Epilepsy & Behavior 153 (2024) 109671

Contents lists available at ScienceDirect

Epilepsy & Behavior

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

Review

Mental health and neurodevelopmental patient-reported outcome measures (PROMs) for children and young people with epilepsy: A systematic review

Bianca De Aveiro [a], Alice Winsor [b], Jessica Davies [a], Timothy R. Nicholson [c], Deb K. Pal [b], Mark P. Richardson [b], Susannah Pick [a,*]

[a] Department of Psychological Medicine, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
[b] Department of Basic and Clinical Neuroscience and MRC Centre for Neurodevelopmental Disorders Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK
[c] Neuropsychiatry Research and Education Group, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK

ARTICLE INFO

Keywords:
Mental health
Neurodevelopment*
Epilepsy
P*ediatric
Child*
Outcome measure

ABSTRACT

Children and young people with epilepsy are at higher risk of mental health disorders and atypical neurodevelopmental outcomes compared to the general population. It is essential to detect such comorbidities early in children with epilepsy and provide appropriate interventions, to improve clinical outcomes. We aimed to identify and evaluate the measurement properties of Patient-Reported Outcome Measures (PROMs) that have been validated specifically to measure mental health and neurodevelopmental outcomes in children and/or young people with epilepsy. We searched Embase, Medline, and PsycINFO in May 2023 for relevant studies. Mental health was defined as psychological symptoms (e.g., anxiety, depression, psychosis) and/or behavioural difficulties (e.g., conduct disorders). Neurodevelopmental outcomes included neurodevelopmental disorder traits such as attention-deficit hyperactivity disorder (ADHD) and autistic spectrum disorders. We assessed methodological quality using Consensus-based Standards for the selection of health Measurement Instruments (COSMIN) guidance. Twelve papers were identified that psychometrically evaluated 13 relevant PROMs (two epilepsy-specific, eleven generic). The appraisal of the PROMs was limited by the availability of only one or two published articles for each, and incomplete psychometric evaluations in some cases. The tool demonstrating the strongest evidence was The Neurological Disorders Depression Inventory-Epilepsy for Youth. The ADHD Rating Scale-IV and The Paediatric Symptom Checklist -17 demonstrated good evidence in favour of at least two measurement properties. This review identified only a small number of mental health and neurodevelopmental PROMs evaluated specifically in paediatric epilepsy. There is a need for further validation of mental health and neurodevelopmental PROMs in children with epilepsy.

1. Introduction

Epilepsy is a chronic neurological condition characterised by the tendency to have recurring seizures [1]. Epilepsy occurs in people of all ages, affecting 65 million people in the world and more specifically 0.9 million children and adolescents in Europe (i.e., prevalence rate of 4.5 per 1000) [2,3]. Epilepsy is one of the most common serious long-term illnesses in young people, with a lifetime prevalence of 1 % [4], with the World Health Organization (WHO) reporting that the risk of premature death is up to three times higher in those with epilepsy than in the general population [5]. Furthermore, when reviewing the global burden of epilepsy, the WHO estimated that up to 70 % of those living with epilepsy could be seizure-free if appropriately diagnosed and treated [5].

One out of every three people with epilepsy experience a comorbid mental health disorder, with mood and anxiety disorders being particularly common and more prevalent in people with epilepsy than in the general population [6]. Children and young people with epilepsy, in particular, are at a higher risk of mental health and neurodevelopmental disorders such as depression, anxiety, intellectual disability and attention deficit hyperactivity disorder (ADHD) compared to the general population [7]. These comorbidities have been associated with reduced tolerance (increased side effects, adverse reactions) to anti-seizure medications (ASMs) in people with epilepsy, which can consequently

* Corresponding author.
E-mail address: susannah.pick@kcl.ac.uk (S. Pick).

https://doi.org/10.1016/j.yebeh.2024.109671
Received 9 November 2023; Received in revised form 26 January 2024; Accepted 28 January 2024
Available online 19 February 2024

interfere with compliance, and as a result, increase risk of seizure recurrence [8,9]. This pattern has also been noted specifically in children with epilepsy [10]. Failing to treat such comorbidities in people with epilepsy can have a substantial negative impact on clinical outcomes, potentially leading to inadequate response to treatment, reduction in quality of life, and increased mortality [9,11]. Despite the high prevalence and negative impact, mental health and neurodevelopmental comorbidities in epilepsy often go undetected and untreated [12]. It is essential to detect such comorbidities early and provide appropriate interventions, to improve long-term clinical outcomes through appropriate intervention and care [7,11].

Patient-reported outcome measures (PROMs) are typically self-report questionnaires or scales which directly measure the subjective experiences of patients in relation to a specific physical or mental health condition [13]. PROMs can measure a range of concepts, including cognitive, physical, emotional and social health perceptions [13]. PROMs were initially developed for use in research; however, they have now been adopted by clinicians to enhance the clinical management of individual patients [14]. PROMs can be implemented into routine clinical practice to understand patients' views of their symptoms, functional status, satisfaction, and health-related quality-of-life (HR-QoL) [14,15]. Previous research has shown that routine use of PROMs has the potential to transform healthcare as they can be used to monitor patients' progress and to inform clinical decision-making regarding their treatment [14].

Proactive screening with relevant PROMS would facilitate timely identification of, and early intervention for, mental health and neurodevelopmental comorbidities in children with epilepsy [16]. Using PROMs such as self-report questionnaires, researchers can take into account the young person's perspective, as well as that of their families and carers, and in turn create a biopsychosocial model of the impact of epilepsy on their lives [17]. This information can then be used to develop or identify biopsychosocial interventions specifically for this population [17].

When selecting PROMs, careful consideration is needed regarding the content of the measure and the relevance to the intended patient group [18]. A measure is considered appropriate when published evidence shows that it is (a) acceptable to patients, (b) reliable, (c) valid and (d) sensitive to change [19]. Furthermore, there must be evidence that these properties have been evaluated in a relevant context with similar types of patients (i.e., age, gender, diagnostic category, cultural context) for whom the PROMs are to be applied [18].

The objectives of this review were to identify mental health and neurodevelopmental PROMs that have been evaluated psychometrically in children with epilepsy, and to critically appraise their measurement properties. We included PROMs that were patient-reported, carer-reported, or both.

2. Methods

A systematic review was carried out to identify relevant studies. The results were grouped according to the mental health symptom/disorder or neurodevelopmental trait/disorder being measured. Methodological quality was assessed using the COSMIN Risk of Bias (RoB) tool [20]. A narrative synthesis was employed to describe and interpret the findings.

2.1. Search strategy and selection criteria

We searched MEDLINE, EMBASE and PsycINFO in December 2021 (BA) and updated the searches in May 2023. The searches were conducted using the following terms: (*"epilepsy" OR "seizures" OR "epileptic") AND ("p?ediatric*" OR "child*" OR "young pe*" OR "adolesc*" OR "teen*" OR "youth") AND ("outcome measure*" OR "PROM*" OR "measurement instrument" OR "scale*" OR "questionnaire")* AND ("exp *mental disorders" OR "depress*" OR "behav*" OR "psych*")*.

Studies were included if they met both of the following criteria:

1. The study explicitly assessed the measurement (psychometric) properties of a mental health or neurodevelopmental PROM in a sample including children with epilepsy (<18 years) and/or their parents/carers. Mental health was defined as any psychological/psychiatric symptom such as anxiety, depression, and psychosis, as well as behavioural problems (e.g., conduct disorders). Neurodevelopmental disorder traits referred to those of any recognised neurodevelopmental disorder, including ADHD, autistic spectrum disorder, and intellectual disability. Measurement properties referred to any psychometric assessment of reliability (e.g., internal consistency, test–retest reliability), validity (e.g., face, construct, criterion), or responsiveness (e.g., pre-post treatment change).

2. The study was a full original research article published in a peer-reviewed journal

Articles were excluded if:

(a) the publication was not written in English,

(b) the source was unpublished (e.g., dissertations/theses, preprints, conference abstracts)

(c) the study evaluated measures that were exclusively interviewed-based or clinician-rated

(d) there was no evidence of psychometric evaluation in paediatric epilepsy patients.

(e) The study reported data on mixed samples including patients with epilepsy and other disorders included (e.g., other neurological diagnoses, functional seizures) without epilepsy-specific data presented separately.

There was no restriction on publication period or search dates.

Following the electronic database search, the Rayyan rating system (https://www.rayyan.ai/) was used to remove duplicates. Rayyan is a web-tool designed for systematic and scoping reviews that screen and select articles according to the authors specific inclusion and exclusion criteria. Remaining studies were then initially screened by the first author (BA) based on title and/or abstract using the pre-determined eligibility criteria. Any articles that did not meet the inclusion criteria were excluded. Two reviewers (BA and AW/SP/JD) independently reviewed all remaining titles and abstracts. Full texts of potentially eligible studies were retrieved. Two reviewers (BA and AW/JD/SP) independently assessed each full text against the exclusion criteria. Any disagreements were resolved by discussion or by a third reviewer.

2.2. Data synthesis and quality assessment

Descriptive data on the eligible PROMS were extracted, including instrument version, original author, intended purpose of the PROM, number of items and domains, age range and respondent (Table 1). Data were also extracted on the details of the studies that psychometrically evaluated the eligible PROMs including instrument version, aim of the study, population, mean age and psychometric properties (Table 2).

