Epilepsy & Behavior 142 (2023) 109154

Contents lists available at ScienceDirect

Epilepsy & Behavior

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

Review

Seizure severity assessment tools for adult epilepsy patients: A systematic review

Rose Aniza Rusli [a,b], Mohd Makmor Bakry [a,*], Noraida Mohamed Shah [a], Tan Hui Jan [c]

[a] Faculty of Pharmacy, Universiti Kebangsaan Malaysia, Jalan Raja Muda Abdul Aziz, 50300 Kuala Lumpur, Malaysia
[b] Hospital Shah Alam, Persiaran Kayangan, 40000 Shah Alam, Selangor, Malaysia
[c] Neurology Unit, Department of Medicine, Universiti Kebangsaan Malaysia Medical Centre, 56000 Kuala Lumpur, Malaysia

ARTICLE INFO

Article history:
Received 21 November 2022
Revised 7 February 2023
Accepted 19 February 2023
Available online 29 March 2023

Keywords:
Epilepsy
Antiseizure medication therapy
Seizure severity
Assessment tool

ABSTRACT

Introduction: Seizure outcomes from antiseizure medication (ASM) therapy can be measured across various domains using assessment tools. The available tools may contain an array of different components or items. Seizure severity assessment, as opposed to seizure frequency count may have been a more accurate measurement in determining the effectiveness of ASM therapy. This study aimed to review studies developing seizure severity assessment tools for adults with epilepsy, describe the development methods and validation, and compare the list of items in these tools.

Methods: The systematic search utilized established databases such as Scopus, Ovid, Web of Science, Medline, Wiley Online, and Cochrane Library. Studies published from inception to December 15, 2022, were selected. Publications describing the development of tools to measure seizure severity among adult epilepsy patients were included. Outcome measures including the tool's content, development methods, validity, and reliability assessments were compared.

Results: The search produced eight publications describing the development of eight seizure severity assessment tools. One of these tools is part of a multidimensional assessment of the overall impact of epilepsy. The frequently used method in the initial development was the qualitative method (n = 6) where two publications reanalyzed the items from previous studies. Face validity was the most common validation test conducted (n = 4). At least one reliability assessment was conducted for each of the tools, most commonly by the test–retest method (n = 6) and inter-rater reliability (n = 5). All of these tools cover the components of pre-ictal (warning/aura), ictal, and postictal (recovery) events.

Conclusion: The identified tools described the assessment of seizure severity using various subscales. The emergence of new methods in quantifying seizure severity unfolds opportunities in discovering more comprehensive assessments of seizure severity in both clinical trials and daily clinical practice.

1. Introduction

Seizure frequency and severity are the two most important outcome variables in evaluating epilepsy treatment with antiseizure medication (ASM) [1]. The reported efficacy of ASM is based on the number of patients who attained more than a 50% reduction in seizure frequency [2]. However, this traditional method led to questions on whether seizure frequency alone is sufficient as an indicator of ASM efficacy. The simple measure of documenting seizure frequency may not accurately determine the overall epilepsy control. It did not take into consideration any changes in the seizure type, for instance, the change from primarily focal impaired awareness to focal aware seizures, the elimination of one or more ictal events, the shortening of the seizure duration, and others [3]. Researchers observed that ASMs act by preventing the progression of focal aware seizures to bilateral tonic-clonic seizures. Thus, according to Baker et al., ASM may affect seizure severity by means of altering seizure types without necessarily reducing the frequency of seizures [4]. Undoubtedly, this warrants a comprehensive method of assessing and measuring the efficacy of drug therapy toward improving seizure outcomes.

Assessment tools or scales are valuable instruments for assigning numerical ratings to events that cannot be measured directly. In this context, the event is seizure severity. The tools or scales comprise groups of items displaying ranks of theoretical variables, which are not evaluated by direct means. The development of the scales requires laborious and systematic research protocols. When justify-

* Corresponding author.
E-mail address: mohdclinpharm@ukm.edu.my (M. Makmor Bakry).

https://doi.org/10.1016/j.yebeh.2023.109154
1525-5050/

ing the development of a new scale, investigators should provide a summary of their existing measurement from the literature review, cite any existing instruments that are similar, and highlight the gap in the existing measurement from the literature by expressing the constructs that existing measures fail to gauge [5–7].

