Metaverse and Pediatrics: measurement methodologies for capturing pediatric patient experience with extended reality

Samantha Jean Kobitowich

Dissertation written under the supervision of
Professor Dr. Henrique Martins.

Dissertation submitted in partial fulfilment of requirements for the International MSc in Management (Strategy and Consulting), at Universidade Católica Portuguesa and for the Master of International Business at Smith School of Business, Queen's University, 06 January 2025.

Abstract

Advanced digital technologies have the capability to reduce or even remove undesired experiences for children in several common medical procedures, encouraging better healthcare engagement as they age. Measuring such experiences remains challenging. The aim of this thesis is to provide a strategic management perspective on the pediatric medical XR industry, capturing how research is conducted, particularly the capture of patient experience, and the challenges the current approaches pose on stakeholders and innovation. Extensive empirical evidence was conducted through the form of a systematic review. Full-text review was completed for 498 studies, and 45 studies met the inclusion criteria for deeper thematic analysis. The review looked at the categorization of medical XR use cases, outcome and experience measurement methods and metrics. Results showed that across the research spectrum there is no consistent, systematic approach when measuring patient experience of XR interactions. There is significant heterogeneity in use cases, the outcome and experience metrics chosen for measurement, and how feedback is collected. This adds uncertainty to companies developing and healthcare providers offering these solutions and makes comparability or meta-analysis for efficacy validation extremely difficult for researchers and regulators alike. Regulators, producers, researchers, clinicians, and patients all benefit from more defined efficacy testing of medical devices and the current variance suggests that more guidance is needed for this to take place. Recommendations include using more consistent and limited sets of outcome measures, harmonize research methods by creating and sharing guidelines across industry and health systems, and expand metrics focussing on experience capture.

Title: Measurement methodologies for capturing pediatric patient experience with advanced patient-facing extended reality (XR) technologies, a systematic review

Author: Samantha Kobitowich

Keywords: Extended Reality, Virtual Reality, User Experience, Quality of Care, Pediatrics, Patient Experience, Experience Measurement

Resumo

As tecnologias digitais avançadas têm a capacidade de reduzir ou até eliminar experiências indesejadas para as crianças em vários procedimentos médicos comuns, promovendo um maior envolvimento com os cuidados de saúde à medida que crescem. Medir essas experiências é um desafio. Esta tese oferece uma perspetiva de gestão estratégica sobre a indústria de soluções baseadas em Realidade Estendida (XR) para a pediatria, abordando a investigação realizada, os desafios da abordagem atual para as partes interessadas e inovações. Foi realizada uma revisão sistemática de 498 estudos, dos quais 45 foram selecionados para análise detalhada, onde foi feita a categorização dos casos de uso, os métodos e métricas de medição de resultados e experiências do doente. Não há uma abordagem consistente para medir a experiência dos pacientes com a XR. Há grande heterogeneidade nos casos de uso, nas métricas escolhidas e na forma como o feedback é recolhido. Isso aumenta a imprevisibilidade para a indústria e as entidades de saúde e dificulta a comparabilidade e a meta-análise para validação da eficácia, tanto para investigadores como para reguladores. Reguladores, produtores, investigadores, clínicos e pacientes beneficiam de uma avaliação mais clara da eficácia dos dispositivos médicos. Recomendações incluem o uso mais consistente de conjuntos de métricas de resultados, a harmonização dos métodos de investigação e expansão das métricas sobre a experiência do doente.

Título: Metodologias de medição para capturar a experiência do paciente pediátrico com tecnologias avançadas de realidade estendida (XR) voltadas para o paciente, uma revisão sistemática

Autor: Samantha Kobitowich

Palavras-chave: Realidade Estendida, Realidade Virtual, Experiência do Usuário, Qualidade do Atendimento, Pediatria, Experiência do Paciente, Medição da Experiência

Table of Contents

List of Abbreviations

Abbreviation	Definition
AI	Artificial Intelligence
AI Act	Artificial Intelligence Act (Regulation [EU] 2024/1689)
AR	Augmented Reality
CENTRAL	Cochrane Central Register of Controlled Trials
CFS	Children's Fear Scale
CONSORT	Consolidated Standards of Reporting Trials
CRD	Centre for Reviews and Dissemination, University of York
CRO	Contract Research Organization
FDA	U.S. Food and Drug Administration
FLACC	Faces, Legs, Activity, Crying, Consolability
MDR	Medical Device Regulation (Regulation [EU] 2017/745)
MR	Mixed Reality
NIHR	U.K. National Institute for Health and Care Research
OECD	Organisation for Economic Co-operation and Development
PaRIS	Patient-Reported Indicator Surveys
PICOS	Population, Intervention, Comparison, Outcomes, Study Design
PICU	Pediatric Intensive Care Unit
PREM	Patient-Reported Experience Measure
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PRO	Patient-Reported Outcome
PROM	Patient-Reported Outcome Measure
PROSPERO	International Prospective Register of Systematic Reviews
RCT	Randomized Controlled Trail
SPIRIT	Standard Protocol Items: Recommendations for Interventional Trials
UX	User Experience
VR	Virtual Reality
VR-CORE	Virtual Reality Clinical Outcomes Research Experts
XR	Extended Reality

List of Figures

List of Tables

1. INTRODUCTION

Needles are the face of pain for kids. Particularly frequent in childhood, vaccines are the most common painful needle procedures throughout the lifespan [46] and injection-related pain and fear are common adverse reactions for children [47]. Setting the tone for future healthcare interactions, painful needle experiences are associated with needle-related fear which can last into adulthood and be a barrier to future care [47, 48]. Advanced digital technologies have the capability to reduce or even remove the experience of pain for children in several common medical procedures, encouraging better healthcare engagement as they age. Given its immersive qualities, Extended Reality (XR) offers solutions to different medical challenges for adults and children alike. While pain and distress from needle-related procedures are common during childhood, there are many other medical procedures and contexts that benefit from interventions supported by XR through different mechanisms, including distraction, which is one of the most widely studied psychological interventions and is known to improve experience of pain [35, 49].

Inconsistency exists across studies using XR for healthcare and is even more severe in pediatrics. Safe use of technology in health settings requires a solid system of comparable evidence to allow for meta-analysis to be undertaken to compare risk versus value of the device or software. Due to the inconsistency of measurement across studies, they are not comparable, meaning cross-learning and meta-analysis cannot be effectively completed, restricting the ability to consolidate evidence.

Since outcome measures and results are not directly comparable, the aim of this thesis is to provide a strategic management perspective on the pediatric XR industry, capturing how research is conducted, challenges the current approach has on stakeholders and innovation, and identifying recommendations for future improvement.

To achieve this, a systematic review of published evidence focusing on the methodologies employed is necessary to uncover best practices and raise the case for the creation of industry guidelines. This is especially important because XR devices and related software with therapeutic or diagnostic uses can be classified as medical devices by regulatory bodies in the US, EU, Canada and increasingly some Asian countries, which can lead to significant costs and

prevent the ability for developers to bring solutions to market quickly. This thesis is relevant for researchers, companies in the XR industry, and regulators who can be inspired to build comparable efficacy and experience testing guidelines which are aligned to shared goals and meet the specific care needs of distinct patient groups.

The objective of this thesis was to identify and assess the metrics and tools used to measure patient experience when XR is applied during medical treatment. The tools used to collect patient experience data in the studies identified were then compared to generate a deeper understanding of which tools provide the greatest insights into experience for patients in their role as technology users, with the intention of identifying best practices for measurement as meta-hospitals increase in prevalence.

The review conducted for this thesis aimed to address the following research questions:

1. For which therapeutic purposes and procedures are XR technologies being applied in pediatrics?

2. What are the metrics used to measure pediatric patient experience and satisfaction?

3. What are the measuring methodologies and modes of delivery being applied in research of the therapeutic uses of XR, including who (patient, caregiver, and/or clinician) reports information?

2. BACKGROUND

2.1 Extended Reality and The Rise of Metaverse-Related Technologies in Healthcare

2.1.1 Definitions of the Types of Extended Reality

Virtual care delivery expanded rapidly during the Covid-19 Pandemic, years ahead of schedule for many countries, which demonstrated the potential for these services to overcome access barriers and sparked further investment [50]. A growing area of innovation is the connection of medicine and the Metaverse, where healthcare will gradually be able to combine virtual and real medical worlds [51]. Notably, XR technologies are a significant part of the several innovative technologies shaping and enabling the future Metaverse [52]. With these interactions increasing, it would be wrong to omit their measurement and doing so could lead to distorted perceptions of quality of care and satisfaction.

XR technology is an umbrella term comprised of the following digitally enhanced immersive experiences:

- **Virtual Reality (VR)** typically creates a new environment which can be similar or completely different from the real world [52]. Non-immersive VR is typically uses a desktop monitor or television, while fully immersive VR uses 3-dimension displays such as head-mounted displays for users to feel as though they exist within the virtual environment [53].

- **Augmented Reality (AR)** combines the user's real-world sight with simulated objects added or superimposed [54]. A key feature is accurate three dimensional registration of virtual and real objects [52].

- **Mixed Reality (MR)** enables "direct interaction between simulations and the physical world [54]". It integrates both virtual and digital environments and users can interact with both [52, 55].

2.1.2 Medical Extended Reality Hardware Types and Sample Applications

Table 1 introduces sample hardware and medical use cases for the three categories of XR. When considering hardware, Shen, et al. 2020 identified that each equipment variety has unique benefits and limitations such as cost, set up, comfort, heaviness, fidelity of graphics, sense of presence, ability to monitor externally, and licenses required [56].

Table 1: Types of XR technologies and their uses in healthcare settings

Intervention	Types of Hardware	Examples in Healthcare	Reference
Virtual Reality (VR)	• Smartphone VR headsets (Google Cardboard) [56] • Tethered PC-based VR headsets (HTC Vive) [56] • Standalone VR headsets (Oculus Quest) [56] • Console VR (PlayStation VR) [56]	• Influence of VR soccer game on walking performance in robotic assisted gait training for children [57] • VR for reducing school impairment resulting from chronic pain [58]	• Brütsch, et al. (2010) • Logan, et al. (2023)
Augmented Reality (AR)	• Untethered headset (Microsoft HoloLens) [8] • Headset (Mira AR Headset) [31]	• 3-dimensional visualization for preoperative assessment of anatomical structures in children with Wilms tumors [59] • Review that AR assistance shows promise within neuro-oncology, neurovascular surgery, spinal neurosurgery, skull-base surgery, and pediatric neurosurgery, but many studies are only in preliminary stages [60]	• Wellens, et al., (2019) • Hey, et al., (2023)
Mixed Reality (MR)	• VR Glasses with Motion Tracker (Leap Motion Tracker), in addition to a mannequin to conduct the training on [61] • Untethered headset (Microsoft HoloLens 2) [62]	• Simulation for newborn life support training [61] • Teaching in a university hospital setting [62]	• Coduri, et al., (2023) • Johnston, et al., (2024)

While benefits from XR are announced widely [63-65], capturing these benefits in a structured and systematic manner is missing and needed to enable continued innovation for medical use.

Measurement is important for capturing the generation of value in healthcare, from better health outcomes and experience to the ability to compare healthcare systems internationally, all of which benefit from collection of patient perspectives. Hospitals are becoming increasingly

digital with XR tools being embedded throughout the patient care journey, yet our approach to experience measurement has not followed. There is a research gap of no technology User Experience (UX) methodology review of XR technologies within the healthcare setting, and pediatric patients offer the most diverse mix of experience measurement methodologies, in part due to the split of response collection between the young patients and their caregivers. Several studies apply XR technologies in hospital settings; however, there is no review which focuses on the metrics used to measure pediatric patient experience and satisfaction when leveraging these therapeutic technologies. Through optimal measurement, hospitals can make pediatric patient experiences more visible, accessing the potential for further value generation with clinical advanced technologies, where even the youngest patients are able to take a more active role in their care journeys.

2.2 Industry Growth and Opportunities

XR has many diverse healthcare applications, many of which have the ability to enable inexpensive and scalable solutions to help address some of the world's greatest healthcare challenges, including capacity with clinical workforce shortages [66]. XR applications include education and training [67], surgery support, monitoring handwashing, remote assistance from specialists [68], virtual tours [69], pain and anxiety management [70], and rehabilitation [71].

VR use alone in the healthcare market is a USD $4.05 billion industry, projected to grow to USD $13.5 billion by 2029 [72]. While increasingly used, sentiment for this technology varies widely from the executive perspective. In a worldwide 2022 survey of healthcare executives conducted by Accenture, sentiments for the influence of the Metaverse on healthcare varied from "minimal" (19%) and "incremental" (32%) to "breakthrough" (34%) and "transformational" (15%) [73].

2.3 Importance of Measurement for the Capture of Value in Hospitals

As policy makers push towards more patient-centred care, there is increasing urgency placed on understanding patient perspectives of care experienced and the influence of care received on overall health, reaching beyond clinician-reported measures [74, 75]. Value in healthcare considers the interests of patients, care providers, suppliers, and payers alike, and often includes

the following high-level categories: quality, safety, patient centricity, and cost management [76]. To improve the quality and effectiveness of healthcare, patient perspectives are highly relevant [74]. The Organisation for Economic Co-operation and Development (OECD) described patient-centred care and the challenges of its measurement in 2017 as the following:

> [A]n objective that is regularly used by policy makers and clinical leaders as defining the way in which they believe health care should develop. However, metrics on whether or not this is being delivered are largely absent, even if there is wide consensus that there are huge benefits to giving providers as well as decision makers such information. [77]

Physicians overwhelmingly believe there should be financial incentives aligned to quality of care, yet few physicians trust the accuracy of their performance metrics [78]. In 2017, the OECD compared the use of Patient-Reported Outcome Measures (PROMs) and Patient-Reported Experience Measures (PREMs) for recommendations of patient-reported indicators of healthcare system performance, emphasizing the need for standardized, validated instruments [77]. Following this, the OECD released the internationally-tested Patient-Reported Indicator Surveys (PaRIS), with the 'Main Survey' released in 2023 following development and testing between 2018-2022 [75]. PaRIS, which includes PROMs and PREMs, focuses on measuring health behaviours, health capabilities, experiences of care, and outcomes [75]; however, there are no questions which directly address the experience of patient interactions with advanced technology despite the increase in interactive technology applications.

The OECD *People-Centred Health Systems Scorecard: Key indicators of health systems' voice, choice, co-production, integrated care, and respectful care* (2021) added two digital technology indicators to their patient-centred care model: Co-production – Use of digital tools for patient engagement, and Integrated Care – Use of digital technology for integration of care [50]; however, the scorecard indicators are focused on patient use of health portals/apps and physician use of computers for prescriptions, referrals, and orders, respectively [50].

2.4 Pediatrics and Heterogeneous Measurement of Value

The pediatric population offers unique aspects which are particularly valuable to review when considering XR experience measurement approach: their digital literacy and limitations of their age. Pediatric patients have been exposed to digital technologies for their entire lives, so they

have a desire for and understanding of how to engage in the virtual environments [79]. Moreover, their age introduces additional opportunities and limitations for measurement; pediatric studies have a mixture of measuring just the patients, just their caregivers, or both [80]. Considering outcome measurement (PROMs), the methods for measuring pain and fear are influenced by age, with the youngest of pediatric patients unable to self-report [46].

Even without considering advancing technologies, experience measurement in pediatrics is weak due to the variance of methods used, many of which do not address key metrics like healthcare quality. There is also no standardized or widely accepted method for measuring satisfaction [81]. Ferreira, et al. (2022) suggests there is a need for more rigor among researchers and pediatric surgical providers in their approach to patient-centered measures of experience and satisfaction [82]. PROMs and PREMs are increasingly recognized as indicators of quality of care, yet many studies exclude PREMs, design measures in-house rather than using validated tools, or measure satisfaction alone, which is considered to be a measure of expectation, rather than quality [82].