Each included paper was assessed by two independent reviewers (BA and AW/JD) for its methodological quality using the COSMIN risk of bias (RoB) checklist for use in the systematic reviews of PROMs (Table 3) (https://www.cosmin.nl/). The quality of each paper was assessed on the basis of the methods used to evaluate the PROM's measurement properties, which included content validity, internal consistency, test–retest reliability, precision, and responsiveness, among others (see [20] for definitions). We used the COSMIN four-point scale: "very good", "adequate", "doubtful", "inadequate" to rate methodological quality for each measurement property in every study. The COSMIN checklist operates a "worst score counts" principle, meaning that a score for each measurement property's quality is determined by selecting the lowest rating among the items within that category [20].

3. Results

3.1. Search results

The search results are presented in Fig. 1. A total of twelve eligible

Table 1
Mental health and neurodevelopmental Patient Reported Outcome Measures (PROMs).

Instrument Version (PROM)	Author	Purpose	Number of items and domains	Age Range	Respondent	Country/ origin
Aberrant Behaviour Checklist	Aman et al (1985) [21]	To assess the severity of intellectual disability	58 items, 5 domains: irritability, hyperactivity/ Noncompliance, lethargy/social withdrawal, stereotypic behaviours, and inappropriate speech	6–54 years	Parent	New Zealand
The ADHD Rating-Scale-IV	DuPaul et al (1998) [22]	To rate severity of ADHD symptoms	18 items, 2 domains: inattention and hyperactivity-impulsivity	6–14 years	Parent	USA
Adolescent Psychosocial Seizure Inventory (APSI)	Batzel et al (1991) [23]	To assess psychosocial problems in adolescents with epilepsy	38 items, 9 domains: Family background, emotional adjustment, interpersonal adjustment, school adjustment, vocational outlook, adjustment to seizures, medical management, antisocial activity and overall psychosocial functioning	12–19 years	Child	USA
Behavioural Assessment Scale for Children - Second Edition (BASC-2)	Reynolds & Kamphaus (2004) [24]	To assess children's emotional, behavioural and social functioning	5 measures: Teacher Rating Scale (100–139 items), Parent Rating Scale (134–160 items), Self-Report of Personality (139–185 items), Structured Developmental History, Student Observation System.16 domains: activities of daily living, functional communication, adaptability, hyperactivity, aggression, leadership, anxiety, learning problems, attention problems, social skills, atypicality, somatisation, conduct disorder, study skills, depression, withdrawal	2–25 years	Teacher, parent or Child	Indianapolis, USA
The Center for Epidemiologic Studies Depression Scale for Children (CES-DC)	Weissman et al (1980) [25]	To rate how many depressive symptoms have been experienced over the past week	20 items	6–17 years	Child	Connecticut, USA
Child Behaviour Checklist (CBCL/6–18)	Achenbach & Rescorla (2001) [26]	To assess emotional and behavioural difficulties in children and young people.	113 items, 2 scales – problem behaviour scale, social competence scale. 8 domains in problem behaviour scale: anxious/depressed, depressed, somatic complaints, social problems, thought problems, attention problems, rule-breaking behaviour and aggressive behaviour	6–18 years	Parent	Vermont, USA
The Child Depression Inventory (CDI)	Kovacs (1985) [27]	To measure the cognitive, affective, and behavioural signs of depression	27 items, 5 subscales: anhedonia, ineffectiveness, interpersonal problems, negative mood, negative self-esteem	7–17 years	Child	Pennsylvania, USA
Conner's Parent Rating Scale (CPRS-48)	(Conners, 1989) [28]	To assess the severity of ADHD symptoms	48 items, 5 domains: conduct problems, learning problems, anxiety, impulsive/hyperactive behaviour, and psychosomatic feelings	3–17 years	Parent	Toronto, Canada
Neurological Disorders Depression Inventory-Epilepsy for Youth (NDDI-E-Y)	Wagner et al (2013, 2016) [29,30]	To screen depressive symptoms in youth with epilepsy	12 items	12–17 years	Child	South Carolina, USA
Paediatric Symptom Checklist – 17	Gardner et al (1999) [31]	To screen for childhood emotional and behavioural problems	17 items, 3 domains: attention, internalizing, and externalizing	4–15 years	Parent	USA
Revised Child Anxiety and Depression Scale (RCADS)	Chorpita et al (2000) [32]	To assess different depressive and anxiety symptoms in children	47 items, 6 subscales: social phobia, panic disorder, separation anxiety disorder, generalised anxiety disorder (GAD), major depressive disorder, obsessive–compulsive disorder	8–18 years	Child or parent	Hawaii, USA
Screen for Children Anxiety Related EmotionalDisorders Scale (SCARED)	Birmaher et al (1997) [33]	To assess anxiety disorders in children	38 items, 5 domains: panic/somatic, generalised anxiety disorder (GAD), separation anxiety, social phobia and school avoidance.	9–18 years	Child and Parent	USA
The Symptom Questionnaire	Kellner (1987) [34]	To assess psychological symptoms	92 items, 5 domains: anxiety, depression, anger/ hostility, somatization and well-being.	No age range	Child	New Mexico, USA

articles were identified, of which 13 questionnaires were reported (Table 1). Of the 13 questionnaires, two were epilepsy-specific and eleven were generic PROMs that had been psychometrically evaluated in paediatric epilepsy populations. Table 1 presents the original articles and summarises the details of the PROMs such as the original author, intended purpose of the PROM, number of items and domains, and age range. Summaries of the characteristics of the identified PROMs and the study populations in which their measurement properties were evaluated are detailed in Table 2. Both epilepsy-specific and generic PROMS were developed or assessed in a variety of geographical locations, including USA, Canada, New Zealand, Europe, India and Oman. Collectively, the identified PROMs were designed for children and young people aged 2 to 25 years of age, with the exception of the Aberrant Behaviour Checklist, that was designed for both paediatric and adult populations (age range 6–54 years).

Some of the scales included subscales that, although may not seem directly relevant to mental health and/or neurodevelopment, are still integral components of broader measurement instruments that aim to provide a comprehensive assessment of psychosocial and neurodevelopmental challenges. For example, subscales focusing on adjustment to seizures are part of measurement instruments that encompass mental health dimensions and can thus contribute to a holistic understanding of the mental health of children and young people with epilepsy.

Table 2
Psychometric evaluations of identified PROMs in paediatric epilepsy samples.

Instrument Version (PROM)	Author	Aim of study	Population	Mean age of children with epilepsy (SD)	Psychometric properties	Country/ origin	Funding
Aberrant Behavior Checklist (ABC)	Kaat et al (2021) [35]	To determine validity of the ABC as an outcome measure for pharmacological and behavioural interventions for young people with Developmental and Epileptic Encephalopathies (DEEs).	122 young people with DEEs (including Dravet and Lennox-Gastaut syndromes) and KCNQ2- SCN2A-, and KCNB1-associated disorders	1–35 yearsMedian: 8 (IQR: 4.25–13.0)	Structural validity (item cluster analysis): α = 0.08—0.96 and β = 0.53 – 0.57 Internal consistency: α α = 0.80—-0.94	Chicago, USA	Supported by the Stanley Manne Children's Research Institute and Ann & Robert H. Lurie Children's Hospital of Chicago under the Precision Medicine Strategic Research Initiative and by a grant from the Paediatric Epilepsy Research Consortium, Dallas, TX.
The ADHD Rating-Scale-IV (ADHD-RS-IV)	Mercier et al (2016) [36]	To investigate some psychometric properties of the French version of the ADHD Rating-Scale IV in children with ADHD and epilepsy.To assess validity of its total and sub scores in this populationTo assess the construct validity, internal consistency, reliability of items and responsiveness	167 children (55 girls and 112 boys) from 10 French neuropediatric units were screened and included. Children aged 6 years − 15 years 11 months diagnosed with epilepsy and diagnosis of ADHD.	9.5 (2.4)	Item reliability: R^2 = 0.137—-0.696 Internal consistency: 0.73 for Inattention and 0.87 for Hyperactivity/ Impulsivity Responsiveness: standardised response mean = 1.19 vs 0.53 for treatment vs control group respectively	France	Funded by a grant from the French Ministry of Health (grant number 27.23) (PHRC 2011).
Adolescent Psychosocial Seizure Inventory (APSI)	Batzel et al (1991) [23]	To develop an assessment of psychosocial problems in adolescents with epilepsy.	120 patients aged 12–19 years, diagnosed with epilepsy.61 males, 59 females.	14.78 (1.86)	Interrater reliability: median coefficient = 0.87 Test-retest reliability: median coefficient = 0.77 Internal consistency: median correlation = 0.73	USA	Supported in part by NIH grants awarded by the National Institute of Neurological Disorders and Stroke
Behavioural Assessment Scale for Children - Parent Rating Scale (BASC-PRS)	Bender et al (2008) [37]	To compare convergent validity of the BASC and CBCL in a paediatric epilepsy population.	60 children and adolescents aged 6–17 years old with a diagnosis of epilepsy, 35 males, 25 females	11.0 (3.4)	Correlations between broadband scales of BASC & CBCL: *r* = 0.71–0.79. Correlations for narrowband scales: *r* = 0.41 to 0.78.	New York, USA	Supported by an Epilepsy Foundation of America (EFA) Behavioural Sciences Student Fellowship.
The Center for Epidemiologic Studies Depression Scale for Children (CES-DC)	Al Kiyumi et al (2021) [38]	To assess the frequency of depressive symptoms in children diagnosed with epilepsy in a tertiary care institution in Oman.	75 children aged between 6 and 12 years with a diagnosis of epilepsy (45 boys, 30 girls)	Not reported	Internal consistency: α = 0.8	Oman	NA
Child Behaviour Checklist (CBCL)	Bender et al (2008) [37]	To compare convergent validity of the BASC and CBCL in a paediatric epilepsy population.	60 children and adolescents aged 6–17 years old with a diagnosis of epilepsy, 35 males, 25 females	11.0 (3.4)	Correlations between broadband scales of BASC & CBCL: *r* = 0.71–0.79. Correlations for narrowband scales: *r* = 0.41 to 0.78.	New York, USA	Supported by an Epilepsy Foundation of America (EFA) Behavioural Sciences Student Fellowship.
The Child Depression Inventory (CDI)	Miniksar et al (2022) [39]	To examine suicide probability, factors affecting suicide, and personality traits of children diagnosed with epilepsy, and to compare their results with those of children without epilepsy.	112 children aged between 11 and 16 years56 children with epilepsy (23 boys, 33 girls)56 children without epilepsy (27 boys, 29 girls)	14 (SD not reported)	Internal consistency: α = 0.82	Turkey	Not reported
Conner's Parent Rating Scale (CPRS-48)	Pal et al (1999) [40]	To validate a version of the CPRS-48 in a rural Bengali dialect for use in a study of anti-epileptic drug side effects in village children.	60 healthy children (30 boys and 30 girls) between the ages of 5 and 14. 63 children between the ages of 6 and 18 years with untreated epilepsy.	Healthy children: 8 years 10 months (2 years 5 months)Not reported	Internal consistency: α α = 0.60 – 0.75 Test-retest reliability: α = 0.84 – 0.99	India	Deb Pal was supported by a Wellcome Trust Research Training Fellowship.