The essence of an outcome assessment in clinical trials and healthcare settings is the patient-reported outcome measure (PROM) [8]. This survey-based assessment method is crucial as it has been associated with greater patient satisfaction and improvement in patient-clinician communication [9]. While it is not without limitation, capturing patients' and carers' own perspectives of their disease and symptoms is fundamental. Limitations of survey-based assessment include weaknesses in the development of the tools, validity, and reliability issues, ambiguous score interpretation, respondents' interpretation and recall bias, and inability to capture objective measures of the disease [10]. In the context of characterizing seizure severity, the three main factors that may be reported by patients and carers are seizure semiology, seizure frequency, and total duration of the seizure event [11]. Details of the ictal phase such as loss of awareness, presence of motor symptoms, disruptive automatisms, falls, or seizure-related injuries are all important elements in seizure assessment.

Previous reviews related to the measurement of seizure severity have highlighted some distinctive features of the existing seizure severity assessment tools, which were used mainly in drug trials [1,12,13]. These reviews included different depths of discussion on the aspect of the initial development method, scale content, and appropriateness of the scoring system. The obvious shortfall of the earlier tools as reported by Cramer and colleagues was the inability to capture changes in treatment effect ("responsiveness" element of the tool). This aspect is important as consistency and reliability tests alone are inadequate in determining the treatment effect. It has also been observed that the content of these tools may or may not cover all the severity components for an individual as the number of items is limited. This is further hampered by the inconsistent and arbitrary scoring system with no clinical interpretation for the final score [1].

There was a notable shift in the seizure severity tools' applicability. The functionality of the tools was not only observed in a controlled environment (drug trials) but also anticipated in a more practical, real-life clinical setting. Consequently, methods of detecting the minimal clinically important difference (MCID) were instituted and this could help clinicians optimize drug therapy based on the slightest yet significant change in the patient's disease status. Additionally, Cramer and French suggested including the anatomic-clinical connections in which electrical events of the brain through electroencephalogram (EEG) reports are integrated with physical events of seizures for severity assessment [1]. On that note, a recent study [10] has demonstrated the association of ictal intracranial EEG (iEEG) with seizure severity. Another possible approach for practical application is to incorporate psychological symptoms, occupational and social functioning in a wholesome psychosocial rating [14]. All of these, and probably many more other dimensions, should be appraised in the pursuit of establishing a gold standard assessment tool for seizure severity.

The current systematic review aimed to identify and compare the available seizure severity assessment tools for adult epilepsy patients in various settings. The comparison was carried out based on original articles published by the developers.

2. Material and methods

2.1. Search strategy and literature search

Articles were searched through the following databases; Scopus, Medline, Wiley Online Library, Web of Science (WOS), Ovid, Cochrane Library, and Google Scholar from the date of inception through December 15, 2022. The Medical Subject Heading (MeSH) and terms that were used to search for relevant publications include "seizure severity," "seizure control," "seizure frequency," "seizure outcomes," "assess*," "measure*," "scale," "score," and "questionnaire." Boolean operators, such as OR and AND, were utilized to narrow the search.

Various types of study designs were explored including clinical trials and observational studies. No limitations were applied with regard to the year of publication. Only articles published in the English language were included. Abstracts and conference proceedings, commentaries, editorials, and columnist and expert opinion articles were excluded.

2.2. Data extraction and analysis

The country of origin, study aim and objectives, study design, development methods, research settings in which the tool was developed, the primary user of the tool, and information on the validity and reliability assessment were analyzed. Information on the application of the tools and their clinical use was also extracted. All the criteria and characteristics of each tool were compared and contrasted from one to another within similar domains.

2.3. Quality assessment

The COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) Risk of Bias (ROB) checklist is employed to evaluate the methodological quality of the measurement tools [15]. The COSMIN ROB checklist comprises 10 criteria, which include PROM development, content validity, structural validity, internal consistency, cross-cultural validity/measurement invariance, reliability, measurement error, criterion validity, hypotheses testing for construct validity, and responsiveness. The checklist was rated as "very good," "adequate," "doubtful," and "inadequate." "The worst score counts" principle was used to determine the overall quality of a study. The response option "NA" (not applicable) was applied for items that were irrelevant to the study. The "NA" rating is not considered the "worst score counts" [16]. It should be aware that COSMIN Checklist is useful mainly as a guide for new tools development, therefore the quality assessment was done only for comparison purposes between the tools.