While measurement of satisfaction when applying XR is of interest for this review, measurement of "satisfaction" alone also does not capture the experience of using XR due to the variety of criteria patients use to measure satisfaction during their interactions in hospital. When evaluating the main criteria underlying satisfaction of pediatric inpatient "customers" (patients), Ferreira, et al. (2021) identified the following themes, none of which have sub-criteria with a technology interaction focus: capacity of providing useful information, facilities, visits, food, doctors, nurses, auxiliary staff, volunteering, diagnosis and treatments, and discharge procedure [83]. Similarly, parent satisfaction may not be sufficient for measuring the patient's experience when applying XR for pain distraction since parent satisfaction has been found to not be statistically different from parent perception of their child's pain [84].

2.5 The Development Process and Regulatory Environment of XR

2.5.1 Relevant Stakeholders

Evans (2022) describes several stakeholders of medical XR, categorized in Table 2 [85]:

Table 2: Stakeholders of Medical XR

Stakeholder Group	Examples
Patients	Patients, as direct users of the XR or benefiting from care team use
Caregivers	Often parents, the adult bringing the child to their appointment
Care Teams	Doctors, nurses, clinicians, healthcare providers
Regulatory Bodies	Governments and associated agencies, regulatory bodies
Research Teams	Scientists, researchers, Contract Research Organizations (CROs)
Industry	Information Technology (IT) professionals, game developers, business and clinical executives, device supply chains, insurance companies
Other Supports	Patient advocacy groups, social workers, support staff

2.5.2 Regulatory Bodies and Processes

XR platforms and devices are regulated by the Medical Device Regulation (MDR, Regulation [EU] 2017/745) in the EU [86], and the United States Food and Drug Administration (FDA) in the US [87]. In both cases, regulation of the devices is organized by risk, leading to different regulatory pathways for approval [87]. According to level of risk (low, medium, or high), more extensive efficacy testing methodologies and funding are required [87]. XR would typically be categorized as low or medium risk; however, the focus of this review is not registration performance. Several stakeholders (Table 2) play a role in the process to regulatory approval. Industry professionals adapt from existing registered devices, research teams influence the outcome measures used by regulators when standards do not already exist, and care teams and patients directly use the technology in later trial stages.

An emerging integration with XR, which did not arise in the included review literature, is Artificial Intelligence (AI). If using AI in the EU, the EU AI Act (Regulation [EU] 2024/1689, 13 June 2024) would apply [88], further increasing regulatory complexity for industry innovators. The applications that the AI Act applies to add legal uncertainty, where it is not always clear to industry leaders what provisions apply to a given use case nor why certain issues have not been covered in the Act. The AI Act is expected to serve as a model for the regulation of AI in other countries and is considered complementary to the MDR [88]; however, according

to Bini, et al. (2023), there are properties of AI software that may oppose the MDR when trying to put a medical device on the market, such as the "assurance for no emerging risks" [86]. In tandem with efficacy of the devices, government bodies are focused on the safety and privacy of XR users [89]. Pediatrics is a particularly sensitive area for this topic since researchers and companies must be even more careful when collecting data from children, and there can be "significant privacy risks" in AI-XR-enabled metaverses that have yet to be addressed [90].

2.5.3 Regulatory Gaps Identified by the FDA

Development of standardized methods and frameworks to evaluate XR safety and effectiveness is needed, especially due to the novelty of its use cases [91]. The FDA has established a Medical Extended Reality Program to begin to address three gaps within the current regulation of XR, which have identified as follows: (1) there are a wide variety of platforms which lack characterization and evaluation methods for different important medical applications; (2) certain hardware, such as sensors and cameras used in the platforms have not been validated for their clinical use contexts; and (3) there are little to no assessment tools for usability related to safety such as cognitive load of the user [92]. These gaps suggest that regulatory bodies do not yet have the methodology for how to best measure device efficacy and safety, yet there is an awareness that there and more devices and platforms used for medical purposes that they must measure.

2.5.4 Development Challenges

Industry experts have shared that the journey from prototype to clinic for XR products has significant technical and regulatory challenges [91], requiring consideration of cost-effectiveness, safety, and the resolution of technical challenges to achieve more widespread clinical adoption [91].

Considering industry stakeholders, such as game designers, the process and cost of approval is a significant barrier to development. Game designers need to balance engagement and effectiveness while adapting to specific feature needs of the users. Part of the design process to determine effectiveness is testing, which involves researchers as well. Espinoza, et al. (2022) shared that over 80% of medical device companies make no sales revenue and have less that 50

employees [87]. When compared to the average size of a pharmaceutical company developing health technologies with standardized, established methods, the companies developing XR software are much smaller and have much less specialized knowledge. These small companies are research-focused and often rely on grants, and the cost to develop and commercialize a low- to moderate-risk device is estimated to be about $31 million USD [87]. The high-cost process is slow and involves intense regulation, yet an increase in standardization of measurement would clarify parts of the process and provide smaller players with more guidance to succeed.

Producers must often adapt devices, not just for procedure-specific needs, but also due to unique demands of pediatric patients. In the context of medical devices, it is common to use devices designed for adults then adapt them for children, and needs and preferences can differ greatly [87]. There are several barriers to developing devices for children, leading to only one quarter of the number of adult devices being available for pediatric use [87]. When creating medical devices for pediatric patients, products need to be adjusted for smaller size, user growth, and possibly longer duration of use, while balancing smaller available sample sizes and significant data safety requirements. There is also a challenge of lower patient reimbursement rates and a lack of coverage standards in the United States, further exacerbating the financial burden [87].

2.6 Lessons of User Experience Research Beyond the Healthcare Setting

2.6.1 The Concept of UX Measurement

To design and effectively measure medical XR experiences, one should consider User Experience (UX) theories more broadly and measurement methodology used for XR studies beyond healthcare settings. While definitions vary, UX is largely considered to be a human–technology interaction, which is highly context-dependent, dynamic, and subjective or user-related [93, 94].

Beauregard & Corriveau (2007) propose that UX measurement should go beyond satisfaction surveys and usability studies by bringing light to the psychological nature of UX. Their interaction-based framework details how UX is accessed through self-report, behavioral observation, and other proxies, such as physiological proxies, of cognitive processes. When considering the concept of interaction between a user and product, there are emotions, thoughts,

and attitudes which arise, influencing intentions and observable interactions with the product. Each individual's interaction is influenced by their unique knowledge/experience, concerns/expectations, skills/abilities, personality, and physical attributes [95], which are the type of factors that should be considered for selecting technology interventions in health settings. To extend this to the pediatric environment when selecting XR elements, patient age is correlated with video game skill/ability and size, influencing patient interest in interacting with a particular game environment and ability to wear equipment [8].

2.6.2 Gap in UX Measurement Agreed Upon Standards

Lallemend, et al. (2015) conducted a survey of 758 international UX practitioners and researchers to assess agreement with statements about UX. When comparing if data collected should be quantitative or qualitative, no clear consensus emerged, with participants identifying quantitative measures as "useful to convince decision makers to modify a problematic design" and qualitative feedback as "easier to derive alternative design ideas from [94]".

A systematic review of empirical research on UX found that of 66 studies reviewed, 50% used qualitative methods, 33% used quantitative methods, and 17% used a combined approach [96], further emphasizing that there is no defined best practice for measuring user experience. Dimensions of UX assessed included emotions and affect (24%), enjoyment (17%), aesthetics (15%), engagement (12%), motivation (8%), and frustration (5%) [96]. Moreover, the data collection methods varied to include questionnaires, semi-structured interviews, live user observations, focus groups, open interviews, body movements, and psychophysiological measures such as heart rate [96]. Expanding the challenge of variation in views of what UX captures, this variation in measurement methodology across user settings makes the studies more difficult to compare and consistently extract high quality experience details.

2.7 The State of Medical Extended Reality Research

Focused on the most common technology intervention in this review, VR, 21 leading VR experts described the state of clinical VR research as the "Wild West" [97]. This group of experts, referred to as VR-CORE (Virtual Reality Clinical Outcomes Research Experts), brought attention to the heterogeneous nature of clinical VR research where clear guidelines

and standards were lacking, and studies were small case reports, retrospective analyses, merely descriptive, or did not employ experimental designs [97]. One of the key recommendations was the following:

> Regardless of the inclusion of a control group, investigators should identify a clinically relevant and validated patient-reported outcome (PRO) to evaluate the evidence of efficacy… Selection of the most appropriate PRO (Patient Reported Outcome) is at the discretion of the research team, but should be carefully justified and capture the most salient features of patient-reported health that might improve with the VR treatment. [97]

Examples VR-CORE listed for PROs included an 11-point numeric rating scale to assess pain and a measurement information system for physical function.

3. METHODOLOGY

This review was conducted to understand the current approaches to outcome and experience measurement when pediatric patients engage with XR and check for consistency and comparability of the tools used. The process involved a large magnitude of empirical research, with over 5300 search results identified for the review. There are a small number of narrative reviews associated with building guidelines for medical applications of XR; however, there is no known research which takes a systematic approach or focuses on the methodology used for pediatric patients as technology users.

3.1 Design of the Review

A systematic review utilizing narrative analysis was undertaken. To facilitate transparency, reproducibility, and minimize potential of reporting bias, the review was registered to PROSPERO, a register which accepts systematic reviews with health-related outcomes, funded by the UK National Institute for Health Research (NIHR) [98]. The protocol for the review was registered on 14 April 2024 with the following PROSPERO ID: CRD42024531293. The process of registry with PROSPERO reduces duplication of systematic reviews by different authors by displaying the title and methodology of ongoing research [98].

Search Strategy

Research databases PubMed, Cochrane Central Register of Controlled Trials (CENTRAL), and Web of Science were searched on 05 April 2024 using the following search string:

> *("Pediatric" OR "Paediatric") AND ("Extended reality" OR "XR" OR "Virtual reality" OR "VR" OR "Augmented reality" OR "AR" OR "Mixed reality" OR "MR") AND ("Experience" OR "Survey" OR "Satisfaction" OR "Measure")*

Only peer-reviewed articles published since 1990 were considered, with no language restriction, reducing bias. The search strategy was constructed in alignment with the PICO method, with details of the Population, Intervention, Comparison, Outcomes, and Study design detailed in this section [99]. Search term strings included a combination of the population (“Pediatric” or “Paediatric”), the relevant technology intervention type (“Extended reality” or “XR” or “Virtual reality” or “VR” or “Augmented reality” or “AR” or “Mixed reality” or “MR”) and elements of measuring experience (“Experience” or “Survey” or “Satisfaction” or “Measure”), which was the main outcome of interest.

3.1.1 Included Study Design

Studies with randomized controlled trials (RCTs) or quasi-experimental design were primarily included in the systematic review; however, qualitative studies were not excluded since they are also able to effectively capture patient experience elements. All studies included an intervention with human participants; systematic reviews and meta-analyses were excluded.

The studies had to include at least one form of measurement for the patient experience related to the technology intervention, such as enjoyment of procedure process or whether the technology was easy to use. There was only a single reviewer completing the screening and extraction steps of the review, but discussions with the supervisor mitigated risks for bias.

3.1.2 Population

The population for review was pediatric patients (21 years of age or younger) undergoing medical procedures. As such, studies with participants over the age of 21 were excluded. Given

the age of participants, studies where parents provided experience-related feedback on behalf of their child were also included, so long as the child was the participant of the XR intervention.

3.1.3 Interventions

Interventions were therapeutic uses of XR by pediatric patients. The specific use cases, such as pain management, distraction, or physical movement, varied. Excluded interventions and uses within the pediatrics area were those that were not patient-focused, such as medical team education or procedure visualization, or for non-therapeutic use, such as XR used for communication during stay but not during a procedure, or a virtual tour before a procedure.

3.1.4 Comparators/Control

The intervention group in each study was compared with the standard/usual care for the given therapeutic use that the study was testing on. Having a control group was not a requirement depending on study method. For the studies with control groups, the control groups underwent the same procedure as the intervention groups, with the typical procedures and standard of care.

3.1.5 Outcomes of the Review

The main outcome of this review was types of measurement methodologies addressing patient experience. The review sought to analyze metrics and tools used, and how ratings and qualitative measures of experience are collected, attempting to identify how advanced patient-facing technology trials are measured. Outcomes from this review include best practices for future trials and whether there is consistency in patient XR experience measurement techniques. The additional outcomes reviewed were the intended patient-reported outcomes of the XR interventions and how they are measured. Some of the outcomes and effects measured include pain, anxiety, and fear, range of motion, and severity of nausea.

3.1.6 Data Extraction (Selection and Coding)

The screening and extraction of article title, abstract, and full-text review took place in Covidence, a software tool for systematic review organization. Adapted based on *CRD's*

guidance for undertaking reviews in healthcare [100], the following information was recorded in Excel then later summarized within the review. Initial data extraction of search results took place on 05 April 2024, recording the following identification features: record number, author(s), article title, country of origin, and citation. From this stage, excluded results were documented following the PRISMA *(Preferred Reporting Items for Systematic Reviews and Meta-Analyses)* 2020 statement [101].

Study characteristics extracted included objectives, study design, inclusion and exclusion criteria, and recruitment and allocation approach. Participant characteristics included age, gender, and disease or procedure that the intervention was for. The number of participants in the intervention and control groups as well as the number of recruited participants that were ineligible were recorded. The setting and purpose of treatment (area of hospital and reason for use) were captured as well as the technology intervention (VR, AR, MR) applied.

The main outcome measured related to study design, so there were extracted details of patient experience measurement, comprised of the following: method or tool/framework(s) used with scales or units where applicable, who was asked to report on experience metrics (participant/parent/practitioner), how reporting was conducted (surveys/interviews), and the questions asked. This component of study design was the main focus of the review and hopes to contribute best practices and/or important metrics to include within future pediatric experience research design when incorporating advanced technology interactions.

Relevant to the intention of the studies themselves, additional outcomes recoded related to the technology interventions; namely, patient-reported outcomes and their measurement, adverse effects of use, and challenges encountered with testing, such as motion sickness or injuries impeding headset use. While the protocol hoped to capture technology cost data, this category was removed since the relevant data was infrequently available in the studies and the XR tools have become much less expensive over time.

3.1.7 Strategy for Data Synthesis

The strategy for data synthesis, given patient experience is qualitative in nature, involved initial descriptive synthesis followed by narrative synthesis. Narrative synthesis of the qualifying

studies included groupings and clusters of the type of intervention (VR and AR, there were no MR studies included) and the purpose of the intervention, translating data through content analysis, and providing graphs and frequency distributions to explore relationships between studies. Frequency distributions were particularly useful to understand the different themes of questions asked to assess experience and their relative popularity.

