judgement due to the prioritization of self-reported measures. Rather than excluding studies based on their overall risk of bias, the sensitivity analysis involved removing studies that received a high risk of bias judgement in more than one domain.

Pain. The effect size was reduced after removing 25 studies with inadequate power (Hedges' $g = 0.67$, 95 % [0.32, 1.01], $z = 3.79$, $p < .001$). Heterogeneity remained high ($I^2 = 93.60\%$). The results were robust to the removal of the three non-randomized studies (Hedges' $g = 0.74$, 95 % [0.42, 1.06], $z = 4.50$, $p < .001$, $I^2 = 94.90\%$) and the 20 studies that

were deemed to have a high risk of bias in two or more individual domains (Hedges' $g = 0.73$, 95 % CI [0.32, 1.14], $z = 3.46$, $p < .001$, $I^2 = 94.37\%$).

Anxiety. After removing the 22 inadequately powered studies, the effect size increased (Hedges' $g = 1.07$, 95 % CI [0.30, 1.83], $z = 2.72$, $p < .001$). Heterogeneity remained high ($I^2 = 98.29\%$). Removing the three non-randomized studies led to a slight reduction in effect size, while heterogeneity was constant (Hedges' $g = 0.85$, 95 % [0.48, 1.23], $z = 4.48$, $p < .001$, $I^2 = 95.72\%$). The mean effect size was slightly elevated when studies with more than one domain at high risk of bias were removed (Hedges' $g = 0.95$, 95 % CI [0.46, 1.45], $z = 3.78$, $p < .001$), while heterogeneity remained constant ($I^2 = 96.54\%$).

Discussion

Summary of Evidence

The aim of this systematic review was to evaluate the evidence regarding the effectiveness of VR on procedural pain and anxiety in children. An overview of the characteristics of VR interventions was provided, as well as the settings and ways in which they were used. Meta-analyses of pain and anxiety outcomes were performed, and the kind of medical procedure, mean patient age, interactivity, and immersion were explored as potential moderators. The strength of evidence was assessed through risk of bias judgements, tests for publication bias, and sensitivity analyses.

Although information about the VR interventions was often lacking, it was clear that they were diverse in terms of technical specifications, level of interactivity, and the media that was displayed. While most VR headsets were fully occlusive and offered auditive stimulation, stereoscopic graphics and head tracking were only used in nearly half of the studies. The screen resolution and field of view also varied greatly. Information regarding the screen refresh rate was often unavailable. Nearly half of the studies used non-interactive simulations

or movies, whereas the interactive group consisted of both minimally interactive simulations (e.g., Aydin & Ozyazicioglu, 2019) and more cognitively taxing games (e.g., Piskorz & Czub, 2018).

Overall, the evidence was deemed at a high risk of bias using the ROB 2.0 tool. This is not surprising, as blinding patients to their allocation to experimental groups was not possible, and self-reported measures were preferred for inclusion in the meta-analysis. The fact that most studies received a high risk of bias judgement does not in itself suggest low methodological quality of studies. However, most studies were deemed to have at least an intermediate risk of bias in several other domains. This raises serious concerns on the validity of study results and their syntheses. For example, studies conducted with lower methodological quality may overestimate treatment effects (Hempel et al., 2011; Moher et al., 1998).

Other reasons to suspect spuriously large treatment effects are the indications of publication bias. Several studies reporting non-significant results are likely lacking from the literature, and there is reason to believe that the true effects are considerably smaller than those observed in the retrieved studies. In conclusion, the meta-analytical findings should thus be interpreted with great caution, and attention should be directed towards the more modest range of estimates suggested by the trim-and-fill method.

Effects of VR on Pain and Anxiety

High levels of heterogeneity were observed in both the pain and anxiety studies, but most studies reported results in favor of VR. Large effects were found for both pain (1.94 points on the Wong-Baker Faces Scale [W-BFS]) and anxiety (1.22 points on the Child Fear Scale [CFS]). Based on estimates adjusted for publication bias, there is however strong reason

to believe that the true effects of VR on pain and anxiety are considerably lower (0.83 - 1.35 points on the W-BFS; 0.65 – 0.79 points on the CFS).

Moderator Analyses

Studies in which VR was used during dental or needle-related procedures reported larger effects on average. The pain and anxiety scores were also lower in the non-interactive VR subgroup. There was a high degree of overlap between these three groups; all the 24 non-interactive VR interventions were used during dental or needle-related procedures among the pain studies, and 20 out of the 24 among the anxiety studies. It is therefore difficult to establish whether it is the medical procedure or the level of interactivity (or neither) that best explains the differences that were observed.

No statistically significant differences in VR effectiveness were found between systems that were highly immersive (i.e., had head tracking, full visual occlusion of the patient's physical surroundings, stereoscopy, and auditive stimuli) and those that lacked at least some of these features (or in which immersion variables could not be confirmed). It should not be concluded based on these results that there is no effect of immersion on VR effectiveness. The immersion variable used in the analysis was based on only four of the many features known to influence presence. They were selected as information regarding other VR features was lacking for several studies. To maintain an acceptable predictor-study ratio, they were used to create a dichotomous variable that only described whether a VR system possessed *all* the four features. Consequently, any potential differences between VR systems with none, some, and all the features were ignored. A more sophisticated approach would involve an assessment of the relative influence of several individual immersion variables. The results of the present analysis should thus only be interpreted as an observed mean difference between studies that had four arbitrary VR features, and that were also

heterogenous in many aspects, such as patient characteristics and medical procedures. The same considerations apply to the statistically significant difference that was observed between interactive and less interactive VR systems. For example, the varying degrees and forms of interactivity were not considered.

No relationship was found between the study-level mean age of participants and the effectiveness of VR. When using aggregate data, rather than individual-level data, only the between-studies variation is analyzed. In this case, it might have concealed any true relationship between the participant's individual age and the effectiveness of VR. This is an example of what is referred to as ecological fallacy (Thompson & Higgins, 2002). It should therefore not be concluded that the age of participants is not related to the effectiveness of VR on pain.

Subgroup and meta-regression analyses are observational in nature and cannot be used to establish causality (Borenstein et al., 2009; Deeks et al., 2020). They are also based on a limited number of studies and are probably not representative of all medical procedures, VR interventions, and patients in hypothetical studies or a clinical setting. Positive results from subgroup- and meta-regression analyses should therefore not be interpreted as conclusive evidence that certain VR systems perform better than others, or that it is more effective in certain settings and patients. Neither should the opposite be inferred from the failure to identify any such differences. In conclusion, the results of the moderator analyses should not be used to draw any definitive conclusions but may inspire new hypotheses and further research on the importance of interactivity and immersion.

Sensitivity Analyses

The overall estimate of the effects on pain was somewhat reduced when inadequately powered studies were removed. Unexpectedly, a slight increase in the effect size estimate for

anxiety was observed when inadequately powered studies were removed. This increase was seemingly caused by a group of studies with narrow confidence intervals and between 50-96 participants that reported effect sizes slightly smaller than the average of studies that were considered adequately powered. As studies are assigned weights proportional to their standard error in a random-effects model, removing these studies likely caused the unexpected increase in the mean effect size estimate. It should also be noted that the power cut-off was based on an arbitrary assumption of a 0.50 effect size.

The summary effect size for both pain and anxiety remained relatively constant after removing the non-randomized studies. This is likely because only a few non-randomized studies were included in each meta-analysis, of which several had wide confidence intervals and thus contributed less to the original summary effect. It should therefore not be concluded that there is no association between the study design and effect sizes.

Only slight or no difference in the mean effects for pain and anxiety was observed when studies with a high risk of bias judgement in more than one domain were removed. However, the retained studies all had at least an intermediate overall risk of bias. This sensitivity analysis should therefore not be interpreted as evidence that bias did not influence the results.

Limitations

The measures obtained for the quantitative synthesis were subjective and thus carry inherent limitations. As pain and anxiety are private, subjective experiences, self-reports were prioritized over observational measures. However, as pointed out by von Baeyer (2009), they should be interpreted with regards to developmental and social factors. Consciously or not, children may underreport or overreport their pain for reasons such as difficulties with understanding the scales or fear of the consequences of reporting certain scores (e.g.,

underreporting pain due to a fear of being subjected to more medical procedures) (O'Brien & Root, 2019, p. 775; von Baeyer, 2009). Furthermore, scales like the Wong-Baker Faces Scale have been criticized for using response options represented by faces that cry, smile, or look angry; if the children themselves do not experience the corresponding emotions, they may avoid selecting these responses even though they most accurately reflect the level of their distress (von Baeyer, 2009). The lack of blinding to the experimental condition may also have introduced bias to the measurement of pain and anxiety. Several other issues related to the measurement of pain and anxiety also apply (see von Baeyer, 2009). An important limitation of the present study is therefore not conducting multiple analyses with reports from several informants or physiological data (e.g., pulse rate).

The validity of results from systematic reviews and meta-analyses is a product of the quality of primary studies (Borenstein et al., 2009, p. 380). For example, methodological issues of primary studies, like flaws in the randomization process and retrospectively registered trials, are also transferred to any syntheses of study results. Updated reviews should therefore be conducted as more trials with larger sample sizes and greater methodological rigor are being published.

Although efforts were made to locate unpublished studies, no studies were identified that did not get published or were in press before the completion of this review. The failure to include any unpublished studies is a significant limitation of the present study, considering the indications of publication bias. Eligible studies may also have been excluded because of language restrictions.

The risk of bias judgements were conducted by only one person in the present study. Although the ROB 2.0 and ROBINS-I tools contain decision algorithms that guide the overall judgements per domain, scoring individual items nevertheless requires at least some

subjective judgements (Higgins et al., 2011). It is therefore recommended that they are conducted by at least two reviewers independently (Higgins et al., 2011).

Another issue to consider at the review-level is the categorization of medical procedures. The categories were created with the intention of describing each included study as accurately as possible while also keeping the number of subgroups low to ensure that they were adequately sized for subgroup analyses. However, the pre-operative and wound care subgroups would preferably have been larger. The medical procedures within each subgroup were certainly not homogenous. For example, while Eijlers, Dierckx, et al. (2019) measured the effect of procedural preparation on post-operative pain, Walther-Larsen et al. (2019) measured the effect of VR distraction on acute pain from intravenous cannulation before surgery. Another example is the needle-related group, which included both lumbar punctures as part of cancer treatment and routine venipuncture in healthy children. It is possible that a different set of categories would have yielded different results and useful insight.

Finally, some issues were not addressed in the present review that are relevant when evaluating the effects of VR in pediatrics. For example, the present review did not systematically collect information regarding the health status of children or whether any pharmacological interventions were used. Furthermore, safety issues and potential side-effects from the use of VR were not discussed. Although common symptoms like nausea and vertigo tend to decline quickly after removing the VR headset, more serious concerns have also been expressed (see Nichols & Patel, 2002).

Conclusions

The results of the present review suggest that VR has beneficial effects on procedural pain in children, compared to other non-VR interventions or no intervention. The direction of the effects is in accordance with previous meta-analyses, but their magnitudes were lower

than those reported in Eijlers, Utens, et al. (2019) and Georgescu et al. (2020). The differences likely reflect the various definitions of VR and immersion and the rapidly developing literature, as well as the inclusion of adult samples in some reviews. However, the strength of evidence is considered weak due to a high risk of bias within and across studies, and it is not possible to draw any definitive conclusions.

The results indicated that non-interactive studies were superior, which contradicts the results of some previously cited studies (e.g., Dahlquist et al., 2004; Hoffman et al., 2004; Hoffman et al., 2006; Wender et al., 2009). Although these results should be interpreted with caution, it is possible that children benefit more from less demanding tasks. Studies like those mentioned above are often conducted on experimentally induced pain and mostly in adult volunteers. It would therefore not be surprising if the findings from these studies do not translate to a clinical setting with pediatric patients. If non-interactive VR interventions truly are more effective, or even equally as effective, as interactive interventions, it would contradict the common assumption that interactive interventions are superior. This would have important implications for VR developers, clinicians, and decision makers. Further research is needed to establish if interactivity could be beneficial, and if so, the optimal level and mode of interactivity for different age groups.

The review has demonstrated the diversity of VR systems in terms of hardware and software. No relationship was found between immersion and treatment effects. However, immersion features were not assessed individually, and their potential role should therefore not be dismissed. VR interventions vary in terms of the content that is displayed. Interestingly, some interventions feature content that is likely to increase arousal (e.g., rollercoaster simulations), whereas some included more relaxing content (e.g., underwater

simulations). The effects of these and other software design decisions would be interesting to address in future studies.

Decision makers should be aware of the differences between VR interventions when considering the implementation of VR in clinical settings. Less immersive and non-interactive technologies may also have additional benefits that were not discussed in the present review. For example, larger screens may be impractical during some procedures (e.g., dental procedures), auditive stimuli may disturb communication with medical personnel, and head tracking may encourage movements of the head and body that could be disruptive to the medical procedure.

In conclusion, the review suggests that VR could be beneficial in pediatrics. However, the results must be seen in context of the limitations of primary studies and the present review. More studies with larger sample sizes and methodological rigor are needed, especially on the effects of using VR for procedural preparation. Researchers should explicitly state their definitions of VR and immersion to avoid confusion. It remains unclear whether VR is more effective than all other interventions, such as non-VR, screen-based interventions. Less interactive VR may be preferable in pediatrics, but more research is needed on the potential differences between various forms and degrees of interactivity. Future studies should also be focused on individual immersion variables and the content that is displayed on the VR headsets.

References

References marked with an asterisk indicate studies included in the meta-analysis.

- Agrawal, S. K., Simon, A. I., Bech, S., Bærentsen, K. B., & Forchhammer, S. (2019). Defining immersion: Literature review and implications for research on immersive audiovisual experiences. *Journal of The Audio Engineering Society*, 68(6), 404-417.
<https://doi.org/10.17743/jaes.2020.0039>
- *Al-Halabi, M. N., Bshara, N., & Al-Nerabieah, Z. (2018). Effectiveness of audio visual distraction using virtual reality eyeglasses versus tablet device in child behavioral management during inferior alveolar nerve block. *Anaesthesia, Pain and Intensive Care*, 22(1), 55-61.
- *Al-Khotani, A., Bello, L. A., & Christidis, N. (2016). Effects of audiovisual distraction on children's behaviour during dental treatment: A randomized controlled clinical trial. *Acta Odontologica Scandinavica*, 74(6), 494-501.
<https://doi.org/10.1080/00016357.2016.1206211>
- *Al-Nerabieah, Z., Alhalabi, M. N., Owayda, A., Alsabek, L., Bshara, N., & Kouchaji, C. (2020). Effectiveness of using virtual reality eyeglasses in the waiting room on preoperative anxiety: A randomized controlled trial. *Perioperative Care and Operating Room Management*, 21, Article 100129.
<https://doi.org/10.1016/j.pcorm.2020.100129>
- *Aminabadi, N. A., Erfanparast, L., Sohrabi, A., Oskouei, S. G., & Naghili, A. (2013). The impact of virtual reality distraction on pain and anxiety during dental treatment in 4-6 year-old children: A randomized controlled clinical trial. *Journal of Dental Research, Dental Clinics, Dental Prospects*, 6(4), 117-124.
<https://doi.org/10.5681/joddd.2012.025>

- The Association of Paediatric Anaesthetists of Great Britain and Ireland. (2012). Good practice in postoperative and procedural pain management (2nd ed.). *Pediatric Anesthesia*, 22(s1), 1-79. <https://doi.org/10.1111/j.1460-9592.2012.03838.x>
- *Asvanund, Y., Mitrakul, K., Juhong, R., & Arunakul, M. (2015). Effect of audiovisual eyeglasses during local anesthesia injections in 5-to 8-year-old children. *Quintessence International*, 46(6), 513-521. <https://doi.org/10.3290/j.qi.a33932>
- *Attar, R., & Baghdadi, Z. (2015). Comparative efficacy of active and passive distraction during restorative treatment in children using an iPad versus audiovisual eyeglasses: a randomised controlled trial. *European Archives of Paediatric Dentistry*, 16(1), 1-8. <https://doi.org/10.1007/s40368-014-0136-x>
- Atzori, B., Grotto, R. L., Giugni, A., Calabrò, M., Alhalabi, W., & Hoffman, H. G. (2018). Virtual reality analgesia for pediatric dental patients. *Frontiers in Psychology*, 9, Article 2265. <https://doi.org/10.3389/fpsyg.2018.02265>
- Atzori, B., Hoffman, H. G., Vagnoli, L., Patterson, D. R., Alhalabi, W., Messeri, A., & Grotto, R. L. (2018). Virtual reality analgesia during venipuncture in pediatric patients with onco-hematological diseases. *Frontiers in Psychology*, 9, Article 2508. <https://doi.org/10.3389/fpsyg.2018.02508>
- *Aydin, A. I., & Ozyazicioglu, N. (2019). Using a virtual reality headset to decrease pain felt during a venipuncture procedure in children. *Journal of Perianesthesia Nursing*, 34(6), 1215-1221. <https://doi.org/10.1016/j.jopan.2019.05.134>
- *Bagattoni, S., D'Alessandro, G., Sadotti, A., Alkhamis, N., & Piana, G. (2018). Effects of audiovisual distraction in children with special healthcare needs during dental restorations: a randomized crossover clinical trial. *International Journal of Paediatric Dentistry*, 28(1), 111-120. <https://doi.org/10.1111/ipd.12304>

- Bailey, J. O., & Bailenson, J. N. (2017). Considering virtual reality in children's lives. *Journal of Children and Media, 11*(1), 107-113.
<https://doi.org/10.1080/17482798.2016.1268779>
- Bailey, B., Gravel, J., & Daoust, R. (2012). Reliability of the visual analog scale in children with acute pain in the emergency department. *Pain, 153*(4), 839-842.
<https://doi.org/10.1016/j.pain.2012.01.006>
- Bentley, J. (2014). Anatomy and Physiology of Pain. In A. Twycross, S. Dowden, & J. Stinson (Eds.), *Managing pain in children: A clinical guide for nurses and healthcare professionals* (2nd ed.). Wiley-Blackwell.
- Birnie, K. A., Chambers, C. T., Fernandez, C. V., Forgeron, P. A., Latimer, M. A., McGrath, P. J., Cummings, E. A., & Finley, G. A. (2014). Hospitalized children continue to report undertreated and preventable pain. *Pain Research & Management, 19*(4), 198-204. <https://doi.org/10.1155/2014/614784>
- Boeldt, D., McMahon, E., McFaul, M., & Greenleaf, W. (2019). Using virtual reality exposure therapy to enhance treatment of anxiety disorders: Identifying areas of clinical adoption and potential obstacles. *Frontiers in Psychiatry, 10*, Article 773.
<https://doi.org/10.3389/fpsy.2019.00773>
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Wiley.
- 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), Article 42. <https://doi.org/10.1007/s11920-017-0788-4>

- *Buldur, B., & Candan, M. (2020). Does virtual reality affect children's dental anxiety, pain, and behaviour? a randomised, placebo-controlled, cross-over trial. *Pesquisa Brasileira em Odontopediatria e Clínica Integrada*, 21, Article 82.
<https://doi.org/10.1590/pboci.2021.002>
- *Caruso, T. J., George, A., Menendez, M., De Souza, E., Khoury, M., Kist, M. N., & Rodriguez, S. T. (2019). Virtual reality during pediatric vascular access: a pragmatic, prospective randomized, controlled trial. *Paediatric Anaesthesia*. 30, 116– 123.
<https://doi.org/10.1111/pan.13778>
- Chan, E., Foster, S., Sambell, R., & Leong, P. (2018). Clinical efficacy of virtual reality for acute procedural pain management: A systematic review and meta-analysis. *Plos One*, 13(7), Article e0200987. <https://doi.org/10.1371/journal.pone.0200987>
- Chan, E. A., Chung, J. W. Y., Wong, T. K. S., Lien, A. S. Y., & Yang, J. Y. (2007). Application of a virtual reality prototype for pain relief of pediatric burn in Taiwan. *Journal of Clinical Nursing*, 16(4), 786-793. <https://doi.org/10.1111/j.1365-2702.2006.01719.x>
- *Chan, E., Hovenden, M., Ramage, E., Ling, N., Pham, J. H., Rahim, A., Lam, C., Liu, L., Foster, S., Sambell, R., & et al. (2019). virtual reality for pediatric needle procedural pain: Two randomized clinical trials. *Journal of Pediatrics*, 209, 160-167.
<https://doi.org/10.1016/j.jpeds.2019.02.034>
- *Chaudhary, S., Showkat, I., & Sinha, A. (2020). Comparative evaluation of the effectiveness of active and passive distraction aids in the management of anxious pediatric dental patients. *International Journal of Scientific Research*, 9(2).
[https://www.worldwidejournals.com/international-journal-of-scientific-research-\(IJSR\)/article/comparative-evaluation-of-the-effectiveness-of-active-and-passive-](https://www.worldwidejournals.com/international-journal-of-scientific-research-(IJSR)/article/comparative-evaluation-of-the-effectiveness-of-active-and-passive-)

[distraction-aids-in-the-management-of-anxious-pediatric-dental-patient/MjQyMzM=?is=1&b1=1137&k=285](https://doi.org/10.1111/jocn.15088)

- *Chen, Y. J., Cheng, S. F., Lee, P. C., Lai, C. H., Hou, I. C., & Chen, C. W. (2020). Distraction using virtual reality for children during intravenous injections in an emergency department: A randomised trial. *Journal of Clinical Nursing*, 29(3-4), 503-510. <https://doi.org/10.1111/jocn.15088>
- Cohen, J. (1998). *Statistical power analysis for the behavioral sciences* (2nd Ed.). Routledge. <https://doi.org/10.4324/9780203771587>
- Cohen, L. L., Blount, R. L., Cohen, R. J., & Johnson, V. C. (2004). Dimensions of pediatric procedural distress: Children's anxiety and pain during immunizations. *Journal of Clinical Psychology in Medical Settings*, 11(1), 41-47. <https://doi.org/10.1023/B:JOCS.0000016268.40662.ed>
- Cohen, L. L., MacLaren, J. E., & Lim, C. S. (2008). Pain and pain management. In R. G. Steele, T. D. Elkin, & M. C. Roberts (Eds.), *Handbook of evidence based therapies for children and adolescents: Bridging science and practice* (pp. 283–296). Springer.
- Cummings, J. J., & Bailenson, J. N. (2016). How immersive is enough? A meta-analysis of the effect of immersive technology on user presence. *Media Psychology*, 19(2), 272-309. <https://doi.org/10.1080/15213269.2015.1015740>
- Curtis, S., Wingert, A., & Ali, S. (2012). The Cochrane Library and procedural pain in children: an overview of reviews. *Evidence-Based Child Health: A Cochrane Review Journal*, 7(5), 1363-1399. <https://doi.org/10.1002/ebch.1864>
- Dahlquist, L. M., McKenna, K. D., Jones, K. K., Dillinger, L., Weiss, K. E., & Ackerman, C. S. (2007). Active and passive distraction using a head-mounted display helmet: Effects

on cold pressor pain in children. *Health Psychology*, 26(6), 794-801.