3. Results

A total of 537 possibly relevant publications were identified in the initial search across all databases. These were evaluated primarily in accordance with the predetermined inclusion and exclusion criteria (Table 1). The PRISMA 2020 statement [17] was used to guide in narrowing the search for eligible publications. A review of the title suitability and the removal of duplicates excluded 390 articles. The remaining 147 publications were individually reassessed for relevance. Studies investigating other epilepsy domains, utilizing other interventions and assessment methods, with populations other than adults as well as animal studies were further excluded (n = 111). From this screening step, 36 articles were found to be relevant. Following re-evaluation and assessment of the abstracts, 24 of these were excluded. Some of these articles may have been excluded on more than one criterion. The details of the excluded articles are described in Fig. 1. References from these remaining articles were screened and two potential studies were found, but later removed due to both studies only adapting and utilizing the existing tools. This reinstated the 12 identified articles

Table 1
Inclusion and exclusion criteria for study selection.

Inclusion criteria	Exclusion criteria
• Original primary studies • English language • Population of study – Adult epilepsy patients, prescribed with one or more ASMs • Developing, modifying, and validating the tools/scale/questionnaire in assessing seizure severity • The assessment of seizure severity should adopt patient/physician reported outcome measures (PROM) method	• Studies that are only utilizing and adapting the tools in clinical trial, other research, or healthcare practice • Investigating exclusively factors, predictors, and the effect of an intervention/disease/condition on seizure severity • The assessment is meant for specific patient conditions – e.g. pregnancy, intellectual disability, behavioral problem, etc • Non-pharmacological intervention instituted in the development process

as the original papers that developed or revised the seizure severity assessment tools. Two of the articles were excluded as these articles only applied, translated, or adapted the existing assessment tools and there were no changes to any part of the tools as reported in the original articles. Further appraisal of the articles found another two of the publications were follow-up studies that result in updated versions of the scale without changing the essential components of the items [18,19]. A pediatric study on the development of the scale was excluded. One study that has a seizure severity component in its tool's content was included during revision. This yielded eight publications for this review. The graphical description of the article search and screening is presented as a PRISMA [17] flowchart (Fig. 1).

3.1. Study background

This systematic review identified eight publications describing eight seizure severity assessment tools for quantitative measurement. Three tools were the updated versions of the two primary ones; the Chalfont Seizure Severity Scale (CSSS) and the first version of the Liverpool Seizure Severity Scale (Original LSSS). The tools were utilized for adult patients ranging from 15 to 80 years of age. The Veteran Administration Scale (VA) was the earliest developed tool which was published in 1983 [20]. It was developed with the initial purpose of measuring and comparing the outcome of ASM therapy in clinical trials [21,22]. The developers of VA have pioneered the method of quantifying seizure severity for the evaluation of drug therapy effectiveness which was then conducted by mere measurement of seizure frequency prior to the publication.

The second tool known as CSSS was published by researchers from the United Kingdom [23]. Later, the tool was further refined, simplified, and renamed the National Hospital Seizure Severity Scale (NHS3) [24]. The next published assessment tool is a patient-based tool called LSSS which was developed by Gus A. Baker et al. [4]. This tool was the foundation for two other tools that were developed later [25,26]. It has been translated into at least 21 languages from all over the world. In this current review, however, the focus was on the original seizure severity assessment tools and their updated versions. In 2002, the principal developers of VA and LSSS collaborated to develop a new seizure severity assessment tool named the Seizure Severity Questionnaire (SSQ) to further improve and increase the sensitivity of the tool [3]. The final tool, which is known as the Personal Impact of Epilepsy Scale – PIES, incorporates the element of seizure severity in its scale with an adverse drug reaction, comorbidities, and quality of life of epilepsy patients [27]. However, since the focus of this review is exclusively on seizure severity assessment, hence, only the seizure domain will be compared and discussed in this study.

For the study setting, five of these tools were developed in specialized epilepsy or neurology clinic (CSSS, NHS3, Revised LSSS, SSQ, and PIES). One was taken from the analysis of a randomized controlled trial (LSSS 2.0) that utilized the original version (Original LSSS) whereas the setting of the development of the latter was not reported. Subjects were mainly recruited from patients followed up under these clinics (Table 2).

3.2. Study quality

Generally, there are mixed rates with regard to the COSMIN checklist [15] for each domain for all of the tools. Most of the tools were rated as at least "adequate" for the reliability and hypothesis testing checklist and "inadequate" for the responsiveness checklist. The details of the rating for each study are described in Table 3.