(*continued on next page*)

Table 2 (*continued*)

Instrument Version (PROM)	Author	Aim of study	Population	Mean age of children with epilepsy (SD)	Psychometric properties	Country/ origin	Funding
Neurological Disorders Depression Inventory-Epilepsy for Youth (NDDI-E-Y)	Wagner et al (2016) [30]	To validate the revised 12 item NDDI-E-Y by establishing internal consistency, reliability and construct validity.	143 youth participants between 12 and 17 years of age diagnosed with epilepsy.99 females, 44 males.	15.1 (1.7)	Convergent validity with Child Depression Inventory (CDI) (r = 0.70, $p < 0.0001$) Criterion validity: Area under the curve = 0.866Sensitivity (0.79)Specificity (0.92) Internal consistency: α = 0.92	South Carolina, USA	Willy's Fund for Childhood Epilepsy Research
Neurological Disorders Depression Inventory-Epilepsy for Youth (NDDI-E-Y)	Viellard et al (2019) [41]	To evaluate the NDDI-E-Y in screening for major depressive disorder (MDD) in French youth with epilepsy. To determine if the NDDI-E-Y offers strengths over the adult NDDI-E.	97 French-speaking adolescents aged 11–17 years diagnosed with epilepsy (with onset of at least 1 year before).52 females, 45 males.	14.9 (1.7)	Convergent validity with Child Depression Inventory (CDI) (r = 0.848, $p < 0.0001$) Criterion validity: Area under the curve = 0.967Sensitivity (1.00)Specificity (0.82) Internal consistency: α α = 0.862	Marseille, France	NA
Paediatric Symptom Checklist − 17(PSC-17)	Wagner et al (2015) [42]	To provide validity and reliability estimates for use of the PSC-17 in a paediatric population with epilepsy.	187 participants from two cohorts of caregivers of youth receiving care in epilepsy centres. 49.7 % girls, 50.3 % boys	10.3 (5.1)	Structural validity: confirmatory factor analysis X^2 (116, N = 187 = 204.54, p < 0.001, CFI = 0.90, TLI = 0.88, RMSEA = 0.064, SRMSR = 0.078). Internal consistency: α = 0.72–0.85 Interscale correlation: r = 0.48 – 0.69	South Carolina, USA	Funding provided by the Epilepsy Foundation "Partnership for Paediatric Epilepsy Research" and Dr Wagner received the William R. Turk Award for Paediatric Epilepsy Research for this study.
Revised Child Anxiety and Depression Scale (RCADS)	Rogac et al (2021) [43]	To evaluate changes in overall cognitive profiles, psychopathological symptoms, and quality of life in newly diagnosed, uncomplicated paediatric epilepsy.	61 participants aged between 8 and 18 years with a diagnosis of epilepsy (35 boys, 26 girls)	12.49 (3.21)	Internal consistency: α = ≥ 0.65 for majority of the scales, GAD scale for parent-report = 0.57, MDD scale for self-report = 0.50.	Serbia	NA
Screen for Children Anxiety Related EmotionalDisorders Scale (SCARED)	Carrozzino et al (2016) [44]	To evaluate the SCARED as a candidate for depression and anxiety screening in adolescent epilepsy by focusing on the extent to which one single item is sufficient in terms of screening.	29 participants with a diagnosis of epilepsy. 29 healthy non-epilepsy participants	12.4 (1.6)13.3 (1.1)	Internal consistency: α = 0.52 – 0.79 Validity – coefficient of homogeneity: 0.22—–0.28	Italy	NA
The Symptom Questionnaire (SQ)	Carrozzino et al (2016) [44]	To evaluate the SQ as a candidate for depression and anxiety screening in adolescent epilepsy by focusing on the extent to which one single item is sufficient in terms of screening.	29 participants with a diagnosis of epilepsy (16 boys, 13 girls).29 healthy non-epilepsy participants (16 boys, 13 girls) .	13.3 (1.1)	Internal consistency: α = 0.80 – 0.90 Validity – coefficient of homogeneity: 0.25—–0.65	Italy	NA

3.2. Quality assessment

Table 3 summarises the results of the quality assessment for each psychometric evaluation study. The most commonly assessed properties across studies were internal consistency and convergent validity, whereas there were very few evaluations of PROM development, test–retest reliability, and criterion validity (Table 3). Furthermore, it was not possible to rate any of the PROMs in relation to content validity, cross-cultural validity, measurement error or content validity, as statistical evaluations were not reported and consequently, study quality and instrument ratings were not performed for these domains. Overall appraisal of the psychometric properties of each PROM was limited by the availability of only one or two published psychometric evaluations for each that were generally incomplete.

3.3. Narrative synthesis

3.3.1. PROMs assessing ADHD

The ADHD Rating Scale-IV (ADHD-RS-IV) is a parent-rated 18-item PROM based on the diagnostic criteria for ADHD as defined in the 4th edition of the *Diagnostic and Statistical Manual of Mental Disorders of the American Psychiatric Association* [22,36]. Mercier and colleagues [36] investigated the psychometric properties of the French version of the ADHD-RS-IV in children with ADHD and epilepsy. The evaluation of

Table 3
Methodological quality of psychometric evaluation studies using the COSMIN Risk of Bias checklist.

		Internal Structure		Reliability		Hypothesis testing for construct validity	Responsiveness	
Instrument & Author	**PROM development**	**Structural Validity**	**Internal Consistency**	**Test-retest reliability**	**Criterion validity**	**Comparison with other measures (convergent validity)**	**Construct approach: (i.e. hypotheses testing: comparison between subgroups)**	**Construct approach: (i.e. hypotheses testing: before and after intervention)**
ABC[1] (Kaat et al, 2021) [35]		Adequate	Doubtful			Very good		
ADHD-RS-IV[2](Mercier et al, 2016) [36]		Adequate	Very good				Doubtful	Very good
APSI[3] (Batzel et al, 1991) [23]	Doubtful		Inadequate	Inadequate				
BASC[4] (Bender et al, 2008) [37]						Very good		
CES-DC[5] (Kiyumi et al, 2021) [38]			Very good					
CBCL[6] (Bender et al, 2008) [37]						Very good		
CDI[7] (Miniksar et al, 2022) [39]			Inadequate					
CPRS-48[8] (Pal et al, 1999) [40]		Inadequate	Very good	Doubtful				
NDDI-E-Y[9](Wagner et al, 2016) [30]	Doubtful	Adequate	Very good		Very good	Very good		
NDDI-E-Y[9](Viellard et al, 2019) [41]			Very good		Very good	Very good		
PSC-17[10] (Wagner et al, 2015) [42]		Very good	Very good					
RCADS[11] (Rogac et al, 2021) [43]			Very good					
SCARED[12](Carrozzino et al, 2016) [44]			Very good					
SQ[13] (Carrozzino et al, 2016) [44]			Very good					

[1] Aberrant Behaviour Checklist; [2] The ADHD Rating Scale-IV; [3] Adolescent Psychosocial Seizure Inventory; [4] Behavioural Assessment Scale for Children; [5] The Center for Epidemiologic Studies Depression Scale for Children; [6] Child Behaviour Checklist; [7] The Child Depression Inventory; [8] Conner's Parent Rating Scale; [9] Neurological Disorders Depression Inventory – Epilepsy for Youth; [10] Paediatric Symptom Checklist; [11] Revised Child Anxiety and Depression Scale; [12] Screen for Children Anxiety Related Emotional Disorders Scale; [13] The Symptom Questionnaire.

responsiveness (in terms of expected differences in changes between subgroups) of the ADHD-RS-IV was rated as *doubtful* due to a minor methodological flaw. Participants in the study by Mercier and colleagues [36] did not undergo systematic IQ testing and thus the relationship between ADHD and IQ could not be evaluated. For this reason, variable IQ levels among participants could have been a confounding factor [36].