<https://doi.org/10.1037/0278-6133.26.6.794>

Das, D. A., Grimmer, K. A., Sparnon, A. L., McRae, S. E., & Thomas, B. H. (2005). The efficacy of playing a virtual reality game in modulating pain for children with acute burn injuries: a randomized controlled trial. *BMC Pediatrics*, 5, Article 1.

<https://doi.org/10.1186/1471-2431-5-1>

Deeks, J. J., Higgins, J. P. T., & Altman, D. G. (2020). Chapter 10: Analysing data and undertaking meta-analyses. In J. P. T. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. J. Page, & V. A. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions version 6.1*. [Updated September 2020]. Cochrane.

www.training.cochrane.org/handbook

*del Castillo, B. T., Torres, J. A. P., Sanchez, L. M., Castellanos, M. E., Fernandez, L. E., Sanchez, M. I. G., & Fernandez, R. R. (2019). Reducing the pain in invasive procedures during paediatric hospital admissions: Fiction, reality or virtual reality? *Anales De Pediatria*, 91(2), 80-87. <https://doi.org/10.1016/j.anpedi.2018.10.019>

DeMore, M., & Cohen, L. L. (2005). distraction for pediatric immunization pain: A critical review. *Journal of Clinical Psychology in Medical Settings*, 12(4), 281-291.

<https://doi.org/10.1007/s10880-005-7813-1>

*Dumoulin, S., Bouchard, S., Ellis, J., Lavoie, K. L., Vézina, M. P., Charbonneau, P., Tardif, J., & Hajjar, A. (2019). A randomized controlled trial on the use of virtual reality for needle-related procedures in children and adolescents in the emergency department.

Games for Health Journal, 8(4), 285-293. <https://doi.org/10.1089/g4h.2018.0111>

- Edward, K.-L., Sanderson, W., & Giandinoto, J.-A. (2015). Mock admissions used to minimise paediatric anxiety before surgery. *ACORN: The Journal of Perioperative Nursing in Australia*, 28(4), 24-26.
- Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ*, 315(7109), 629–634.
<https://doi.org/10.1136/bmj.315.7109.629>
- *Eijlers, R., Dierckx, B., Staals, L. M., Berghmans, J. M., van der Schroeff, M. P., Strabbing, E. M., Wijnen, R. M. H., Hillegers, M. H. J., Legerstee, J. S., & Utens, E. (2019). Virtual reality exposure before elective day care surgery to reduce anxiety and pain in children: a randomised controlled trial. *European Journal of Anaesthesiology*, 36(10), 728-737. <https://doi.org/10.1097/EJA.0000000000001059>
- Eijlers, R., Utens, E., Staals, L. M., de Nijs, P. F. A., Berghmans, J. M., Wijnen, R. M. H., Hillegers, M. H. J., Dierckx, B., & Legerstee, J. S. (2019). Systematic review and meta-analysis of virtual reality in pediatrics: Effects on pain and anxiety. *Anesthesia and Analgesia*, 129(5), 1344-1353. <https://doi.org/10.1213/ane.0000000000004165>
- El-Sharkawi, H. F. A., El-Housseiny, A. A., & Aly, A. M. (2012). Effectiveness of new distraction technique on pain associated with injection of local anesthesia for children. *Pediatric Dentistry*, 34(2), 142-145.
<https://www.ingentaconnect.com/contentone/aapd/pd/2012/00000034/00000002/art00016>
- Elbourne, D. R., Altman, D. G., Higgins, J. P., Curtin, F., Worthington, H. V., & Vail, A. (2002). Meta-analyses involving cross-over trials: Methodological issues. *International Journal of Epidemiology*, 31(1), 140-149.
<https://doi.org/10.1093/ije/31.1.140>

- *Fakhrudin, K. S., Hisham, E. B., & Gorduysus, M. O. (2015). Effectiveness of audiovisual distraction eyewear and computerized delivery of anesthesia during pulp therapy of primary molars in phobic child patients. *European Journal of Dentistry*, 9(4), 470-475. <https://doi.org/10.4103/1305-7456.172637>
- Freeman, D., Reeve, S., Robinson, A., Ehlers, A., Clark, D., Spanlang, B., & Slater, M. (2017). Virtual reality in the assessment, understanding, and treatment of mental health disorders. *Psychological Medicine*, 47(14), 2393-2400. <https://doi.org/10.1017/S003329171700040X>
- Friedrichsdorf, S. J., & Goubert, L. (2020). Pediatric pain treatment and prevention for hospitalized children. *PAIN Reports*, 5(1), e804. <https://doi.org/10.1097/pr9.0000000000000804>
- Fuchs, P. (2019). VR headsets. In P. Fuchs (Ed.), *Virtual reality headsets - a theoretical and pragmatic approach* (pp. 55-65). CRC Press.
- Gandhi, M., Thomson, C., Lord, D., & Enoch, S. (2010). Management of pain in children with burns. *International Journal of Pediatrics*, 2010. <https://doi.org/10.1155/2010/825657>
- Garrocho-Rangel, A., Ibarra-Gutiérrez, E., Rosales-Bérber, M., Esquivel-Hernández, R., Esparza-Villalpando, V., & Pozos-Guillén, A. (2018). A video eyeglasses/earphones system as distracting method during dental treatment in children: A crossover randomised and controlled clinical trial. *European Journal of Paediatric Dentistry*, 19(1), 74-79. <https://doi.org/10.23804/ejpd.2018.19.01.14>
- Georgescu, R., Fodor, L. A., Dobrea, A., & Cristea, I. A. (2020). Psychological interventions using virtual reality for pain associated with medical procedures: a systematic review

and meta-analysis. *Psychological Medicine*, 50(11), 1795-1807.

<https://doi.org/10.1017/s0033291719001855>

*Gerceker, G. Ö., Ayar, D., Özdemir, E. Z., & Bektaş, M. (2020). Effects of virtual reality on pain, fear and anxiety during blood draw in children aged 5–12 years old: A randomised controlled study. *Journal of Clinical Nursing*, 29(7-8), 1151-1161.

<https://doi.org/10.1111/jocn.15173>

*Gerçeker, G., Binay, Ş., Bilsin, E., Kahraman, A., & Yılmaz, H. B. (2018). Effects of virtual reality and external cold and vibration on pain in 7- to 12-year-old children during phlebotomy: A randomized controlled trial. *Journal of Perianesthesia Nursing*, 33(6),

981-989. <https://doi.org/10.1016/j.jopan.2017.12.010>

*Gerceker, G. O., Bektas, M., Aydinok, Y., Ören, H., Ellidokuz, H., & Olgun, N. (in press). The effect of virtual reality on pain, fear, and anxiety during access a port with Huber needle in pediatric hematology-oncology patients: Randomized controlled trial.

European Journal of Oncology Nursing.

Gershon, J., Zimand, E., Pickering, M., Rothbaum, B. O., & Hodges, L. (2004). A pilot and feasibility study of virtual reality as a distraction for children with cancer. *Journal of the American Academy of Child & Adolescent Psychiatry*, 43(10), 1243-1249.

<https://doi.org/10.1097/01.chi.0000135621.23145.05>

Gigante, M. A. (1993). 1 - Virtual Reality: Definitions, History and Applications. In R. A. Earnshaw, M. A. Gigante, & H. Jones (Eds.), *Virtual Reality Systems* (pp. 3-14).

Academic Press. <https://doi.org/10.1016/B978-0-12-227748-1.50009-3>

*Gold, J. I., Kim, S. H., Kant, A. J., Joseph, M. H., & Rizzo, A. S. (2006). Effectiveness of virtual reality for pediatric pain distraction during I.V. placement. *Cyberpsychology & Behavior*, 9(2), 207-212. <https://doi.org/10.1089/cpb.2006.9.207>

- *Gold, J. I., & Mahrer, N. E. (2018). Is virtual reality ready for prime time in the medical space? A randomized control trial of pediatric virtual reality for acute procedural pain management. *Journal of Pediatric Psychology*, 43(3), 266-275.
<https://doi.org/10.1093/jpepsy/jsx129>
- *Goldman, R. D., & Behboudi, A. (2020). Virtual reality for intravenous placement in the emergency department - a randomized controlled trial. *European Journal of Pediatrics*. Advance online publication. <https://doi.org/10.1007/s00431-020-03771-9>
- Gupta, A., Scott, K., & Dukewich, M. (2017). Innovative technology using virtual reality in the treatment of pain: Does it reduce pain via distraction, or is there more to it? *Pain Medicine*, 19(1), 151-159. <https://doi.org/10.1093/pm/pnx109>
- Gutiérrez-Maldonado, J., Gutiérrez-Martínez, O., & Cabas-Hoyos, K. (2011). Interactive and passive virtual reality distraction: effects on presence and pain intensity. *Studies in Health Technology and Informatics*, 167, 69-73. <https://doi.org/10.3233/978-1-60750-766-6-69>
- Gutiérrez-Martínez, O., Gutiérrez-Maldonado, J., & Loreto-Quijada, D. (2011). Control over the virtual environment influences the presence and efficacy of a virtual reality intervention on pain. *Annual Review of Cybertherapy and Telemedicine*, 9, 90-93.
<https://doi.org/10.3233/978-1-60750-766-6-111>
- Han, S. H., Park, J. W., Choi, S. I., Kim, J. Y., Lee, H., Yoo, H. J., & Ryu, J. H. (2019). Effect of immersive virtual reality education before chest radiography on anxiety and distress among pediatric patients: A randomized clinical trial. *JAMA Pediatrics*, 173(11), 1026-1031. <https://doi.org/10.1001/jamapediatrics.2019.3000>
- Hardin, A. P., & Hackell, J. M. (2017). Age limit of pediatrics. *Pediatrics*, 140(3), Article e20172151. <https://doi.org/10.1542/peds.2017-2151>

- *Hashimoto, Y., Chaki, T., Hirata, N., Tokinaga, Y., Yoshikawa, Y., & Yamakage, M. (2020). Video glasses reduce preoperative anxiety compared with portable multimedia player in children: A randomized controlled trial. *Journal of Perianesthesia Nursing*, 35(3), 321-325. <https://doi.org/10.1016/j.jopan.2019.10.001>
- Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, 6(2), 107-128. <https://doi.org/10.2307/1164588>
- Hempel, S., Suttorp, M. J., Miles, J. N., Wang, Z., Maglione, M., Morton, S., Johnsen, B., Valentine, D., & Shekelle, P. G. (2011). *Empirical evidence of associations between trial quality and effect size*. Agency for Healthcare Research and Quality (US). <https://www.ncbi.nlm.nih.gov/books/NBK56925/>
- Hicks, C. L., von Baeyer, C. L., Spafford, P. A., van Korlaar, I., & Goodenough, B. (2001). The Faces Pain Scale-Revised: toward a common metric in pediatric pain measurement. *Pain*, 93(2), 173-183. [https://doi.org/10.1016/s0304-3959\(01\)00314-1](https://doi.org/10.1016/s0304-3959(01)00314-1)
- Higgins, J., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M., & Welch, V. (Eds.). (2020). *Cochrane handbook for systematic reviews of interventions version 6.1* [Updated September 2020]. The Cochrane Collaboration. www.training.cochrane.org/handbook
- Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ*, 327(7414), 557-560. <https://doi.org/10.1136/bmj.327.7414.557>
- Hoffman, H. G., Rodriguez, R. A., Gonzalez, M., Bernardy, M., Peña, R., Beck, W., Patterson, D. R., & Meyer, W. J., III. (2019). Immersive virtual reality as an adjunctive non-opioid analgesic for pre-dominantly latin american children with large

severe burn wounds during burn wound cleaning in the intensive care unit: a pilot study. *Frontiers in Human Neuroscience*, 13, Article 262.

<https://doi.org/10.3389/fnhum.2019.00262>

Hoffman, H. G., Seibel, E. J., Richards, T. L., Furness, T. A., Patterson, D. R., & Sharar, S. R. (2006). Virtual reality helmet display quality influences the magnitude of virtual reality analgesia. *Journal of Pain*, 7(11), 843-850.

<https://doi.org/10.1016/j.jpain.2006.04.006>

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

*Hoge, M. A., Howard, M. R., Wallace, D. P., & Allen, K. D. (2012). Use of video eyewear to manage distress in children during restorative dental treatment. *Pediatric Dentistry*, 34(5), 378-382.

<https://www.ingentaconnect.com/contentone/aapd/pd/2012/00000034/00000005/art00004>

*Hua, Y., Qiu, R., Yao, W.-y., Zhang, Q., & Chen, X.-l. (2015). The effect of virtual reality distraction on pain relief during dressing changes in children with chronic wounds on lower limbs. *Pain Management Nursing*, 16(5), 685-691.

<https://doi.org/10.1016/j.pmn.2015.03.001>

Hugues, O. (2019). Functional and technical characteristics of VR headsets. In P. Fuchs (Ed.), *Virtual reality headsets - a theoretical and pragmatic approach* (pp. 79-96). CRC Press.

Iannicelli, A. M., Vito, D., Dodaro, C. A., De Matteo, P., Nocerino, R., Sepe, A., & Raia, V. (2019). Does virtual reality reduce pain in pediatric patients? A systematic review.

Italian Journal of Pediatrics, 45(1), Article 171. <https://doi.org/10.1186/s13052-019-0757-0>

*Inangil, D., Sendir, M., & Buyukyilmaz, F. (2020). Efficacy of cartoon viewing devices during phlebotomy in children: A randomized controlled trial. *Journal of Perianesthesia Nursing*, 35(4), 407-412. <https://doi.org/10.1016/j.jopan.2020.01.008>

Indovina, P., Barone, D., Gallo, L., Chirico, A., De Pietro, G., & Giordano, A. (2018). Virtual reality as a distraction intervention to relieve pain and distress during medical procedures. *The Clinical Journal of Pain*, 34(9), 858-877. <https://doi.org/10.1097/ajp.0000000000000599>

The International Association for the Study of Pain (IASP). (2011). *IASP Terminology*. Retrieved August 25, 2019, from <https://www.iasp-pain.org/Education/Content.aspx?ItemNumber=1698>

*Isong, I. A., Rao, S. R., Holifield, C., Iannuzzi, D., Hanson, E., Ware, J., & Nelson, L. P. (2014). Addressing dental fear in children with autism spectrum disorders: A randomized controlled pilot study using electronic screen media. *Clinical Pediatrics*, 53(3), 230-237. <https://doi.org/10.1177/0009922813517169>

James, K. H., Humphrey, G. K., Vilis, T., Corrie, B., Baddour, R., & Goodale, M. A. (2002). “Active” and “passive” learning of three-dimensional object structure within an immersive virtual reality environment. *Behavior Research Methods, Instruments, & Computers*, 34(3), 383-390. <https://doi.org/10.3758/BF03195466>

*Jeffs, D., Dorman, D., Brown, S., Files, A., Graves, T., Kirk, E., Meredith-Neve, S., Sanders, J., White, B., & Swearingen, C. J. (2014). Effect of virtual reality on adolescent pain during burn wound care. *Journal of Burn Care & Research*, 35(5), 395-408. <https://doi.org/10.1097/bcr.0000000000000019>

- Johnson, M. H. (2005). How does distraction work in the management of pain? *Current Pain and Headache Reports*, 9(2), 90-95. <https://doi.org/10.1007/s11916-005-0044-1>
- Johnson, S., & Coxon, M. (2016). Sound can enhance the analgesic effect of virtual reality. *Royal Society Open Science*, 3(3), Article 150567. <https://doi.org/10.1098/rsos.150567>
- Jaaniste, T., Hayes, B., & Von Baeyer, C. L. (2007). Providing children with information about forthcoming medical procedures: A review and synthesis. *Clinical Psychology: Science and Practice*, 14(2), 124-143. <https://doi.org/10.1111/j.1468-2850.2007.00072.x>
- *Jung, M. J., Libaw, J. S., Ma, K., Whitlock, E. L., Feiner, J. R., & Sinskey, J. L. (2020). Pediatric Distraction on Induction of Anesthesia With Virtual Reality and Perioperative Anxiolysis: a Randomized Controlled Trial. *Anesthesia and Analgesia*. Advance online publication. <https://doi.org/10.1213/ANE.0000000000005004>
- Kain, Z. N., Mayes, L. C., Cicchetti, D. V., Bagnall, A. L., Finley, J. D., & Hofstadter, M. B. (1997). The Yale Preoperative Anxiety Scale: How does it compare with a "gold standard"? *Anesthesia & Analgesia*, 85(4), 783-788. <https://doi.org/10.1097/00000539-199710000-00012>
- Kao, G. S., & Schwartz, E. R. (2019). Distraction techniques for pediatric pain management. In A. C. Adler, A. Chandrakantan, & R. S. Litman (Eds.), *Case studies in pediatric anesthesia* (pp. 260-263). Cambridge University Press. <https://doi.org/10.1017/9781108668736.058>
- Kennedy, R. M., Luhmann, J., & Zempsky, W. T. (2008). Clinical implications of unmanaged needle-insertion pain and distress in children. *Pediatrics*, 122(3, Suppl.), S130. <https://doi.org/10.1542/peds.2008-1055e>

- Kerimoglu, B., Neuman, A., Paul, J., Stefanov, D. G., & Twersky, R. (2013). Anesthesia induction using video glasses as a distraction tool for the management of preoperative anxiety in children. *Anesthesia & Analgesia*, *117*(6), 1373-1379.
<https://doi.org/10.1213/ane.0b013e3182a8c18f>
- *Khadra, C., Ballard, A., Paquin, D., Cotes-Turpin, C., Hoffman, H. G., Perreault, I., Fortin, J. S., Bouchard, S., Th roux, J., & Le May, S. (2020). Effects of a projector-based hybrid virtual reality on pain in young children with burn injuries during hydrotherapy sessions: A within-subject randomized crossover trial. *Burns*, *46*(7), 1571-1584.
<https://doi.org/10.1016/j.burns.2020.04.006>
- Khan, S., Rao, D., Jasuja, P., Malik, S., Al Yami, S. M. H., Al Makrami, M. H. T., & Al Milaq, F. H. S. (2019). Passive distraction: A technique to maintain children's behavior undergoing dental treatment. *Indo American Journal of Pharmaceutical Sciences*, *6*(2), 4043-4048. <https://doi.org/10.5281/zenodo.2566813>
- Kipping, B., Rodger, S., Miller, K., & Kimble, R. M. (2012). Virtual reality for acute pain reduction in adolescents undergoing burn wound care: A prospective randomized controlled trial. *Burns*, *38*(5), 650-657. <https://doi.org/10.1016/j.burns.2011.11.010>
- Koticha, P., Katge, F., Shetty, S., & Patil, D. P. (2019). Effectiveness of virtual reality eyeglasses as a distraction aid to reduce anxiety among 6–10-year-old children undergoing dental extraction procedure. *International Journal of Clinical Pediatric Dentistry*, *12*(4), 297. <https://dx.doi.org/10.5005%2Fjfp-journals-10005-1640>
- Kourtesis, P., Collina, S., Doumas, L. A. A., & MacPherson, S. E. (2019). Technological competence is a pre-condition for effective implementation of virtual reality head mounted displays in human neuroscience: A technological review and meta-analysis.