3.3. Methods of tool development

The qualitative research method has been applied in the initial development of the identified tools. The two main approaches of the qualitative method were interviews with target individuals and a reanalysis of previous studies. The interviews were conducted with either patients or relatives and were reviewed by expert panels. Three of the studies clearly mentioned the involvement of expert panels who were directly involved in the clinical practice to devise the tool (Original LSSS, LSSS 2.0, SSQ). Five studies (CSSS, NHS3, Revised LSSS, LSSS 2.0, PIES) involved patients and relatives/observers. The LSSS 2.0 included a patient-focus group discussion to clarify the components of the tool.

Variables obtained from the epilepsy patients' interviews were utilized by CSSS and LSSS 2.0. In the interview, the perception of patients on the severity of their seizures was documented and analyzed. These variables were further synthesized to form the items in the scales. In contrast to the development of the Original LSSS, neurologists with a special interest in epilepsy were recruited to devise a scale based on their clinical experience. The Revised LSSS was also the result of the modification and further validity assessment of the existing scale (Original LSSS). The main difference between the Revised and Original LSSS was the elimination of four items which was considered redundant in assessing symptom descriptions. The NHS3, on the other hand, was developed as a result of qualitative observations of CSSS in clinical and ASM trials.

The VA was the product of pre-existing data reanalysis. The investigators reviewed all clinical trials on ASM therapy effectiveness, critically analyzed the variety of methods utilized, and listed all the relevant variables. A mechanism for calculating a single numerical composite rating that accounts for all of the variables was developed. The tool assigned absolute numerical values to specific problem areas. However, for certain items, the sensitivity thresholds can be adjusted. As for SSQ, it was first developed by a thorough review of the existing seizure severity assessment tools. The investigators aligned the flow of seizure activity and composed it to become a series of questionnaires. Each of the items in the scales was weighted and underwent further sensitivity evaluation for the seizure severity. Finally, the most current tool, PIES, correlated the items in its preliminary questions (the draft PIES items included questions from SSQ) with the comparator instrument, NHS3, resulting in a total score of $R = 0.2886$ ($n = 35$, $p = 0.05$).

The two most recent studies (SSQ and PIES) have instituted MCID analysis in their follow-up studies [19,28]. Both studies calculated the minimally important change (MIC) using the standard anchor-based method. The pronounced difference between these two studies was the study setting; SSQ conducted the analysis during a phase 3 lacosamide trial [29] whereas PIES conducted the analysis in a Specialized Epilepsy Center. The