3.2 Systematic Process for Selecting Included Studies

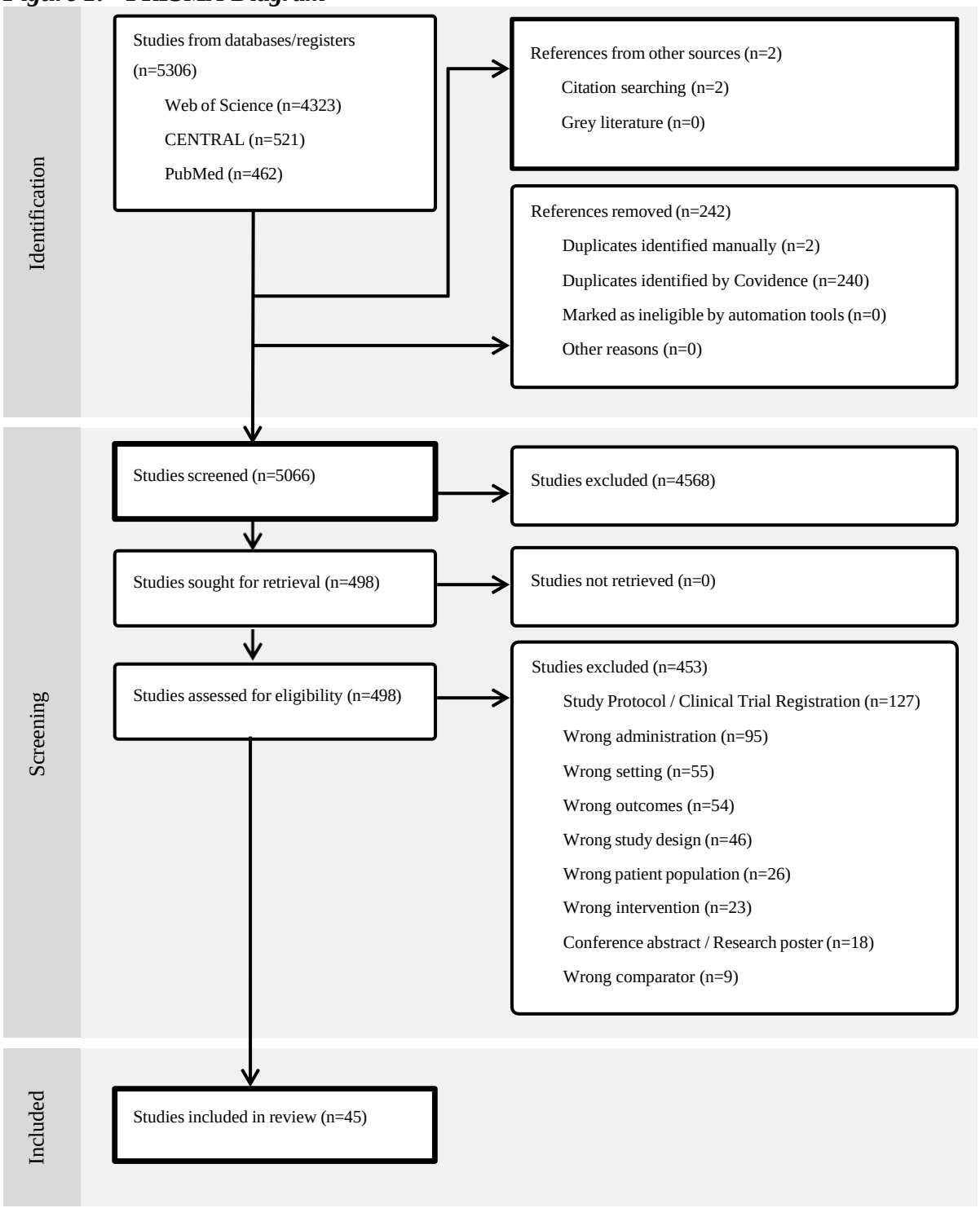
3.2.1 Identification, Screening, and Retrieval of Relevant Literature

The *PRISMA 2020 Statement* [101] was used to track the process, supported by the systematic review management tool, Covidence. Figure 1 details the process of gathering and screening relevant literature. The initial search, as defined in the PROSPERO registry and methodology detailed earlier, returned 5306 results. Two papers were added from citation searching, where a protocol registry was included in search results, but not the related complete study. Following the removal of 242 duplicate studies, 5066 studies were screened by title and abstract. Of the studies screened, 498 were retrieved for full-text review against the defined inclusion criteria.

3.2.2 Assessment of Eligibility

Of the 453 results excluded at the full-text review stage, 127 were protocols or trial registrations rather than full studies (full studies were searched for and many already included in the initial database search), 95 had the wrong administration of technology, 55 took place in excluded settings (i.e. homes, dental offices, schools), 54 focused on the wrong outcomes (often only PROMs, not UX), 46 had the wrong study design (for example, only a focus group on XR feasibility [58]), 26 had the wrong population (above age 21), 23 used the wrong intervention (not a type of XR, such as watching television), 18 were conference abstracts or research posters, and 9 had the wrong comparator (comparing types of XR interventions, not having an appropriate standard of care as a control). Ultimately, 45 studies met all inclusion criteria for the review.

Figure 1: PRISMA Diagram



3.3 Rating and Review of Included Studies

Table 3 details the 45 papers included in the results, used to address research questions and build recommended practices. The type of XR intervention applied and category of procedure that the application was applied to are detailed in the table.

Following the engagement rating process defined below, a rating of low, medium, or high was assigned to each included paper for the quality and breadth of engagement measurement, as shown in the right-most column of Table 3, to focus recommendations on methods which provide the strongest patient insights. The following three categories were used for scoring:

Low: A couple of questions, only the caregiver provided insight, and/or the insights generated were very superficial

Medium: In depth insights from only one group (patient, caregiver, or clinician), and/or superficial insights from multiple points of view

High: Used at least one deep instrument to capture multiple views of patient experience

The mix of studies by rating is described in Figure 2, showing the engagement of 12 studies (26.7%) were assigned a low engagement rating, 22 (48.9%) were medium, and 11 (24.4%) were high.

Figure 2: Mix of Included Studies by Experience Engagement Rating

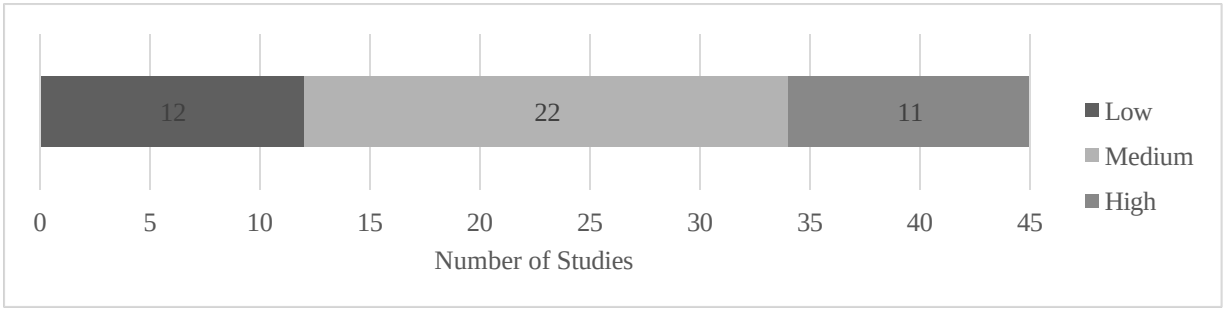


Table 3: Summary of References Included in Review [1-45]

#	Publication Reference	Intervention	Procedure Category*	Engagement
1	Acar et al., (2016)	Non-immersive VR	Rehabilitation	Medium
2	Agrawal et al., (2019)	Immersive VR	Cancer and Blood Diseases	Medium
3	Alrimy et al., (2023)	Non-immersive VR	Burn	Medium
4	Ammann-Reiffer et al., (2022)	Immersive VR	Rehabilitation	High
5	Atzori et al., (2018)	Immersive VR	Cancer and Blood Diseases	Medium
6	Atzori et al., (2022)	Immersive VR	Needle-Related	Medium
7	Caruso et al., (2020)	Immersive VR	Needle-Related	Medium
8	Chamberland et al., (2024)	Augmented Reality	Anesthesia	Low
9	Chan et al., (2019)	Immersive VR	Needle-Related	Medium
10	Clerc et al., (2021)	Immersive VR	Surgery	Low
11	Dumoulin et al., (2019)	Immersive VR	Needle-Related	Medium
12	Dunn et al., (2019)	Immersive VR	Needle-Related	High
13	Ellerton et al., (2023)	Immersive VR	Needle-Related	Low
14	Gold et al., (2006)	Immersive VR	Needle-Related	Low
15	Gold & Mahrer, (2018)	Immersive VR	Needle-Related	High
16	Gold et al., (2021)	Immersive VR	Needle-Related	Medium
17	Goldman & Behboudi, (2021b)	Immersive VR	Needle-Related	Low
18	Goldman & Behboudi, (2021a)	Immersive VR	Surgery	Low
19	Hernández et al., (2018)	Non-immersive VR	Rehabilitation	Medium
20	Hoffman et al., (2020)	Immersive VR	Burn	Medium
21	Jain et al., (2024)	Immersive VR	Burn	High
22	Jeffs et al., (2014)	Immersive VR	Burn	Medium
23	Jivraj et al., (2020)	Immersive VR	Cast Removal	Low
24	Jung et al., (2021)	Immersive VR	Anesthesia	Medium
25	Jyskä, et al., (2023a)	Non-immersive VR	Needle-Related	Low
26	Jyskä, et al., (2023b)	Non-immersive VR	Needle-Related	Low
27	Kucher et al., (2020)	Immersive VR	Surgery	Low
28	Le May et al., (2021)	Immersive VR	Surgery	Medium
29	Lee Wong et al., (2022)	Immersive VR	Cancer and Blood Diseases	High
30	Litwin et al., (2021)	Immersive VR	Needle-Related	Medium
31	Madill et al., (2020)	Augmented Reality	Surgery	Medium
32	Mohamed & Mohamed, (2023)	Non-immersive VR	Needle-Related	Medium
33	Nilsson et al., (2009)	Non-immersive VR	Needle-Related	High
34	Olivieri et al., (2013)	Non-immersive VR	Rehabilitation	Low
35	Osmanlliu et al., (2021)	Immersive VR	Needle-Related	Medium
36	Phelan et al., (2021)	Immersive VR	Rehabilitation	High
37	Richey et al., (2022)	Immersive VR	Cast Removal	Medium
38	Schmitt et al., (2011)	Immersive VR	Burn	Medium
39	Schneider & Workman, (2000)	Immersive VR	Cancer and Blood Diseases	Medium
40	Shen et al., (2022)	Immersive VR	Rehabilitation	High
41	Taylor et al., (2021)	Immersive VR	Surgery	Medium
42	Tennant et al., (2020)	Immersive VR	Cancer and Blood Diseases	High
43	Walther-Larsen et al., (2019)	Immersive VR	Needle-Related	Low
44	Winkels et al., (2013)	Non-immersive VR	Rehabilitation	High
45	Xiang et al., (2021)	Immersive VR	Burn	High

* Rehabilitation Category includes the following: Cerebral Palsy, Upper Limb Injury, TBI, and Hemiparesis

4. RESULTS

4.1 Analysis of Patient-Reported Outcome Measures (PROMs) Methodologies

Table 4, organized by outcome into subsections as 4.1-4.5, showcases that there are a lot of different PROMs in XR-based interventions in pediatrics with varying degrees of reliability and validity used to measure patient outcomes. One technique for assessing PROs is the comparison of before and after, which was used to assess the efficacy of XR in reducing pain [3,7,15,16,18,31,33,35,37] and anxiety [3,18,26,37]. There was, however, a great degree of variety within the pain and anxiety PROMs, which ranged from reliable and valid scales to self-made scales and questions, which are challenging to compare across studies.

4.1.1 Pain

For pain measurement, the Faces, Legs, Activity, Crying, Consolability PROM contrasts the Atzori et al. (2018) use of a Visual Analog Scale to evaluate components of pain: cognitive (time spent thinking about pain), affective (pain unpleasantness) and sensory (worst pain) [5]. While not directly comparable, several pain PROs are highly correlated for measuring pain intensity and unpleasantness; however, there are distinct differences in preference for pediatric patients which should be considered by researchers [102]. In addition to the contribution that the Numeric Rating Scale shows evidence of construct validity and sensitivity for measurement of pain intensity and unpleasantness, Pagé et al. (2012) shared that children find faces scales easiest to use while Verbal Rating Scales are liked least and considered the hardest to use [102].

Table 4.1: Frameworks and tools used to gather pain outcome insights

Outcome	PROM Examples and Sources
Pain 34 (75.6%)	• WongBaker Faces Pain Scale Rating [11,14,27,32] • Faces Pain Scale – Revised [7,9,10,13,14,15,16,17,18] • GRS (Graphic Rating Scale) [3,20,38] • Numeric Rating Scale [28,30,31,37] • Verbal Numeric Pain Scale Rating [6,35,41] • Visual Analog Scale [5, 11,13,15,21,43,45] • Adapted Visual Analog Scale [26] • Coloured Analog Scale [15,33] • Adolescent Pediatric Pain Tool, Word Graphic Scale [22] • Facial Affective Scale [33] • Adjusted Pediatric Pain Tool [2]

	• Faces, Legs, Activity, Crying, Consolability (FLACC) [3,33,43]
	• FLACC – Revised [21,45]
	• OCCEB-BECCO –similar to FLACC but specifically for burn patients [28]
	• Children's Emotional Manifestation Scale [23]
	• Poker Chip Tool (self-rated pain score) [13]
	• Pain Catastrophizing Scale for Adults [41]
	
	• Questionnaire/survey developed to include some of the following constructs:
	  ○ Time spent thinking about pain [5,21,45]
	  ○ Pain unpleasantness [5]
	  ○ Worst pain [5]
	  ○ Expectation of pain [11,26]
	  ○ Pain intensity [11,26]
	  ○ How techniques impacted pain [12]
	  ○ Memory/recall of pain [28,35]

4.1.2 Anxiety

Anxiety PROMs show even more variability, comparing the objective measure of heartrate [29,30,33] and short-form modified Yale Preoperative Anxiety Scale [8,24] to the use of an unvalidated, made-for-study "novel nature-based anxiety scale" [27].

Table 4.2: Frameworks and tools used to gather anxiety outcome insights

Outcome	**PROM Examples and Sources**

Anxiety
22 (48.9%)
	• Heartrate [29,30,33]
	• Blood pressure [29]
	• Numerical Rating Scale [30]
	• Procedure Behaviours Checklist [3]
	• Graphic Rating Scale [3]
	• Subjective Anxiety Scale [3]
	• Short form modified Yale Preoperative Anxiety Scale [8, 24]
	• Visual Analog Scale [9,12,16]
	• Adapted Visual Analog Scale [26]
	• Venham Situational Anxiety Scale [10,17]
	• Childhood Anxiety Sensitivity Index [15,16]
	• Children's Emotional Manifestation Scale [23]
	• Pre-intervention survey [21,45]
	• Post-procedure questionnaire [22]
	• Spielberger State-Trait Anxiety Inventory for Children [22,24]
	• Short Spielberger State-Trait Anxiety Inventory for Children [23]
	• Penn State Worry Questionnaire for Children [23]
	• Novel Nature Based Anxiety Scale – own unvalidated framework [27]
	• Children's Anxiety Meter – State [37,41]
	• Brief Behaviours Distress Scale [41]
	
	• Specific questions:
	  ○ Effectiveness of technology of distraction from anxiety [12]
	  ○ Satisfaction with anxiety management [17]

4.1.3 Fear

Within the 45 papers, fear was the only PRO that consistently used the same PROM: the Children's Fear Scale (CFS). The scale is an adaptation from the McKinley et al (2003) Faces Anxiety Scale [103], designed for children undergoing painful experiences, and is a self-reported numerical scale showing five faces with expressions from neutral representing no fear (0) to extremely fearful (4) [28]. As should be the foundation of all PROMs, the CFS shows good evidence of test-retest and interrater reliability [28]. In addition to papers where fear was self-rated [7,32,35], there were also studies assessing fear where the CFS was rated by the clinician [13,32] and the caregiver [7,32,35].

Table 4.3: Frameworks and tools used to gather fear outcome insights

Outcome	PROM Examples and Sources
Fear 10 (22.2%)	• Children's Fear Scale [7,13,14,28,30,31,32,35,37,41] • Recall of fear [28]

4.1.4 Individual Improvement

Individual improvement has a rational level of variation due to the need for measuring several types of behaviours and physical movements. Almost all studies with this type of PRO were for "rehabilitation", though there was one "burn" study measuring range of motion.

Table 4.4: Frameworks and tools used to gather improvement outcome insights

Outcome	PROM Examples and Sources
Individual Improvement 8 (17.8%)	• Quality of Upper Extremity Skills Test [1] • Jebsen Taylor Hand Function Test [1] • ABILHAND-Kids Test [1,44] • Pediatric Functional Independence Measure [1] • Gillette Functional Assessment Questionnaire [4] • Functional Mobility Scale [4] • Movement Behaviours [4] • Level of physical exertion [40] • Canadian Occupational Performance Measure [19] • Melbourne Assessment for Upper Limb Movement [34,44] • Ashworth Scale [34] • Range of Motion [36,38]; Arm's Passive Range of Motion [34] • Conners' Continuous Performance Test 3rd Edition [40] • Pediatric Quality of Life Core Scale [40] • Child Behaviour Checklist [40] • Behavior Rating Inventory of Executive Function 2 [40]

4.1.5 Other Outcomes

The most common additional outcomes measured involved motion and/or cyber sickness [4,40] and duration [12], both of which relate to how patients evaluate experience.