Frontiers in Human Neuroscience, 13, 342-342.

<https://doi.org/10.3389/fnhum.2019.00342>

Lambert, V., Boylan, P., Boran, L., Hicks, P., Kirubakaran, R., Devane, D., & Matthews, A.

(2020). Virtual reality distraction for acute pain in children. *Cochrane Database of*

Systematic Reviews. <https://doi.org/https://doi.org/10.1002/14651858.CD010686.pub2>

Laver, K. E., Lange, B., George, S., Deutsch, J. E., Saposnik, G., & Crotty, M. (2017). Virtual

reality for stroke rehabilitation. *Cochrane Database of Systematic Reviews*.

<https://doi.org/10.1002/14651858.CD008349.pub4>

Lavoie, K. (2013). Anxiety. In M. D. Gellman & J. R. Turner (Eds.), *Encyclopedia of*

Behavioral Medicine (pp. 106-108). Springer. <https://doi.org/10.1007/978-1-4419->

[1005-9_1090](https://doi.org/10.1007/978-1-4419-1005-9_1090)

Li, A., Montaña, Z., Chen, V. J., & Gold, J. I. (2011). Virtual reality and pain management:

Current trends and future directions. *Pain Management*, 1(2), 147-157.

<https://doi.org/10.2217/pmt.10.15>

Linton, S. J., & Shaw, W. S. (2011). Impact of psychological factors in the experience of

pain. *Physical Therapy*, 91(5), 700-711. <https://doi.org/10.2522/ptj.20100330>

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and

behavioral treatment. Confirmation from meta-analysis. *American Psychologist*,

48(12), 1181-1209. <https://doi.org/10.1037//0003-066x.48.12.1181>

Liszio, S., Graf, L., Basu, O., & Masuch, M. (2020). Pengonaut trainer: A playful VR app to

prepare children for MRI examinations: In-depth game design analysis. *IDC '20:*

Proceedings of the Interaction Design and Children Conference, 470–482.

Association for Computing Machinery. <https://doi.org/10.1145/3392063.3394432>

- *Litwin, S., Nguyen, C., Hundert, A., Stuart, S., Liu, D., Maguire, B., Matava, C., & Stinson, J. (2020). Virtual reality to reduce procedural pain during iv insertion in the pediatric emergency department: A pilot randomized controlled trial. *The Clinical Journal of Pain*. Advance online publication. <https://doi.org/10.1097/ajp.0000000000000894>
- Luo, D., Wan, X., Liu, J., & Tong, T. (2018). Optimally estimating the sample mean from the sample size, median, mid-range, and/or mid-quartile range. *Statistical Methods in Medical Research*, 27(6), 1785-1805. <https://doi.org/10.1177/0962280216669183>
- Malloy, K. M., & Milling, L. S. (2010). The effectiveness of virtual reality distraction for pain reduction: A systematic review. *Clinical Psychology Review*, 30(8), 1011-1018. <https://doi.org/10.1016/j.cpr.2010.07.001>
- Mathews, L. (2011). Pain in Children: Neglected, unaddressed and mismanaged. *Indian Journal of Palliative Care*, 17(4), 70-73. <https://doi.org/10.4103/0973-1075.76247>
- McCaul, K. D., & Malott, J. M. (1984). Distraction and coping with pain. *Psychological Bulletin*, 95(3), 516-533. <https://doi.org/10.1037/0033-2909.95.3.516>
- McGuinness, L. A., & Higgins, J. P. T. (2020). Risk-of-bias VISualization (robvis): An R package and Shiny web app for visualizing risk-of-bias assessments. *Research Synthesis Methods*. Advance online publication. <https://doi.org/https://doi.org/10.1002/jrsm.1411>
- McMurtry, C. M., Noel, M., Chambers, C. T., & McGrath, P. J. (2011). Children's fear during procedural pain: preliminary investigation of the Children's Fear Scale. *Health Psychol*, 30(6), 780-788. <https://doi.org/10.1037/a0024817>
- McMurtry, C. M., Pillai Riddell, R., Taddio, A., Racine, N., Asmundson, G. J. G., Noel, M., Chambers, C. T., Shah, V., & HELPinKids&Adults Team. (2015). Far From “Just a Poke”: Common Painful Needle Procedures and the Development of Needle Fear. *The*

Clinical Journal of Pain, 31(10, Suppl.), S3-S11.

<https://doi.org/10.1097/ajp.0000000000000272>

Melzack, R. (1999). From the gate to the neuromatrix. *Pain*, 82 (Suppl. 1), S121-S126.

[https://doi.org/10.1016/s0304-3959\(99\)00145-1](https://doi.org/10.1016/s0304-3959(99)00145-1)

Melzack, R., & Wall, P. D. (1965). Pain Mechanisms: A New Theory. *Science*, 150(3699),

971. <https://doi.org/10.1126/science.150.3699.971>

*Mitrakul, K., Asvanund, Y., Arunakul, M., & Paka-Akekaphat, S. (2015). Effect of audiovisual eyeglasses during dental treatment in 5-8 year-old children. *European Journal of Paediatric Dentistry*, 16(3), 239-245.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The, P. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLOS*

Medicine, 6(7), Article e1000097. <https://doi.org/10.1371/journal.pmed.1000097>

Moher, D., Pham, B., Jones, A., Cook, D. J., Jadad, A. R., Moher, M., Tugwell, P., & Klassen, T. P. (1998). Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet*, 352(9128), 609-613.

[https://doi.org/10.1016/s0140-6736\(98\)01085-x](https://doi.org/10.1016/s0140-6736(98)01085-x)

Neri, S. G., Cardoso, J. R., Cruz, L., Lima, R. M., de Oliveira, R. J., Iversen, M. D., & Carregaro, R. L. (2017). Do virtual reality games improve mobility skills and balance measurements in community-dwelling older adults? Systematic review and meta-analysis. *Clinical Rehabilitation*, 31(10), 1292-1304.

<https://doi.org/10.1177/0269215517694677>

Nichols, S., & Patel, H. (2002). Health and safety implications of virtual reality: a review of empirical evidence. *Applied Ergonomics*, 33(3), 251-271.

[https://doi.org/10.1016/S0003-6870\(02\)00020-0](https://doi.org/10.1016/S0003-6870(02)00020-0)

- *Niharika, P., Reddy, N. V., Srujana, P., Srikanth, K., Daneswari, V., & Geetha, K. S. (2018). Effects of distraction using virtual reality technology on pain perception and anxiety levels in children during pulp therapy of primary molars. *Journal of the Indian Society of Pedodontics & Preventive Dentistry*, 36(4), 364-369. https://doi.org/10.4103/jisppd.jisppd_1158_17
- Niki, K., Okamoto, Y., Maeda, I., Mori, I., Ishii, R., Matsuda, Y., Takagi, T., & Uejima, E. (2019). A novel palliative care approach using virtual reality for improving various symptoms of terminal cancer patients: A preliminary prospective, multicenter study. *Journal of Palliative Medicine*, 22(6), 702-707. <https://doi.org/10.1089/jpm.2018.0527>
- Nilsson, N., Nordahl, R., & Serafin, S. (2016). Immersion revisited: A review of existing definitions of immersion and their relation to different theories of presence. *Human Technology*, 12, 108-134. <https://doi.org/10.17011/ht/urn.201611174652>
- *Nunna, M., Dasaraju, R. K., Kamatham, R., Mallineni, S. K., & Nuvvula, S. (2019). Comparative evaluation of virtual reality distraction and counter-stimulation on dental anxiety and pain perception in children. *Journal of Dental Anesthesia and Pain Medicine*, 19(5), 277-288.
- *Nuvvula, S., Alahari, S., Kamatham, R., & Challa, R. R. (2014). Effect of audiovisual distraction with 3D video glasses on dental anxiety of children experiencing administration of local analgesia: A randomised clinical trial. *European Archives of Paediatric Dentistry*, 16(1), 43-50. <https://doi.org/10.1007/s40368-014-0145-9>
- O'Brien, J. H., & Root, M. C. (2019). Pediatric pain: Knowing the child before you. In B. R. Ferrell & J. A. Paice (Eds.), *Oxford textbook of palliative nursing* (5th ed.). Oxford University Press. <https://doi.org/10.1093/med/9780190862374.003.0064>

- O'Regan, J. K., & Noë, A. (2001). A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5), 939-973.
<https://doi.org/10.1017/S0140525X01000115>
- *Osmanliu, E., Trottier, E. D., Bailey, B., Lagacé, M., Certain, M., Khadra, C., Sanchez, M., Thériault, C., Paquin, D., & May, S. L. (in press). Distraction in the Emergency department using Virtual reality for INtravenous procedures in Children to Improve comfort (DEVINCI): A pilot pragmatic randomized controlled trial. *Canadian Journal of Emergency Medicine*.
- Pan, X., & Hamilton, A. F. C. (2018). Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape. *British Journal of Psychology*, 109(3), 395-417. <https://doi.org/10.1111/bjop.12290>
- Pate, J. T., Blount, R. L., Cohen, L. L., & Smith, A. J. (1996). Childhood medical experience and temperament as predictors of adult functioning in medical situations. *Children's Health Care*, 25(4), 281-298. https://doi.org/10.1207/s15326888chc2504_4
- Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, 61(10), 991-996.
<https://doi.org/https://doi.org/10.1016/j.jclinepi.2007.11.010>
- *Piskorz, J., & Czub, M. (2018). Effectiveness of a virtual reality intervention to minimize pediatric stress and pain intensity during venipuncture. *Journal for Specialists in Pediatric Nursing*, 23(1), Article e12201. <https://doi.org/10.1111/jspn.12201>
- *Piskorz, J. E., Czub, M., Šulžickaja, B., & Kiliś-Pstrusińska, K. (2020). Mobile virtual reality distraction reduces needle pain and stress in children. *Cyberpsychology*, 14(1), Article 3. <https://doi.org/10.5817/CP2020-1-3>

- Reid, K., Twycross, A., & Tuterra, D. (2014). Management of painful procedures. In A. Twycross, S. Dowden, & J. Stinson (Eds.), *Managing Pain in Children: A Clinical Guide for Nurses and Healthcare Professionals*. Wiley-Blackwell.
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2006). Publication bias in meta-analysis. In H. R. Rothstein, M. Borenstein, & A. J. Sutton (Eds), *Publication bias in meta-analysis: Prevention, assessment and adjustments*. Wiley.
<https://doi.org/10.1002/0470870168>
- *Ryu, J. H., Oh, A. Y., Yoo, H. J., Kim, J. H., Park, J. W., & Han, S. H. (2019). The effect of an immersive virtual reality tour of the operating theater on emergence delirium in children undergoing general anesthesia: A randomized controlled trial. *Paediatric Anaesthesia*, 29(1), 98-105. <https://doi.org/10.1111/pan.13535>
- *Ryu, J. H., Park, J. W., Nahm, F. S., Jeon, Y. T., Oh, A. Y., Lee, H. J., Kim, J. H., & Han, S. H. (2018). The effect of gamification through a virtual reality on preoperative anxiety in pediatric patients undergoing general anesthesia: A prospective, randomized, and controlled trial. *Journal of Clinical Medicine*, 7(9).
<https://doi.org/10.3390/jcm7090284>
- *Ryu, J. H., Park, S. J., Park, J. W., Kim, J. W., Yoo, H. J., Kim, T. W., Hong, J. S., & Han, S. H. (2017). Randomized clinical trial of immersive virtual reality tour of the operating theatre in children before anaesthesia. *British Journal of Surgery*, 104(12), 1628-1633. <https://doi.org/10.1002/bjs.10684>
- *Sander Wint, S., Eshelman, D., Steele, J., & Guzzetta, C. E. (2002). Effects of distraction using virtual reality glasses during lumbar punctures in adolescents with cancer. *Oncology Nursing Forum*, 29(1), e8-e15. <https://doi.org/10.1188/02.onf.e8-e15>

Schechter, N. L., Zempsky, W. T., Cohen, L. L., McGrath, P. J., McMurtry, C. M., & Bright, N. S. (2007). Pain reduction during pediatric immunizations: Evidence-based review and recommendations. *Pediatrics*, *119*(5), e1184-e1198.

<https://doi.org/10.1542/peds.2006-1107>

*Schlechter, A. K., Whitaker, W., Iyer, S., Gabriele, G., & Wilkinson, M. (2020). Virtual reality distraction during pediatric intravenous line placement in the emergency department: A prospective randomized comparison study. *American Journal of Emergency Medicine*. <https://doi.org/10.1016/j.ajem.2020.04.009>

*Schmitt, Y. S., Hoffman, H. G., Blough, D. K., Patterson, D. R., Jensen, M. P., Soltani, M., Carrouger, G. J., Nakamura, D., & Sharar, S. R. (2011). A randomized, controlled trial of immersive virtual reality analgesia, during physical therapy for pediatric burns. *Burns*, *37*(1), 61-68. <https://doi.org/10.1016/j.burns.2010.07.007>

Schünemann, H., Vist, G., Higgins, J., Santesso, N., Deeks, J., Glasziou, P., Akl, E., & Guyatt, G. (2020). Chapter 15: Interpreting results and drawing conclusions. In J. Higgins, J. Thomas, J. Chandler, M. Cumpston, T. Li, M. Page, & V. Welch (Eds.), *Cochrane handbook for systematic reviews of interventions version 6.1* [Updated September 2020]. The Cochrane Collaboration. www.training.cochrane.org/handbook

*Shah, U., & Bhatia, R. (2018). Effectiveness of audiovisual distraction eyeglass method compared to tell-play-do technique among 4–7-year-old children: A randomized controlled trial. *International Journal of Oral Care Research*, *6*, 1-7.

*Shetty, V., Suresh, L. R., & Hegde, A. M. (2019). Effect of virtual reality distraction on pain and anxiety during dental treatment in 5 to 8 year old children. *Journal of Clinical Pediatric Dentistry*, *43*(2), 97-102. <https://doi.org/10.17796/1053-4625-43.2.5>

- Shi, J., Luo, D., Weng, H., Zeng, X.-T., Lin, L., Chu, H., & Tong, T. (2020). Optimally estimating the sample standard deviation from the five-number summary. *Research Synthesis Methods, 11*(5), 641-654. <https://doi.org/https://doi.org/10.1002/jrsm.1429>
- Shi, L., & Lin, L. (2019). The trim-and-fill method for publication bias: Practical guidelines and recommendations based on a large database of meta-analyses. *Medicine, 98*(23), Article e15987. <https://doi.org/10.1097/md.0000000000015987>
- Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1535), 3549-3557. <https://doi.org/doi:10.1098/rstb.2009.0138>
- Slater, M. (2018). Immersion and the illusion of presence in virtual reality. *British Journal of Psychology, 109*(3), 431-433. <https://doi.org/10.1111/bjop.12305>
- Slater, M., & Sanchez-Vives, M. V. (2016). Enhancing our lives with immersive virtual reality. *Frontiers in Robotics and AI, 3*(74). <https://doi.org/10.3389/frobt.2016.00074>
- Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators & Virtual Environments, 6*(6), 603-616. <https://doi.org/10.1162/pres.1997.6.6.603>
- Slifer, K. J. (2013). *A clinician's guide to helping children cope and cooperate with medical care: An applied behavioral approach*. Johns Hopkins University Press.
- Snider, L., Majnemer, A., & Darsaklis, V. (2010). Virtual reality as a therapeutic modality for children with cerebral palsy. *Developmental Neurorehabilitation, 13*(2), 120-128. <https://doi.org/10.3109/17518420903357753>
- StataCorp. (2019). *Stata Statistical Software: Release 16* [Computer software]. StataCorp LLC. <https://www.stata.com/>

- Sterne, J. A., Becker, B. J., & Egger, M. (2006). The funnel plot. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis* (pp. 73-98).
<https://doi.org/10.1002/0470870168.ch5>
- Sterne, J. A., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis* (pp. 99-110).
<https://doi.org/https://doi.org/10.1002/0470870168.ch6>
- Sterne, J. A., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., Henry, D., Altman, D. G., Ansari, M. T., Boutron, I., Carpenter, J. R., Chan, A.-W., Churchill, R., Deeks, J. J., Hróbjartsson, A., Kirkham, J., Jüni, P., Loke, Y. K., Pigott, T. D., . . . Higgins, J. P. (2016). ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*, 355(8080), Article i4919.
<https://doi.org/10.1136/bmj.i4919>
- Sterne, J. A. C., Savović, J., Page, M. J., Elbers, R. G., Blencowe, N. S., Boutron, I., Cates, C. J., Cheng, H. Y., Corbett, M. S., Eldridge, S. M., Emberson, J. R., Hernán, M. A., Hopewell, S., Hróbjartsson, A., Junqueira, D. R., Jüni, P., Kirkham, J. J., Lasserson, T., Li, T., . . . Higgins, J. P. T. (2019). RoB 2: A revised tool for assessing risk of bias in randomised trials. *BMJ*, 366, Article l4898. <https://doi.org/10.1136/bmj.l4898>
- Stevens, B. J., Abbott, L. K., Yamada, J., Harrison, D., Stinson, J., Taddio, A., Barwick, M., Latimer, M., Scott, S. D., Rashotte, J., Campbell, F., & Finley, G. A. (2011). Epidemiology and management of painful procedures in children in Canadian hospitals. *CMAJ*, 183(7), E403-410. <https://doi.org/10.1503/cmaj.101341>

- Sullivan, C., Schneider, P. E., Musselman, R. J., Dummett Jr, C. O., & Gardiner, D. (2000). The effect of virtual reality during dental treatment on child anxiety and behavior. *Journal of Dentistry for Children*, 67(3), 193-196.
- Taddio, A., Katz, J., Ilersich, A. L., & Koren, G. (1997). Effect of neonatal circumcision on pain response during subsequent routine vaccination. *The Lancet*, 349(9052), 599-603. [https://doi.org/https://doi.org/10.1016/S0140-6736\(96\)10316-0](https://doi.org/https://doi.org/10.1016/S0140-6736(96)10316-0)
- Thompson, S. G., & Higgins, J. P. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine*, 21(11), 1559-1573. <https://doi.org/10.1002/sim.1187>
- Tuena, C., Serino, S., Dutriaux, L., Riva, G., & Piolino, P. (2019). Virtual enactment effect on memory in young and aged populations: A systematic review. *Journal of Clinical Medicine*, 8(5), Article 620. <https://doi.org/10.3390/jcm8050620>
- Twycross, A., Parker, R., Williams, A., & Gibson, F. (2015). Cancer-related pain and pain management: Sources, prevalence, and the experiences of children and parents. *Journal of Pediatric Oncology Nursing*, 32(6), 369-384. <https://doi.org/10.1177/1043454214563751>
- Veroniki, A. A., Jackson, D., Viechtbauer, W., Bender, R., Bowden, J., Knapp, G., Kuss, O., Higgins, J. P. T., Langan, D., & Salanti, G. (2016). Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research Synthesis Methods*, 7(1), 55-79. <https://doi.org/10.1002/jrsm.1164>
- *Walther-Larsen, S., Petersen, T., Friis, S. M., Aagaard, G., Drivenes, B., & Opstrup, P. (2019). Immersive virtual reality for pediatric procedural pain: A randomized clinical trial. *Hospital Pediatrics*, 9(7), 501-507. <https://doi.org/10.1542/hpeds.2018-0249>

- Wan, X., Wang, W., Liu, J., & Tong, T. (2014). Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Medical Research Methodology*, *14*(1), Article 135. <https://doi.org/10.1186/1471-2288-14-135>
- Weisman, S. J., Bernstein, B., & Schechter, N. L. (1998). Consequences of inadequate analgesia during painful procedures in children. *Archives of Pediatrics & Adolescent Medicine*, *152*(2), 147-149. <https://doi.org/10.1001/archpedi.152.2.147>
- Wender, R., Hoffman, H. G., Hunner, H. H., Seibel, E. J., Patterson, D. R., & Sharar, S. R. (2009). Interactivity influences the magnitude of virtual reality analgesia. *Journal of cyber therapy and rehabilitation*, *2*(1), 27-33. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2853033/>
- Wilson-Smith, E. M. (2011). Procedural Pain Management in Neonates, Infants and Children. *Reviews in Pain*, *5*(3), 4-12. <https://doi.org/10.1177/204946371100500303>
- Wilson, D. B. (n.d.). *Practical meta-analysis effect size calculator*. The Campbell Collaboration. <https://campbellcollaboration.org/research-resources/effect-size-calculator.html>
- Wismeijer, A. A., & Vingerhoets, A. J. (2005). The use of virtual reality and audiovisual eyeglass systems as adjunct analgesic techniques: A review of the literature. *Annals of Behavioral Medicine*, *30*(3), 268-278. https://doi.org/10.1207/s15324796abm3003_11
- *Wolitzky, K., Fivush, R., Zimand, E., Hodges, L., & Rothbaum, B. O. (2005). Effectiveness of virtual reality distraction during a painful medical procedure in pediatric oncology patients. *Psychology and Health*, *20*(6), 817-824. <https://doi.org/10.1080/14768320500143339>

Wong-Baker FACES Foundation. (2018). *Wong-Baker FACES® Pain Rating Scale*.