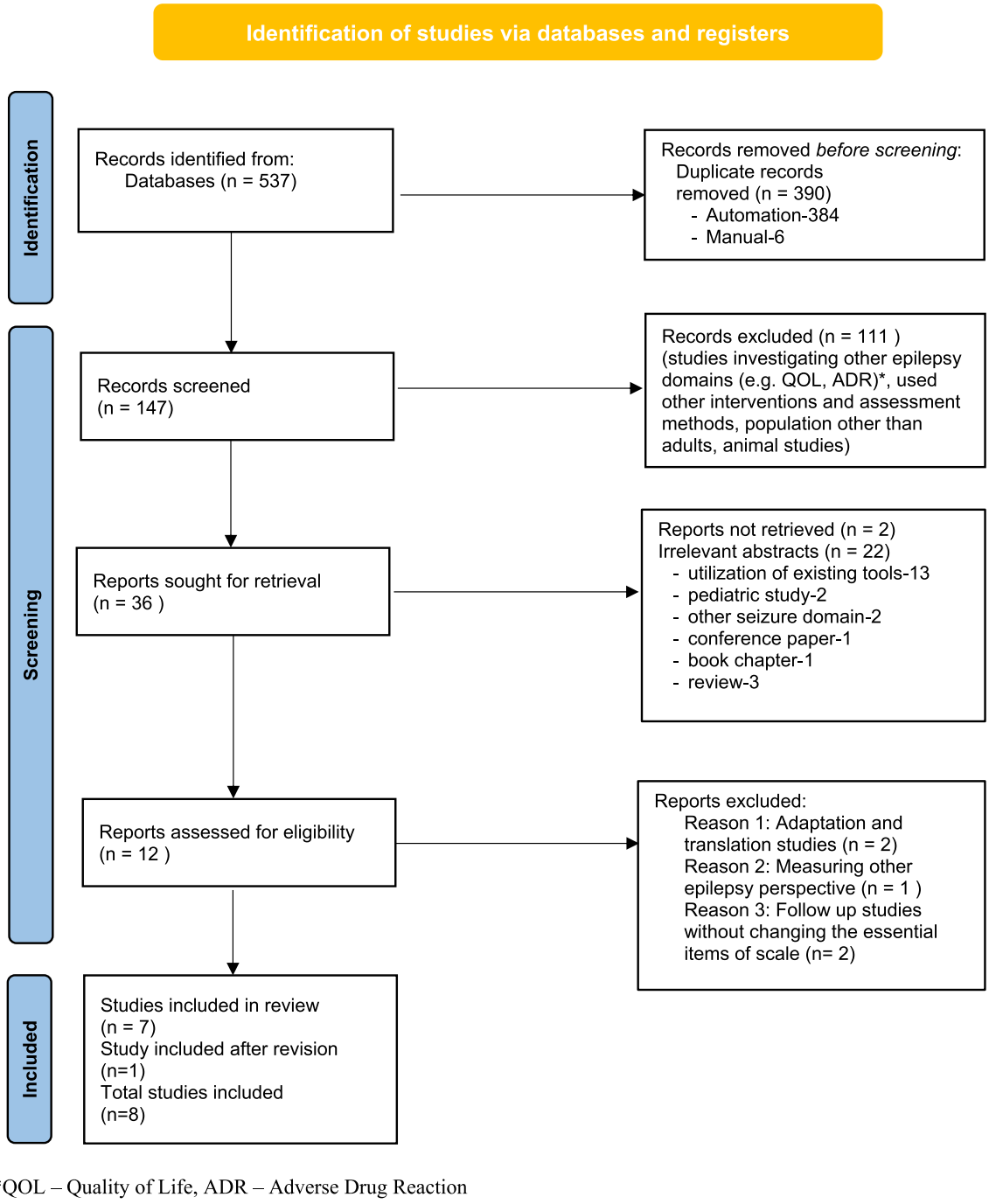

Fig. 1. PRISMA 2020 flow diagram for the systematic review, *QOL – Quality of Life, ADR – Adverse Drug Reaction.

SSQ utilized the Patient Global Impression of Change (PGIC) score with the categories of "much improved," "minimally improved," "much worsened," and "minimally worsened" being applied [19]. PIES compared the mean of change in anchor questions, dubbed as "improved," "same," and "worse," between two groups (clinic patients and blinded raters). A 0.48-point change in the total score (MIC threshold range: 0.34 to 0.50) represents a clinically meaningful change in seizure severity from the patient's perspective as measured using the SSQ. For PIES, the 8% MCID reduction that was achieved from the study demonstrated a meaningful improvement in patient function on any of the PIES subscales or the total score.

3.4. Validity and reliability tests

Various validity tests have been used to develop the tools such as the face, content, construct, and criterion validity tests. All of the tools included in this study have been tested with at least one of the established validity methods (Table 4). The exception was the VA, whereby the validity assessment was not clearly reported and only addressed as inter-rater validation. The CSSS and SSQ have undergone face and content validities. The investigators of CSSS recruited patients, their relatives, and panels of medical and nursing professionals who are involved in the management of epilepsy to complete the scales. Good concurrence was demonstrated for

Table 2
Characteristics of included studies.