Table 4.5: Frameworks and tools used to gather other outcome insights

Outcome	PROM Examples and Sources

Other
16 (35.6%)
	• Simulator/cyber sickness, nausea [2,4,5,6,11,14,15,20,21,25,27,29,38,40,42,45]
	  ○ Malaise [15], BARF Scale [23], Nausea GRS [20], observe for signs [25], simulator sickness questionnaire [40]
	• Time/duration/length of procedure [5,9,10,12,17,18,20,21,22,24,41,43,45]
	• Procedural Compliance [7,31,41]
	  ○ Modified Induction Compliance Checklist [7]
	• Need for restraint [9]
	• Number of needle attempts/successes [9]
	• Need for sedation [9,41] or administration of anesthetic [10]
	• Parental anxiety [24]
	• Completion of treatment sessions [34]

4.2 Addressing Research Questions

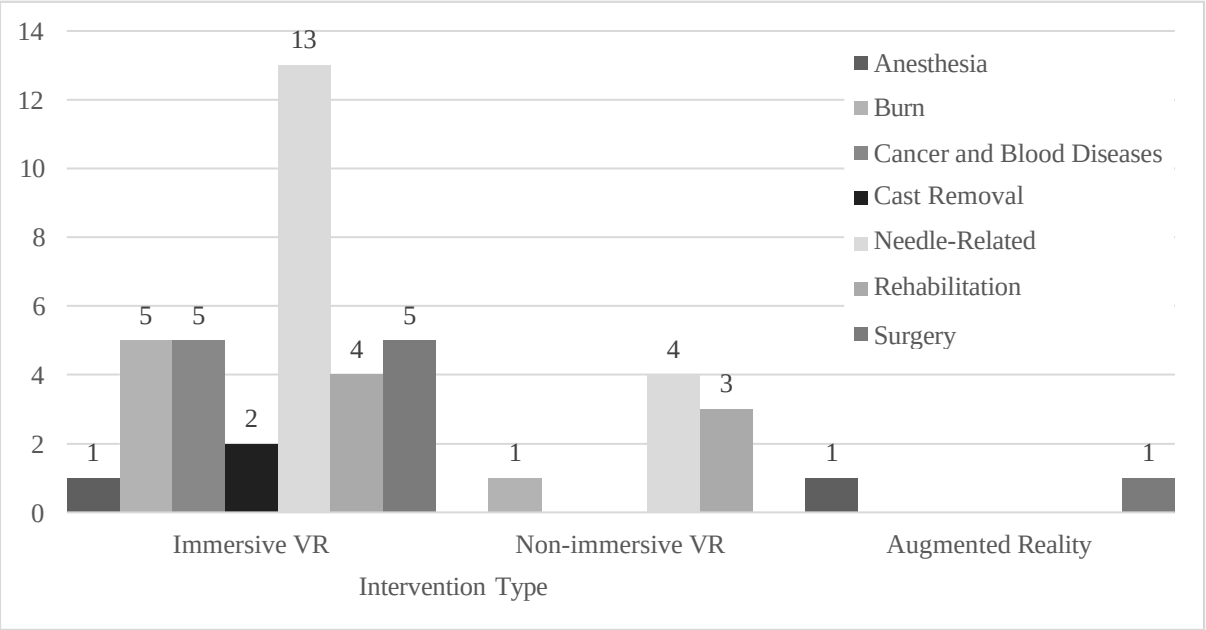
4.2.1 Procedure Types by XR Intervention Category

To begin to address the measurement of UX it is important to understand the mix of procedures and XR interventions included in the review, which are detailed in Figure 3. Immersive VR was the most common intervention (77.8%), used for all procedure types, while AR (4.4%) was only represented in one surgery and one anesthesia study. Non-immersive VR (17.8%) was used for burn, needle, and rehabilitation studies, and at a lower frequency than Immersive VR. Needle-related procedures represented the most common procedure studied (37.8%) while anesthesia was studied least.

MR applications were not represented in the studies reviewed since none met the inclusion criteria determined in the registered methodology. It can be found from the excluded studies that MR uses tend to be for users beyond the patient. Outcomes and uses in the excluded MR studies were tele-consent through increased access to information and explanation of inter-operative findings for caregivers [104], more accurate diagnostics for clinician use, and more effective training received for students since MR extends beyond the limitation of VR where

manual skills cannot be simulated in a realistic [61]. These MR uses show there are far more possible outcomes to measure for XR when looking beyond pediatric patients as users.

Figure 3: Frequency of Procedures by Intervention Type



4.2.2 Metrics Used to Measure Pediatric Patient Experience

Narrative analysis was used to categorize the user experience constructs contained within the 45 studies and have been summarized in Table 5. Measuring fear and anxiety as PROs is unique because they are not mutually exclusive from UX metrics [105], with emotion and affect being relevant for PREMs as well. Moreover, anxiety can be measured in several ways, with studies considering a mix of trait anxiety [24], state anxiety [14, 23], anticipatory anxiety [24], and psychophysiological markers of anxiety (heartrate [29,30,33] and blood pressure [29]).

Since measurement of anxiety and fear could impact the UX dimension of "Affect, Emotion", "anxiety" and "fear" were excluded from the PREMs defined below since they were captured as PROs. In other words, while anxiety and fear measures would also influence "Affect, Emotion", the PREMs tracked for frequency were kept within the Table 4 constructs to separate the UX measurement from the primary patient-reported outcomes in the included literature.

Table 5: Metrics Analysis – Categorization of UX Constructs

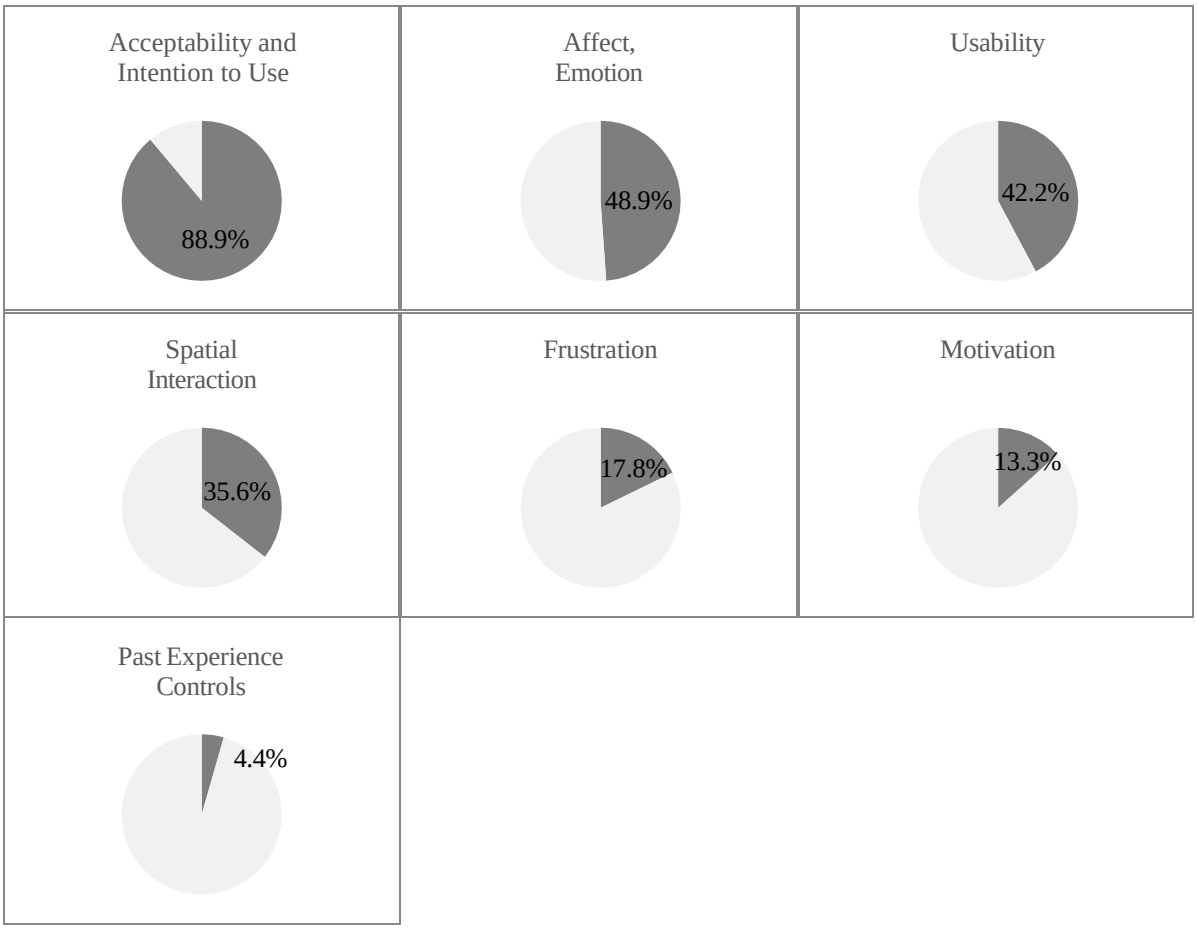
Experience Dimensions	Included Constructs		
Acceptability and Intention to Use*	• Helpful • Worth implementing • Supported procedure • Efficacy of distraction	• Use again • Overall experience/opinion • Acceptability	• General or procedure-based satisfaction • General sentiment
Motivation**	• Need for external motivation	• Motivation	
Affect, Emotion**	• Joy • Fun • Engaging	• Boredom • Enjoyment	• Relaxed • Enthusiasm
Usability (Abilities and Physical Attributes)	• Difficulty • Ability • Cognitive Ability • Comfort	• Ease of use • Physical Equipment • Adapted • Suitable	• Keep wearing • Safety • Removal of equipment
Spatial Interaction***	• Realism • Immersion	• Presence • "Went inside"	• Ability to manipulate objects
Experience Controls	• Expectation of Helpfulness	• Expectation of Fun	• Previous XR experiences
Frustration**	• Disadvantages • Suggestions	• Challenges	• Dislikes

* Intention to Use is in reference to the concept of behavioural intention [105]
** UX Dimensions [96]; "Affect, Emotion" excludes pain, anxiety, and fear since covered in PROMs
*** Influenced by the 3 domains of immersion: sense of involvement, perceived realism of the VR experience, and sense of transportation into the experience [16]

From the categorization in Table 5, Figure 4 showcases the frequency of use of each dimension. The vast majority (88.9%) of studies included a construct of acceptability of the technology or the desire to use it again; however, the remaining categories were less common. Almost half (48.9%) of studies included questions around affect or emotion beyond anxiety and fear, yet only 17.8% considered frustrations and 13.3% considered motivation. When considering technology experience, only 4.4% of studies had questions regarding patients' previous use of XR technology which were used to control for novelty of the technology use.

An excluded MR study added constructs of "Spatial Interaction", which would be relevant for measuring MR UX in the future, capturing hologram experience with depth perception and morphology understanding, to quality of tools and ability to move through the model [106].

Figure 4: Proportion of Studies which Include Identified UX Dimensions



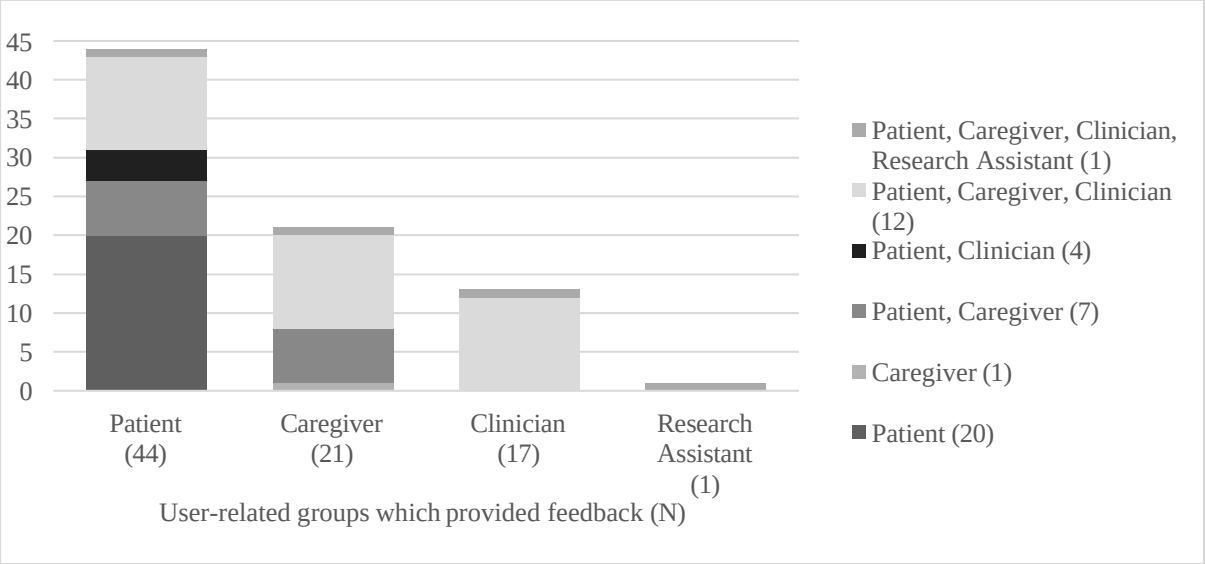
Measurement varies extensively in terms of scales used. Examples include 5-Point Likert [40,44], 6- and 7-Point Likert [42], Yes/No [37,45], rating from 1-100 [43,45], rating from 1-10 [36], and open-ended questions with results allocated as positive, negative, or neutral [26]. See the Appendix for more comprehensive examples of the types of questions asked for each respective category, covering various levels of depth. While not apparent in the table, motivation is predominantly used in the context of rehabilitation procedures due to the nature of repeat use for sessions which is less common for the other categories. Though not sufficiently compared due to the small sample of AR studies, yet supported by literature, AR research tends to use the same type of research methods as VR (often the same types of subjective questionnaires); however, there are technical differences that should be evaluated [107].

4.2.3 Measurement Methodologies and Modes of Delivery Applied

When considering measurement methodology and modes of delivery of UX metrics, the key questions are "who?" and "how?".

First addressing who is asked to report on pediatric patient experiences with XR, 44 of 45 studies (97.8%) asked the patient at least one experience question, with the remaining 1 asking caregivers on behalf of the child to report on the experience. Caregivers, often parents, reported their own experience of having their child interact with XR in 21 of 45 studies (46.7%) and clinicians reported their experience having the patient interacting with the technology in 17 of 45 studies (37.8%). Though the pediatric patients were the direct users of the XR technologies in all cases, Figure 5 describes the combinations of who was asked to report on experience in each study and the frequency with which that combination was used in the legend on the right-hand side. The bars total the number of times each group (patients, caregivers, clinicians, and research assistants) were asked to report on the XR user experience as they each represent a relevant user in the pediatric care context.

Figure 5: Parties from which UX Data was Gathered From, Grouped by Combination



Next addressing how experience measures were collected, Table 6 displays the frequency at which each collection method (survey/questionnaire, semi-structured or open-ended interview, free-text response, or sketch) was used to capture experience insights. Collection methods were

based on the UX methods reported by Bargas-Avila and Hornbæk (2011) [96]*;* however, live user observations and psychophysiological metrics were not captured here as these were outcome measures for pain and anxiety (i.e. FLACC and heartrate).

Table 6: User Experience Data Collection Methods by Rating of Engagement

Collection Method	Rating = High (n=11)	Rating = Medium (n=22)	Rating = Low (n=12)	Total (n=45)
Survey / Questionnaire	9 (81.8%)	19 (86.4%)	8 (66.7%)	36 (80.0%)
Interview (Semi-Structured)	6 (54.5%)	5 (22.7%)	2 (16.7%)	13 (28.9%)
Interview (Open-ended)	3 (27.3%)	3 (13.6%)	2 (16.7%)	8 (17.8%)
Free-text Response	1 (9.1%)	2 (9.1%)	2 (16.7%)	5 (11.1%)
Sketch	0 (0.0%)	0 (0.0%)	1 (8.3%)	1 (2.2%)

Surveys/questionnaires were the most common method for capturing experience sentiments across all engagement rating levels; however, what is distinct about the column with the high engagement rating is the more frequent use of interviews (both semi-structured and open-ended). These methods make it easier to capture a wider range of insights such as areas of improvement, yet there is a trade-off of taking longer to analyze and possibly more difficulty to compare these insights.