<https://wongbakerfaces.org/>

*Wong, C. L., Li, C. K., Chan, C. W. H., Choi, K. C., Chen, J., Yeung, M. T., & Chan, O. N. (2020). Virtual reality intervention targeting pain and anxiety among pediatric cancer patients undergoing peripheral intravenous cannulation: A randomized controlled trial. *Cancer Nursing*. Advance online publication.

<https://doi.org/10.1097/NCC.0000000000000844>

Yang, S. N., Schlieski, T., Selmins, B., Cooper, S. C., Doherty, R. A., Corriveau, P. J., & Sheedy, J. E. (2012). Stereoscopic viewing and reported perceived immersion and symptoms. *Optometry and Vision Science*, *89*(7), 1068-1080.

<https://doi.org/10.1097/OPX.0b013e31825da430>

Young, K. D. (2005). Pediatric procedural pain. *Annals of Emergency Medicine*, *45*(2), 160-171. <https://doi.org/10.1016/j.annemergmed.2004.09.019>

Öhman, A. (2008). Fear and anxiety: Overlaps and dissociations. In L. F. Barrett, M. Lewis, & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (3rd ed., pp. 709-729). Guilford Press.

*Özkan, T. K., & Polat, F. (2020). The effect of virtual reality and kaleidoscope on pain and anxiety levels during venipuncture in children. *Journal of Perianesthesia Nursing*, *35*(2), 206-211. <https://doi.org/10.1016/j.jopan.2019.08.010>

Appendix A

Characteristics and results of included studies

Kind of medical procedure	Author(s) (year)	Study design	Participants		Measures		Procedural setting	Purpose of VR	Comparison group(s)	Key findings
			<i>n</i>	Age Range <i>M</i> (<i>SD</i>)	Pain	Anxiety				
Dental	Al-Halabi et al. (2018)	RCT parallel	101	6-10 7.40 (-)	W-B Faces	N/A	Pediatric dentistry	Distraction during inferior alveolar nerve block	Control (basic behavior guidance techniques); tablet distraction	No differences in pain scores between the groups
	Al-Khotani et al. (2016)	RCT parallel	56	7-9 8.20 (0.80)	N/A	FIS	Pediatric dentistry	Distraction during restorative treatment	Control (no distraction)	No difference in anxiety scores according to FIS. Lower anxiety scores in AV group according to MVARs
	Al-Nerabieah et al. (2020)	RCT parallel	64	6-10 7.50 (1.30)	W-B Faces	mYPAS-SF	Pediatric dentistry	Distraction in the waiting room before dental procedures	No distraction	Lower pain and anxiety scores in the VR group
	Aminabadi et al. (2012)	RCT cross-over	120	4-6 5.42 (0.73)	W-B Faces	MCDAS(f)	Pediatric dentistry	Distraction during restorative treatment	No VR distraction	Lower pain and anxiety scores during VR
	Asvanund et al. (2015)	RCT cross-over	49	5-8 7.00 (0.87)	FPS-R	N/A	Pediatric dental clinic	Distraction during restorative treatment	Behavior management techniques	Lower pain scores during VR
	Attar & Baghdadi, 2015	RCT cross-over	39	4-8* 6.27 (1.24)	W-B Faces	N/A	Pediatric dental clinic	Distraction during administration of local anesthesia	iPad distraction	Higher pain scores during VR

Appendix A (cont.)

Kind of medical procedure	Author(s) (year)	Study design	Participants		Measures		Procedural setting	Purpose of VR	Comparison group(s)	Key findings
			<i>n</i>	Age Range <i>M (SD)</i>	Pain	Anxiety				
	Atzori, Grotto, et al. (2018)	RCT cross-over	5	7-17 13.20 (2.39)	GRS	N/A	Private dental clinic	Distraction during dental fillings or tooth extraction	No VR	Lower sensory and affective pain scores during VR
	Bagattoni et al. (2018)	RCT cross-over	48	5-10 7.30 (1.50)	FPS-R	N/A	Special needs/ pediatric dentistry	Distraction during restorative treatment	Conventional behavior management (protective eyeglasses)	No difference between the groups on study day I, but lower pain scores for audiovisual group on study day II
	Buldur & Candan (2020)	RCT cross-over	76	7-11 9.02 (1.39)	W-B Faces	FIS	Pediatric dentistry	Distraction during routine dental treatment	Control/placebo (protective eyeglasses)	No difference between the groups on any self-reported measures
	Chaudhary et al. (2020)	RCT parallel	60	-	W-B Faces	N/A	Pediatric dentistry	Distraction during administration of inferior alveolar nerve block (IANB)	Control (behavior management techniques; no distraction)	Lower pain scores in VR group
	El-Sharkawi et al. (2012)	RCT cross-over	48	5-7 (-)	FPS	N/A	Pediatric dentistry	Distraction during administration of inferior alveolar nerve block (IANB)	Control (tell-show-do technique; topical anesthesia)	Lower pain scores during VR/AV-distraction

Appendix A (cont.)

Kind of medical procedure	Author(s) (year)	Study design	Participants		Measures		Procedural setting	Purpose of VR	Comparison group(s)	Key findings
			<i>n</i>	Age Range <i>M (SD)</i>	Pain	Anxiety				
	Fakhruddin et al. (2015)	RCT cross-over	60	4-7 5.24 (1.20)	W-B Faces	MCDAS(f)	Dental hospital	Distraction during pulp therapy	Non-VR distraction (projector display)	Lower pain scores in AV group. AV-glasses during visit I reduced pre-operative anxiety before visit II
	Garrocho-Rangel et al. (2018)	RCT cross-over	36	5-8 6.20 (1.30)	FLACC	N/A	Pediatric dentistry	Distraction during dental treatment	Control (behavior management techniques)	No difference in pain scores between the AV and control groups
	Hoge et al.	RCT parallel	128	4-16 9.31 (2.79)	FPS-R	N/A	Dental clinic	Distraction during restorative treatment	Sunglasses	No difference in pain scores between the AV and control groups
	Isong et al. (2014)	RCT parallel	40	7-17* 9.95 (2.80)	N/A	VARs	Pediatric dental clinic	Distraction during preventative treatment	SOC (no intervention)	Lower anxiety scores in AV group
	Khan et al. (2019)	RCT parallel	100	4-10 6.36 (-)	N/A	FIS	Pediatric/preventive dentistry	Distraction during restorative dental treatment	SOC	No difference between VR and non-VR group
	Koticha et al. (2019)	RCT cross-over	60	7-17 13.20 (2.39)	N/A	VPT	Pediatric/preventive dentistry	Distraction during tooth extraction	No VR	No difference in self-reported anxiety between groups
	Mitrakul et al. (2015)	RCT cross-over	42	7-17* 10.92 (2.64)	FPS-R	N/A	Pediatric dentistry	Distraction during restorative treatment	Behavior management techniques	AV glasses reduced pain during dental treatment

Appendix A (cont.)

Kind of medical procedure	Author(s) (year)	Study design	Participants		Measures		Procedural setting	Purpose of VR	Comparison group(s)	Key findings
			<i>n</i>	Age Range <i>M (SD)</i>	Pain	Anxiety				
	Niharika et al. (2018)	RCT cross-over	40	4-8 7.23 (0.31)	W-B Faces	MCDAS(f)	Pediatric dentistry	Distraction during pulp therapy	No VR	VR reduced pain and anxiety scores
	Nunna et al. (2019)	RCT parallel	70	7-11 8.86 (1.41)	VAS	VCARS ^a	Pediatric dentistry	Distraction during administration of local anesthesia	Counter-stimulation Control (standard behavior techniques);	No difference in pain scores, but higher anxiety scores in VR group AV-glasses reduced anxiety
	Nuvvula et al. (2014)	RCT parallel	90	7-10 8.40 (-)	N/A	MCDAS(f)	Pediatric/preventative dentistry	Distraction during administration of local anesthesia	Music (music distraction + standard behavior techniques)	AV-glasses reduced anxiety
	Shah et al. (2018)	RCT parallel	50	4-7 (-)	N/A	FIS	Pediatric/preventative dentistry	Distraction during restorative treatment	Tell-play-do technique	No difference in anxiety scores between groups
	Shetty et al. (2019)	RCT parallel	120	5-8 6.76 (1.03)	W-B Faces	MCDAS(f)-r	Pediatric dentistry	Distraction during pulp therapy	Conventional behavior management techniques	Virtual reality reduced pain and anxiety
	Sullivan et al. (2000)	W-GD N-R	30	5-7 (-)	N/A	KRS	Dental clinic	Distraction during administration of local anesthesia	No VR	No difference between groups

Appendix A (cont.)

Kind of medical procedure	Author(s) (year)	Study design	Participants		Measures		Procedural setting	Purpose of VR	Comparison group(s)	Key findings
			<i>n</i>	Age Range <i>M (SD)</i>	Pain	Anxiety				
Needle-related procedures	Atzori, Hoffman, et al. (2018)	RCT cross-over	15	7-17 10.92 (2.64)	VAS	N/A	Pediatric onco-hematological setting	Distraction during venipuncture	SOC (No VR; non-medical conversation with nurse)	Lower affective, cognitive, and sensory pain scores during VR
	Aydin & Ozyazicioglu (2019)	RCT parallel	120	9-12 10.40 (1.13)	VAS; W-B Faces	N/A	Pediatric hospital	Distraction during phlebotomy	No intervention	Lower pain scores in VR group
	Caruso et al. (2019)	RCT parallel	220	7-18 13.6 (3.10)	FPS-R	CFS	Pediatric hospital	Distraction during vascular access	Various standard coping methods	No significant differences in pain or anxiety scores
	Chan et al. (2019) (study I, Emergency department)	RCT parallel	123	4-11 8.06 (2.42)	FPS-R	VAT	Emergency department in pediatric hospital	Distraction during venipuncture or cannulation	SOC (various distraction techniques)	VR reduced pain
	Chan et al. (2019) (study II, pathology)	RCT parallel	129	4-11 7.8 (2.33)	FPS-R	VAT	Pathology (outpatient) in pediatric hospital	Distraction during venipuncture	SOC (various distraction techniques)	Lower change from baseline scores in VR group
	Chen et al. (2020)	RCT parallel	136	7-12 9.13 (1.71)	W-B Faces	CFS	Emergency department	Distraction during intravenous injections	Verbal comforting	Lower pain and fear scores in VR group
	del Castillo et al. (2019)	B-GD (Q-E)	34	4-15 9.86 (5.01)	W-B Faces	CFS	Pediatric ward	Distraction during various	No intervention	Lower pain and anxiety scores in VR group

Appendix A (cont.)

Kind of medical procedure	Author(s) (year)	Study design	Participants		Measures		Procedural setting	Purpose of VR	Comparison group(s)	Key findings
			<i>n</i>	Age Range <i>M (SD)</i>	Pain	Anxiety				
	Dumoulin et al. (2019)	RCT parallel	59	8-17 13.37 (2.94)	VAS	VAS	Pediatric emergency department	needle-related procedures Distraction during various needle-related procedures	TV distraction; SOC (Child Life)	Larger reduction in fear scores in VR group compared to TV and SOC. No difference in pain scores
	Gerceker et al. (2018)	RCT parallel	121	7-12 9.44 (1.50)	W-B Faces	N/A	Pediatric phlebotomy unit	Distraction during phlebotomy	Buzzy device; Control (No intervention)	Lower pain scores in the VR and Buzzy groups compared to the control group
	Gerceker et al. (2020)	RCT parallel	136	5-12 (-)	W-B Faces	CAMS-S; CFS	Phlebotomy unit	Distraction during phlebotomy	No distraction	Lower pain and anxiety scores in VR group
	Gerceker et al. (in press)	RCT parallel	42	6-17 11.40 (3.10)	W-B Faces	CAMS-S; CFS	Pediatric hematology-oncology (two hospitals)	Distraction during various needle-related	SOC (information about the procedure)	Lower pain and anxiety scores in VR group
	Gershon et al. (2004)	RCT parallel	59	7-19 12.70 (-)	VAS; CHEOPS	VAS	Outpatient oncology	Distraction during port access	Non-VR distraction (computer); SOC (no distraction)	No difference in pain or anxiety scores according to child and parent reports (VAS). Lower pain scores in VR group compared to no distraction group, according to nurse

Appendix A (cont.)

Kind of medical procedure	Author(s) (year)	Study design	Participants		Measures		Procedural setting	Purpose of VR	Comparison group(s)	Key findings
			<i>n</i>	Age Range <i>M (SD)</i>	Pain	Anxiety				
										reports (VAS), as well as some pain behaviors (CHEOPS)
	Gold et al. (2006)	RCT parallel	20	8-12 10.20 (1.44)	FPS-R	N/A	Pediatric radiology	Distraction during IV placement	SOC (no distraction)	No difference between VR and non-VR groups
	Gold & Mahrer (2018)	RCT parallel	143	10-21 15.43 (3.13)	VAS; CAS	VAS; FAS	Pediatric hospital	Distraction during venipuncture	SOC	VR reduced pain and anxiety scores compared to SOC
	Goldman & Behboudi (2020)	RCT parallel	66	6-16 10.10 (4.28)	FPS-R	VPT	Pediatric emergency department	Distraction during intravenous catheterization	SOC	Lower pain scores in VR group. No difference in anxiety scores
	Inangil et al. (2020)	RCT parallel	120	7-12 9.10 (1.70)	W-B Faces	CFS	Private university hospital	Distraction during phlebotomy	Tablet distraction; control (no intervention)	Lower pain and anxiety scores in VR group
	Litwin et al. (2020)	RCT parallel	48	8-17 12.48 (2.68)	11-point numerical scale	CFS	Pediatric emergency department	Distraction during IV insertion	Tablet distraction	VR reduced pain scores, but no difference in fear scores
	Osmanlliu et al. (in press)	RCT parallel	62	7-17 11.70 (2.99)	VNRS	CFS	Pediatric emergency department	Distraction during vascular access	SOC (pharmacological analgesia; parental presence; non-VR distraction)	Lower anxiety scores in VR group. No difference in pain scores

Appendix A (cont.)

Kind of medical procedure	Author(s) (year)	Study design	Participants		Measures		Procedural setting	Purpose of VR	Comparison group(s)	Key findings
			<i>n</i>	Age Range <i>M (SD)</i>	Pain	Anxiety				
	Piskorz & Czub (2018)	B-GD (N-R)	38	7-12 11.26 (2.92)	VAS	VAS	Pediatric nephrology clinic	Distraction during phlebotomy	No VR	Lower pain and stress scores in VR group
	Piskorz et al. 2020	B-GD (N-R)	57	7-12 12.00 (3.09)	VAS	VAS	Pediatric nephrology clinic	Distraction during phlebotomy	No VR	Lower pain and stress scores in VR group
	Sander Wint et al. (2002)	RCT parallel	30	10-19 13.81 (2.25)	VAS	N/A	Oncology clinic in pediatric hospital	Distraction during lumbar puncture	SOC	Tendency for lower pain scores in VR group
	Schlechter et al. (2020)	RCT parallel	115	4-17 11.00 (4.19)	FPS-R	3-point likert-type scale	Pediatric emergency department	Distraction during intravenous line placement	SOC (in absence of child life specialists)	Changes in anxiety and pain scores were similar across groups
	Wolitzky et al. (2005)	RCT parallel	20	7-14 10.50 (2.33)	CHEOPS	N/A	Pediatric hospital	Distraction during port access	No VR	VR reduced pain and distress
	Wong et al. (2020)	RCT parallel	108	6-17 10.40 (3.60)	FPS-R	CSAS-C (short version)	Pediatric cancer center in hospital	Distraction during peripheral intravenous cannulation	SOC (verbal explanation)	Children in the VR group reported significantly less procedural pain and
	Özkan & Polat, 2020	RCT parallel	135	4-10 9.28 (0.93)	VAS	CFS	Hospital	Distraction during venipuncture	Control; kaleidoscope distraction	Pain scores were lower in the VR and kaleidoscope group compared to control group

Appendix A (cont.)

Kind of medical procedure	Author(s) (year)	Study design	Participants		Measures		Procedural setting	Purpose of VR	Comparison group(s)	Key findings
			<i>n</i>	Age Range <i>M (SD)</i>	Pain	Anxiety				
Pre-operative	Eijlers, Dierckx, et al. (2019)	RCT parallel	191	4-12 8.00 (3.62)	FPS-R	mYPAS	Pediatric hospital	Preparation for elective surgery/ anesthesia	SOC (recommendation to watch informative film about general anesthesia at home)	anxiety, compared with the control group No differences between groups in self-reported anxiety or pain
	Hashimoto et al. (2020)	RCT parallel	58	4-12* 5.18 (1.24)	N/A	mYPAS	Hospital	Preparation for elective surgery/general anesthesia	Cartoons viewed on a portable media player	Lower pre-operative anxiety scores in the AV-glasses group
	Kerimoglu et al. (2013)	RCT parallel	96	4-9 6.24 (2.19)	N/A	mYPAS	Hospital	Distraction during induction of general anesthesia (inhaled) before surgery	Pharmacological intervention only (midazolam)	No difference in anxiety scores between the groups
	Jung et al. (2020)	RCT parallel	70	5-12 8.00 (2.3)	N/A	mYPAS	Pediatric hospital	Distraction during induction of general anesthesia (inhaled) before surgery	SOC (no audiovisual distraction)	Lower anxiety scores in the VR group, compared to the control group

Appendix A (cont.)