The structural validity appraisal was rated *adequate* as although a confirmatory factor analysis (CFA) was performed, the aforementioned methodological flaw lowered the overall rating. On the other hand, the evaluation of responsiveness (in terms of expected magnitude of change following an intervention) of the ADHD-RS-IV was rated as *very good* as the Standardised Response Mean (SRM) was computed and yielded a value of 1.19 for the treatment group which is similar to SRM reported by other authors [45,46]. The two domains (Inattention and Hyperactivity/Impulsivity) of the ADHD-RS-IV had evidence of acceptable internal consistency ($\alpha = 0.73 – 0.87$) [36]. Item reliability was found to be lower for Inattention items compared to Hyperactivity/Impulsivity items ($R^2 = 0.137$ vs 0.696); however, this is consistent with previous studies using the English version [44]. Based on these results Mercier and colleagues [36] concluded that the ADHD-RS-IV is an appropriate objective tool to assess behaviour in children with ADHD and epilepsy.

The Conner's Parent Rating Scale-48 (CPRS-48) [28,40] is a parent-rated 48-item PROM that is designed to assess the severity of ADHD symptoms in general populations. Pal et al. [40] aimed to validate a Bengali adaptation of the CPRS-48 in paediatric epilepsy. Test-retest reliability was rated *doubtful* due to a lack of detail regarding the stability of participants during the interim period. However, most subscales demonstrated excellent test–retest reliability ($\alpha = 0.95 – 0.99$), except the conduct problem scale which demonstrated good test–retest reliability ($\alpha = 0.84$). Whilst the measure was rated methodologically *very good* due to meeting all criteria on the COSMIN internal consistency tool, most subscales had Cronbach's alpha within a questionable range ($\alpha = 0.60 – 0.69$), aside from the hyperactivity index scale [40]. The authors concluded that the level of internal consistency and good test–retest reliability confirmed the validity and stability of the Bengali version of the CPRS-48 in paediatric epilepsy [40].

3.3.2. PROMs assessing behaviour

The Aberrant Behaviour Checklist (ABC) is a 58-item parent behaviour rating scale originally developed to measure behavioural problems of children and adults with intellectual disability [21]. The ABC was evaluated in a sample of young people with Developmental and Epileptic Encephalopathies (DEEs) by Kaat et al. [35]. The study was rated methodologically *adequate,* providing evidence for structural

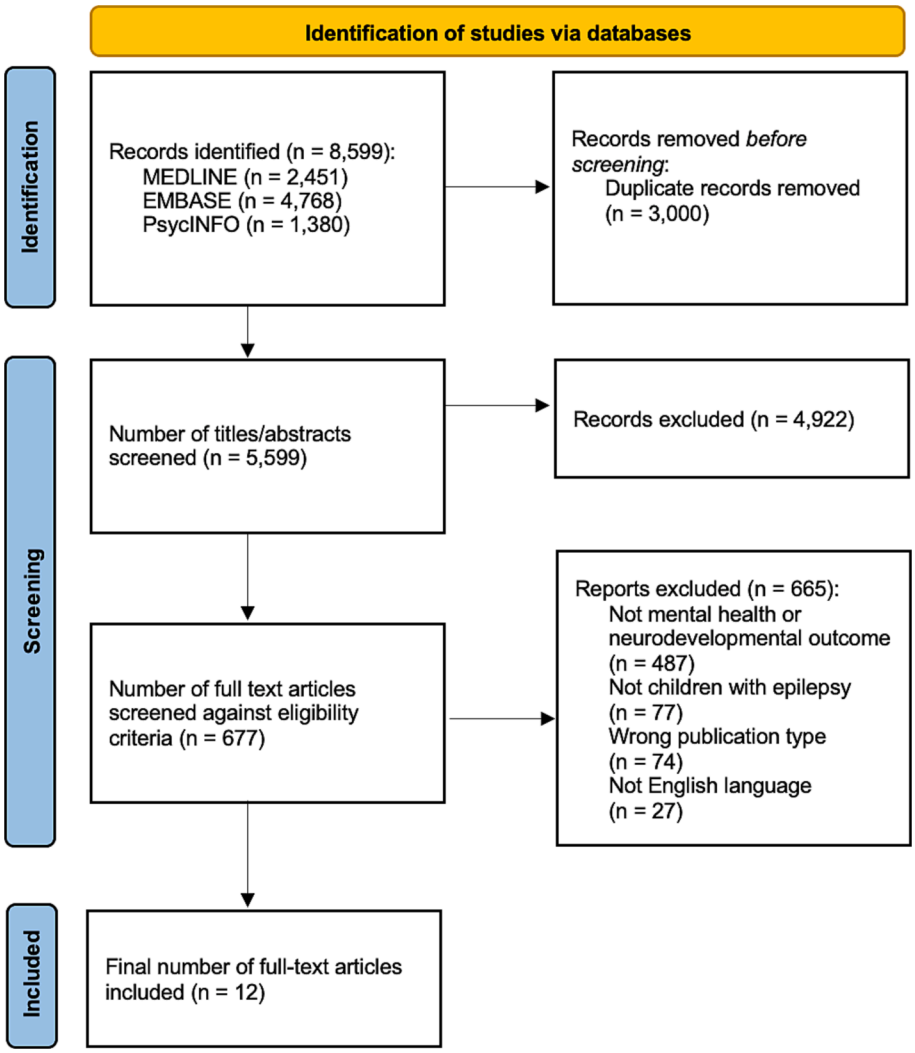

Fig. 1. PRISMA flowchart illustrating identification and selection of eligible studies.

validity. Although the study did not use factor analysis to assess structural validity, it did use a combination approach using Cronbach's alpha (split-half reliability) and coefficient beta, which may be more methodologically appropriate than a factor analysis given the study's small sample size [35]. The study was also rated *very good* for its risk of bias concerning construct validity, as it obtained correlations for the ABC with other validated measures, thus demonstrating the convergent validity of the scale.

The Behavioural Assessment Scale for Children (BASC) and The Child Behaviour Checklist (CBCL) were assessed for convergent validity by Bender and colleagues [37]. In this study, the BASC Parent Rating Scale (BASC-PRS) was used to assess adaptive functioning and behaviour [24]. The BASC-PRS has three developmentally appropriate versions; preschool (ages 2–5 years), child (ages 6–11 years) and adolescent (ages 12–21 years). Parents of children with epilepsy answer between 134 and 160 items by rating the behaviour frequency of their child on a 4-point scale (ranging from "never" to "almost always"). Similarly, The Child Behaviour Checklist (CBCL) is a 113-item parent-reported behaviour rating scale that also has different developmentally appropriate versions [26]. The study by Bender et al. [37] utilised the CBCL - Ages 6–18 version and asked parents of children with epilepsy to rate the behaviour frequency of their child on a 3-point scale (e.g., 0 = "not true", 1 = "somewhat true" and 2 = "very true"). The study was rated as *very good* for its risk of bias as it met all criteria for convergent validity by comparing the BASC to the CBCL, a gold standard tool for screening behavioural problems in children [47]. Bender et al. [37] also evaluated broadband and narrowband behaviour rating scales of the BASC and CBCL. Significant correlations were reported for broadband scales (r = 0.71—0.79), whereas narrowband correlations were more variable (range = 0.41–0.78). In general, correlations were higher for scales measuring externalizing behaviours (r = 0.58–0.78) compared to scales measuring internalising behaviours (r = 0.43–0.65).

The Paediatric Symptom Checklist −17 (PSC-17) is a 17-item parent-rated PROM [31]. The PSC-17 is a shortened version of the Paediatric Symptom Checklist, which is a 35-item screening tool used for the early detection of behavioural and emotional problems in children [48]. Structural validity was assessed using confirmatory factor analysis with a paediatric epilepsy population and the method was rated *very good* due to the sample size (n = 187) being large enough to meet the required COSMIN criteria (sample size over 100, and 7 times the number of items in the scale [20]) [42]. Supporting evidence was found for good internal consistency of the overall scale (α = 0.85) and acceptable internal consistency for the three sub scales (α α = 0.72 – 0.78) which is consistent with previous research [31]. Lastly, moderate inter-scale correlations were found (r = 0.48 – 0.69), similar to previous studies [31]. Thus, the PSC-17 demonstrated good psychometrics when used in a sample of caregivers of children and adolescents with epilepsy, consistent with findings in healthy children and those with chronic illness [49].

3.3.3. PROMs assessing anxiety or depression

The Child Depression Inventory (CDI) is a self-report scale for children aged 7–17 years to understand the depressive symptoms they may be experiencing at a given point in time [27]. In a study by Miniksar et al. [39], a Turkish adaptation of the CDI was used to help assess suicide probability and risks in children diagnosed with epilepsy. The study has been rated methodologically *inadequate* for its internal consistency risk of bias. This is because, despite publishing a Cronbach's alpha for the whole scale, no internal consistency scores were provided for the five subscales of the CDI, which violates the design requirements of the COSMIN checklist.