The sum of percentages by column is also interesting, given that the greater the sum of percentages, the more tools used on average. While ratings of low and medium have similar percentage sums at 125.0% and 131.8% respectively, studies that capture the greatest depth of insights have a sum of 172.7%. This means that the most insightful papers leverage an average of about 1.73 methods for capturing experience, while the lowest scoring only include an average of 1.25.

An excluded MR example suggests similar use of experience collection methodology and UX metrics across XR types. Yun, et al. (2022) had caregivers use MR to follow surgery findings and provide real-time consent during operations; they measured the experience of caregivers and clinicians using questionnaires which captured communication, satisfaction with MR glasses, ease of using MR glasses, previous experience and familiarity with MR, and overall satisfaction [104].

4.3 Additional Findings

4.3.1 Heterogeneity in Comparators and Study Design

While PRO results are not the primary outcome of this thesis, this additional observation critiques the comparator element of trial design. Studies were inconsistent in how they defined the standard of care for a given procedure within their practice, ranging from little to no distraction: pain medication [3], no own distraction devices [5], and no music, videogames, or tablets [8], to larger variety of distraction options: could be a mixture of tv or non-procedural talk [7], none, tv, phone, tablet, child life, toy, or book [9], or music, video games, tablets, or parent [10]. This variance was study dependant, where different hospital facilities applied their own standards of care. This is notable because distraction alternatives such as video games had instances of making pain reduction as an outcome measure not statistically different from the standard of care [10,12], meaning more clear guidelines of the standard comparator is needed to effectively conduct a meta-analysis on future studies.

4.3.2 Conflicting Experience Results

As introduced above, there were conflicting findings between studies. While the majority of studies found statistically significant improvements in outcome measures and positive experiences, there were instances where that did not hold. Jain, et al. (2024) stated the following unexpected findings showing the possibility that experience measures may not be independent from one another nor the intended outcome:

> Fun presented with a negative direct (-0.18), indirect (-0.03), and total (-0.20) effect on self-reported pain; however, none of these effects were statistically significant. VR features also exhibit significant interrelationships–a positive bidirectionality between fun and engagement (0.63) and game realism's positive effect on fun (0.49) and engagement (0.35). [21]

Further supporting the idea certain metrics can be dependent, Walther-Larsen, et al. (2019) found that satisfaction expressed by patients is a strong indicator of efficient pain and anxiety measures during procedural pain [43], suggesting that outcome metrics can be related to the more simplified experience questions such as rating overall satisfaction.

4.3.3 Positive Experience Sentiments and Perceived Technology Advantages

In spite of the variability in testing and findings, there is optimism from parents, caregivers, and clinicians alike, contributing to the demand for further innovation. In a study of pediatric cancer patients using VR while receiving their first chemotherapy treatments, one of the parents shared, "Kids these days are so smart, and they are curious about new technology. VR is so attractive to them that they have completely forgotten about chemotherapy [29]." The nurses in the same study shared that VR could make patients feel more relaxed and less focused on the procedure to reduce resistance and might support allocation of hospital resources by relieving side effects of chemotherapy and therefore duration of stay [29].

Looking to the more recently explored area of AR, the following benefit, which reaches beyond the capabilities of VR, was shared. "AR offers a potential advantage, utilizing distracting holographic images when patients [can] maintain eye contact with parents [31]." This quotation adds to the argument that there are differences between the different XR technologies that should be considered in the questions asked to patients, yet the two included AR studies in the review paralleled the metrics used in the VR studies. In some ways, seeing a parent while wearing an AR headset could be comforting, in others, less sense of escapism from the hospital could maintain levels of anxiety.

5. DISCUSSION

5.1 Heterogeneity Found is Consistent with the Clinical VR Landscape

The results of the review support the sentiment in the background literature: there is a lack of agreement across stakeholders on standardized measurement tools for XR experiences [107].

When considering all the PROMs for each outcome in Table 4, it is clear there are far too many measurement tools being used to measure the same outcome, making metrics that should be reasonably comparable confusing. Pain and anxiety could be limited to 2-3 PROMs, rather than almost 20 options each. The most interesting measure is the individual improvement category, since it requires far more variation. Individual improvement measures could be condensed; however, this is the one PRO category which will likely continue to need several options for

researchers to use simply because there are unique scales for specific joints and behaviours depending on patient need.

Ultimately, the lack of consistency in Table 4 and the Appendix suggest that the description of clinical VR as the "Wild West" by VR-CORE [97] should be extended to apply to the pediatric XR landscape, to encompass the heterogeneity of measurement of therapeutic applications. While the search methodology excluded many of the poor study designs described by VR-CORE, Table 4 draws attention to the variety of PROs and unnecessarily large selection of measurement tools for capture.

VR-CORE provided guidance to build consensus and best practices for the future of VR research, yet there were many papers reviewed in this thesis that were designed following the release of these recommendations and fail to meet them. Moreover, there exists additional opportunity to strengthen the research landscape to capture value as the technology continues to advance.

The example PROs recommended by VR-CORE are some of the PROs captured in Table 4. To reach beyond this VR-CORE recommendation, especially given the volume of needle-related VR studies where the PRO is an evaluation of pain, it would be beneficial if researchers used PROs with more consistent measurement methods and scales for similar procedures and outcomes. Table 4 shows this opportunity through the wide variety of ways pain, among the other patient-reported outcomes, was measured, which varies widely rather than capturing a few streamlined, more easily comparable approaches. Unfortunately, this is a common challenge that reaches across clinical research domains. Weldring & Smith (2013) describe that "[t]raditionally, the choices of PROMs are often based on professional judgment versus strong conceptual models creating issues with grouping and scoring items into domains [108]."

5.2 Connecting Results to Existing Literature

5.2.1 Purposes and Procedures

Pain and anxiety reduction were the most common outcomes reviewed and VR for pain was the most popular type of study. Pain was measured in the majority of studies (75.6%), and the most

represented type of procedure was needle-related, representing 17 (37.8%) of the studies. This is consistent with the introduction which explains how needle-related procedures, such as vaccinations, are very common in childhood and are known to cause pain.

The heterogeneity of PROMs measuring pain in a systematic review of animal therapy show similar variation to the variety displayed in Table 4.1. The PROMS in Correale, et al., 2022 included the following: Wong-Baker Scale, Numerical Rating Scale, Visual Analog Scale, and PedsQL Present Functioning Scales [109]. This variety meant that only qualitative analysis, tracking whether pain or anxiety increased, decreased, or stayed the same could be conducted. Had the tools aligned, the effect size would be quantifiable.

Figure 3, which detailed the frequencies of procedures by intervention type, provides a mapping of distinct uses of XR in pediatrics that have been trialed and begins to communicate the slow timeline from prototype to RCT with real patients. Several medical applications of MR and AR are in development, meaning there would be more of these types to categorize into Figure 3 if repeated in the future, yet at the time of submission, VR represented 43 of 45 studies.

The identified issues of inconsistency mirror those of older therapeutic interventions. In a systematic review of RCTs studying the effectiveness of music interventions in pediatric healthcare, limitations which "precluded formal aggregation of the results and completion of a meta-analysis" therefore limiting the ability to determine the effectiveness of the intervention drew several parallels to the challenges identified throughout this thesis. Limitations included outcome and measurement heterogeneity, within and across diagnostic groups, and lack of standardization of both interventions and appropriate controls [110]. The challenge was summarized as "not simply a lack of research but rather a lack of high quality research," suggesting the need for adherence to strict methodological quality and transparent reporting guidelines, such as CONSORT (Consolidated Standards of Reporting Trials) [110].

It is recommended that the categories in Figure 3 be used as a starting template for how to section PROM guidelines for efficacy measurement. The order of designing guidelines could be approached by popularity, and as such, it is recommended to begin with needle-related procedures and the PROs of pain and anxiety.

5.2.2 UX Metrics

The words used when defining questions to ask patients matter and need to be catered for the age group, even if at a high level they map to the same dimensions, just as faces scales are a good fit for young children. An interesting example is "fun." When Schmitt, et al. (2011) assessed the "magnitude of fun" experienced by participants, they noticed the following:

> The concept of 'fun' is related to mood/affect, but unlike more formal tools used to assess mood/affect in adults, is readily and easily understood in pediatric subjects spanning the wide age range in this study. [38]

Most studies include a measure of acceptability (88.9%), and almost half measure emotions, affect (48.9%) or usability (42.2%). While only documented in 17.8% of studies, to ensure patients remain engaged with the technology and have as pleasant of an experience as possible, frustration is a useful dimension for developers to capture. Having measurement of frustrations allows the clinicians, researchers, and game designers to better adapt to the needs of the patients, and learn about preferences, dislikes, and challenges. This category was seen in the form of open-ended questions, both written and verbal.

Even less common (4.4%) but valuable as XR continues to grow in popularity is consideration for involving a control for past experience, since when measuring emotional aspects such as "fun", researchers do not want to accidentally capture "novelty" which will reduce overtime.

The greatest problem is the constructs used to measure each UX dimension varies widely by study, as shown in Table 5. For Acceptability, there could be one simple question about whether patients would like to use the technology again, or a list of questions all tied to the overall theme of acceptability. This is a significant challenge for comparing results.

Experience measures for older therapeutic areas, such as animal therapy in pediatric hospital care, also measure constructs in the categories of acceptability and intention to use, affect, emotion, and outcomes including pain [111], but a contrasting factor from both animal and music therapy examples [109-111] to XR is that there are not comparable digital technology-related experience measures that can be applied from general UX research. The included studies have captured experience measures from patients, caregivers, and clinicians [109]; however, there is not an example of combining pediatric PREMs to include digital or immersive

experience since this is a novel condition of extended reality that has not previously needed measurement in the medical setting, especially the spatial or past XR use dimensions.

5.2.3 Measurement Modes of Delivery

Patients were almost always involved in providing feedback of experience directly (44 of 45 studies) and PROMs such as FACES scales should be used to ensure they can continue to provide feedback in an age-appropriate way. Caregivers provided feedback almost half of the time (46.7%) and clinicians were occasionally asked as well (37.8%). The benefit of incorporating clinician feedback is to understand the benefit of XR compared to how these experts would otherwise expect the procedure to unfold; however, repeating the feedback process for each patient studied is costly in time and likely would not have significant results between patients, given the clinicians' opinions would be unlikely to fluctuate drastically. Selecting more groups to provide feedback around the patient interaction with XR creates a more comprehensive understanding of the care experience. For example, the findings of one study detailed several inter-connected positive results related to experience: patients felt happy; parents believed their children looked happy; nurses felt the intervention was helpful; patients showed interest and felt as though they were in the environment; and the XR helped reduce patient distress when they had felt anxious before [29].

The usefulness of parent/caregiver feedback is influenced by the types of questions asked and there may be quickly diminishing returns to adding questions for this group. As mentioned earlier, parent satisfaction may not be sufficient for measuring the patient's experience when applying XR for pain distraction since parent satisfaction has been found to not be statistically different from parent perception of their child's pain or anxiety [84]. In contrast, Dunn et al. (2019) built exemplary experience questions that were more specific for different parties while being directly comparable and incorporated objective observations. Examples include the following measures for the usability and likeability of the XR product:

- If a participant wore the VR equipment: (1) during the entire procedure; (2) part of the procedure; or (3) only prior to the procedure
- VAS/FACES scale to answer "How easy was it for you/your child/your patient to use the VR equipment?"
- VAS/FACES scale to assess the VR likeability by answering "How much would you/your child/your patient like to use VR for future IV procedures?" [12]

Overall, further research is needed to clarify whether researchers should involve caregivers and clinicians for experience surveys and which experience metrics should be used for these groups, especially given there can be measures which are not statistically different from one another. Metrics of patient UX to measure should be lower for these groups; for example, spatial interaction would be illogical to include for anyone but the patient.

5.2.4 Feedback Collection Methods

The mix of methodological preference was consistent with Lallemend et al. (2015), where there was no clear consensus or preference for quantitative or qualitative measures [78]; however, the most depth in experience findings were captured when there was a combination of quantitative and qualitative methods used. The most popular method was surveys/questionnaires, which were used to gather experience feedback in 80.0% of studies. Interviews were the most popular qualitative method, providing the most depth of experience details, yet it is important to balance the cost and benefit of adding such measures. Qualitative feedback is most useful for developers to iterate solutions, so this feedback is most useful in earlier study stages where adjustments to the games or virtual environments can still be made. While interviews benefit production and adaptation of games, it is understood that they are more expensive and often do not produce quantitative data. For this reason, qualitative feedback should remain optional in future design.

Kammin, et al. (2024) found that in the context in palliative music therapy, stakeholder experience perspectives and feedback also often underrepresented [112]. 3 of 5 studies included in the review used qualitative measurement for stakeholder experience review and the remaining 2 of 5 used a mixed methods approach [112]. The authors used thematic synthesis in a systematic review to build greater understanding of caregiver and clinician perspectives and found unique benefits of use included supporting child and family well-being, suggesting there is a case for the benefits of measuring stakeholders beyond the pediatric patient.

Whether adding feedback from more perspectives or qualitative results tracking, adding measurement elements to studies creates additional costs for development, time to review, and possibilities of incomplete data if the commitment to provide feedback becomes too large [9].

5.2.5 Customization for Relevant Users

Differences of pediatric users, from physical and psychological needs [20] to gender stereotypes of game preferences [33] emerged in design discussions of reviewed studies. For example, the creation of a popular VR world, SnowWorld, which was originally designed for pediatric burn patients, exemplifies the importance of considering the psychological processes of user experience:

> SnowWorld was designed to be the antithesis of fire, to help [burn] patients avoid their pain during wound care, and to help patients avoid thinking about fire during wound care, in a simple environment that is easy to render, attention grabbing (e.g., interactive) but non-nauseogenic with passive navigation, and canyon walls that discourage wild changes in viewpoint. [113]

Common patient attributes, as described above with burn patients, should be considered for effective design of XR products to encourage interaction, and by extension immersion, which can then be measured with more specific metrics.

5.3 Impact of Heterogeneity on Relevant Stakeholders

Based on the stakeholders of medical XR identified (Table 2), the impact of the identified heterogeneity is discussed by stakeholder category [114].

5.3.1 Research Teams

For researchers, the main challenges are comparability and choosing the best outcome and experience measures. Results of other studies in the same intended outcome domain can be unnecessarily difficult to interpret because the studies are not comparable, that is, there are too many ways of measuring the same intended outcome when applying the same technology. Additionally, there is ambiguity in designing metrics to include because of the multidisciplinary nature of medical XR. There is not a clearly defined process, which is particularly difficult when considering UX is not typically the specialty of medical researchers. Clear guidelines save researchers time when designing studies because measures can be targeted to be more comparable based on intended outcomes. The current lack of comparability hinders the ability to conduct meta-analyses; with harmonization this would be possible and should be undertaken.

5.3.2 Regulatory Bodies

Since regulators must ensure devices being tested are both safe and effective for patients, the lack of uniformity of outcome measures can slow the approval process, especially when a selected PROM is not widely accepted to demonstrate the intended clinical benefit. Regulators particularly benefit from harmonization since it makes it easier to interpret evidence of the efficacy of technology. Moreover, not having clear guidelines has left no current consensus on how to best evaluate medical XR. This is important because, as described earlier, many XR platforms lack characterization and evaluation methods for the FDA to currently grant approval. Looking forward beyond initial approval, the large variation in PROM use can make monitoring real-world effectiveness beyond initial trials more difficult since the measures may not remain consistent.

5.3.3 Industry

Industry challenges include small developer firm sizes, generating acceptance from deployers (healthcare institutions), targeting intended patient outcomes, and capturing improvement opportunities. Small firms face significant barriers to entry for registration and acceptance due to regulatory requirements. Innovation is hindered by heterogeneity because the process from prototype to clinical adoption is inconsistent, yet firms encounter the same efficacy requirements as large pharmaceutical companies with significantly greater resources. When outcome metrics measured differ between studies, there may also be additional time and cost allocated to adapting games to meet the different methods without tangible benefit to patients. Considering experience measures, there was an opportunity identified to include feedback regarding challenges or frustrations faced in more studies, since most studies did not include this metric. This feedback targets specific areas for game designers to improve the offering for distinct needs and preferences.