Kind of medical procedure	Author(s) (year)	Study design	Participants		Measures		Procedural setting	Purpose of VR	Comparison group(s)	Key findings
			<i>n</i>	Age Range <i>M (SD)</i>	Pain	Anxiety				
	Ryu et al. (2017)	RCT parallel	69	4-10 6.04 (1.99)	N/A	mYPAS	Hospital	Preparation for elective surgery under general anesthesia (inhaled or intravenous)	Conventional education about preoperative process	Lower anxiety scores in the VR group
	Ryu et al. (2018)	RCT parallel	80	4-10 6.18 (1.95)	N/A	mYPAS,	Hospital	Preparation for elective surgery under general anesthesia (inhaled)	Conventional education about preoperative process	Lower anxiety scores in the VR group
	Ryu et al. (2019)	RCT parallel	69	4-10 6.40 (1.74)	N/A	mYPAS	Hospital	Preparation for elective surgery under general anesthesia	Conventional education about the preoperative process	Lower anxiety scores in the VR group
	Walther-Larsen et al. (2019)	RCT parallel	59	7-16 10.48 (2.51)	VAS	N/A	Anesthetic department in hospital	Distraction during IV cannulation for anesthesia before elective surgery	SOC (topical numbing cream, positioning, and smartphone distraction)	No difference in pain scores
Wound care	Chan et al. (2007)	RCT cross-over	8	- 6.54 (2.27)	FPS	N/A	Burn facility	Distraction during burn wound care	SOC	VR reduced pain during and after the procedure

Appendix A (cont.)

Kind of medical procedure	Author(s) (year)	Study design	Participants		Measures		Procedural setting	Purpose of VR	Comparison group(s)	Key findings
			<i>n</i>	Age Range <i>M (SD)</i>	Pain	Anxiety				
	Das et al. (2005)	RCT cross-over	7	5-18 10.00 (3.90)	Faces Scale	N/A	Women's and Children Hospital	Distraction during burn wound dressing change	No VR (pharmacological analgesia only)	VR reduced pain score
	Hoffman et al. (2019)	RCT cross-over	48	6-17 12.00 (-)	GRS	N/A	Intensive care unit	Distraction during burn wound care	No VR (pharmacological analgesia only)	VR reduced pain intensity scores
	Hua et al. (2015)	RCT parallel	65	4-16 8.72 (3.36)	W-B Faces	N/A	Pediatric center in hospital	Distraction during chronic wound care	Standard distraction methods (e.g.,	VR distraction reduced pain and anxiety scores during dressing changes
	Jeffs et al. (2014)	RCT parallel	28	10-17 13.50 (2.30)	WGRS (APPT)	N/A	Outpatient burn clinic of children's hospital	Distraction during burn wound care	Passive distraction (television); SOC	Higher pain scores in VR group
	Khadra et al. (2020)	RCT cross-over	38	0.5-7 1.83 (1.33)	FLACC; NRS-obs	N/A	Surgical-trauma burn unit in pediatric hospital	Distraction during hydrotherapy for burn wounds	Standard pharmacological analgesia	VR reduced pain scores according to FLACC, but not the NRS (nurse-rated)
	Kipping et al. (2012)	RCT parallel	41	11-17 13.08 (1.6)	VAS	N/A	Two burn units	Distraction during burn wound care	SOC (access to non-VR distraction)	No difference between VR and control group, except for nurses' ratings
	Schmitt et al. (2011)	RCT cross-over	54	6-19 12 (3.9)	GRS	N/A	Burn center	Distraction during post-burn physical therapy	Control (standard pharmacological analgesia, no VR)	Patients reported significantly lower pain intensity when using VR

Note. High scores indicate high levels of pain/anxiety in all measures. W-GD = Within-groups design; N-R = Non-randomized; B-GD = Between-groups design; Q-E = Quasi-experimental design; W-B Faces = Wong Baker Faces Scale; FIS = Facial Image Scale; mYPAS-SF = Modified Yale Preoperative Anxiety Scale - Short Form; MCDAS(f) = Faces version of the Modified Child Dental Anxiety Scale; FPS-R = Faces Pain Scale-Revised; GRS = Graphic Rating Scale; FPS = Faces Pain Scale; FLACC = Face, Legs, Activity, Cry, Consolability Scale; VARS = Venham Anxiety Rating Scale; VPT = Venham Picture Test; VAS = Visual Analogue Scale; CAS = Colored Analogue Scale; FAS = Faces Affective Scale; VCARS = Venham Clinical Anxiety Rating Scale; MCDAS(f)-r = Revised Faces version of the Modified Child Dental Anxiety scale; KRS = Koppitz Rating Scale; CFS = Child Fear Scale; VAT = Visual Analogue Thermometer; CAMS-S = Children's Anxiety Meter Scale; CHEOPS = Children's Hospital of Eastern Ontario Pain Scale; VNRS = Verbal Numerical Rating Scale; CSAS-C= The short form of the Chinese version of the State Anxiety Scale for Children; mYPAS = modified Yale Preoperative Anxiety Scale; WGRS (APPT) = Word Graphic Rating Scale from The Adolescent Pediatric Pain Tool; NRS-obs = Observational Numerical Rating Scale. SOC = Standard of care

* reflects the study inclusion criteria and may differ from the actual age range of participants

Appendix B

Characteristics of VR Systems and Interventions

Kind of medical procedure	Author(s) (year)	Name of device(s)	Screen			Stereoscopy/3D	Head tracking	Full visual occlusion	Non-visual stimuli	Interactivity	Media description
			Resolution	Field of view	Refresh rate						
Dental	Al-Halabi et al. (2018)	VR Box + Asus Zenfone 2 Deluxe	1920 x 1080	62°* (Stereoscopic field of view)	-	Yes	Yes	Yes	Audio	No	Cartoon: Self-selected
	Al-Khotani et al. (2016)	Merlin i-theatre	-	-	-	No	-	No	Audio	No	Cartoon: Self-selected
	Al-Nerabieah et al. (2020)	Samsung Gear VR + Samsung S9+	2960 x 1440	≈ 96°*	60 Hz	Yes	-	Yes	Audio	No	Cartoon
	Aminabadi et al. (2012)	Ilixco i-glasses 920 HR Shenzhen	640 x 480	35° (diagonal)	-	-	No	Yes	Audio	No	Cartoon: Tom & Jerry
	Asvanund et al. (2015)	Longway Vision Technology Coolvision 3	-	-	-	-	No	-	Audio	No	Cartoons: Self-selected
	Attar & Baghdadi, 2015	Koolertron AV glasses JVE-3107G	-	-	-	-	-	-	-	-	(Unknown)
	Atzori, Grotto, et al. (2018)	Oculus Rift DK2 or Oculus Rift CV1	960 x 1080 (DK2) or 1080 x 1200 (CV1)	100° (DK2, unspecified)	75 Hz (DK2)	Yes	No	Yes	Audio	Yes	Game: SnowWorld

Appendix B (cont.)

Kind of medical procedure	Author(s) (year)	Name of device(s)	Screen			Stereoscopy/3D	Head tracking	Full visual occlusion	Non-visual stimuli	Interactivity	Media description
			Resolution	Field of view	Refresh rate						
			(per eye)								
	Bagattoni et al. (2018)	Unidexx video eyeglasses	-	-	-	-	-	-	Audio	No	Cartoon
	Buldur & Candan (2020)	PlayStation 4 VR	1920 x 1080	≈ 100° (unspecified)	120 or 90 Hz	-	-	Yes	Audio	No	Cartoon: Self-selected
	Chaudhary et al. (2020)	Virtual Reality Box + undisclosed smartphone	-	-	-	-	-	Yes	Audio	Both interactive and non-interactive	Video games (n=20), cartoon (n=20)
	El-Sharkawi et al. (2012)	Eyewear Cinema IMV260	320 x 240	26° (diagonal)	-	No	No	No	Audio	No	Cartoon: Self-selected
	Fakhruddin et al. (2015)	Vuzix Wrap 310XL	428 x 240 (per eye)	26° (diagonal)	60 Hz	-	No	No	Audio	No	Cartoon: Self-selected
	Garrocho-Rangel et al. (2018)	Chinavision Virtual Private Theater Video Glasses	-	-	-	-	-	-	Audio	No	Cartoon: Self-selected
	Hoge et al. (2012)	(Unknown)	-	-	-	No	No	Yes	Audio	No	Movies: Self-selected
	Isong et al. (2014)	Vuzix (Unknown model)	-	-	-	No	-	-	-	No	Movie

Appendix B (cont.)

Kind of medical procedure	Author(s) (year)	Name of device(s)	Screen			Stereoscopy/3D	Head tracking	Full visual occlusion	Non-visual stimuli	Interactivity	Media description
			Resolution	Field of view	Refresh rate						
		Visual									
	Khan et al. (2019)	Reality Glasses 3D Box + Apple iPhone 7	1334 x 750	-	60 Hz	-	No	Yes	Audio	No	Cartoon: Self-selected
	Koticha et al. (2019)	BlackBug Virtual Reality Glasses 3D + Apple iPhone 6 Shenzhen	1334 x 750	90° (horizontal)	60 Hz	No	Yes	Yes	Audio	No	Cartoon
	Mitrakul et al. (2015)	Longway Vision Technology Coolvision 3	-	-	-	-	No	No	Audio	No	Cartoon: Self-selected
	Niharika et al. (2018)	Google VR Box and Anti Tank Virtual Reality 3D Glasses + undisclosed mobile device	-	-	-	-	-	Yes	Audio	No	Cartoon
	Nunna et al. (2019)	ANTVR Phone Glass T2 (model: PA15LF53A) + Lenovo Vibe K4 Note	1920 x 1080	100°* (unspecified)	-	Yes	-	Yes	Audio	No	Cartoon: Self-selected
	Nuvvula et al. (2014)	Vuzix Wrap 920	640 × 480 (unspecified)	-	-	Yes	No	No	Audio	No	Movie

Appendix B (cont.)

Kind of medical procedure	Author(s) (year)	Name of device(s)	Screen			Stereoscopy/3D	Head tracking	Full visual occlusion	Non-visual stimuli	Interactivity	Media description
			Resolution	Field of view	Refresh rate						
Needle-related	Shah et al. (2018)	(Unknown)	-	-	-	-	-	No	Audio	No	Cartoons
	Shetty et al. (2019)	Ilixco i-glasses 920 HR	640 x 480	35° (diagonal)	-	No	No	Yes	Audio	No	Cartoon: Self-selected
	Sullivan et al. (2000)	(Unknown)	-	-	-	Yes	-	-	Audio	No	Movie
	Atzori, Hoffman et al. (2018)	Sony Personal 3D Viewer HMZ T-2 Fiit VR 3D glasses	1280 x 720	45° (unspecified)	-	-	No	Yes	Audio	Yes	Game: SnowWorld
	Aydin & Özyazicioglu (2019)	+ General Mobile Discovery Air Samsung Gear VR	720 x 1280	120°* (unspecified)	60 Hz	Yes	Yes	Yes	No	Yes	Simulation: Submarine journey
	Caruso et al. (2019)	+ Samsung S7 or S8	2560 × 1440 or 2220 x 1080	≈ 96°* (unspecified)	60 Hz	Yes	Yes	Yes	Audio	Yes	Various VR experiences
	Chan et al. (2019) (study I and II)	+ Google Daydream Google Pixel XL smartphone	1440 x 2560	100°, reduced to 45° during needle insertion (unspecified)	Average FPS: 45	Yes	Yes	Yes	Audio	Yes	Simulation: Interactive underwater environment

Appendix B (cont.)

Kind of medical procedure	Author(s) (year)	Name of device(s)	Screen			Stereoscopy/3D	Head tracking	Full visual occlusion	Non-visual stimuli	Interactivity	Media description
			Resolution	Field of view	Refresh rate						
	Chen (2020)	BoboVR Z4 + iPhone 6s+	1920 x 1080	120°* (unspecified)	60 Hz	Yes	Yes	Yes	Audio	No	Simulations
	del Castillo et al. (2019)	Woxter Neo VR1 glasses + undisclosed smartphone	-	-	-	-	-	Yes	Audio	No	Video: Self-selected
	Dumoulin et al. (2019)	eMagin Z800 3DVISOR	800 × 600	39.5° (diagonal)	60 Hz	Yes	Yes	No	No	Yes	Game: Shooting flies
	Gerceker et al. (2018)	Samsung Gear + Galaxy S5 Note	2560 x 1440	≈ 96°* (unspecified)	60 Hz	Yes	Yes	Yes	Audio	No	Cartoon: Self-selected
	Gerceker et al. (2020)	Samsung Gear + Samsung Galaxy 5 Note	2560 x 1440	≈ 96°* (unspecified)	60 Hz	Yes	Yes	Yes	Audio	Both interactive and non-interactive	Simulations: Underwater adventure (<i>n</i> =45), rollercoaster (<i>n</i> =45)
	Gerceker et al. (in press)	Samsung Gear + Samsung Galaxy S7 Edge	2560 x 1440	≈ 96°* (unspecified)	60 Hz	Yes	Yes	Yes	Audio	Both interactive and non-interactive	Simulation: Self-selected
	Gershon et al. (2004)	(Unknown)	-	-	-	-	-	-	Audio	Yes	Simulation: Gorilla habitat
	Gold et al. (2006)	5DT HMD 800 (unspecified edition)	800 x 600	-	-	Yes	Yes	Yes	Audio; tactile	Yes	Game: Street Luge

Appendix B (cont.)

Kind of medical procedure	Author(s) (year)	Name of device(s)	Screen			Stereoscopy/3D	Head tracking	Full visual occlusion	Non-visual stimuli	Interactivity	Media description
			Resolution	Field of view	Refresh rate						
		+ Intersense Inertia Cube 2 Samsung Gear + Samsung Galaxy S6 or Merge VR + Google Pixel VOX+ Z3 3D + Asus Zenfone 2 ZE551ML Samsung Gear + undisclosed smartphone		-	-						
	Gold & Mahrer (2018)		2560 x 1440	≈ 96°* (Samsung Gear) (unspecified)	60 Hz (Samsung Galaxy S6)	Yes	Yes	Yes	Audio	Yes	Game: Bear Blast
	Goldman & Behboudi (2020)		1920 x 1080	80°* (unspecified)	-	Yes	Yes	Yes	Audio	No	Simulation: Roller Coaster
	Inangil et al. (2020)		-	≈ 96°* (unspecified)	-	Yes	Yes	Yes	Audio	No	Cartoon: Self-selected
	Litwin et al. (2020)	Samsung Gear + Samsung Galaxy S6	2560 x 1440	≈ 96°* (unspecified)	60 Hz	Yes	Yes	Yes	Audio	Yes	Simulation: Underwater environment
	Osmanlliu et al. (in press)	Oculus Rift	1080 x 1200	-	90 Hz	Yes	Yes	Yes	Audio	Yes	Game: Dreamland
	Piskorz & Czub (2018)	Oculus Rift DK2	960 x 1080 (per eye)	100° (unspecified)	75 Hz	Yes	Yes	Yes	Audio	Yes	Game: Multiple Object Tracking

Appendix B (cont.)

Kind of medical procedure	Author(s) (year)	Name of device(s)	Screen			Stereoscopy/3D	Head tracking	Full visual occlusion	Non-visual stimuli	Interactivity	Media description
			Resolution	Field of view	Refresh rate						
	Piskorz et al. (2020)	Samsung Gear + Galaxy S7	1280 + 1440 (per eye)	≈ 96°* (unspecified)	60 Hz	Yes	Yes	Yes	Audio	Both interactive and non-interactive	Game (n=19): Multiple Object Tracking task. Movie (n=17): Images from the game
	Sander Wint et al. (2002)	(Unknown)	-	-	-	Yes	-	-	Audio	No	Video
	Schlechter et al. (2020)	Merge 360° + iPod Touch 6th generation	1136 x 640	≈ 96°* (unspecified)	-	Yes	Yes	Yes	Audio	Yes	Game: Narwhal swimming through hoops
	Walther-Larsen et al. (2019)	Samsung Gear + Samsung Galaxy S6 + controller	2560 x 1440	≈ 96°* (unspecified)	60 Hz	Yes	No	Yes	Audio	Yes	Game: Seagull Splash
	Wolitzky et al. (2005)	(Unknown)	-	-	-	-	Yes	-	Audio	Yes	Simulation: Gorilla habitat
	Wong et al. (2020)	Google Cardboard + Xiaomi Mi 9 VR Box	2340 x 1080	-	60 Hz	Yes	Yes	Yes	Audio	No	Cartoon: Self-selected
	Özkan & Polat (2020)	+ iPhone 6+ HTC Vive (unknown model)	1920 x 1080	-	60 Hz	Yes	Yes	Yes	Audio	No	Cartoon: Self-selected
Pre-operative	Eijlers, Dierckx, et al. (2019)		2160 x 1200	110° (unspecified)	90 Hz	Yes	Yes	Yes	Audio	Yes	Interactive simulation of operating theatre

Appendix B (cont.)

Kind of medical procedure	Author(s) (year)	Name of device(s)	Screen			Stereoscopy/3D	Head tracking	Full visual occlusion	Non-visual stimuli	Interactivity	Media description
			Resolution	Field of view	Refresh rate						
	Hashimoto et al. (2020)	IVS Video glasses	480 x 240	-	-	-	No	No	Audio	No	Movie: Self-selected
	Jung et al. (2020)	Customized Samsung Gear + Samsung Galaxy S8	2960 x 1440	≈ 96°* (unspecified)	60 Hz	Yes	Yes	Yes	Audio	Yes	Game
	Kerimoglu et al. (2013)	Vuzix video eyeglasses	-	-	-	-	No	No	Audio	No	TV shows: Self-selected
	Ryu et al. (2017)	Samsung Gear + Samsung Galaxy S6	2560 x 1440	≈ 96°* (unspecified)	60 Hz	Yes	Yes	Yes	Audio	No	Virtual tour of the operating theatre
	Ryu et al. (2018)	Oculus Rift + Leap Motion controller	1080 x 1200	-	90 Hz	Yes	Yes	Yes	Audio	Yes	Gamified virtual tour of the operating theatre
	Ryu et al. (2019)	Samsung Gear + Samsung Galaxy S6 i-glasses (unspecified model)	2560 x 1440	≈ 96°* (unspecified)	60 Hz	Yes	Yes	Yes	Audio	No	Virtual tour of the operating theatre
Wound care	Chan et al. (2007)	i-glasses (unspecified model)	-	35° (unspecified)	-	-	-	-	Audio	Yes	Game: Shoot foxes with ice cream
	Das et al. (2005)	IOGlasses Head Mount Display	800 x 600	-	-	-	Yes	-	-	Yes	Game: Shoot monsters

Appendix B (cont.)

Kind of medical procedure	Author(s) (year)	Name of device(s)	Screen			Stereoscopy/3D	Head tracking	Full visual occlusion	Non-visual stimuli	Interactivity	Media description
			Resolution	Field of view	Refresh rate						
	Hoffman et al. (2019)	NVIS MX90 (mounted on a custom arm)	1280 × 1024 (per eye)	90° (diagonal, per eye)	-	-	No	No	Audio	Yes	Game: SnowWorld
	Hua et al. (2015)	eMagin Z800 3DVISOR	800 x 600	39.5° (diagonal)	60 Hz	Yes	Yes	No	Audio	Yes	Game: Ice Age 2: The Meltdown
	Jeffs et al. (2014)	Kaiser Optics SR80a (mounted on a tripod)	1280 x 1024	80° (unspecified)	-	Yes	No	Yes	Audio	Yes	Game: SnowWorld
	Khadra et al. (2020)	Projector-based hybrid VR dome environment	1920 x 1080 (projector)	-	-	No	No	No	Audio	Yes	Game: Bubbles
	Kipping et al. (2012)	eMagin Z800 3DVisor	800 x 60	39.5° (diagonal)	60 Hz	Yes	Yes	No	Audio	Yes	Games: Chicken Little or Need for Speed (depending on patient's age)
	Schmitt et al. (2011)	nVisor SX or VR-1280 or ProView XL 50 or ProView SR 80	Minimum 1024 x 1280 (per eye)	Minimum 50° (diagonal)	-	Yes	Yes	Yes	Audio	Yes	Game: SnowWorld

Appendix B (cont.)