The Neurological Disorders Depression Inventory-Epilepsy for Youth (NDDI-E-Y) is a 12-item self-reported PROM designed to assess depressive symptomology in young people with epilepsy [30]. The original version of the NDDI-E-Y was an 11-item questionnaire based on the 6-item NDDI-E for adults [29]. The evaluation of the PROM development was rated as *doubtful* due to a lack of qualitative data. Two studies provided support of satisfactory convergent validity by comparing the NDDI-E-Y with the Children's Depression Inventory (CDI) [30,41], a gold standard for screening major depressive disorder (MDD) in children [50]. The convergent validity for both studies was rated as *very good* as both tools measure the same construct (i.e., depression). Wagner and colleagues [30] found a strong positive correlation between the NDDI-E-Y and the CDI-2 ($r = 0.70$, $p < 0.0001$), and found that twenty-five percent of participants who also completed the CDI-2 had T-scores of ≥ 65, which signal clinical levels of depression [51]. Similarly, Viellard and colleagues [41] found a strong positive correlation between the NDDI-E-Y and the CDI ($r = 0.0848$, $p < 0.0001$). Criterion validity was assessed in both studies by calculating the Area Under the Curve (AUC), as well as sensitivity and specificity. A total score of 0.866 [30] and 0.848 [41] for AUC for the CDI was reported, indicating the NDDI-E-Y can distinguish participants with high/low total scores on the CDI. Viellard et al. [41] reported sensitivity of 1.00 and specificity of 0.82 using a cut-off score of 23, and Wagner et al. [30] reported sensitivity (0.79) and specificity (0.92) with a cut-off score of 32. Both studies demonstrated good internal consistency ($\alpha = 0.862$ and $\alpha = 0.92$) [30,41]. Additionally, the study by Wagner and colleagues [30] reported an exploratory factor analysis (EFA), therefore receiving a rating of *adequate* in terms of structural validity. COSMIN criteria require a CFA to be performed to achieve a rating of *very good* in this category.

The Screen for Children Anxiety Related Emotional Disorders Scale (SCARED) is a 38-item self-rating scale to assess anxiety disorders [33]. One study evaluated the individual panic attack item in the SCARED as a screening candidate for adolescent epilepsy [44]. The study was rated methodologically *very good* for its risk of bias concerning its internal structure; however, only two of the five domains showed acceptable internal consistency (Generalised Anxiety Disorder: $\alpha = 0.73$, panic disorder: α $\alpha = 0.79$), whereas the other three (separation anxiety, school anxiety, social anxiety) domains demonstrated poor/questionable internal consistency ($\alpha = 0.52 – 0.63$). In addition, the evaluation of responsiveness (within subgroup comparison) was rated as *doubtful* because although the Mann-Whitney test was used to adequately assess comparisons between clinical and control groups, there were minor methodological flaws in the study.

The Revised Child Anxiety and Depression Scale (RCADS) is a 47-item self- or parent-report used to assess different depressive and anxiety symptoms in children [32]. A study by Rogac and colleagues [43] utilised this PROM to evaluate psychopathological changes in children with newly diagnosed epilepsy. The study was rated *very good* for its methodological internal structure, due to appropriate use of Cronbach's alpha to generate internal consistency scores for the subscales of the PROM. Despite this, the generalised anxiety disorder parent-report and the major depressive disorder self-report subscales both displayed poor internal consistency.

The Center for Epidemiologic Studies Depression Scale for Children (CES-DC) is a 20-item self-report scale used to rate depressive symptoms experienced over the previous week [25]. A study by Al Kiyumi and colleagues [38] sought to provide an Arabic translation of the scale for use in their study. They followed a validated translation method and produced a scale with a Cronbach's alpha of 0.8, which can be interpreted as good. The methodological risk of bias score for internal structure of this study was rated *very good*, due to appropriate Cronbach's alpha calculations for the entirety of the scale.

3.3.4. PROMs assessing psychological symptoms

The Adolescent Psychosocial Seizure Inventory (APSI) is a 38-item self-reported PROM, based on the Washington Psychosocial Seizure Inventory, a scale used to evaluate psychosocial problems in adults with epilepsy [23]. The APSI was developed to screen for these problems in adolescents with epilepsy and demonstrated good correlations for inter-rater reliability (median coefficient = 0.87) and acceptable internal consistency (median coefficient = 0.73). However, a rating of *inadequate* was given for the internal structure of the study, because the study failed to report the Cronbach's alpha for the APSI, which is a required statistic in COSMIN criteria. Furthermore, the study was rated as *doubtful* for its risk of bias due to limited description of the measure development methods. For example, it was unclear if the qualitative interviewers were trained for their work in measure development, and whether the recordings of the interviews were transcribed. Acceptable test–retest reliability was reported (median coefficient = 0.77); however, the methodology was rated as *inadequate* as the similarity of the testing conditions was not confirmed. Only 22 participants out of a possible 120 completed the questionnaire a second time (30 days after first administration). The initial administration of the inventory was at the first clinic visit or before a seizure-related hospitalisation, but the conditions of the second administration were not explicitly stated. However, the authors acknowledged that changes in medical management were likely, therefore potentially confounding the test–retest reliability.

The Symptom Questionnaire (SQ) is a 92-item self-rating scale that covers five mental health-related symptom subscales (anxiety, depression, anger hostility, somatisation, and well-being) [34]. One study evaluated the single depression item of the SQ as a screening candidate for adolescents with epilepsy [44]. The study was rated *very good* as it demonstrated good internal consistency across all 5 sub scales ($\alpha = 0.80 – 0.90$) [44]. However, as with the SCARED scale, a rating of *doubtful* was given for the risk of bias concerning responsiveness, due to minor methodological flaws in the study.

4. Discussion

This systematic review identified 13 PROMs that have been psychometrically evaluated for the assessment of mental health or neurodevelopmental outcomes in paediatric epilepsy and evaluated their measurement properties using the COSMIN checklist [20]. The COSMIN checklist is a standard developed to optimise and standardise the evaluation of PROM measurement properties [50]. Numerous reviews have utilised COSMIN methodology to evaluate PROM assessment across a variety of diseases such as atrial fibrillation [50] cystic fibrosis [52], multiple sclerosis [53], Parkinson's disease [53], and cancer [54]. COSMIN standards motivate improvements in the methodology of development and validation of PROMs and allow clinicians to have greater confidence in their use [50].

The PROMs assessed in this review were reflective by nature, meaning that they comprise items that reflect a single underlying construct. COSMIN criteria delineate between these and formative PROMs, which are composed of items that collectively define the construct [20]. Consequently, in formative PROMs, the correlation between individual items is not required, leading to COSMIN criteria stating that evaluating structural validity or internal consistency methodologies for these measures is deemed unnecessary [20]. By focusing on reflective PROMs, our review acknowledges the multifaceted nature of paediatric epilepsy, recognising that various symptoms often

intertwine and collectively influence the wellbeing of these young people. This approach to outcome measurement allows researchers and clinicians to grasp the holistic impact of these interconnected symptoms, portraying them not as isolated issues but as integral components within a broader construct of paediatric epilepsy experiences. To our knowledge, this is the first review to systematically evaluate the methodological quality of mental health and neurodevelopmental reflective PROMS in paediatric epilepsy patients, synthesising literature from numerous countries, whilst assessing the measures with rigorous COSMIN criteria.

One of the key findings was the lack of robust and comprehensive psychometric testing for any of the currently available instruments, particularly content validity, cross-cultural validity, and measurement error. Content validity is considered a critical measurement property, as it is important that a PROM measures what it intends to measure and that it is comprehensible to the target population [1]. None of the PROMs met standard criteria for all measurement properties with deficiencies found in responsiveness and criterion validity for some instruments. Notable weaknesses in test–retest reliability and PROM development were observed across measures, which raise questions regarding the use of these instruments in research and daily clinical practice, suggesting further evaluation is required.

Despite the lack of evidence for some measurement properties, the ADHD-RS-IV, NDDI-E-Y and PSC-17 have good evidence in favour of at least two measurement properties being rated *very good* for methodological quality. There is more evidence for the NDDI-E-Y questionnaire, with two studies evaluating its use in comparison to other PROMs. The NDDI-E-Y was assessed across five different measurement properties, demonstrating strong evidence for internal consistency, criterion validity, and convergent validity. However, it was rated *doubtful* for PROM development due to a lack of qualitative data. Furthermore, Wagner et al. [30] found an optimal cut-off of ≥ 32 for the detection of clinical depressive symptoms whereas Viellard et al. [41] found a lower cut-off of 23. The authors noted that if a cut-off of 32 were to be used in this study, almost a quarter of MDD cases would have gone undetected. Viellard et al. [41] opted to use the CDI to test the sensitivity of the NDDI-E-Y, whilst Wagner et al. [30] used the CDI-2, which uses a different cutoff score. This could explain why the large discrepancy in cutoff scores existed. Also, it is possible that sociocultural differences may have played a role in this discrepancy, although validation studies of the NDDI-E obtained similar results between French and American populations [55]. However, due to this discrepancy in cut-off scores between the two populations, the NDDI-E-Y cannot currently be considered an equivalent to the well-established adult NDDI-E, which is recommended for routine screening by The International League Against Epilepsy [56].

The ADHD-RS-IV demonstrated strong evidence for internal consistency and was one of the few measures to demonstrate evidence for responsiveness. The PSC-17 also had strong evidence for internal consistency, as well as structural validity. Furthermore, synthesis of results showed that the CPRS-48, SQ, SCARED, RCADS and CES-DC all met standard criteria for internal consistency and were rated methodologically *very good.* Neither cross-cultural validity, content validity nor measurement error were appropriately assessed in any of the studies. Lastly there were limited data for responsiveness, and thus there is limited information about important changes seen with these instruments.