5.3.4 Patients

Ultimately, the cost of game production, testing, and approval is higher in this heterogenous state than if all the relevant groups shared the same standards where products were designed in a way that was closely aligned to how they would be measured and assessed for years to come.

Without alignment between stakeholders of more specific best practices, patients suffer because there is less innovation, contributing to a worsened ability to have improved health experiences with XR. This suggests that quality of care is lower when metrics and measurement methodology are inconsistent across the stakeholders.

5.4 Recommended Practices for Gathering More Insightful Data and Innovating

5.4.1 #1: Focus on VR methodology as a foundation to advance AR and MR research

VR is the most developed category of XR, having been released years earlier than AR and MR, and should act as the foundation for future research. For instance, there were 3 papers reviewed with VR applications published between 2000 and 2009 [14, 33, 39], while the two AR trials were in 2020 [31] and 2024 [8]. The 45 papers reviewed only included two AR applications and no MR applications, even though the intention was to capture measurement methodology across the three types of XR. From the articles assessed for inclusion, AR and MR were applied in pediatrics when used by medical teams for applications such as surgery planning or visualizations, but MR was never applied as an intervention with the patient as the user.

While the implication of this is over-representation of VR in the sample and thus concern around applying recommendations to the categories of AR and MR, this highlights a need for further research as AR and MR continue to develop. At the same time, the learnings of needing more specific guidelines based on the intended purpose of the technologies would benefit AR and MR categories, even though they represented a very small portion of the reviewed works.

Target Stakeholders: Research Teams

5.4.2 #2: Improve the digital patient experience through PREM adjustments

Support game designers by directing innovation in areas with “pain-points” and opportunity for greater engagement. When designing experience questions, researchers should go beyond satisfaction or general interest in using the technology, by adding questions in the categories of frustration and spatial interaction (Table 5, see Appendix for examples) and using open-ended questions when possible. Open-ended questions can inspire designers to build new options, by

uncovering content preferences. For example, in Tennant et al. (2020), the most popular content was exploring nature (27.9%), followed by sports (20.6%), theme parks (16.2%), animals (14.7%) and travel (14.7%) [42].

Since the OECD does not currently have PREMs which specifically measure patient interactions with advanced technologies like XR, UX dimension such as spatial interaction should be added as either their own category or into in themes of Co-production – Use of digital tools for patient engagement or Integrated Care – Use of digital technology for integration of care on the OECD scorecard [50].

Target Stakeholders: Research Teams

5.4.3 #3: Limit recommended selection of validated PROMs for comparability

Reduce the recommended PROMs for researchers to design their studies with to encourage more comparable results including for meta-analyses. To limit PROMs from VR-CORE guidance of any validated measure of the researcher's choice, it is advised that 1-3 options be given for pain and anxiety measurement depending on the procedure being measured.

For pain, some of the most popular and validated tools proposed include the Faces Pain Scale and Faces, Legs, Activity, Crying, Consolability (FLACC). For anxiety, there was both state and trait anxiety PROMs, so the Spielberg State-Trait Anxiety Inventory should be used for older elementary school children and older and either a Visual Analog Scale or physiological measurements such as heartrate could be used for younger patients. Fear is already captured by the CFS across studies. Individual improvement PROMs vary based on characteristics such as mobility and other specific purposes, but overall, the list of options could be condensed.

Target Stakeholders: Regulatory Bodies, Research Teams, and Industry

5.4.4 #4: Enhance knowledge sharing and comparability by harmonizing research

For each category of procedures (Figure 3), develop clear instructions for measurement to harmonize efforts between researchers and regulators. There is no best methodology unless

guidance is narrowed down to smaller areas of XR applications, such as pediatric pain, mental health, or improvement of mobility. The measures needed to assess the efficacy of XR to support surgeons are completely different from the measures needed to assess the efficacy of XR applied for patient distraction, but within these groups, the studies should be comparable. Guideline building should begin with the most popular procedure type for which reducing pain can have a large influence on patients' healthcare experiences for years to come, which are needle-related, specifically vaccines.

There are several guidelines for how to discuss and report real world evidence including SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials) for evidence-based recommendations of the minimum items for clinical trials to address [115], CONSORT built in tandem with SPIRIT for reporting standards [115], and European Society for Medical Oncology (ESMO) Guidance for Reporting Oncology real-World evidence (GROW) for oncology research [116]. More recently, there has been development of SPIRIT and CONSORT guideline extensions to have AI-specific factors like the human-AI interaction, referred to as SPIRIT-AI and CONSORT-AI [115, 117]. Baba, et al. (2024), have also called to attention the need for extensions specific to research on children, referred to as SPIRIT-Children and CONSORT-Children, for which development has now started [118].

It is recommended based on the heterogeneity found in this analysis that similar guidance be established for distinct pediatric XR use cases, especially the most common of pediatric pain management, as SPIRIT and CONSORT extensions. The purpose of this effort would be to build harmony, providing more robust guidance within the industry. Regulators should signal the type of methodology that should be used for the given topic, and with this guidance, there would be consensus and alignment across stakeholders.

Target Stakeholders: Care Teams, Regulatory Bodies, Research Teams, Industry, and Other Supports

5.5 The Benefits of Methodology Harmonization Efforts

The value of harmonization is increased efficiency, providing clarity for regulators and bringing the industry producers closer to the consumers (patients and their care teams). Harmonization

decreases associated transaction costs [119], such as market entry costs for newer products [120], which in turn encourages innovation to help more children, especially for small firms.

In contrast, heterogeneity brings market uncertainty. Consumers lose due to higher prices and limited customization for differing care needs. Producers face barriers to getting approvals, making the market less appealing and discourage the rate of iterating solutions for new use cases. Finally, regulators and researchers lose because there is not clarity on how to regulate or compare findings to prior works, adding time and expense to try to solve the same challenges. In summary, the cost of not having guidelines to harmonize research and regulation hurts all relevant stakeholders, harming the ability to provide the best care to patients.

5.6 Summary of Recommendations

1. Leverage VR research methodology as a foundation for AR and MR research, which are in earlier stages of development and use compared to VR.

2. PREMs should reach beyond concepts of acceptability, usability, and affect to include innovation-driving metrics. Specifically, frustration and spatial interaction questions should be included when designing future XR studies, using open-ended questions when possible.

3. Create a limited selection of recommended PROMs, based on the historic use (Table 4), so researchers consistently use comparable collection methods to enable meta-analyses.

4. Build guidelines for the measurement of XR effects for specific procedure types, to create alignment between how the medical devices or platforms are developed and researched to how they are approved and monitored by regulatory bodies.

Together, the recommendations benefit regulators, producers (industry), researchers, care teams, and patients through the development of more XR use cases, reduced costs, faster innovation, more study design and regulatory clarity, safer access, and overall quality of care.

6. CONCLUSION

Expanding beyond current literature, this thesis provides a strategic management perspective to the medical XR industry by identifying the impact the results of the research questions have on relevant stakeholders and market dynamics such as transaction costs, speed of regulatory approval, and the ability to design products catered to the health needs of patients. There is not currently a consistent, systematic approach to measuring value from a user experience perspective when Extended Reality (XR) technology is used by pediatric patients, yet the harmonization of metrics would increase efficiency and reduce costs across relevant stakeholders, while improving quality of care for patients.

- Pediatric patients engage with medical XR for anesthesia, burns, blood and cancer diseases, cast-removal, needles, rehabilitation, and surgery with intended outcomes of reducing pain, anxiety, and fear, and/or creating improvement such as increasing range of motion. The methods used to measure these outcomes are inconsistent and vary more widely than the types of procedures they are for.

- Narrative analysis of PREMs showed some alignment to those used in broader UX research. The most measured metrics were acceptability and intention to use, affect, emotion, and usability. Frustration, spatial interaction, and controls for past experience offer significant insights to support innovation for developers yet are often omitted.

- Patients most commonly provide feedback on their own experiences, though caregivers provided feedback about half of the time. Surveys/questionnaires are routinely used to assess experience; however, many of the studies which produced the most depth of insights also engaged open-ended or semi-structured interviews.

To improve patient care through more effective XR innovation, researchers should use current VR research as a foundation for further development of AR and MR studies and include PREMs which capture frustrations and spatial interaction assessment. Developing cross-functionally with several XR stakeholders (Table 2), it is advised to limit the recommended PROMs for future XR research and efficacy testing and build specific harmonized guidelines for how to measure and discuss XR effects.

APPENDIX

Metrics Analysis – UX Measurement by Dimension

UX/PREM Dimension	Examples and Sources	N (%)
Acceptability and Intention to Use	• VAS/FACES scale to answer "How easy was it for you/your child/your patient to use the VR equipment?" and "How much would you/your child/your patient like to use VR for future IV procedures? [12] • Yes/No for "My child was satisfied with the way he/she went to sleep" [24] • Yes/No for "I would like to wear the glasses again…" [8]	40 (88.9%)
Affect, Emotion	• 0-5, ranking on the "Joy Scale" where zero = no joy and 5 = ecstatic [3] • Likert scale from 0-6 for "The application was boring for me" [25] • 0 = No fun at all, 10 = extremely fun, answering "How much FUN did you have during wound care?" [20]	22 (48.9%)
Usability (Abilities and Physical Attributes)	• Rank level of agreement from totally disagree to totally agree on "The VR game was adapted/suitable to the age group of children," and "The VR device was adapted/suitable to the clinic's environment" [3] • Yes/No and Comfort Rating Scale for "Did the headset hurt? Was not seeing the normal environment a problem for you? Was not seeing your body a problem for you? How well could you move with the headset on?" [4] • Whether a participant wore the VR equipment: (1) during the entire procedure; (2) part of the procedure; or (3) only prior to the procedure [12] • Semi-structured interview, comment on physical qualities of the equipment [39]	19 (42.2%)
Spatial Interaction	• Score of 1-3 on Gold-Rizzo Immersion and Presence (GRIP) Inventory [121] which has 16-items and covers 3 domains of immersion [16] • Rated from "completely fake'' to ''indistinguishable from a real object'', "How real did the objects in the virtual world seem to you?" [38]	16 (35.6%)
Frustration	• Open-ended opportunity to discuss opinion of tool inclusive of disadvantages and problems [4] • Semi-structured discussion where outputs included dislikes and recommendations [33]	8 (17.8%)
Motivation	• Rating from 1-5 of questions such as "I liked it that the game was getting more difficult" [1, 44] • Quantitative ranking and open-ended feedback to capture motivation such as whether patients were more motivated to attend follow up sessions [40]	6 (13.3%)
Past Experience Controls	• Yes/No of "Use of VR before, rating of knowledge of VR, and rating of self-confidence with operating VR technology" [42]	2 (4.4%)

REFERENCES

[1] G. Acar, G. P. Altun, S. Yurdalan, and M. G. Polat, "Efficacy of neurodevelopmental treatment combined with the Nintendo® Wii in patients with cerebral palsy," *JOURNAL OF PHYSICAL THERAPY SCIENCE,* vol. 28, no. 3, pp. 774-780, 2016, doi: 10.1589/jpts.28.774.

[2] A. K. Agrawal *et al.*, "Virtual reality as complementary pain therapy in hospitalized patients with sickle cell disease," *Pediatr Blood Cancer,* vol. 66, no. 2, 2019, Art no. e27525, doi: 10.1002/pbc.27525.

[3] T. Alrimy *et al.*, "Desktop Virtual Reality Offers a Novel Approach to Minimize Pain and Anxiety during Burn Wound Cleaning/Debridement in Infants and Young Children: a Randomized Crossover Pilot Study," *Journal of Clinical Medicine,* vol. 12, no. 15, 2023, doi: 10.3390/jcm12154985.

[4] C. Ammann-Reiffer, A. Kläy, and U. Keller, "Virtual Reality as a Therapy Tool for Walking Activities in Pediatric Neurorehabilitation: Usability and User Experience Evaluation," *JMIR SERIOUS GAMES,* vol. 10, no. 3, 2022, doi: 10.2196/38509.

[5] B. Atzori *et al.*, "Virtual Reality Analgesia During Venipuncture in Pediatric Patients With Onco-Hematological Diseases," *FRONTIERS IN PSYCHOLOGY,* vol. 9, 2018, doi: 10.3389/fpsyg.2018.02508.

[6] B. Atzori *et al.*, "An Exploratory Study on the Effectiveness of Virtual Reality Analgesia for Children and Adolescents with Kidney Diseases Undergoing Venipuncture," *Int J Environ Res Public Health,* vol. 19, no. 4, 2022, doi: 10.3390/ijerph19042291.

[7] T. J. Caruso *et al.*, "Virtual reality during pediatric vascular access: A pragmatic, prospective randomized, controlled trial," *Paediatr Anaesth,* vol. 30, no. 2, pp. 116-123, 2020, doi: 10.1111/pan.13778.

[8] C. Chamberland, M. Bransi, A. Boivin, S. Jacques, J. Gagnon, and S. Tremblay, "The effect of augmented reality on preoperative anxiety in children and adolescents: A randomized controlled trial," *Paediatr Anaesth,* vol. 34, no. 2, pp. 153-159, 2024, doi: 10.1111/pan.14793.

[9] E. Chan *et al.*, "Virtual Reality for Pediatric Needle Procedural Pain: two Randomized Clinical Trials," *Journal of Pediatrics,* vol. 209, pp. 160-167.e4, 2019, doi: 10.1016/j.jpeds.2019.02.034.

[10] P. G. B. Clerc, J. S. Arneja, C. M. Zwimpfer, A. Behboudi, and R. D. Goldman, "A Randomized Controlled Trial of Virtual Reality in Awake Minor Pediatric Plastic Surgery Procedures," *Plast Reconstr Surg,* vol. 148, no. 2, pp. 400-408, 2021, doi: 10.1097/PRS.0000000000008196.

[11] S. Dumoulin *et al.*, "A Randomized Controlled Trial on the Use of Virtual Reality for Needle-Related Procedures in Children and Adolescents in the Emergency Department," *Games Health J,* vol. 8, no. 4, pp. 285-293, 2019, doi: 10.1089/g4h.2018.0111.

[12] A. Dunn *et al.*, "A Novel Clinician-Orchestrated Virtual Reality Platform for Distraction During Pediatric Intravenous Procedures in Children With Hemophilia: Randomized Controlled Trial," *JMIR SERIOUS GAMES,* vol. 7, no. 1, 2019, Art no. e10902, doi: 10.2196/10902.

[13] K. Ellerton *et al.*, "Virtual Reality for IMMunisation (VRIMM) pain in young children: Results of a randomised controlled trial in general practice," *Aust J Gen Pract,* vol. 52, no. 10, pp. 704-710, 2023, doi: 10.31128/AJGP-08-22-6537.