Note. Specifications for two studies using the same VR equipment may differ due to study authors' configurations (e.g., disabling head tracking function or displaying two-dimensional graphics on a headset that is capable of stereoscopy). Manufacturer information: AsusTek Computer Inc. (Beitou District, Taipei, Taiwan); Merlin Soft Magic Systems LLC (Al Ain Center, Dubai, UAE); Samsung Electronics (Suwon, Korea); Ilixco Inc. (Menlo Park, CA, USA); Shenzhen Longway Vision Technology LLT (Shenzhen, China); Oculus VR LLC (Menlo Park, CA, USA); Sony Corporation (Sony City, Minato, Tokyo, Japan); Estar Display Tech Co, Ltd (Shenzhen, China); Vuzix Corporation (Rochester, NY, USA); Chinavision (Kowloon, Hong Kong, China); Apple Inc. (Cupertino, CA, USA); ANTVR Technology Co. LTD (Beijing, China); Lenovo Group Limited (Beijing, China); General Mobile Corporation (Songsshan District, Taipei, Taiwan); Google LLC (Mountain View, CA, USA); Quatrotec Electrónica, S.L. (Leganés, Madrid, Spain); eMagin Corporation (Bellevue, DC, United States); 5DT (Orlando, FL, USA); Intersense/Thales Defense & Security, Inc. (MA, USA); Xiaomi Corporation (Haidian District, Beijing, China); HTC Corporation (Xindian, New Taipei, Taiwan); NVIS Inc. (Reston VA); Kaiser Electro-Optics (Carlsbad, CA, USA); Panasonic Corporation (Kadoma, Osaka, Japan); Virtual Research Systems (Aptos, CA, USA).

- indicates that the information is not available

* Field of view may vary slightly depending on the size of the smartphone screen

Appendix C

Summary Statistics of the Studies Included in the Meta-Analysis

Author(s), year	Pain						Anxiety					
	VR			Non-VR			VR			Non-VR		
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>
Al-Halabi et al., 2018	-0.75	0.98	33	-0.83	1.07	68						
Al-Khotani et al., 2016							-1.32	0.67	28	-1.46	0.69	28
Al-Nerabieah et al., 2020	-1.56	1.16	32	-4.09	0.85	32	-45.89	12.96	32	-78.96	8.24	32
Aminabadi et al., 2013	-1.89	0.65	58	-3.05	0.60	59	-12.58	1.01	58	-18.25	1.02	59
Asvanund et al., 2015	-2.23	2.29	21	-3.04	3.08	23						
Aydin & Ozyazicioglu, 2019	-3.07	2.86	60	-3.23	3.05	60						
Bagattoni et al., 2018	-3.60	2.40	24	-2.70	2.20	24						
Buldur & Candan, 2020	-3.52	1.65	38	-4.11	2.13	38	-2.44	0.92	38	-2.86	1.15	38
Caruso et al., 2019	-0.72	1.32	106	-0.89	1.68	114	-0.31	0.66	106	-0.38	0.72	114
Chan et al., 2019 (study I)	-3.08	3.29	64	-4.20	3.59	59	-3.66	3.42	64	-4.73	3.43	59
Chan et al., 2019 (study II)	-2.41	3.11	63	-4.21	3.76	66	-3.16	3.16	63	-4.79	3.72	66
Chaudhary et al., 2020	-3.40	2.87	40	-6.40	2.30	20						
Chen et al., 2020	-3.35	2.38	68	-4.35	2.95	68	-1.32	1.19	68	-1.78	1.40	68
del Castillo et al., 2019	-1.18	1.22	16	-3.64	0.80	18	-0.50	0.81	16	-3.77	1.41	18

Appendix C (cont.)

Author(s), year	Pain						Anxiety					
	VR			Non-VR			VR			Non-VR		
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>
Dumoulin et al., 2019	-21.75	20.92	20	-31.55	30.08	39	-19.75	21.18	20	-33.08	35.05	39
Eijlers, Dierckx, et al., 2019	-2.00	2.58	94	-1.85	1.93	97	-45.80	21.75	94	-45.52	22.49	97
Fakhruddin et al., 2015	-1.59	0.56	30	-4.01	0.60	30	-7.87	2.06	30	-11.51	2.10	30
Gerceker et al., 2018	-1.50	0.20	40	-3.53	1.59	81						
Gerceker et al., 2020	-1.10	1.89	90	-4.10	3.50	46	-0.80	2.13	90	-6.30	3.60	46
Gerceker et al., in press	-2.40	1.80	21	-5.30	1.80	21	-2.90	2.00	21	-5.40	2.00	21
Gold & Mahrer, 2018	-1.31	1.59	70	-1.93	2.22	73	-1.90	2.22	70	-2.48	2.07	73
Gold et al., 2006	-1.80	2.40	10	-2.40	1.84	10						
Goldman & Behboudi, 2020	-2.00	3.09	35	-4.00	3.11	31						
Hashimoto et al., 2020							-23.91	1.33	29	-33.62	16.3	29
Hua et al., 2015	-2.42	1.85	33	-4.19	2.12	32						
Inangil et al., 2020	-1.30	2.15	40	-4.75	3.53	80	-0.65	0.92	40	-2.40	1.45	80
Isong et al., 2014							-1.7	1.8	20	-2.3	1.6	20
Jeffs et al., 2014	-58.25	31.75	8	-33.60	29.95	20						
Jung et al., 2020							-28.30	7.75	33	-45.00	18.05	37
Khadra et al., 2020	-2.37	2.65	18	-2.5	2.62	17						
Litwin et al., 2020	-2.60	2.10	24	-3.80	2.10	24	-1.40	1.10	23	-1.00	1.00	23

Appendix C (cont.)

Author(s), year	Pain						Anxiety					
	VR			Non-VR			VR			Non-VR		
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>
Mitrakul et al., 2015	-1.90	2.93	21	-1.62	2.94	21						
Niharika et al., 2018	-2.56	1.65	18	-5.44	2.90	18	-14.72	0.84	18	-19.56	0.88	18
Nunna et al., 2019	-2.94	1.77	35	-3.06	2.31	35	-0.57	0.61	35	-0.80	0.68	35
Nuvvula et al., 2014							-8.30	2.50	30	-17.50	6.80	60
Osmanlliu et al., in press	-1.35	2.32	31	-2.00	2.46	31	-0.35	0.92	31	-0.55	1.03	31
Piskorz & Czub, 2018	-15.16	20.51	19	-37.05	30.66	19	-11.16	18.58	19	-41.89	40.89	19
Piskorz et al., 2020	-1.11	1.40	36	-3.34	2.21	21	-1.20	1.89	36	-3.08	2.57	21
Ryu et al., 2017							-30.92	11.30	34	-47.50	27.05	35
Ryu et al., 2018							-29.51	10.37	34	-43.14	15.46	35
Ryu et al., 2019							-37.45	21.24	41	-47.84	23.05	39
Sander Wint et al., 2002	-12.50	13.36	17	-16.56	17.63	13						
Schlechter et al., 2020	-5.02	3.38	58	-5.08	3.51	57	-1.68	0.83	58	-1.95	0.83	57
Schmitt et al., 2011	-36.27	25.44	22	-54.52	28.42	29						
Shetty et al., 2019	-2.42	1.47	60	-5.60	1.22	60	-11.28	3.51	58	-16.47	3.48	60
Walther-Larsen et al., 2019	-22.50	16.50	28	-17.50	14.90	31						
Wolitzky et al., 2005	-4.90	0.99	10	-8.30	2.41	10						
Wong et al., 2020	-1.94	1.73	54	-4.00	3.53	54	-14.81	2.93	54	-17.83	4.69	54

Appendix C (cont.)

Author(s), year	Pain						Anxiety					
	VR			Non-VR			VR			Non-VR		
	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>
Özkan & Polat, 2020	-1.97	1.20	46	-4.81	2.81	89	-0.43	0.50	46	-1.83	1.37	89

Note: Effect size for Shah & Bhatia (2018) was calculated from the *t*-statistic: $t(48) = 1.114, p = .271$

Appendix D

Figure D1: RoB 2 Judgements per Domain (Pain Studies, Part 1/2)

Study	Risk of bias domains					Overall
	D1	D2	D3	D4	D5	
Al-Halabi et al. (2018)	-	-	+	X	-	X
Al-Nerabieah et al. (2020)	-	-	+	X	+	X
Aminabadi et al. (2012)	-	-	+	X	-	X
Asvanund et al. (2015)	-	X	X	X	-	X
Attar & Baghdadi (2015)	-	X	-	X	-	X
Atzori, Grotto, et al. (2018)	-	-	+	X	-	X
Atzori, Hoffman, et al. (2018)	-	-	-	X	-	X
Aydin & Ozyazicioglu (2019)	X	X	+	X	-	X
Bagattoni et al. (2018)	-	X	+	X	-	X
Buldur & Candan (2020)	X	X	+	X	-	X
Caruso et al. (2019)	+	-	X	X	-	X
Chan et al. (2007)	X	-	+	X	-	X
Chan et al. (2019, study I)	+	-	+	X	+	X
Chan et al. (2019, study II)	+	-	+	X	+	X
Chaudhary et al. (2020)	-	X	X	X	-	X
Chen et al. (2020)	+	-	+	X	-	X
Das et al. (2005)	-	-	+	X	-	X
Dumoulin et al. (2015)	+	-	+	X	-	X
Eijlers et al. (2019)	-	-	+	X	+	X
El-Sharkawi et al. (2012)	X	X	X	X	-	X
Fakhrudin et al. (2015)	-	X	X	X	-	X
Garraco-Rangel et al. (2018)	-	-	X	X	-	X
Gerceker et al. (2018)	-	-	X	X	-	X
Gerceker et al. (2020)	-	-	+	X	-	X

Domains:
 D1: Bias arising from the randomization process
 D2: Bias due to deviations from intended intervention.
 D3: Bias due to missing outcome data.
 D4: Bias in measurement of the outcome.
 D5: Bias in selection of the reported result.




Judgement
 High
 Some concerns
 Low

Figure D2: RoB 2 Judgements per Domain (Pain Studies, Part 2/2)

Study	Risk of bias domains					Overall
	D1	D2	D3	D4	D5	
Gerceker et al. (in press)	-	-	+	X	-	X
Gershon et al. (2004)	-	X	X	X	-	X
Gold & Mahrer (2018)	-	-	+	X	-	X
Gold et al. (2006)	-	X	+	X	-	X
Goldman & Behboudi (2020)	+	-	+	X	-	X
Hoffman et al. (2019)	-	-	+	X	-	X
Hoge et al. (2012)	-	X	X	X	-	X
Hua et al. (2015)	-	X	X	X	-	X
Inangil et al. (2020)	+	-	+	X	-	X
Jeffs et al. (2014)	+	-	+	X	-	X
Khadra et al. (2020)	+	-	+	X	-	X
Kipping et al. (2012)	+	-	+	X	-	X
Koticha et al. (2019)	-	X	-	X	-	X
Litwin et al. (2020)	+	-	+	X	-	X
Mitrakul et al. (2015)	-	X	+	X	X	X
Niharika et al. (2018)	-	-	+	X	-	X
Nunna et al. (2019)	-	-	+	X	-	X
Osmanliu et al. (in press)	+	-	+	X	-	X
Sander Wint et al. (2002)	-	X	X	X	-	X
Schlechter et al. (2020)	-	-	+	X	-	X
Schmitt et al. (2011)	-	-	+	X	-	X
Shetty et al. (2019)	-	X	+	X	-	X
Walther-Larsen et al. (2020)	-	-	X	X	-	X
Wolitzky et al. (2005)	-	X	+	X	X	X
Wong et al. (2020)	+	-	+	X	-	X
Özkan & Polat (2020)	+	-	X	X	-	X

Domains:
 D1: Bias arising from the randomization process
 D2: Bias due to deviations from intended intervention.
 D3: Bias due to missing outcome data.
 D4: Bias in measurement of the outcome.
 D5: Bias in selection of the reported result.




Judgement
 High
 Some concerns
 Low

Figure D3: ROBINS-I Judgements per Domain (Pain Studies)

Study	Risk of bias domains							Overall
	D1	D2	D3	D4	D5	D6	D7	
del Castillo et al. (2019)								
Piskorz & Czub (2018)								
Piskorz et al. (2020)								

Domains:
 D1: Bias due to confounding.
 D2: Bias due to selection of participants.
 D3: Bias in classification of interventions.
 D4: Bias due to deviations from intended interventions.
 D5: Bias due to missing data.
 D6: Bias in measurement of outcomes.
 D7: Bias in selection of the reported result.

Judgement
 Serious
 Moderate
 Low
 No information

Appendix E

Figure E1: RoB 2 Judgements per Domain (Anxiety Studies, Part 1/2)

Study	Risk of bias domains					Overall
	D1	D2	D3	D4	D5	
Al-Khotani et al. (2016)	⊗	-	+	⊗	-	⊗
Al-Nerabieah et al. (2020)	-	-	+	+	+	-
Aminabadi et al. (2013)	-	-	+	⊗	-	⊗
Attar & Baghdadi (2015)	-	⊗	-	⊗	-	⊗
Buldur & Candan (2020)	⊗	-	+	⊗	-	⊗
Caruso et al. (2019)	+	-	⊗	⊗	-	⊗
Chan et al. (2019, study I)	+	-	+	⊗	+	⊗
Chan et al. (2019, study II)	+	-	+	⊗	+	⊗
Chen et al. (2020)	+	-	+	⊗	-	⊗
Dumoulin et al. (2015)	+	-	+	⊗	-	⊗
Eijlers et al. (2019)	-	-	+	+	+	-
Fakhrudin et al. (2015)	-	⊗	⊗	⊗	-	⊗
Gerceker et al. (2020)	-	-	+	⊗	-	⊗
Gerceker et al. (in press)	-	-	+	⊗	-	⊗
Gershon et al. (2004)	-	⊗	⊗	⊗	-	⊗
Gold & Mahrer (2018)	-	-	+	⊗	-	⊗
Goldman & Behboudi (2020)	+	-	+	⊗	-	⊗
Hashimoto et al. (2020)	-	-	+	⊗	-	⊗

Domains:

- D1: Bias arising from the randomization process.
- D2: Bias due to deviations from intended intervention.
- D3: Bias due to missing outcome data.
- D4: Bias in measurement of the outcome.
- D5: Bias in selection of the reported result.

Judgement

- ⊗ High
- Some concerns
- +

Figure E2: RoB 2 Judgements per Domain (Anxiety Studies, Part 2/2)

Study	Risk of bias domains					Overall
	D1	D2	D3	D4	D5	
Inangil et al. (2020)	+	-	+	X	-	X
Isong et al. (2014)	-	-	X	X	-	X
Jung et al. (2020)	+	-	+	X	+	X
Kerimoglu et al. (2013)	X	X	X	X	-	X
Khan et al. (2019)	-	X	+	X	-	X
Koticha et al. (2019)	-	X	+	X	-	X
Litwin et al. (2020)	+	-	+	X	-	X
Niharika et al. (2018)	-	-	+	X	-	X
Nunna et al. (2019)	-	-	+	X	-	X
Nuvvula et al. (2015)	+	-	+	X	-	X
Osmanlliu et al. (in press)	+	-	+	X	-	X
Ryu et al. (2017)	+	-	+	+	-	-
Ryu et al. (2018)	+	-	+	+	-	-
Ryu et al. (2019)	+	-	+	+	-	-
Schlechter et al. (2020)	-	-	+	X	-	X
Shah et al. (2018)	-	X	+	X	-	X
Shetty et al. (2019)	-	X	+	X	-	X
Wong et al. (2020)	+	-	+	X	-	X
Özkan & Polat (2020)	+	-	X	X	-	X

Domains:
D1: Bias arising from the randomization process.
D2: Bias due to deviations from intended intervention.
D3: Bias due to missing outcome data.
D4: Bias in measurement of the outcome.
D5: Bias in selection of the reported result.

Judgement
 High
 Some concerns
 Low

Figure E3: ROBINS-I Judgements per Domain (Anxiety Studies)

Study	Risk of bias domains							Overall
	D1	D2	D3	D4	D5	D6	D7	
del Castillo et al. (2019)								
Piskorz & Czub (2018)								
Piskorz et al. (2020)								
Sullivan et al. (2000)								

Domains:
D1: Bias due to confounding.
D2: Bias due to selection of participants.
D3: Bias in classification of interventions.
D4: Bias due to deviations from intended interventions.
D5: Bias due to missing data.
D6: Bias in measurement of outcomes.
D7: Bias in selection of the reported result.

Judgement
 Serious
 Moderate
 Low
 No information

Appendix F**Stata Output: Tests for Publication Bias (Pain)**

The output of the trim-and-fill procedures are presented in the following order: (a) random-random effects (L_0 , R_0 , Q_0), random-fixed-effects (L_0 , R_0 , Q_0), fixed-fixed effects (L_0 , R_0 , Q_0), and fixed-random effects (L_0 , R_0 , Q_0).

Regression-based Egger test for small-study effects

Random-effects model

Method: REML

H0: beta1 = 0; no small-study effects

beta1 =	5.57
SE of beta1 =	1.345
z =	4.14
Prob > z =	0.0000

.

. meta trimfill, estimator(linear) itermethod(reml) poolmethod(reml)

Effect-size label: Effect Size
 Effect size: effektstr
 Std. Err.: standardfeil

Nonparametric trim-and-fill analysis of publication bias
 Linear estimator, imputing on the left

Iteration Number of studies = 42
 Model: Random-effects observed = 42
 Method: REML imputed = 0

Pooling
 Model: Random-effects
 Method: REML

Studies	Effect Size	[95% Conf. Interval]	
Observed	0.789	0.480	1.098
Observed + Imputed	0.789	0.480	1.098

. meta trimfill, estimator(run) itermethod(reml) poolmethod(reml)

Effect-size label: Effect Size
 Effect size: effektstr
 Std. Err.: standardfeil

Nonparametric trim-and-fill analysis of publication bias
 Run estimator, imputing on the left

Iteration Number of studies = 46
 Model: Random-effects observed = 42
 Method: REML imputed = 4

Pooling
 Model: Random-effects
 Method: REML

Studies	Effect Size	[95% Conf. Interval]	
Observed	0.789	0.480	1.098
Observed + Imputed	0.552	0.163	0.941

. meta trimfill, estimator(quadratic) itermethod(reml) poolmethod(reml)

Effect-size label: Effect Size
 Effect size: effektstr
 Std. Err.: standardfeil

Nonparametric trim-and-fill analysis of publication bias
 Quadratic estimator, imputing on the left

Iteration Number of studies = 42
 Model: Random-effects observed = 42
 Method: REML imputed = 0

Pooling
 Model: Random-effects
 Method: REML

Studies	Effect Size	[95% Conf. Interval]	
Observed	0.789	0.480	1.098
Observed + Imputed	0.789	0.480	1.098

.