Another key finding of this review is that no paediatric epilepsy validation studies were found for the Strengths and Difficulties Questionnaire (SDQ), which is often used in paediatric epilepsy populations (see [57,58] as examples). The SDQ is a screening tool that is used globally to assess both child and adolescent mental health problems [59]. The search results revealed a total of 41 studies which mentioned the SDQ in children with epilepsy; however, none of these were psychometric evaluation studies and therefore did not meet inclusion criteria for the review. Other authors have noted the limited literature investigating the SDQ's measurement equivalence invariance between treatment groups [59]. The imbalance of research on the included PROMs does not mean that they should be dismissed, as they may still be robust and useful, but require more extensive validation in paediatric epilepsy samples for clinicians to reliably detect at-risk children. Although one must be cautious when extrapolating data from other illnesses to epilepsy, previous studies have supported the use of some of these PROMs for other childhood illnesses, for example, paediatric oncology [60,61], diabetes [62], and asthma [63]. The use of these PROMs in other conditions may provide clinicians or researchers with reassurance that these are well-used and validated measures. Furthermore, for studies where quality was rated as *doubtful or inadequate,* this does not mean that the instrument was designed or carried out poorly, but rather means that the evidence was limited. For example, the APSI met most standard criteria for PROM development; however, it was not clear if the interviewers were trained sufficiently for aiding PROM development, nor was it clear if the recordings were transcribed and thus reduced the rating to *doubtful.*

We defined mental health as any psychological or psychiatric symptom such as anxiety, depression, and psychosis, as well as including any behavioural problems and neurodevelopmental disorder traits. Of the 13 questionnaires identified, only two were epilepsy-specific (NDDI-E-Y and APSI) with the remaining being generic questionnaires. Generic measures are useful to establish overall mental health in the population or to compare between diseases and general populations [50]. However, they may not be as sensitive to the effects of epilepsy on mental health as disease-specific questionnaires which contain domains that are more relevant to this condition, and thus may provide more focused information that can inform shared-decision making [50]. For example, the study by Carrozzino and colleagues [44] used the single item in the SCARED for panic attacks and found 24.1 % of patients with epilepsy had panic disorder compared to 0 % in the control group [44]. However, when the conventional SCARED subscale was used, 72 % of patients with epilepsy had panic disorder compared to 24 % of the control group, highlighting the problem of overdiagnosis [44]. Therefore, from a clinical point of view, it is important to improve detection of anxiety in epilepsy by differentiating anxiety symptoms from seizure symptoms [44]. Disease-specific measures are argued to have greater responsiveness to changes in a patient's condition and thus are more appropriate for measuring treatment outcomes within specific clinical populations. Therefore, generic and disease-specific PROMs could be used concurrently to obtain a more accurate assessment [64].

Many of the mental health PROMs identified in this review are parent-rated; however, when evaluating the mental health of children and young people with epilepsy, it is important that there is the opportunity for the child to rate their own mental health alongside their parent or caregiver [1]. Mental health is a subjective matter and there may be factors that parents or caregivers deem less important or are unaware about. Parents in non-clinical samples may not be particularly sensitive to mood symptoms and disorders reported by children themselves, thus potentially underestimating levels of mood and behavioural problems [65]. On the other hand, parents of children with epilepsy may be especially sensitive to mood and behaviour problems, leading to overestimated levels of mood and behavioural problems. Furthermore, parent-proxy reports may be influenced by factors other than the child themselves [65]. For example, a study with a sample of children with newly diagnosed epilepsy found that parental stress correlated with negative parent-proxy assessments regarding the child's quality of life [65]. A study by Eom and colleagues [66] compared case-control differences in behaviour using parent-proxy and self-reports and found substantially higher levels of behavioural problems for cases compared to controls [66]. However, when results were adjusted for measures of parental emotional impact, there was no longer a difference in scores between case and control groups [66]. Furthermore, in comparing self-report measures completed by adolescents or young adults, results showed that there was no evidence of higher behavioural burden in

epilepsy vs control groups [66]. Therefore, these results demonstrate how parent-proxy reports can be influenced by other factors which in turn distorts assessment of subjective outcomes [65]. This in turn influences the understanding of the association between childhood epilepsy and psychological-behavioural problems [65]. Therefore, whilst parent-proxy reports can provide an independent parent perspective of the child's mental health, they should ideally be used to complement the child self-report measure. Future studies aiming to validate mental health PROMs in paediatric epilepsy should also consider and assess parent proxy reliability.

The majority of studies identified here included both young children and adolescents. This may be an issue to consider as mental health in young children may present differently compared to mental health in adolescents [42]. Many of these studies reported total results and thus it is unknown if there is a difference in scores between lower and upper ends of the age range. The BASC and the CBCL were the only PROMs identified in this review with developmentally appropriate versions. Furthermore, the study by Wagner et al. [42] was the only study to examine the association between scale scores, chronological age of youth and age of seizure onset. Results showed that as age increased, internalising symptoms increased [42]. This is consistent with previous literature providing evidence that adolescents with epilepsy are more vulnerable to anxiety and depression compared to children with epilepsy [67]. This follows trends seen in the general population, for example, in England, 1 in 10 5–10-year-olds were estimated to have a mental health disorder, whilst it was 1 in 6 for 17–19-year-olds [68]. However, there may also be factors that are specific to people with epilepsy. Neurodevelopmental changes such as fluctuating hormones during puberty may increase seizure activity, in turn increasing vulnerability for adolescents [69]. Furthermore, studies have found that higher age of epilepsy onset is correlated with higher internalising symptoms [67]. Thus, psychosocial and neuropathophysiological stressors may contribute to the higher internalising symptoms seen in older youth with epilepsy [42]. Furthermore, there is also the possibility that the PSC-17 may be better as detecting internalising symptoms in older children than younger children [42]. Future research should consider using large sample designs to investigate the association between scale scores and chronological age of youth.

Finally, to improve the use of PROMs in clinical practice, it may be beneficial to obtain consensus amongst professionals, which can be aided through the development of a core outcome set (COS). These provide a list of standardised outcomes that, as a minimum, should be reported when conducting clinical research into a specific population [70]. One key COS developed for use in paediatric epilepsy research identified a total of 39 outcomes across 10 domains, including outcomes such as pain, self-esteem, and friendships, among others [71]. These can therefore help to inform PROM development, and to ensure that they capture the most meaningful information for those with paediatric epilepsy.

4.1. Limitations

A limitation of the current review was that some of the included instruments were either developed or psychometrically evaluated before the COSMIN checklist became broadly available and this may have had an impact on scores given for methodological quality. It was found that many of the studies did not report sufficient detail and as a result, using the COSMIN checklist to appraise the PROMs proved to be challenging at times. For example, the APSI [23] provided limited information regarding the development of the PROM, in turn leading to harsh ratings of methodological risk of bias. Therefore, it is important that PROM developers or any researchers aiming to validate PROMs consider the possible methodological risks of bias and report sufficient details regarding the evaluation of measurement properties. The COSMIN methodology has a notable advantage in identifying potential flaws in a measurement method, as it doesn't allow compensating for these flaws with higher scores in other areas [72]. However, a drawback of this approach is that even a single low rating can result in an overall poor rating for a measurement property, requiring all aspects to be rated as good or excellent to be considered as such [73]. Therefore, alternative methods could be explored, such as moving away from the "worst score counts" principle, to arrive at a more suitable and reliable scoring approach. Finally, it should be noted that the overall appraisal of each questionnaire in this review was limited by the availability of only one or two published articles for each PROM.

4.2. Future research directions

With the high prevalence of mental health symptoms and neurodevelopmental disorders in paediatric epilepsy, and the burden it causes for individuals with epilepsy, their families and caregivers, there is a clear need for routine mental health screening and assessment in this population. The assessment of methodological quality reported here suggested that the available questionnaires be treated with some caution and that more robust validation of frequently used measures is needed. Future studies may seek to address the responsiveness of PROMs regarding changes in patient symptoms before and after treatment, for example. Nonetheless, acceptable measurement properties are just one aspect that determines if a PROM is useful in research and clinical practice [50]. Other important considerations include the cost, time taken, ease of administration and patient acceptability [50]. Future research may also wish to address the broader spectrum of domains that may affect those with paediatric epilepsy. Topics beyond the scope of our inclusion criteria such as social determinants of health, or sleep disturbances, may further reflect the interconnectedness of symptoms affecting outcomes in paediatric epilepsy.

4.3. Conclusion

Only a small number of mental health and neurodevelopmental PROMs have been evaluated psychometrically in paediatric epilepsy, with few having sufficient evidence of robust measurement properties to justify recommending their routine use, according to COSMIN criteria. The existing tool demonstrating the most evidence to date was the NDDI-Y-E. Development and validation of new PROMs for measurement of mental health and neurodevelopmental outcomes in paediatric epilepsy therefore appears to be warranted. Nevertheless, whilst awaiting future PROMs developed specifically for this population, researchers and clinicians should not be dissuaded from using existing measures, if they have been well-validated in other paediatric populations [60–63].

Funding

This work was supported by the Maudsley Charity (AW, MPR, DKP) and the Medical Research Council (SP).