[14] J. I. Gold, S. H. Kim, A. J. Kant, M. H. Joseph, and A. S. Rizzo, "Effectiveness of virtual reality for pediatric pain distraction during i.v. placement," *Cyberpsychol Behav,* vol. 9, no. 2, pp. 207-12, 2006, doi: 10.1089/cpb.2006.9.207.
[15] J. I. Gold and N. E. Mahrer, "Is Virtual Reality Ready for Prime Time in the Medical Space? A Randomized Control Trial of Pediatric Virtual Reality for Acute Procedural Pain Management," *J Pediatr Psychol,* vol. 43, no. 3, pp. 266-275, 2018, doi: 10.1093/jpepsy/jsx129.
[16] J. I. Gold, M. SooHoo, A. M. Laikin, A. S. Lane, and M. J. Klein, "Effect of an Immersive Virtual Reality Intervention on Pain and Anxiety Associated With Peripheral Intravenous Catheter Placement in the Pediatric Setting: a Randomized Clinical Trial," *JAMA Network Open,* vol. 4, no. 8, p. e2122569, 2021, doi: 10.1001/jamanetworkopen.2021.22569.
[17] R. D. Goldman and A. Behboudi, "Virtual reality for intravenous placement in the emergency department-a randomized controlled trial," *Eur J Pediatr,* vol. 180, no. 3, pp. 725-731, 2021, doi: 10.1007/s00431-020-03771-9.
[18] R. D. Goldman and A. Behboudi, "Pilot Randomized Controlled Trial of Virtual Reality vs. Standard-of-Care During Pediatric Laceration Repair," *Journal of Child & Adolescent Trauma,* vol. 14, no. 2, pp. 295-298, 2021, doi: 10.1007/s40653-021-00350-4.
[19] H. A. Hernández, A. Khan, L. Fay, J. S. Roy, and E. Biddiss, "Force Resistance Training in Hand Grasp and Arm Therapy: Feasibility of a Low-Cost Videogame Controller," *GAMES FOR HEALTH JOURNAL,* vol. 7, no. 4, pp. 277-287, 2018, doi: 10.1089/g4h.2017.0193.
[20] H. G. Hoffman, D. R. Patterson, R. A. Rodriguez, R. Pena, W. Beck, and W. J. Meyer, "Virtual Reality Analgesia for Children With Large Severe Burn Wounds During Burn Wound Debridement," *FRONTIERS IN VIRTUAL REALITY,* vol. 1, 2020, doi: 10.3389/frvir.2020.602299.
[21] S. Jain *et al.*, "Features of virtual reality impact effectiveness of VR pain alleviation therapeutics in pediatric burn patients: a randomized clinical trial," *PLOS Digital Health,* vol. 3, no. 1, 2024, Art no. e0000440, doi: 10.1371/journal.pdig.0000440.
[22] D. Jeffs *et al.*, "Effect of virtual reality on adolescent pain during burn wound care," *Journal of Burn Care & Research,* vol. 35, no. 5, pp. 395-408, 2014, doi: 10.1097/BCR.0000000000000019.
[23] B. A. Jivraj *et al.*, "The use of virtual reality in reducing anxiety during cast removal: a randomized controlled trial," *Journal of Children's Orthopaedics,* vol. 14, no. 6, pp. 574-580, 2020, doi: 10.1302/1863-2548.14.200090.
[24] M. J. Jung, J. S. Libaw, K. Ma, E. L. Whitlock, J. R. Feiner, and J. L. Sinskey, "Pediatric Distraction on Induction of Anesthesia With Virtual Reality and Perioperative Anxiolysis: A Randomized Controlled Trial," *Anesth Analg,* vol. 132, no. 3, pp. 798-806, 2021, doi: 10.1213/ANE.0000000000005004.
[25] I. Jyskä *et al.*, "Design and User Experience of VirNE Application: Deep Breathing Exercise in a Virtual Natural Environment to Reduce Treatment Anxiety in Pediatrics," *HEALTHCARE,* vol. 11, no. 24, 2023, doi: 10.3390/healthcare11243129.
[26] I. Jyskä *et al.*, "Effects of Using Guided Deep Breathing Exercises in a Virtual Natural Environment to Reduce Stress during Pediatric Treatment," *HEALTHCARE,* vol. 11, no. 24, 2023, doi: 10.3390/healthcare11243140.
[27] N. Kucher, J. M. Larson, G. Fischer, M. Mertaugh, L. Peterson, and L. A. Gershan, "3-Dimensional nature-based therapeutics in pediatric patients with total

pancreatectomy and islet auto-transplant," *COMPLEMENTARY THERAPIES IN MEDICINE,* vol. 48, 2020, doi: 10.1016/j.ctim.2019.102249.
[28] S. Le May *et al.*, "Decreasing Pain and Fear in Medical Procedures with a Pediatric Population (DREAM): A Pilot Randomized Within-Subject Trial," *Pain Manag Nurs,* vol. 22, no. 2, pp. 191-197, 2021, doi: 10.1016/j.pmn.2020.10.002.
[29] C. Lee Wong *et al.*, "Effects of immersive virtual reality for managing anxiety, nausea and vomiting among paediatric cancer patients receiving their first chemotherapy: An exploratory randomised controlled trial," *Eur J Oncol Nurs,* vol. 61, 2022, Art no. 102233, doi: 10.1016/j.ejon.2022.102233.
[30] S. P. Litwin *et al.*, "Virtual Reality to Reduce Procedural Pain During IV Insertion in the Pediatric Emergency Department: A Pilot Randomized Controlled Trial," *Clin J Pain,* vol. 37, no. 2, pp. 94-101, 2021, doi: 10.1097/AJP.0000000000000894.
[31] M. Madill *et al.*, "Augmented reality as a tool to reduce fear and promote cooperation during pediatric nasal endoscopy," *Laryngoscope,* vol. 130, no. 5, p. 86, 2020, doi: 10.1213/01.ane.0000662056.84859.62.
[32] E. W. A. Mohamed and H. I. Mohamed, "Effect of flippits versus virtual reality on pain, fear, and satisfaction during phlebotomy among children," *JOURNAL OF THE PAKISTAN MEDICAL ASSOCIATION,* vol. 73, no. 4, pp. S215-S220, 2023, doi: 10.47391/JPMA.EGY-S4-43.
[33] S. Nilsson, B. Finnström, E. Kokinsky, and K. Enskär, "The use of Virtual Reality for needle-related procedural pain and distress in children and adolescents in a paediatric oncology unit," *EUROPEAN JOURNAL OF ONCOLOGY NURSING,* vol. 13, no. 2, pp. 102-109, 2009, doi: 10.1016/j.ejon.2009.01.003.
[34] I. Olivieri, M. Chiappedi, P. Meriggi, M. Mazzola, A. Grandi, and L. Angelini, "Rehabilitation of children with hemiparesis: a pilot study on the use of virtual reality," *Biomed Res Int,* vol. 2013, 2013, Art no. 695935, doi: 10.1155/2013/695935.
[35] E. Osmanlliu *et al.*, "Distraction in the Emergency department using Virtual reality for INtravenous procedures in Children to Improve comfort (DEVINCI): a pilot pragmatic randomized controlled trial," *CJEM,* vol. 23, no. 1, pp. 94-102, 2021, doi: 10.1007/s43678-020-00006-6.
[36] I. Phelan *et al.*, "Immersive virtual reality in children with upper limb injuries: Findings from a feasibility study," *JOURNAL OF PEDIATRIC REHABILITATION MEDICINE,* vol. 14, no. 3, pp. 401-414, 2021, doi: 10.3233/PRM-190635.
[37] A. E. Richey *et al.*, "Virtual Reality Reduces Fear and Anxiety During Pediatric Orthopaedic Cast Room Procedures: A Randomized Controlled Trial," *J Pediatr Orthop,* vol. 42, no. 10, pp. 600-607, 2022, doi: 10.1097/BPO.0000000000002250.
[38] Y. S. Schmitt *et al.*, "A randomized, controlled trial of immersive virtual reality analgesia, during physical therapy for pediatric burns," *Burns,* vol. 37, no. 1, pp. 61-8, 2011, doi: 10.1016/j.burns.2010.07.007.
[39] S. M. Schneider and M. L. Workman, "Virtual reality as a distraction intervention for older children receiving chemotherapy," *Pediatr Nurs,* vol. 26, no. 6, pp. 593-7, 2000.
[40] J. B. Shen *et al.*, "VR-Based Cognitive Rehabilitation for Children With Traumatic Brain Injuries: Feasibility and Safety," *REHABILITATION PSYCHOLOGY,* vol. 67, no. 4, pp. 474-483, 2022, doi: 10.1037/rep0000458.
[41] J. S. Taylor *et al.*, "Small surgeries, big smiles: using virtual reality to reduce the need for sedation or general anesthesia during minor surgical procedures," *Pediatr Surg Int,* vol. 37, no. 10, pp. 1437-1445, 2021, doi: 10.1007/s00383-021-04955-6.
[42] M. Tennant, J. McGillivray, G. J. Youssef, M. C. McCarthy, and T. J. Clark, "Feasibility, Acceptability, and Clinical Implementation of an Immersive Virtual

Reality Intervention to Address Psychological Well-Being in Children and Adolescents With Cancer," *J Pediatr Oncol Nurs,* vol. 37, no. 4, pp. 265-277, 2020, doi: 10.1177/1043454220917859.
[43] S. Walther-Larsen, T. Petersen, S. M. Friis, G. Aagaard, B. Drivenes, and P. Opstrup, "Immersive Virtual Reality for Pediatric Procedural Pain: A Randomized Clinical Trial," *Hosp Pediatr,* vol. 9, no. 7, pp. 501-507, 2019, doi: 10.1542/hpeds.2018-0249.
[44] D. G. M. Winkels, A. I. R. Kottink, R. A. J. Temmink, J. M. M. Nijlant, and J. H. Buurke, "Wii™-habilitation of upper extremity function in children with Cerebral Palsy. An explorative study," *DEVELOPMENTAL NEUROREHABILITATION,* vol. 16, no. 1, pp. 44-51, 2013, doi: 10.3109/17518423.2012.713401.
[45] H. Xiang *et al.*, "Efficacy of Smartphone Active and Passive Virtual Reality Distraction vs Standard Care on Burn Pain Among Pediatric Patients: A Randomized Clinical Trial," *JAMA Netw Open,* vol. 4, no. 6, 2021, Art no. e2112082, doi: 10.1001/jamanetworkopen.2021.12082.
[46] C. M. McMurtry *et al.*, "Far From "Just a Poke": Common Painful Needle Procedures and the Development of Needle Fear," (in eng), *Clin J Pain,* vol. 31, no. 10 Suppl, pp. S3-11, Oct. 2015, doi: 10.1097/ajp.0000000000000272.
[47] A. Taddio *et al.*, "Prevalence of pain and fear as barriers to vaccination in children – Systematic review and meta-analysis," *Vaccine,* vol. 40, no. 52, pp. 7526-7537, Dec 12 2022, doi: 10.1016/j.vaccine.2022.10.026.
[48] A. M. McCarthy, C. Kleiber, K. Hanrahan, M. B. Zimmerman, N. Westhus, and S. Allen, "Factors explaining children's responses to intravenous needle insertions," (in eng), *Nurs Res,* vol. 59, no. 6, pp. 407-16, Nov-Dec 2010, doi: 10.1097/NNR.0b013e3181f80ed5.
[49] K. A. Birnie, M. Noel, C. T. Chambers, L. S. Uman, and J. A. Parker, "Psychological interventions for needle-related procedural pain and distress in children and adolescents," (in eng), *Cochrane Database Syst Rev,* vol. 10, no. 10, Oct. 4 2018, Art no. Cd005179, doi: 10.1002/14651858.CD005179.pub4.
[50] OECD, *Health for the People, by the People* (OECD Health Policy Studies). Paris: OECD Publishing, 2021.
[51] M. Sun, C. Bai, and D. Yang, "Exploring the metaverse hospital beyond your imagination," *Clinical eHealth,* vol. 6, pp. 53-59, Dec. 2023, doi: 10.1016/j.ceh.2023.07.004.
[52] S. K. Jagatheesaperumal, K. Ahmad, A. Al-Fuqaha, and J. Qadir, "Advancing Education Through Extended Reality and Internet of Everything Enabled Metaverses: Applications, Challenges, and Open Issues," *IEEE Transactions on Learning Technologies,* vol. 17, pp. 1120-1139, 2024, doi: 10.1109/TLT.2024.3358859.
[53] E. Kim *et al.*, "Examining the Academic Trends in Neuropsychological Tests for Executive Functions Using Virtual Reality: Systematic Literature Review," *JMIR Serious Games,* vol. 9, no. 4, Nov. 24 2021, Art no. e30249, doi: 10.2196/30249.
[54] T. Ong, H. Wilczewski, S. R. Paige, H. Soni, B. M. Welch, and B. E. Bunnell, "Extended Reality for Enhanced Telehealth During and Beyond COVID-19: Viewpoint," *JMIR Serious Games,* vol. 9, no. 3, Jul. 26 2021, Art no. e26520, doi: 10.2196/26520.
[55] A. Logeswaran, C. Munsch, Y. J. Chong, N. Ralph, and J. McCrossnan, "The role of extended reality technology in healthcare education: Towards a learner-centred approach," *Future Healthcare Journal,* vol. 8, no. 1, pp. e79-e84, Mar. 2021, doi: 10.7861/fhj.2020-0112.

[56] J. Shen *et al.*, "Virtual Reality–Based Executive Function Rehabilitation System for Children With Traumatic Brain Injury: Design and Usability Study," *JMIR Serious Games*, vol. 8, no. 3, Aug. 25 2020, Art no. e16947, doi: 10.2196/16947.
[57] K. Brütsch *et al.*, "Influence of virtual reality soccer game on walking performance in robotic assisted gait training for children," (in eng), *J Neuroeng Rehabil*, vol. 7, p. 15, Apr. 22 2010, doi: 10.1186/1743-0003-7-15.
[58] D. E. Logan, K. Khanna, E. Randall, S. O'Donnell, T. Reks, and L. McLennan, "Centering Patient and Clinician Voices in Developing Tools to Address Pain Related School Impairment: A Phase I Study of a Virtual Reality School Simulation for Children and Adolescents with Chronic Pain," *Children*, vol. 10, no. 10, doi: 10.3390/children10101644.
[59] L. M. Wellens *et al.*, "Comparison of 3-Dimensional and Augmented Reality Kidney Models With Conventional Imaging Data in the Preoperative Assessment of Children With Wilms Tumors," *JAMA Network Open*, vol. 2, no. 4, 2019, Art no. e192633, doi: 10.1001/jamanetworkopen.2019.2633.
[60] G. Hey, M. Guyot, A. Carter, and B. Lucke-Wold, "Augmented Reality in Neurosurgery: A New Paradigm for Training," *Medicina*, vol. 59, no. 10, doi: 10.3390/medicina59101721.
[61] M. Coduri, A. Calandrino, G. Addiego Mobilio, M. Casadio, and S. Ricci, "RiNeo MR: A mixed reality simulator for newborn life support training," *PLOS ONE*, vol. 18, no. 12, 2023, Art no. e0294914, doi: 10.1371/journal.pone.0294914.
[62] M. Johnston *et al.*, "The feasibility and usability of mixed reality teaching in a hospital setting based on self-reported perceptions of medical students," *BMC Medical Education*, vol. 24, no. 1, p. 701, Jun. 27 2024, doi: 10.1186/s12909-024-05591-z.
[63] C. Andrews, M. K. Southworth, J. N. A. Silva, and J. R. Silva, "Extended Reality in Medical Practice," (in eng), *Curr Treat Options Cardiovasc Med*, vol. 21, no. 4, p. 18, Mar 30 2019, doi: 10.1007/s11936-019-0722-7.
[64] N. J. Rowan, "Digital technologies to unlock safe and sustainable opportunities for medical device and healthcare sectors with a focus on the combined use of digital twin and extended reality applications: A review," *Science of The Total Environment*, vol. 926, May 20 2024, Art no. 171672, doi: 10.1016/j.scitotenv.2024.171672.
[65] Y. M. Tang, K. Y. Chau, A. P. K. Kwok, T. Zhu, and X. Ma, "A systematic review of immersive technology applications for medical practice and education - Trends, application areas, recipients, teaching contents, evaluation methods, and performance," *Educational Research Review*, vol. 35, Feb. 2022, Art no. 100429, doi: 10.1016/j.edurev.2021.100429.
[66] M. C. Davis, D. D. Can, J. Pindrik, B. G. Rocque, and J. M. Johnston, "Virtual Interactive Presence in Global Surgical Education: International Collaboration Through Augmented Reality," *World Neurosurgery*, vol. 86, pp. 103-111, Feb. 2016, doi: 10.1016/j.wneu.2015.08.053.
[67] Y. Lee, S. K. Kim, and M.-R. Eom, "Usability of mental illness simulation involving scenarios with patients with schizophrenia via immersive virtual reality: A mixed methods study," (in English), *PLoS ONE*, Report vol. 15, Sept. 16 2020, doi: 10.1371/journal.pone.0238437.
[68] H. U. Khan, Y. Ali, F. Khan, and M. A. Al-antari, "A comprehensive study on unraveling the advances of immersive technologies (VR/AR/MR/XR) in the healthcare sector during the COVID-19: Challenges and solutions," *Heliyon*, vol. 10, no. 15, 2024, doi: 10.1016/j.heliyon.2024.e35037.