Effect-size label: Effect Size
 Effect size: effektstr
 Std. Err.: standardfeil

Nonparametric trim-and-fill analysis of publication bias
 Linear estimator, imputing on the left

Iteration Number of studies = 42
 Model: Random-effects observed = 42
 Method: REML imputed = 0

Pooling
 Model: Fixed-effects
 Method: Inverse-variance

Studies	Effect Size	[95% Conf. Interval]	
Observed	0.568	0.497	0.638
Observed + Imputed	0.568	0.497	0.638

. meta trimfill, estimator(run) itermethod(reml) poolmethod(fixed)

Effect-size label: Effect Size
 Effect size: effektstr
 Std. Err.: standardfeil

Nonparametric trim-and-fill analysis of publication bias
 Run estimator, imputing on the left

Iteration Number of studies = 46
 Model: Random-effects observed = 42
 Method: REML imputed = 4

Pooling
 Model: Fixed-effects
 Method: Inverse-variance

Studies	Effect Size	[95% Conf. Interval]	
Observed	0.568	0.497	0.638
Observed + Imputed	0.499	0.430	0.569

. meta trimfill, estimator(quadratic) itermethod(reml) poolmethod(fixed)

Effect-size label: Effect Size
 Effect size: effektstr
 Std. Err.: standardfeil

Nonparametric trim-and-fill analysis of publication bias
 Quadratic estimator, imputing on the left

Iteration Number of studies = 42
 Model: Random-effects observed = 42
 Method: REML imputed = 0

Pooling
 Model: Fixed-effects
 Method: Inverse-variance

Studies	Effect Size	[95% Conf. Interval]	
Observed	0.568	0.497	0.638
Observed + Imputed	0.568	0.497	0.638

Effect-size label: Effect Size
 Effect size: effektstr
 Std. Err.: standardfeil

Nonparametric trim-and-fill analysis of publication bias
 Linear estimator, imputing on the left

Iteration Number of studies = 46
 Model: Fixed-effects observed = 42
 Method: Inverse-variance imputed = 4

Pooling
 Model: Fixed-effects
 Method: Inverse-variance

Studies	Effect Size	[95% Conf. Interval]	
Observed	0.568	0.497	0.638
Observed + Imputed	0.496	0.427	0.566

. meta trimfill, estimator(run) itermethod(fixed) poolmethod(fixed)

Effect-size label: Effect Size
 Effect size: effektstr
 Std. Err.: standardfeil

Nonparametric trim-and-fill analysis of publication bias
 Run estimator, imputing on the left

Iteration Number of studies = 48
 Model: Fixed-effects observed = 42
 Method: Inverse-variance imputed = 6

Pooling
 Model: Fixed-effects
 Method: Inverse-variance

Studies	Effect Size	[95% Conf. Interval]	
Observed	0.568	0.497	0.638
Observed + Imputed	0.411	0.343	0.479

. meta trimfill, estimator(quadratic) itermethod(fixed) poolmethod(fixed)

Effect-size label: Effect Size
 Effect size: effektstr
 Std. Err.: standardfeil

Nonparametric trim-and-fill analysis of publication bias
 Quadratic estimator, imputing on the left

Iteration Number of studies = 46
 Model: Fixed-effects observed = 42
 Method: Inverse-variance imputed = 4

Pooling
 Model: Fixed-effects
 Method: Inverse-variance

Studies	Effect Size	[95% Conf. Interval]	
Observed	0.568	0.497	0.638
Observed + Imputed	0.496	0.427	0.566

Effect-size label: Effect Size
 Effect size: effektstr
 Std. Err.: standardfeil

Nonparametric trim-and-fill analysis of publication bias
 Linear estimator, imputing on the left

Iteration Number of studies = 46
 Model: Fixed-effects observed = 42
 Method: Inverse-variance imputed = 4

Pooling
 Model: Random-effects
 Method: REML

Studies	Effect Size	[95% Conf. Interval]	
Observed	0.789	0.480	1.098
Observed + Imputed	0.543	0.148	0.938

. meta trimfill, estimator(run) itermethod(fixed) poolmethod(reml)

Effect-size label: Effect Size
 Effect size: effektstr
 Std. Err.: standardfeil

Nonparametric trim-and-fill analysis of publication bias
 Run estimator, imputing on the left

Iteration Number of studies = 48
 Model: Fixed-effects observed = 42
 Method: Inverse-variance imputed = 6

Pooling
 Model: Random-effects
 Method: REML

Studies	Effect Size	[95% Conf. Interval]	
Observed	0.789	0.480	1.098
Observed + Imputed	0.453	0.050	0.855

. meta trimfill, estimator(quadratic) itermethod(fixed) poolmethod(reml)

Effect-size label: Effect Size
 Effect size: effektstr
 Std. Err.: standardfeil

Nonparametric trim-and-fill analysis of publication bias
 Quadratic estimator, imputing on the left

Iteration Number of studies = 46
 Model: Fixed-effects observed = 42
 Method: Inverse-variance imputed = 4

Pooling
 Model: Random-effects
 Method: REML

Studies	Effect Size	[95% Conf. Interval]	
Observed	0.789	0.480	1.098
Observed + Imputed	0.543	0.148	0.938

.

Stata Output: Subgroup Analyses (Pain)

. meta summarize, subgroup(medical_procedure)

Effect-size label: Effect Size
 Effect size: effektstr
 Std. Err.: standardfeil
 Study label: author_year

Subgroup meta-analysis summary Number of studies = 42
 Random-effects model
 Method: REML
 Group: medical_procedure

Study	Effect Size	[95% Conf. Interval]	% Weight
Group: Dental			
Al-Halabi et al., 2018	0.076	-0.337 0.489	2.47
Al-Nerabieah et al., 2020	2.458	1.813 3.102	2.33
Aminabadi et al., 2013	1.843	1.412 2.274	2.46
Asvanund et al., 2015	0.291	-0.293 0.875	2.37
Bagattoni et al., 2018	-0.385	-0.946 0.177	2.38
Buldur & Candan, 2020	0.307	-0.141 0.754	2.45
Chaudhary et al., 2020	1.098	0.533 1.663	2.38
Fakhruddin et al., 2015	4.116	3.226 5.006	2.13
Hoge et al., 2012	-0.141	-0.491 0.209	2.50
Mitrakul et al., 2015	-0.094	-0.687 0.500	2.36
Niharika et al., 2018	5.080	3.744 6.416	1.74
Nunna et al., 2019	0.058	-0.406 0.521	2.45
Shetty et al., 2019	2.339	1.877 2.802	2.45
theta	1.262	0.353 2.172	
Group: Needle-related proce-s			
Aydin & Ozyazicioglu, 2019	0.054	-0.302 0.409	2.50
Caruso et al., 2019	0.112	-0.152 0.375	2.54
Chan et al., 2019 (study I)	0.324	-0.030 0.678	2.50
Chan et al., 2019 (study II)	0.517	0.169 0.866	2.51
Chen et al., 2020	0.371	0.034 0.708	2.51
Del-Castillo et al., 2019	2.358	1.494 3.222	2.15
Dumoulin et al., 2019	0.353	-0.182 0.889	2.40
Gerceker et al., 2018	1.541	1.118 1.965	2.47
Gerceker et al., 2020	1.171	0.792 1.551	2.49
Gerceker et al., in press	1.581	0.898 2.264	2.30
Gold & Mahrer, 2018	0.318	-0.010 0.647	2.51
Gold et al., 2006	0.269	-0.575 1.112	2.17
Goldman & Behboudi, 2020	0.638	0.148 1.128	2.43
Inangil et al., 2020	1.091	0.690 1.493	2.48
Litwin et al., 2020	0.562	-0.006 1.130	2.38
Osmanliu et al., in press	0.268	-0.225 0.762	2.43
Piskorz & Czub, 2018	0.822	0.172 1.471	2.32
Piskorz et al., 2020	1.265	0.686 1.844	2.37
Sander Wint et al., 2002	0.258	-0.448 0.963	2.28
Schlechter et al., 2020	0.017	-0.346 0.380	2.50
Wolitzky et al., 2005	1.768	0.765 2.770	2.03
Wong et al., 2020	0.736	0.349 1.123	2.49
Özkan & Polat, 2020	1.182	0.801 1.563	2.49
theta	0.717	0.486 0.947	
Group: Pre-operative			
Eijlers et al., 2019	-0.066	-0.348 0.217	2.53
Walther-Larsen et al., 2019	-0.315	-0.822 0.193	2.42
theta	-0.125	-0.372 0.122	
Group: Wound care			
Hua et al., 2015	0.880	0.376 1.384	2.42
Jeffs et al., 2014	-0.786	-1.608 0.036	2.19
Khadra et al., 2020	0.048	-0.600 0.696	2.32
Schmitt et al., 2011	0.661	0.101 1.222	2.38
theta	0.251	-0.446 0.949	
Overall			
theta	0.789	0.480 1.098	

Heterogeneity summary

Group	df	Q	P > Q	tau2	% I2	H2
Dental	12	266.96	0.000	2.695	97.50	39.99
Needle-relat-s	22	120.70	0.000	0.251	89.69	6.13
Pre-operative	1	0.71	0.401	0.000	0.00	1.00
Wound care	3	13.44	0.004	0.402	80.32	5.08
Overall	41	441.46	0.000	0.960	94.61	18.56

Test of group differences: $Q_b = \text{chi2}(3) = 27.83$

Prob > $Q_b = 0.000$

```
. meta summarize, subgroup(highimmersion)

Effect-size label: Effect Size
Effect size: effektstr
Std. Err.: standardfeil
Study label: author_year

Subgroup meta-analysis summary      Number of studies =    42
Random-effects model
Method: REML
Group: highimmersion
```

Study	Effect Size	[95% Conf. Interval]		% Weight
Group: 0				
Al-Halabi et al., 2018	0.076	-0.337	0.489	2.47
Aminabadi et al., 2013	1.843	1.412	2.274	2.46
Asvanund et al., 2015	0.291	-0.293	0.875	2.37
Aydin & Ozyazicioglu, 2019	0.054	-0.302	0.409	2.50
Bagattoni et al., 2018	-0.385	-0.946	0.177	2.38
Buldur & Candan, 2020	0.307	-0.141	0.754	2.45
Chaudhary et al., 2020	1.098	0.533	1.663	2.38
Del-Castillo et al., 2019	2.358	1.494	3.222	2.15
Dumoulin et al., 2019	0.353	-0.182	0.889	2.40
Fakhruddin et al., 2015	4.116	3.226	5.006	2.13
Gold et al., 2006	0.269	-0.575	1.112	2.17
Hoge et al., 2012	-0.141	-0.491	0.209	2.50
Hua et al., 2015	0.880	0.376	1.384	2.42
Jeffs et al., 2014	-0.786	-1.608	0.036	2.19
Khadra et al., 2020	0.048	-0.600	0.696	2.32
Mitrakul et al., 2015	-0.094	-0.687	0.500	2.36
Miharika et al., 2018	5.080	3.744	6.416	1.74
Nunna et al., 2019	0.058	-0.406	0.521	2.45
Piskorz et al., 2020	1.265	0.686	1.844	2.37
Sander Wint et al., 2002	0.258	-0.448	0.963	2.28
Shetty et al., 2019	2.339	1.877	2.802	2.45
Walther-Larsen et al., 2019	-0.315	-0.822	0.193	2.42
Wolitzky et al., 2005	1.768	0.765	2.770	2.03
theta	0.857	0.303	1.411	
Group: 1				
Al-Nerabieah et al., 2020	2.458	1.813	3.102	2.33
Caruso et al., 2019	0.112	-0.152	0.375	2.54
Chan et al., 2019 (study I)	0.324	-0.030	0.678	2.50
Chan et al., 2019 (study II)	0.517	0.169	0.866	2.51
chen et al., 2020	0.371	0.034	0.708	2.51
Eijlers et al., 2019	-0.066	-0.348	0.217	2.53
Gerceker et al., 2018	1.541	1.118	1.965	2.47
Gerceker et al., 2020	1.171	0.792	1.551	2.49
Gerceker et al., in press	1.581	0.898	2.264	2.30
Gold & Mahrer, 2018	0.318	-0.010	0.647	2.51
Goldman & Behboudi, 2020	0.638	0.148	1.128	2.43
Inangil et al., 2020	1.091	0.630	1.493	2.48
Litwin et al., 2020	0.562	-0.006	1.130	2.38
Osmanlliu et al., in press	0.268	-0.225	0.762	2.43
Piskorz & Czub, 2018	0.822	0.172	1.471	2.32
Schlechter et al., 2020	0.017	-0.346	0.380	2.50
Schmitt et al., 2011	0.661	0.101	1.222	2.38
Wong et al., 2020	0.736	0.349	1.123	2.49
Özkan & Polat, 2020	1.182	0.801	1.563	2.49
theta	0.724	0.452	0.996	
Overall				
theta	0.789	0.480	1.098	

Heterogeneity summary

Group	df	Q	P > Q	tau2	% I2	H2
0	22	305.53	0.000	1.725	95.75	23.53
1	18	135.93	0.000	0.314	88.47	8.67
Overall	41	441.46	0.000	0.960	94.61	18.56

Test of group differences: Q_b = chi2(1) = 0.18 Prob > Q_b = 0.673

```
. meta summarize, subgroup(interactivity)

Effect-size label: Effect Size
Effect size: effektstr
Std. Err.: standardfeil
Study label: author_year

Subgroup meta-analysis summary          Number of studies =    42
Random-effects model
Method: REML
Group: interactivity
```

Study	Effect Size	[95% Conf. Interval]	% Weight
Group: no			
Al-Halabi et al., 2018	0.076	-0.337 0.489	2.73
Al-Nerabieah et al., 2020	2.458	1.813 3.102	2.57
Aminabadi et al., 2013	1.843	1.412 2.274	2.72
Asvanund et al., 2015	0.291	-0.293 0.875	2.62
Bagattoni et al., 2018	-0.385	-0.946 0.177	2.64
Buldur & Candan, 2020	0.307	-0.141 0.754	2.71
Chen et al., 2020	0.371	0.034 0.708	2.78
Del-Castillo et al., 2019	2.358	1.494 3.222	2.38
Fakhruddin et al., 2015	4.116	3.226 5.006	2.36
Gerceker et al., 2018	1.541	1.118 1.965	2.73
Goldman & Behboudi, 2020	0.638	0.148 1.128	2.69
Hoge et al., 2012	-0.141	-0.491 0.209	2.77
Inangil et al., 2020	1.091	0.690 1.493	2.74
Mitrakul et al., 2015	-0.094	-0.687 0.500	2.61
Miharika et al., 2018	5.080	3.744 6.416	1.93
Nunna et al., 2019	0.058	-0.406 0.521	2.70
Sander Wint et al., 2002	0.258	-0.448 0.963	2.52
Shetty et al., 2019	2.339	1.877 2.802	2.70
Wong et al., 2020	0.736	0.349 1.123	2.75
Özkan & Polat, 2020	1.182	0.801 1.563	2.75
theta	1.156	0.560 1.752	
Group: yes			
Aydın & Özyazıcıoğlu, 2019	0.054	-0.302 0.409	2.77
Caruso et al., 2019	0.112	-0.152 0.375	2.81
Chan et al., 2019 (study I)	0.324	-0.030 0.678	2.77
Chan et al., 2019 (study II)	0.517	0.169 0.866	2.77
Dumoulin et al., 2019	0.353	-0.182 0.889	2.65
Eijlers, Dierckx, et al., 2019	-0.066	-0.348 0.217	2.80
Gold & Mahrer, 2018	0.318	-0.010 0.647	2.78
Gold et al., 2006	0.269	-0.575 1.112	2.40
Hua et al., 2015	0.880	0.376 1.384	2.68
Jeffs et al., 2014	-0.786	-1.608 0.036	2.42
Khadra et al., 2020	0.048	-0.600 0.696	2.57
Litwin et al., 2020	0.562	-0.006 1.130	2.63
Osmanliu et al., in press	0.268	-0.225 0.762	2.68
Piskorz & Czub, 2018	0.822	0.172 1.471	2.57
Schlechter et al., 2020	0.017	-0.346 0.380	2.76
Schmitt et al., 2011	0.661	0.101 1.222	2.64
Walther-Larsen et al., 2019	-0.315	-0.822 0.193	2.67
Wolitzky et al., 2005	1.768	0.765 2.770	2.25
theta	0.277	0.104 0.451	
Overall	theta	0.742 0.402 1.081	

Heterogeneity summary

Group	df	Q	P > Q	tau2	% I2	H2
no	19	304.91	0.000	1.758	96.72	30.49
yes	17	43.43	0.000	0.078	61.37	2.59
Overall	37	412.24	0.000	1.056	95.17	20.72

Test of group differences: Q_b = chi2(1) = 7.69 Prob > Q_b = 0.006

Stata Output: Meta-Regression (Pain)

```
. meta regress age_m interactivity highimmersion
```

```
Effect-size label: Effect Size
Effect size: effektstr
Std. Err.: standardfeil
```

```
Random-effects meta-regression
Method: REML
```

```
Number of obs = 38
Residual heterogeneity:
    tau2 = .9237
    I2 (%) = 94.35
    H2 = 17.69
R-squared (%) = 12.54
Wald chi2(3) = 7.95
Prob > chi2 = 0.0471
```

_meta_es	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age_m	-.0911651	.0688672	-1.32	0.186	-.2261424	.0438122
interactivity	-.6543144	.3560705	-1.84	0.066	-1.3522	.043571
highimmersion	.1652802	.3454901	0.48	0.632	-.511868	.8424284
_cons	1.835253	.6079273	3.02	0.003	.6437374	3.026769

```
Test of residual homogeneity: Q_res = chi2(34) = 331.88 Prob > Q_res = 0.0000
```

Stata Output: Sensitivity Analyses (Pain)

The Stata command ‘meta summarize if ntotal >= 100’ was used to remove studies with less than 100 participants, the ‘meta summarize if design >= 1’ command was used to remove non-randomized studies, and the ‘morethanonehighrisk == 0’ removed studies that had a high risk of bias in two or more domains.

```
. meta summarize if ntotal >=100
```

```
Effect-size label: Hedges' g
Effect size: effektstr
Std. Err.: standardfeil
Study label: author_year
```

```
Meta-analysis summary
Random-effects model
Method: REML
```

```
Number of studies = 17
Heterogeneity:
tau2 = 0.4896
I2 (%) = 93.60
H2 = 15.63
```

Study	Hedges' g	[95% Conf. Interval]		% Weight
Al-Halabi et al., 2018	0.076	-0.337	0.489	5.79
Aminabadi et al., 2013	1.843	1.412	2.274	5.75
Aydin & Özyazicioglu, 2019	0.054	-0.302	0.409	5.91
Caruso et al., 2019	0.112	-0.152	0.375	6.09
Chan et al., 2019 (study I)	0.324	-0.030	0.678	5.92
Chan et al., 2019 (study II)	0.517	0.169	0.866	5.93
Chen et al., 2020	0.371	0.034	0.708	5.95
Eijlers et al., 2019	-0.066	-0.348	0.217	6.06
Gerceker et al., 2018	1.541	1.118	1.965	5.76
Gerceker et al., 2020	1.171	0.792	1.551	5.86
Gold & Mahrer, 2018	0.318	-0.010	0.647	5.97
Hoge et al., 2012	-0.141	-0.491	0.209	5.93
Inangil et al., 2020	1.091	0.690	1.493	5.81
Schlechter et al., 2020	0.017	-0.346	0.380	5.90
Shetty et al., 2019	2.339	1.877	2.802	5.67
Wong et al., 2020	0.736	0.349	1.123	5.85
Özkan & Polat, 2020	1.182	0.801	1.563	5.86
theta	0.666	0.321	1.010	

```
Test of theta = 0: z = 3.79
Test of homogeneity: Q = chi2(16) = 210.95
```

```
Prob > |z| = 0.0002
Prob > Q = 0.0000
```

```
.
```

. meta summarize if design >= 1

Effect-size label: Hedges' g
 Effect size: effektstr
 Std. Err.: standardfeil
 Study label: author_year

Meta-analysis summary
 Random-effects model
 Method: REML

Number of studies = 39
 Heterogeneity:
 tau2 = 0.9775
 I2 (%) = 94.90
 H2 = 19.61

Study	Hedges' g	[95% Conf. Interval]		% Weight
Al-Halabi et al., 2018	0.076	-0.337	0.489	2.65
Al-Werabieah et al., 2020	2.458	1.813	3.102	2.50
Aminabadi et al., 2013	1.843	1.412	2.274	2.64
Asvanund et al., 2015	0.291	-0.293	0.875	2.54
Aydin & Özyazicioglu, 2019	0.054	-0.302	0.409	2.68
Bagattoni et al., 2018	-0.385	-0.946	0.177	2.56
Buldur & Candan, 2020	0.307	-0.141	0.754	2.63
Caruso et al., 2019	0.112	-0.152	0.375	2.72
Chan et al., 2019 (study I)	0.324	-0.030	0.678	2.68
Chan et al., 2019 (study II)	0.517	0.169	0.866	2.69
Chaudhary et al., 2020	1.098	0.533	1.663	2.56
Chen et al., 2020	0.371	0.034	0.708	2.69
Dumoulin et al., 2019	0.353	-0.182	0.889	2.58
Eijlers et al., 2019	-0.066	-0.348	0.217	2.72
Fakhruddin et al., 2015	4.116	3.226	5.006	2.29
Gerceker et al., 2018	1.541	1.118	1.965	2.65
Gerceker et al., 2020	1.171	0.792	1.551	2.67
Gerceker et al., in press	1.581	0.898	2.264	2.47
Gold & Mahrer, 2018	0.318	-0.010	0.647	2.70
Gold et al., 2006	0.269	-0.575	1.112	2.33
Goldman & Behboudi, 2020	0.638	0.148	1.128	2.61
Hoge et al., 2012	-0.141	-0.491	0.209	2.69
Hua et al., 2015	0.880	0.376	1.384	2.60
Inangil et al., 2020	1.091	0.690	1.493	2.66
Jeffs et al., 2014	-0.786	-1.608	0.036	2.35
Khadra et al., 2020	0.048	-0.600	0.696	2.50
Litwin et al., 2020	0.562	-0.006	1.130	2.55
Mitrakul et al., 2015	-0.094	-0.687	0.500	2.54
Niharika et al., 2018	5.080	3.744	6.416	1.88
Nunna et al., 2019	0.058	-0.406	0.521	2.62
Osmanlliu et al., in press	0.268	-0.225	0.762	2.61
Sander Wint et al., 2002	0.258	-0.448	0.963	2.45
Schlechter et al., 2020	0.017	-0.346	0.380	2.68
Schmitt et al., 2011	0.661	0.101	1.222	2.56
Shetty et al., 2019	2.339	1.877	2.802	2.62
Walther-Larsen et al., 2019	-0.315	-0.822	0.193	2.60
Wolitzky et al., 2005	1.768	0.765	2.770	2.19
Wong et al., 2020	0.736	0.349	1.123	2.67
Özkan & Polat, 2020	1.182	0.801	1.563	2.67
theta	0.741	0.418	1.064	