CRediT authorship contribution statement

Bianca De Aveiro: Writing – review & editing, Writing – original draft, Project administration, Methodology, Formal analysis, Data curation, Conceptualization. **Alice Winsor:** Writing – review & editing, Writing – original draft, Formal analysis, Data curation. **Jessica Davies:** Writing – review & editing, Formal analysis, Data curation. **Timothy R. Nicholson:** Writing – review & editing, Methodology, Funding acquisition. **Deb K. Pal:** Writing – review & editing, Methodology, Funding acquisition, Conceptualization. **Mark P. Richardson:** Writing – review & editing, Methodology, Funding acquisition, Conceptualization. **Susannah Pick:** Writing – review & editing, Writing – original draft, Supervision, Project administration, Methodology.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] Crudgington H, Collingwood A, Bray L, Lyle S, Martin R, Gringras P, et al. Mapping epilepsy-specific patient-reported outcome measures for children to a proposed core outcome set for childhood epilepsy. Epilepsy Behav 2020;112:107372. https://doi.org/10.1016/j.yebeh.2020.107372.

[2] Ngugi AK, Bottomley C, Kleinschmidt I, Sander JW, Newton CR. Estimation of the burden of active and life-time epilepsy: a meta-analytic approach. Epilepsia 2010; 51:883–90. https://doi.org/10.1111/j.1528-1167.2009.02481.x.

[3] Forsgren L, Beghi E, Öun A, Sillanpää M. The epidemiology of epilepsy in Europe – a systematic review. Eur J Neurol 2005;12:245–53.

[4] Russ SA, Larson K, Halfon N. A National Profile of Childhood Epilepsy and Seizure Disorder. Pediatrics 2012;129:256–64. https://doi.org/10.1542/peds.2010-1371.

[5] Beghi E, Giussani G, Nichols E, Abd-Allah F, Abdela J, Abdelalim A, et al. Global, regional, and national burden of epilepsy, 1990–2016: A systematic analysis for the global burden of disease study 2016. Lancet Neurol 2019;18:357–75. https://doi.org/10.1016/s1474-4422(18)30454-x.

[6] Kanner AM. Lennox-Lombroso Lecture, 2013: Psychiatric Comorbidities through the Life of the Seizure Disorder: A Complex Relation with a Not So Complex Solution. Epilepsy Currents 2014;14:323–8. https://doi.org/10.5698/1535-7597-14.6.323.

[7] Lee SY, Park J-H, Park S-J, Kim Y, Lee KY. Cognitive Function and Neuropsychological Comorbidities in Children with Newly Diagnosed Idiopathic Epilepsy. J Korean Med Sci 2018;33.

[8] Kanner AM, Barry JJ, Gilliam F, Hermann B, Meador KJ. Depressive and anxiety disorders in epilepsy: Do they differ in their potential to worsen common antiepileptic drug–related adverse events? Epilepsia 2012;53:1104–8.

[9] Kanner AM. Management of psychiatric and neurological comorbidities in epilepsy. Nat Rev Neurol 2016;12:106–16. https://doi.org/10.1038/nrneurol.2015.243.

[10] Russo A, Posar A, Conti S, Parmeggiani A. Prognostic factors of drug-resistant epilepsy in childhood: An Italian study. Pediatrics Int 2015;57:1143–8. https://doi.org/10.1111/ped.12705.

[11] Patel RS, Elmaadawi A, Mansuri Z, Kaur M, Shah K, Nasr S. Psychiatric Comorbidities and Outcomes in Epilepsy Patients: An Insight from a Nationwide Inpatient Analysis in the United States. Cureus 2017.

[12] Tellez-Zenteno JF, Patten SB, Jetté N, Williams J, Wiebe S. Psychiatric Comorbidity in Epilepsy: A Population-Based Analysis. Epilepsia 2007;48:2336–44.

[13] Sadeghi S, Fayed N, Ronen GM. Patient-reported outcome measures in pediatric epilepsy: A content analysis using World Health Organization definitions. Epilepsia 2014;55:1431–7. https://doi.org/10.1111/epi.12740.

[14] Black N. Patient reported outcome measures could help transform healthcare. BMJ 2013;346:f167–.

[15] Ghimire P, Hasegawa H, Kalyal N, Hurwitz V, Ashkan K. Patient-Reported Outcome Measures in Neurosurgery: A Review of the Current Literature. Neurosurgery 2018;83:622–30.

[16] Guilfoyle SM, Wagner JL, Smith G, Modi AC. Early screening and identification of psychological comorbidities in pediatric epilepsy is necessary. Epilepsy Behav 2012;25:495–500.

[17] Corrigan FM, Broome H, Dorris L. A systematic review of psychosocial interventions for children and young people with epilepsy. Epilepsy Behav 2016; 56:99–112.

[18] Powell J, Powell S, Robson A. A systematic review of patient-reported outcome measures in paediatric otolaryngology. J Laryngol Otol 2018;132:2–7.

[19] Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technology Assessment (Winchester, England) 1998;2:i–iv, 1–74.

[20] Mokkink LB, Boers M, van der Vleuten CPM, Bouter LM, Alonso J, Patrick DL, et al. COSMIN Risk of Bias tool to assess the quality of studies on reliability or measurement error of outcome measurement instruments: a Delphi study. BMC Med Res Method 2020;20:293. https://doi.org/10.1186/s12874-020-01179-5.

[21] Aman MG, Singh NN, Stewart AW, Field CJ. The aberrant behavior checklist: a behavior rating scale for the assessment of treatment effects. Am J Ment Defic 1985;89:485–91.

[22] DuPaul GJ, Power TJ, Anastopoulos AD, Reid R. ADHD Rating Scale—IV: Checklists, norms, and clinical interpretation. The Guilford Press; 1998.

[23] Batzel LW, Dodrill CB, Dubinsky BL, Ziegler RG, Connolly JE, Fréeman RD, et al. An Objective Method for the Assessment of Psychosocial Problems in Adolescents with Epilepsy. Epilepsia 1991;32:202–11. https://doi.org/10.1111/j.1528-1157.1991.tb05245.x.

[24] Reynolds CR. Behavior Assessment System for Children. The Corsini Encyclopedia of Psychology. Wiley 2010:1–2. https://doi.org/10.1002/9780470479216.corpsy0114.

[25] Weissman MM, Orvaschel H, Padian N. Children's Symptom and Social Functioning Self-Report Scales Comparison of Mothers' and Children's Reports. J Nerv Ment Dis 1980;168:736–40. https://doi.org/10.1097/00005053-198012000-00005.

[26] Achenbach TM, Rescorla L. Manual for the ASEBA School-Age Forms and Profiles. ASEBA 2001.

[27] Kovacs M. The Children's Depression, Inventory (CDI). Psychopharmacol Bull 1985;21:995–8.

[28] Conners CK. Conners' Rating Scale Manual. New York: Multi-Health Systems; 1989.

[29] Wagner JL, Smith G, Ferguson PL, Fedele DA. Preliminary Psychometrics of the Neurological Disorders Depression Inventory for Epilepsy-Youth. J Child Neurol 2013;28:1392–9. https://doi.org/10.1177/0883073813483367.

[30] Wagner JL, Kellermann T, Mueller M, Smith G, Brooks B, Arnett A, et al. Development and validation of the NDDI-E-Y: a screening tool for depressive symptoms in pediatric epilepsy. Epilepsia 2016;57:1265–70. https://doi.org/10.1177/0883073813483367.

[31] Gardner W, Murphy MB, Childs GE, Kelleher KJ, Pagano M, Jellinek MS, et al. The PSC-17: A brief pediatric symptom checklist with psychosocial problem subscales. A report from PROS and ASPN. Ambulatory Child Health 1999;5:225–36.

[32] Chorpita BF, Yim L, Moffitt C, Umemoto LA, Francis SE. Assessment of symptoms of DSM-IV anxiety and depression in children: a revised child anxiety and depression scale. Behav Res Ther 2000;38:835–55. https://doi.org/10.1016/S0005-7967(99)00130-8.

[33] Birmaher B, Khetarpal S, Brent D, Cully M, Balach L, Kaufman J, et al. The Screen for Child Anxiety Related Emotional Disorders (SCARED): Scale Construction and Psychometric Characteristics. J Am Acad Child Adolesc Psychiatry 1997;36: 545–53. https://doi.org/10.1097/00004583-199704000-00018.

[34] Kellner R. A symptom questionnaire. J Clin Psychiatry 1987;48:268–74.

[35] Kaat AJ, Zelko F, Wilkening G, Berg AT. Evaluation of the Aberrant Behavior Checklist for Developmental and Epileptic Encephalopathies. Epilepsy Behav 2021; 119:107958. https://doi.org/10.1016/j.yebeh.2021.107958.

[36] Mercier C, Roche S, Gaillard S, Kassai B, Arzimanoglou A, Herbillon V, et al. Partial validation of a French version of the ADHD-rating scale IV on a French population of children with ADHD and epilepsy. Factorial structure, reliability, and responsiveness. Epilepsy Behav 2016;58:1–6. https://doi.org/10.1016/j.yebeh.2016.02.016.

[37] Bender HA, Auciello D, Morrison CE, MacAllister WS, Zaroff CM. Comparing the convergent validity and clinical utility of the Behavior Assessment System for Children-Parent Rating Scales and Child Behavior Checklist in children with epilepsy. Epilepsy Behav 2008;13:237–42. https://doi.org/10.1016/j.yebeh.2008.03.007.