[69] L. Vogt, M. Klasen, R. Rossaint, U. Goeretz, P. Ebus, and S. Sopka, "Virtual Reality Tour to Reduce Perioperative Anxiety in an Operating Setting Before Anesthesia: Randomized Clinical Trial," *J Med Internet Res,* vol. 23, no. 9, Sept. 2021, Art no. e28018, doi: 10.2196/28018.
[70] Z. Bashir, C. Misquith, P. Has, and S. Bukhari, "Effectiveness of virtual reality on anxiety and pain management in patients undergoing cardiac procedures: a protocol for systematic review and meta-analysis," (in eng), *Open Heart,* vol. 10, no. 1, Jun 2023, doi: 10.1136/openhrt-2023-002305.
[71] A. Denche-Zamorano *et al.*, "Rehabilitation Is the Main Topic in Virtual and Augmented Reality and Physical Activity Research: A Bibliometric Analysis," (in eng), *Sensors (Basel),* vol. 23, no. 6, Mar. 9 2023, doi: 10.3390/s23062987.
[72] Mordor Intelligence, "Virtual reality (VR) value in healthcare market worldwide in 2024 and forecast for 2029 (in billion U.S. dollars)," in *Statista*, ed, 2023.
[73] Accenture, "Feelings towards the positive impact of the metaverse on healthcare worldwide," in *Statista*, ed, 2022.
[74] S. Dean, F. Al Sayah, and J. A. Johnson, "Measuring value in healthcare from a patients' perspective," (in eng), *J Patient Rep Outcomes,* vol. 5, no. Suppl 2, p. 88, Oct. 12 2021, doi: 10.1186/s41687-021-00364-4.
[75] C. Kendir, M. van den Berg, F. Guanais, J. M. Valderas, and N. Klazinga, "Lessons from early implementation of the OECD's Patient-Reported Indicator Surveys (PaRIS) in primary care: making the case for co-development and adaptation to national contexts," *IJQHC Communications,* vol. 3, no. 1, 2023, Art no. lyad003, doi: 10.1093/ijcoms/lyad003.
[76] S. Hlongwane and S. S. Grobbelaar, "A Practical Framework for Value Creation in Health Information Systems From an Ecosystem Perspective: Evaluated in the South African Context," (in English), *Frontiers in Psychology,* Original Research vol. 13, Jun. 2 2022, doi: 10.3389/fpsyg.2022.637883.
[77] OECD, *Recommendations to OECD Ministers of Health from the High-Level Reflection Group on the Future of Health Statistics: Strengthening the International Comparison of Health System Performance through Patient-Reported Indicators* (OECD Health Policy Studies). Paris: OECD Publishing, 2017.
[78] L. A. Petersen, "Pay for performance: what we measure matters," *Virtual Mentor,* vol. 15, no. 7, pp. 570-5, Jul. 1 2013, doi: 10.1001/virtualmentor.2013.15.7.ecas2-1307.
[79] OECD, *Children and digital technologies: Trends and outcomes*. 2019.
[80] A. G. Espinel, R. K. Shah, M. E. McCormick, P. R. Krakovitz, and E. F. Boss, "Patient Satisfaction in Pediatric Surgical Care: A Systematic Review," *Otolaryngology–Head and Neck Surgery,* vol. 150, no. 5, pp. 739-749, 2014, doi: 10.1177/0194599814527232.
[81] K. A. Calabro, M. V. Raval, and D. H. Rothstein, "Importance of patient and family satisfaction in perioperative care," *Seminars in Pediatric Surgery,* vol. 27, no. 2, pp. 114-120, Apr. 2018, doi: 10.1053/j.sempedsurg.2018.02.009.
[82] J. Ferreira *et al.*, "Patient experience or patient satisfaction? A systematic review of child- and family-reported experience measures in pediatric surgery," *Journal of Pediatric Surgery,* vol. 58, no. 5, pp. 862-870, 2023, doi: 10.1016/j.jpedsurg.2023.01.015.
[83] D. C. Ferreira, R. C. Marques, A. M. Nunes, and J. R. Figueira, "Customers satisfaction in pediatric inpatient services: A multiple criteria satisfaction analysis," *Socio-Economic Planning Sciences,* vol. 78, Dec. 2021, Art no. 101036, doi: 10.1016/j.seps.2021.101036.

[84] H. N. Lee *et al.*, "Virtual reality environment using a dome screen for procedural pain in young children during intravenous placement: A pilot randomized controlled trial," (in eng), *PLoS One,* vol. 16, no. 8, 2021, Art no. e0256489, doi: 10.1371/journal.pone.0256489.
[85] J. Evans, "The IEEE Global Initiative on Ethics of Extended Reality (XR) Report--Extended Reality (XR) Ethics in Medicine," *The IEEE Global Initiative on Ethics of Extended Reality (XR) Report--Extended Reality (XR) Ethics in Medicine,* pp. 1-31, 2022.
[86] F. Bini, M. Franzò, A. Maccaro, D. Piaggio, L. Pecchia, and F. Marinozzi, "Is medical device regulatory compliance growing as fast as extended reality to avoid misunderstandings in the future?," *Health and Technology,* vol. 13, no. 5, pp. 831-842, Sept. 2023, doi: 10.1007/s12553-023-00775-x.
[87] J. Espinoza, P. Shah, G. Nagendra, Y. Bar-Cohen, and F. Richmond, "Pediatric Medical Device Development and Regulation: Current State, Barriers, and Opportunities," *Pediatrics,* vol. 149, no. 5, 2022, Art no. e2021053390, doi: 10.1542/peds.2021-053390.
[88] C. Bublitz, F. Molnár-Gábor, and S. R. Soekadar, "Implications of the novel EU AI Act for neurotechnologies," *Neuron,* vol. 112, no. 18, pp. 3013-3016, Sept. 25 2024, doi: 10.1016/j.neuron.2024.08.011.
[89] E. Hine, I. N. Rezende, H. Roberts, D. Wong, M. Taddeo, and L. Floridi, "Safety and Privacy in Immersive Extended Reality: An Analysis and Policy Recommendations," *Digital Society,* vol. 3, no. 2, p. 33, Jul. 3 2024, doi: 10.1007/s44206-024-00114-1.
[90] M. Alkaeed, A. Qayyum, and J. Qadir, "Privacy preservation in Artificial Intelligence and Extended Reality (AI-XR) metaverses: A survey," *Journal of Network and Computer Applications,* vol. 231, Nov. 2024, Art no. 103989, doi: 10.1016/j.jnca.2024.103989.
[91] O. Freyer and S. Gilbert, "Bridging between hype and implementation in medical extended reality," *npj Digital Medicine,* vol. 6, no. 1, p. 227, Dec. 2023, doi: 10.1038/s41746-023-00972-y.
[92] FDA, "Medical Extended Reality Program: Research on Medical Extended Reality-Based Medical Devices," Mar. 24 2021. [Online]. Available: https://www.fda.gov/medical-devices/medical-device-regulatory-science-research-programs-conducted-osel/medical-extended-reality-program-research-medical-extended-reality-based-medical-devices.
[93] M. Hassenzahl, S. Diefenbach, and A. Göritz, "Needs, affect, and interactive products – Facets of user experience," *Interacting with Computers,* vol. 22, no. 5, pp. 353-362, 2010, doi: 10.1016/j.intcom.2010.04.002.
[94] C. Lallemand, G. Gronier, and V. Koenig, "User experience: A concept without consensus? Exploring practitioners' perspectives through an international survey," *Computers in Human Behavior,* vol. 43, pp. 35-48, Feb. 2015, doi: 10.1016/j.chb.2014.10.048.
[95] R. Beauregard and P. Corriveau, "User Experience Quality: A Conceptual Framework for Goal Setting and Measurement," in *Digital Human Modeling,* Berlin, Heidelberg, V. G. Duffy, Ed., 2007: Springer Berlin Heidelberg, pp. 325-332.
[96] J. A. Bargas-Avila and K. Hornbæk, "Old wine in new bottles or novel challenges: a critical analysis of empirical studies of user experience," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems,* pp. 2689–2698, 2011, doi: 10.1145/1978942.1979336.

[97] B. Birckhead *et al.*, "Recommendations for Methodology of Virtual Reality Clinical Trials in Health Care by an International Working Group: Iterative Study," *JMIR Ment Health,* vol. 6, no. 1, Jan. 2019, Art no. e11973, doi: 10.2196/11973.
[98] M. J. Page, L. Shamseer, and A. C. Tricco, "Registration of systematic reviews in PROSPERO: 30,000 records and counting," *Systematic Reviews,* vol. 7, no. 1, p. 32, Feb. 20 2018, doi: 10.1186/s13643-018-0699-4.
[99] A. M. Methley, S. Campbell, C. Chew-Graham, R. McNally, and S. Cheraghi-Sohi, "PICO, PICOS and SPIDER: a comparison study of specificity and sensitivity in three search tools for qualitative systematic reviews," (in eng), *BMC Health Serv Res,* vol. 14, p. 579, Nov. 21 2014, doi: 10.1186/s12913-014-0579-0.
[100] U. o. Y. Centre for Reviews and Dissemination, *Systematic Reviews: CRD's guidance for undertaking reviews in health care* (Systematic Reviews). York Publishing Services Ltd, 2009.
[101] M. J. Page *et al.*, "The PRISMA 2020 statement: an updated guideline for reporting systematic reviews," *BMJ,* vol. 372, 2021, Art no. n71, doi: 10.1136/bmj.n71.
[102] M. G. Pagé, J. Katz, J. Stinson, L. Isaac, A. L. Martin-Pichora, and F. Campbell, "Validation of the Numerical Rating Scale for Pain Intensity and Unpleasantness in Pediatric Acute Postoperative Pain: Sensitivity to Change Over Time," *The Journal of Pain,* vol. 13, no. 4, pp. 359-369, 2012, doi: 10.1016/j.jpain.2011.12.010.
[103] S. McKinley, K. Coote, and J. Stein-Parbury, "Development and testing of a Faces Scale for the assessment of anxiety in critically ill patients," *Journal of Advanced Nursing,* vol. 41, no. 1, pp. 73-79, Jan. 2003, doi: 10.1046/j.1365-2648.2003.02508.x.
[104] W.-G. Yun *et al.*, "Tele-consent using mixed reality glasses (NREAL) in pediatric inguinal herniorrhaphy: a preliminary study," *Scientific Reports,* vol. 12, no. 1, p. 3105, Feb. 24 2022, doi: 10.1038/s41598-022-06653-2.
[105] K. Hornbæk and M. Hertzum, "Technology Acceptance and User Experience: A Review of the Experiential Component in HCI," *ACM Trans. Comput.-Hum. Interact.,* vol. 24, no. 5, p. Article 33, 2017, doi: 10.1145/3127358.
[106] H. Brun *et al.*, "Mixed reality holograms for heart surgery planning: first user experience in congenital heart disease," *European Heart Journal - Cardiovascular Imaging,* vol. 20, no. 8, pp. 883-888, 2019, doi: 10.1093/ehjci/jey184.
[107] D. Alexandrovsky *et al.*, "Evaluating User Experiences in Mixed Reality," presented at the Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems, Yokohama, Japan, 2021.
[108] T. Weldring and S. M. S. Smith, "Article Commentary: Patient-Reported Outcomes (PROs) and Patient-Reported Outcome Measures (PROMs)," *Health Services Insights,* vol. 6, Jan. 2013, Art no. HSI.S11093, doi: 10.4137/HSI.S11093.
[109] C. Correale *et al.*, "Improving the Emotional Distress and the Experience of Hospitalization in Children and Adolescent Patients Through Animal Assisted Interventions: A Systematic Review," (in English), *Frontiers in Psychology,* Systematic Review vol. 13, Mar. 4 2022, doi: 10.3389/fpsyg.2022.840107.
[110] K. T. Naylor, S. Kingsnorth, A. Lamont, P. McKeever, and C. Macarthur, "The effectiveness of music in pediatric healthcare: a systematic review of randomized controlled trials," (in eng), *Evid Based Complement Alternat Med,* vol. 2011, 2011, Art no. 464759, doi: 10.1155/2011/464759.
[111] M. Lindström Nilsson, E.-L. Funkquist, A. Edner, and G. Engvall, "Children report positive experiences of animal-assisted therapy in paediatric hospital care," *Acta Paediatrica,* vol. 109, no. 5, pp. 1049-1056, Jun. 2020, doi: 10.1111/apa.15047.

[112] V. Kammin, L. Fraser, K. Flemming, and J. Hackett, "Experiences of music therapy in paediatric palliative care from multiple stakeholder perspectives: A systematic review and qualitative evidence synthesis," (in eng), *Palliat Med*, vol. 38, no. 3, pp. 364-378, Mar. 2024, doi: 10.1177/02692163241230664.
[113] H. G. Hoffman, D. R. Patterson, R. A. Rodriguez, R. Peña, W. Beck, and W. J. Meyer, "Virtual Reality Analgesia for Children With Large Severe Burn Wounds During Burn Wound Debridement," (in eng), *Front Virtual Real*, vol. 1, Dec. 2020, doi: 10.3389/frvir.2020.602299.
[114] B. M. R. Spiegel *et al.*, "What Is Medical Extended Reality? A Taxonomy Defining the Current Breadth and Depth of an Evolving Field," (in eng), *J Med Ext Real*, vol. 1, no. 1, pp. 4-12, Jan. 1 2024, doi: 10.1089/jmxr.2023.0012.
[115] S. Cruz Rivera *et al.*, "Directrices para los protocolos de ensayos clínicos de intervenciones con inteligencia artificial: la extensión SPIRIT-AI," (in es), *Revista Panamericana de Salud Pública*, Journal articles vol. 48, p. 17 p., 2024, doi: 10.26633/RPSP.2024.12.
[116] L. Castelo-Branco *et al.*, "ESMO Guidance for Reporting Oncology real-World evidence (GROW)," *ESMO Real World Data and Digital Oncology*, vol. 1, 2023, doi: 10.1016/j.esmorw.2023.10.001.
[117] X. Liu, S. Cruz Rivera, D. Moher, M. J. Calvert, A. K. Denniston, and S.-A. I. y. C.-A. I. Grupo de Trabajo, "Directrices para presentación de informes de ensayos clínicos sobre intervenciones con inteligencia artificial: extensión CONSORT-AI," (in es), *Revista Panamericana de Salud Pública*, vol. 48, p. 15 p., 2024, doi: 10.26633/RPSP.2024.13.
[118] A. Baba, M. Smith, B. K. Potter, A.-W. Chan, D. Moher, and M. Offringa, "Guidelines for reporting pediatric and child health clinical trial protocols and reports: study protocol for SPIRIT-Children and CONSORT-Children," *Trials*, vol. 25, no. 1, p. 96, Jan. 2024, doi: 10.1186/s13063-024-07948-7.
[119] R. A. Stiles, S. S. Mick, and C. G. Wise, "The Logic of Transaction Cost Economics in Health Care Organization Theory," *Health Care Management Review*, vol. 26, no. 2, pp. 85-92, 2001. [Online]. Available: http://www.jstor.org/stable/44951021.
[120] A. Sertkaya, R. DeVries, A. Jessup, and T. Beleche, "Estimated Cost of Developing a Therapeutic Complex Medical Device in the US," (in eng), *JAMA Netw Open*, vol. 5, no. 9, Sept. 1 2022, Art no. e2231609, doi: 10.1001/jamanetworkopen.2022.31609.
[121] J. I. Gold, M. SooHoo, A. M. Laikin, A. S. Lane, and M. J. Klein, "Effect of an Immersive Virtual Reality Intervention on Pain and Anxiety Associated With Peripheral Intravenous Catheter Placement in the Pediatric Setting: A Randomized Clinical Trial," (in eng), *JAMA Netw Open*, vol. 4, no. 8, Aug. 2 2021, Art no. e2122569, doi: 10.1001/jamanetworkopen.2021.22569.