Test of theta = 0: z = 4.50

Prob > |z| = 0.0000

Test of homogeneity: Q = chi2(38) = 418.30

Prob > Q = 0.0000

. meta summarize if morethanonehighrisk == 0

Effect-size label: Hedges' g
 Effect size: effektstr
 Std. Err.: standardfeil
 Study label: author_year

Meta-analysis summary
 Random-effects model
 Method: REML

Number of studies = 22
 Heterogeneity:
 tau2 = 0.8927
 I2 (%) = 94.37
 H2 = 17.75

Study	Hedges' g	[95% Conf. Interval]		% Weight
Al-Halabi et al., 2018	0.076	-0.337	0.489	4.73
Al-Nerabieah et al., 2020	2.458	1.813	3.102	4.42
Aminabadi et al., 2013	1.843	1.412	2.274	4.71
Chan et al., 2019 (study I)	0.324	-0.030	0.678	4.79
Chan et al., 2019 (study II)	0.517	0.169	0.866	4.79
Chen et al., 2020	0.371	0.034	0.708	4.80
Dumoulin et al., 2019	0.353	-0.182	0.889	4.58
Eijlers et al., 2019	-0.066	-0.348	0.217	4.85
Gerceker et al., 2020	1.171	0.792	1.551	4.76
Gerceker et al., in press	1.581	0.898	2.264	4.37
Gold et al., 2006	0.269	-0.575	1.112	4.11
Goldman & Behboudi, 2020	0.638	0.148	1.128	4.64
Inangil et al., 2020	1.091	0.690	1.493	4.74
Jeffs et al., 2014	-0.786	-1.608	0.036	4.14
Khadra et al., 2020	0.048	-0.600	0.696	4.42
Litwin et al., 2020	0.562	-0.006	1.130	4.53
Niharika et al., 2018	5.080	3.744	6.416	3.26
Nunna et al., 2019	0.058	-0.406	0.521	4.67
Osmanlliu et al., in press	0.268	-0.225	0.762	4.63
Schlechter et al., 2020	0.017	-0.346	0.380	4.78
Schmitt et al., 2011	0.661	0.101	1.222	4.54
Wong et al., 2020	0.736	0.349	1.123	4.75
theta	0.729	0.317	1.141	

Test of theta = 0: z = 3.46 Prob > |z| = 0.0005
 Test of homogeneity: Q = chi2(21) = 192.85 Prob > Q = 0.0000

Stata Output: Tests for Publication Bias (Anxiety)

The output of the trim-and-fill procedures are presented in the following order: (a) random-random effects (L_0 , R_0 , Q_0), random-fixed-effects (L_0 , R_0 , Q_0), fixed-fixed effects (L_0 , R_0 , Q_0), and fixed-random effects (L_0 , R_0 , Q_0).

```
> ed)
meta bias, egger
```

```
Effect-size label: Effect Size
Effect size: effektstr
Std. Err.: standardfeil
```

```
Regression-based Egger test for small-study effects
Random-effects model
Method: REML
```

```
H0: beta1 = 0; no small-study effects
      beta1 =      10.05
SE of beta1 =      1.722
          z =       5.84
Prob > |z| =      0.0000
```

Nonparametric trim-and-fill analysis of publication bias
 Linear estimator, imputing on the left

Iteration Number of studies = 44
 Model: Random-effects observed = 44
 Method: REML imputed = 0

Pooling
 Model: Random-effects
 Method: REML

Studies	Hedges' g	[95% Conf. Interval]	
Observed	0.656	0.581	0.731
Observed + Imputed	0.656	0.581	0.731

. meta trimfill, estimator(run) itermethod(reml) poolmethod(reml)

Effect-size label: Hedges' g
 Effect size: effektstr
 Std. Err.: standardfeil

Nonparametric trim-and-fill analysis of publication bias
 Run estimator, imputing on the left

Iteration Number of studies = 56
 Model: Random-effects observed = 44
 Method: REML imputed = 12

Pooling
 Model: Random-effects
 Method: REML

Studies	Hedges' g	[95% Conf. Interval]	
Observed	0.656	0.581	0.731
Observed + Imputed	0.575	0.501	0.649

. meta trimfill, estimator(quadratic) itermethod(reml) poolmethod(reml)

Effect-size label: Hedges' g
 Effect size: effektstr
 Std. Err.: standardfeil

Nonparametric trim-and-fill analysis of publication bias
 Quadratic estimator, imputing on the left

Iteration Number of studies = 44
 Model: Random-effects observed = 44
 Method: REML imputed = 0

Pooling
 Model: Random-effects
 Method: REML

Studies	Hedges' g	[95% Conf. Interval]	
Observed	0.656	0.581	0.731
Observed + Imputed	0.656	0.581	0.731

.

Effect-size label: Hedges' g
 Effect size: effektstr
 Std. Err.: standardfeil

Nonparametric trim-and-fill analysis of publication bias
 Linear estimator, imputing on the left

Iteration Number of studies = 44
 Model: Random-effects observed = 44
 Method: REML imputed = 0

Pooling
 Model: Fixed-effects
 Method: Inverse-variance

Studies	Hedges' g	[95% Conf. Interval]	
Observed	0.656	0.581	0.731
Observed + Imputed	0.656	0.581	0.731

. meta trimfill, estimator(run) itermethod(reml) poolmethod(fixed)

Effect-size label: Hedges' g
 Effect size: effektstr
 Std. Err.: standardfeil

Nonparametric trim-and-fill analysis of publication bias
 Run estimator, imputing on the left

Iteration Number of studies = 56
 Model: Random-effects observed = 44
 Method: REML imputed = 12

Pooling
 Model: Fixed-effects
 Method: Inverse-variance

Studies	Hedges' g	[95% Conf. Interval]	
Observed	0.656	0.581	0.731
Observed + Imputed	0.575	0.501	0.649

. meta trimfill, estimator(quadratic) itermethod(reml) poolmethod(fixed)

Effect-size label: Hedges' g
 Effect size: effektstr
 Std. Err.: standardfeil

Nonparametric trim-and-fill analysis of publication bias
 Quadratic estimator, imputing on the left

Iteration Number of studies = 44
 Model: Random-effects observed = 44
 Method: REML imputed = 0

Pooling
 Model: Fixed-effects
 Method: Inverse-variance

Studies	Hedges' g	[95% Conf. Interval]	
Observed	0.656	0.581	0.731
Observed + Imputed	0.656	0.581	0.731

.

. meta summarize, subgroup(highimmersion)

Effect-size label: Effect Size
 Effect size: effektstr
 Std. Err.: standardfeil
 Study label: author_year

Subgroup meta-analysis summary Number of studies = 35
 Random-effects model
 Method: REML
 Group: highimmersion

Study	Effect Size	[95% Conf. Interval]		% Weight	
Group: 0					
Al-Khotani et al., 2016	0.203	-0.315	0.721	2.85	
Aminabadi et al., 2013	5.549	4.752	6.346	2.64	
Buldur & Candan, 2020	0.399	-0.050	0.849	2.90	
Del-Castillo et al., 2019	2.734	1.810	3.658	2.52	
Dumoulin et al., 2019	0.423	-0.115	0.960	2.84	
Fakhruddin et al., 2015	1.727	1.140	2.315	2.81	
Hashimoto et al., 2020	0.828	0.299	1.358	2.85	
Isong et al., 2014	0.345	-0.267	0.958	2.79	
Kerimoglu et al., 2013	0.318	-0.105	0.741	2.91	
Niharika et al., 2018	1.291	0.586	1.996	2.71	
Nunna et al., 2019	0.352	-0.115	0.819	2.89	
Nuvvula et al., 2014	1.586	1.094	2.079	2.87	
Piskorz et al., 2020	0.858	0.304	1.411	2.83	
Shah & Bhatia, 2018	-0.310	-0.860	0.240	2.83	
Shetty et al., 2019	1.475	1.070	1.880	2.92	
theta	1.163	0.459	1.868		
Group: 1					
Al-Nerabieah et al., 2020	3.008	2.297	3.720	2.71	
Caruso et al., 2019	0.101	-0.163	0.365	2.99	
Chan et al., 2019 (study I)	0.310	-0.043	0.664	2.95	
Chan et al., 2019 (study II)	0.469	0.121	0.816	2.95	
Chen et al., 2020	0.352	0.015	0.689	2.96	
Eijlers et al., 2019	-0.013	-0.295	0.270	2.98	
Gerceker et al., 2020	2.015	1.588	2.442	2.91	
Gerceker et al., in press	1.226	0.578	1.875	2.76	
Gold & Mahrer, 2018	0.269	-0.059	0.597	2.96	
Inangil et al., 2020	1.339	0.925	1.752	2.92	
Jung et al., 2020	1.166	0.663	1.668	2.86	
Litwin et al., 2020	-0.374	-0.947	0.199	2.82	
Osmanliu et al., in press	0.203	-0.290	0.696	2.87	
Piskorz & Czub, 2018	0.947	0.289	1.605	2.75	
Ryu et al., 2017	0.787	0.302	1.271	2.88	
Ryu et al., 2018	1.021	0.524	1.518	2.87	
Ryu et al., 2019	-0.464	-0.904	-0.024	2.90	
Schlechter et al., 2020	0.323	-0.042	0.689	2.94	
Wong et al., 2020	0.767	0.379	1.155	2.93	
Özkan & Polat, 2020	1.210	0.828	1.592	2.94	
theta	0.713	0.371	1.056		
Overall	theta	0.902	0.547	1.258	

Heterogeneity summary

Group	df	Q	P > Q	tau2	% I2	H2
0	14	218.99	0.000	1.852	96.13	25.87
1	19	199.76	0.000	0.556	92.94	14.15
Overall	34	437.69	0.000	1.080	95.43	21.88

Test of group differences: Q_b = chi2(1) = 1.27 Prob > Q_b = 0.260

. meta summarize, subgroup(interactivity)

Effect-size label: Hedges' g
 Effect size: effektstr
 Std. Err.: standardfeil
 Study label: author_year

Subgroup meta-analysis summary Number of studies = 35
 Random-effects model
 Method: REML
 Group: interactivity

Study	Hedges' g	[95% Conf. Interval]	% Weight
Group: no			
Al-Khotani et al., 2016	0.203	-0.315 0.721	3.12
Al-Nerabieah et al., 2020	3.008	2.297 3.720	2.96
Aminabadi et al., 2013	5.549	4.752 6.346	2.88
Buldur & Candan, 2020	0.399	-0.050 0.849	3.17
Chen et al., 2020	0.352	0.015 0.689	3.23
del Castillo et al., 2019	2.734	1.810 3.658	2.75
Fakhruddin et al., 2015	1.727	1.140 2.315	3.07
Hashimoto et al., 2020	0.828	0.299 1.358	3.11
Inangil et al., 2020	1.339	0.925 1.752	3.19
Isong et al., 2014	0.345	-0.267 0.958	3.05
Kerimoglu et al., 2013	0.318	-0.105 0.741	3.18
Niharika et al., 2018	1.291	0.586 1.996	2.97
Nunna et al., 2019	0.352	-0.115 0.819	3.15
Nuvvula et al., 2014	1.586	1.094 2.079	3.14
Ryu et al., 2017	0.787	0.302 1.271	3.14
Ryu et al., 2019	-0.464	-0.904 -0.024	3.17
Shah & Bhatia, 2018	-0.310	-0.860 0.240	3.10
Shetty et al., 2019	1.475	1.070 1.880	3.19
Wong et al., 2020	0.767	0.379 1.155	3.20
Özkan & Polat, 2020	1.210	0.828 1.592	3.21
theta	1.151	0.574 1.729	
Group: yes			
Caruso et al., 2019	0.101	-0.163 0.365	3.27
Chan et al., 2019 (study I)	0.310	-0.043 0.664	3.22
Chan et al., 2019 (study II)	0.469	0.121 0.816	3.23
Dumoulin et al., 2019	0.423	-0.115 0.960	3.10
Eijlers, Dierckx, et al., 2019	-0.013	-0.295 0.270	3.26
Gold & Mahrer, 2018	0.269	-0.059 0.597	3.24
Jung et al., 2020	1.166	0.663 1.668	3.13
Litwin et al., 2020	-0.374	-0.947 0.199	3.08
Osmanlliu et al., in press	0.203	-0.290 0.696	3.14
Piskorz & Czub, 2018	0.947	0.289 1.605	3.01
Ryu et al., 2018	1.021	0.524 1.518	3.13
Schlechter et al., 2020	0.323	-0.042 0.689	3.22
theta	0.378	0.150 0.607	
Overall	theta	0.860 0.478 1.242	

Heterogeneity summary

Group	df	Q	P > Q	tau2	% I2	H2
no	19	304.11	0.000	1.659	96.43	28.04
yes	11	36.70	0.000	0.115	74.27	3.89
Overall	31	393.19	0.000	1.149	95.79	23.74

Test of group differences: Q_b = chi2(1) = 5.95 Prob > Q_b = 0.015

Stata Output: Meta-Regression (Anxiety)

```
. meta regress age_m interactivity highimmersion
```

```
Effect-size label: Hedges' g
Effect size: effektstr
Std. Err.: standardfeil
```

```
Random-effects meta-regression
Method: REML
```

```
Number of obs =      31
Residual heterogeneity:
    tau2 = 1.022
    I2 (%) = 95.19
    H2 = 20.78
R-squared (%) = 10.69
Wald chi2(3) = 6.51
Prob > chi2 = 0.0894
```

_meta_es	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age_m	-.1055984	.0860619	-1.23	0.220	-.2742767	.0630799
interactivity	-.3595948	.5104343	-0.70	0.481	-1.360028	.640838
highimmersion	-.3048104	.4522132	-0.67	0.500	-1.191132	.5815111
_cons	2.154221	.7304636	2.95	0.003	.7225382	3.585903

```
Test of residual homogeneity: Q_res = chi2(27) = 309.19 Prob > Q_res = 0.0000
```

Stata Output: Sensitivity Analyses (Anxiety)

The Stata command ‘meta summarize if ntotal >= 100’ was used to remove studies with less than 100 participants, the ‘meta summarize if design >= 1’ command was used to remove non-randomized studies, and the ‘morethanonehighrisk == 0’ removed studies that had a high risk of bias in two or more domains.

```
. meta summarize if ntotal >= 100

Effect-size label: Effect Size
Effect size: effektstr
Std. Err.: standardfeil
```

```
Meta-analysis summary
Random-effects model
Method: REML
```

```
Number of studies = 13
Heterogeneity:
tau2 = 1.9484
I2 (%) = 98.29
H2 = 58.55
```

Study	Effect Size	[95% Conf. Interval]		% Weight
Study 3	5.549	4.752	6.346	7.25
Study 6	0.101	-0.163	0.365	7.79
Study 7	0.310	-0.043	0.664	7.73
Study 8	0.469	0.121	0.816	7.74
Study 9	0.352	0.015	0.689	7.75
Study 12	-0.013	-0.295	0.270	7.78
Study 14	2.015	1.588	2.442	7.68
Study 17	0.269	-0.059	0.597	7.75
Study 20	1.339	0.925	1.752	7.69
Study 36	0.323	-0.042	0.689	7.73
Study 38	1.475	1.070	1.880	7.70
Study 40	0.767	0.379	1.155	7.71
Study 41	1.210	0.828	1.592	7.71
theta	1.066	0.299	1.833	

```
Test of theta = 0: z = 2.72
```

```
Prob > |z| = 0.0065
```

```
Test of homogeneity: Q = chi2(12) = 273.34
```

```
Prob > Q = 0.0000
```

. meta summarize if design >= 1

Effect-size label: Effect Size
 Effect size: effektstr
 Std. Err.: standardfeil

Meta-analysis summary Number of studies = 32
 Random-effects model Heterogeneity:
 Method: REML tau2 = 1.0983
 I2 (%) = 95.72
 H2 = 23.34

Study	Effect Size	[95% Conf. Interval]		% Weight
Study 1	0.203	-0.315	0.721	3.11
Study 2	3.008	2.297	3.720	2.95
Study 3	5.549	4.752	6.346	2.87
Study 5	0.399	-0.050	0.849	3.15
Study 6	0.101	-0.163	0.365	3.25
Study 7	0.310	-0.043	0.664	3.21
Study 8	0.469	0.121	0.816	3.21
Study 9	0.352	0.015	0.689	3.22
Study 11	0.423	-0.115	0.960	3.09
Study 12	-0.013	-0.295	0.270	3.24
Study 13	1.727	1.140	2.315	3.05
Study 14	2.015	1.588	2.442	3.17
Study 15	1.226	0.578	1.875	3.00
Study 17	0.269	-0.059	0.597	3.22
Study 19	0.828	0.299	1.358	3.10
Study 20	1.339	0.925	1.752	3.17
Study 21	0.345	-0.267	0.958	3.03
Study 22	1.166	0.663	1.668	3.12
Study 23	0.318	-0.105	0.741	3.17
Study 26	-0.374	-0.947	0.199	3.06
Study 27	1.291	0.586	1.996	2.96
Study 28	0.352	-0.115	0.819	3.14
Study 29	1.586	1.094	2.079	3.12
Study 30	0.203	-0.290	0.696	3.12
Study 33	0.787	0.302	1.271	3.13
Study 34	1.021	0.524	1.518	3.12
Study 35	-0.464	-0.904	-0.024	3.16
Study 36	0.323	-0.042	0.689	3.20
Study 37	-0.310	-0.860	0.240	3.08
Study 38	1.475	1.070	1.880	3.18
Study 40	0.767	0.379	1.155	3.19
Study 41	1.210	0.828	1.592	3.19
theta	0.853	0.479	1.226	

Test of theta = 0: z = 4.48 Prob > |z| = 0.0000
 Test of homogeneity: Q = chi2(31) = 416.70 Prob > Q = 0.0000

```
. meta summarize if morethanonehighrisk == 0
```

```
Effect-size label: Effect Size
Effect size: effektstr
Std. Err.: standardfeil
```

```
Meta-analysis summary          Number of studies =    23
Random-effects model          Heterogeneity:
Method: REML                   tau2 = 1.4032
                                I2 (%) = 96.54
                                H2 = 28.93
```

Study	Effect Size	[95% Conf. Interval]		% Weight
Study 2	3.008	2.297	3.720	4.16
Study 3	5.549	4.752	6.346	4.07
Study 7	0.310	-0.043	0.664	4.44
Study 8	0.469	0.121	0.816	4.45
Study 9	0.352	0.015	0.689	4.45
Study 11	0.423	-0.115	0.960	4.32
Study 12	-0.013	-0.295	0.270	4.48
Study 14	2.015	1.588	2.442	4.40
Study 15	1.226	0.578	1.875	4.22
Study 17	0.269	-0.059	0.597	4.46
Study 19	0.828	0.299	1.358	4.32
Study 20	1.339	0.925	1.752	4.41
Study 22	1.166	0.663	1.668	4.34
Study 26	-0.374	-0.947	0.199	4.29
Study 27	1.291	0.586	1.996	4.16
Study 28	0.352	-0.115	0.819	4.37
Study 29	1.586	1.094	2.079	4.35
Study 30	0.203	-0.290	0.696	4.35
Study 33	0.787	0.302	1.271	4.36
Study 34	1.021	0.524	1.518	4.35
Study 35	-0.464	-0.904	-0.024	4.39
Study 36	0.323	-0.042	0.689	4.44
Study 40	0.767	0.379	1.155	4.42
theta	0.954	0.459	1.449	

```
Test of theta = 0: z = 3.78          Prob > |z| = 0.0002
Test of homogeneity: Q = chi2(22) = 343.88      Prob > Q = 0.0000
```