[38] Kiyumi HA, Al-Huseini S, Mirza H, Balushi NA, Al-Risi K, Toubi AA, et al. Depressive Symptoms and Its Correlate Among Children with Epilepsy at Single-center Study in Oman. Oman Med J 2021;36:e329–. https://doi.org/10.5001/omj.2021.112.

[39] Miniksar DY, Kılıç B, Kılıç M, Miniksar ÖH, Topçu Y, Aydın K. Evaluation of suicide probability in children and adolescents with epilepsy. Pediatr Int 2022;64. https://doi.org/10.1111/ped.15130.

[40] Pal DK, Chaudhury G, Das T, Sengupta S. Validation of a Bengali adaptation of the Conners' Parent Rating Scale (CPRS-48). Br J Med Psychol 1999;72:525–33. https://doi.org/10.1348/000711299160211.

[41] Viellard M, Villeneuve N, Milh M, Lépine A, Micoulaud-Franchi J-A, McGonigal A. Screening for depression in youth with epilepsy: Psychometric analysis of NDDI-E-Y and NDDI-E in a French population. Epilepsy & Behavior: E&B 2019;98:19–26. https://doi.org/10.1016/j.yebeh.2019.06.012.

[42] Wagner JL, Guilfoyle SM, Rausch J, Modi AC. Psychometric validation of the Pediatric Symptom Checklist-17 in a pediatric population with epilepsy: A methods study. Epilepsy Behav 2015;51:112–6. https://doi.org/10.1016/j.yebeh.2015.06.027.

[43] Rogač Ž, Stevanović D, Bečanović S, Božić L, Dimitrijević A, Bogićević D, et al. Cognitive profile, psychopathological symptoms, and quality of life in newly diagnosed pediatric epilepsy: A six-month, naturalistic follow-up study. Epilepsy Res 2022;179:106844. https://doi.org/10.1016/j.eplepsyres.2021.106844.

[44] Carrozzino D, Marchetti D, Laino D, Minna M, Verrocchio MC, Fulcheri M, et al. Anxiety in adolescent epilepsy. A clinimetric analysis. Nord J Psychiatry 2016;70: 424–9. https://doi.org/10.3109/08039488.2016.1143029.

[45] Faries DE, Yalcin I, Harder D, Heiligenstein JH. Validation of the ADHD Rating Scale as a clirlician administered and scored instrument. J Atten Disord 2001;5: 107–15. https://doi.org/10.1177/108705470100500204.

[46] Zhang S, Faries DE, Vowles M, Michelson D. ADHD rating scale IV: psychometric properties from a multinational study as clinician-administered instrument. Int J Methods Psychiatr Res 2005;14:186–201. https://doi.org/10.1002/mpr.7.

[47] Perrin EC, Stein REK, Drotar D. Cautions in Using the Child Behavior Checklist: Observations Based on Research About Children with a Chronic Illness. J Pediatr Psychol 1991;16:411–21.

[48] Jellinek MS, Murphy JM, Burns BJ. Brief psychosocial screening in outpatient pediatric practice. J Pediatr 1986;109:371–8. https://doi.org/10.1016/S0022-3476(86)80408-5.

[49] Stoppelbein L, Greening L, Jordan SS, Elkin TD, Moll G, Pullen J. Factor Analysis of the Pediatric Symptom Checklist with a Chronically Ill Pediatric Population. J Dev Behav Pediatr 2005;26:349–55. https://doi.org/10.1097/00004703-200510000-00002.

[50] Kotecha D, Ahmed A, Calvert M, Lencioni M, Terwee CB, Lane DA. Patient-Reported Outcomes for Quality of Life Assessment in Atrial Fibrillation: A Systematic Review of Measurement Properties. PLoS One 2016;11:e0165790.

[51] Kovacs M. Children's depression inventory (cdi2) : technical manual (2nd ed.). Multi-Health Systems; 2011.

[52] Ratnayake I, Ahern S, Ruseckaite R. A systematic review of patient-reported outcome measures (PROMs) in cystic fibrosis. BMJ Open 2020;10:e033867.

[53] Elbers RG, Rietberg MB, van Wegen EEH, Verhoef J, Kramer SF, Terwee CB, et al. Self-report fatigue questionnaires in multiple sclerosis, Parkinson's disease and

stroke: a systematic review of measurement properties. Qual Life Res 2012;21:925–44.
[54] Wheelwright S, Darlington A-S, Hopkinson JB, Fitzsimmons D, White A, Johnson CD. A systematic review of health-related quality of life instruments in patients with cancer cachexia. Support Care Cancer 2013;21:2625–36.
[55] Micoulaud-Franchi J-A, Barkate G, Fonseca A-T-D, Vaugier L, Gavaret M, Bartolomei F, et al. One step closer to a global tool for rapid screening of major depression in epilepsy: Validation of the French NDDI-E. Epilepsy Behav 2015;44:11–6.
[56] Kerr MP, Mensah S, Besag F, de Toffol B, Ettinger A, Kanemoto K, et al. International consensus clinical practice statements for the treatment of neuropsychiatric conditions associated with epilepsy. Epilepsia 2011;52:2133–8.
[57] Reilly C, Atkinson P, Das KB, Chin RF, Aylett SE, Burch V, et al. Screening for mental health disorders in active childhood epilepsy: Population-based data. Epilepsy Res 2014;108:1917–26.
[58] Turky A, Beavis JM, Thapar AK, Kerr MP. Psychopathology in children and adolescents with epilepsy: An investigation of predictive variables. Epilepsy Behav 2008;12:136–44.
[59] Hall CL, Guo B, Valentine AZ, Groom MJ, Daley D, Sayal K, et al. The validity of the Strengths and Difficulties Questionnaire (SDQ) for children with ADHD symptoms. PLoS One 2019;14:e0218518.
[60] Myers RM, Balsamo L, Lu X, Devidas M, Hunger SP, Carroll WL, et al. A prospective study of anxiety, depression, and behavioral changes in the first year after a diagnosis of childhood acute lymphoblastic leukemia. Cancer 2014;120:1417–25.
[61] Wolfe-Christensen C, Mullins LL, Stinnett TA, Carpentier MY, Fedele DA. Use of the Behavioral Assessment System for Children 2nd Edition: Parent Report Scale in Pediatric Cancer Populations. J Clin Psychol Med Settings 2009;16:322–30. https://doi.org/10.1007/s10880-009-9174-7.
[62] Holmes CS, Respess D, Greer T, Frentz J. Behavior Problems in Children With Diabetes: Disentangling Possible Scoring Confounds on the Child Behavior Checklist. J Pediatr Psychol 1998;23:179–85. https://doi.org/10.1093/jpepsy/23.3.179.
[63] Goldbeck L, Koffmane K, Lecheler J, Thiessen K, Fegert JM. Disease severity, mental health, and quality of life of children and adolescents with asthma. Pediatr Pulmonol 2007;42:15–22.
[64] Churruca K, Pomare C, Ellis LA, Long JC, Henderson SB, Murphy LED, et al. Patient-reported outcome measures (PROMs): A review of generic and condition-specific measures and a discussion of trends and issues. Health Expect 2021;24:1015–24. https://doi.org/10.1111/hex.13254.
[65] Berg AT, Altalib HH, Devinsky O. Psychiatric and behavioral comorbidities in epilepsy: A critical reappraisal. Epilepsia 2017;58:1123–30. https://doi.org/10.1111/epi.13766.
[66] Eom S, Caplan R, Berg AT. Behavioral Problems and Childhood Epilepsy: Parent vs Child Perspectives. J Pediatr 2016;179:233–239.e5. https://doi.org/10.1016/j.jpeds.2016.08.096.
[67] Wagner JL, Wilson DA, Smith G, Malek A, Selassie AW. Neurodevelopmental and mental health comorbidities in children and adolescents with epilepsy and migraine: a response to identified research gaps. Dev Med Child Neurol 2015;57:45–52. https://doi.org/10.1111/dmcn.12555.
[68] Sadler K, Vizard T, Ford T, Marcheselli F, Pearce N, Mandalia D, et al. Mental Health of Children and Young People in England. NHS Digital 2017.
[69] Baram TZ. The Brain, Seizures and Epilepsy Throughout Life: Understanding a Moving Target. Epilepsy Currents 2012;12:7–12. https://doi.org/10.5698/1535-7511-12.4s.7.
[70] Williamson PR, Altman DG, Blazeby JM, et al. Developing core outcome sets for clinical trials: issues to consider. Trials 2012;13:132. https://doi.org/10.1186/1745-6215-13-132.
[71] Crudgington H, Rogers M, Bray L, Carter B, Currier J, Dunkley C, et al. Core Health Outcomes in Childhood Epilepsy (CHOICE): Development of a core outcome set using systematic review methods and a Delphi survey consensus. Epilepsia 2019;60(5):857–71. https://doi.org/10.1111/epi.14735.
[72] Terwee CB, Mokkink LB, Knol DL, Ostelo RWJG, Bouter LM, de Vet HCW. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res 2012;21:651–7. https://doi.org/10.1007/s11136-011-9960-1.
[73] Denman D, Speyer R, Munro N, Pearce WM, Chen Y-W, Cordier R. Psychometric Properties of Language Assessments for Children Aged 4–12 Years: A Systematic Review. Front Psychol 2017;8. https://doi.org/10.3389/fpsyg.2017.01515.