Author & year of publication (country or origin) (Ref. no)	Aim	Name of Tool	Initial Scale Development Method (brief steps in development)	Study subjects (total, n)	Study Setting	Finalized tools appended in the publication? explicit/implicit
Cramer et al., 1983(USA) [20]	Developing a scale to assess frequency and severity of seizures for evaluation of antiepileptic therapy	Veteran Administration (VA)	Criteria formulated by investigators based on clinical experience and pilot study	>800 patients participated in VA Cooperative Study	Clinical trials of ASM treatment	Yes / Explicit
Duncan & Sander 1991(United Kingdom) [23]	Developing a seizure severity scale for AED evaluation and routine clinical practice	The Chalfont Seizure Severity Scale (CSSS)	– Open Interviews with patients and relatives to identify factors – Each factor was assigned with weighted scale – Piloted and tested to targeted study	Inter-rater reliability n = 93 Test-retest reliability n = 101	– Specialised Epilepsy Clinic – Inpatient Assessment Unit	Yes / Explicit
O'Donoghue et al., 1996 (United Kingdom) [24]	Refining and simplifying the Chalfont Seizure Severity Scale	The National Hospital Seizure Severity Scale (NHS3)	– Qualitative, observational (Interview, Visual Analogue Scale) – Construct Validity	Inter-observer reliability n = 87 Test-retest reliability n = 18 Validity n = 80	Routine Epilepsy Clinic	Yes / Explicit
Baker et al., 1991(United Kingdom) [25]	Developing a patient-based seizure severity scale as an outcome measure in evaluation of treatment in intractable Epilepsy	Liverpool Seizure Severity Scale (Original LSSS)	– Expert panels of the field (neurologists, neurophysiologists) created a scale on a basis of clinical experience.	Total n = 159	Not reported	No / Implicit
Baker et al., 1998(United Kingdom) [26]	Examining the psychometric properties of Revised Liverpool Seizure Severity Scale	Revised Liverpool Seizure Severity Scale (Revised LSSS)	– Modification of the original (LSSS Four items were added to improve the content validity of the scale)	Total n = 97	Specialist Epilepsy Clinic	No / Implicit
Scott-Lennox et al., 2001 (United Kingdom) [27]	Examining the reliability, validity, and responsiveness of the revised scoring system for the LSSS	Liverpool Seizure Severity Scale 2.0 (LSSS 2.0)	– Qualitative (Patient Focus group and expert panels) – Modified LSSS items – Revised scoring system of both versions of LSSS – Construct Analysis – Retrospective Re-analysis	n = 944 (n = 804 from ALERT study, n = 140 from LAM 30/31)	Analysis from RCT that utilized LSSS	No / Explicit
Cramer et al., 2002(USA & UK) [28]	Developing a new seizure severity scale that will be beneficial in clinical trials	Seizure Severity Questionnaires (SSQ)	– Review of the existing tools – Series of questions were reviewed by patients and physicians, and pilot testing – Panel of experts reviews the items – Items were formatted as a structured interview	Pilot study: Patient, n = 33 Observer n = 28 Reliability & validity assessment: n = 91	Tertiary Hospital Neurology Clinic	No / Explicit
Fisher et al., 2015 (USA) [29]	Developing a multidimensional PRO scale to quantify the overall impact of epilepsy on a single-scale (incorporating the impact of seizures, side effects, comorbidities and quality of life)	The Personal Impact of Epilepsy Scale (PIES)	– Open interview with 8 epilepsy patients to generate themes – Questionnaire constructed from the theme (152 items) – The drafted questionnaire was given to patients – NHS3 was completed and compared to the questionnaire	Initial interview n = 8 Respondent to draft questionnaire n = 50, final questionnaire n = 46	Epilepsy Clinic	No / Implicit

the scores obtained on different seizure types for both patients (and relatives) and healthcare professional perceptions.

The SSQ involved a panel of international epileptologists to review the revised questionnaires for receptiveness in a clinical setting. The investigators assessed the validity of SSQ using the construct validity method. Other tools that used construct validity for validity assessment were NHS3, LSSS 2.0, and PIES. Unlike NHS3, LSSS 2.0 tested the correlations between the instruments based on known relationships. This was executed on data from previous studies [30,31].

The most common types of reliability assessment that have been utilized include Test-retest, (CSSS, NHS3, Original and Revised LSSS, SSQ, PIES), inter-rater reliability (VA, CSSS, Original LSSS, SSQ), and Internal Consistency (NHS3, all LSSS versions). Differences that have been noted among these tools were whether the reliability assessment was carried out on the whole tool (VA, CSSS, NHS3, PIES) or tested by subscales (Original LSSS, Revised LSSS, LSSS 2.0, and SSQ). In the LSSS 2.0 development study, the seizure status was categorized as "minor," "major," and "most severe" and the reliability assessment was done individually. Due to the inabil-

Table 3
COSMIN Risk of Bias Checklist assessing methodological quality of each study per measurement property.

Tool, year (Ref. no.)	PROM development	Content Validity	Structural Validity	Internal Consistency	Reliability	Measurement Error	Criterion Validity	Hypotheses Testing	Responsive -ness
VA, 1983 [20]	–	D	I	I	I	I	I	I	I
Chalfont Scale, 1991 [23]	A	A	D	I	A	D	–	A	I
Original LSSS, 1991 [25]	A	I	I	I	A	I	I	A	I
NHS3, 1996 [24]	VG	A	–	VG	VG	VG	–	VG	I
Revised LSSS, 1998 [26]	D	D	D	D	A	I	I	A	I
LSSS 2.0, 2001 [27]	A	A	A	VG	A	I	–	A	A
SSQ, 2002 [28,19]	VG	VG	A	VG	VG	–	VG	VG	VG
PIES, 2015 [29,30]	VG	VG	A	VG	VG	I	VG	VG	VG

VG = Very Good, A = Adequate, D = Doubtful, I = Inadequate, – = Not applicable.

ity of capturing change in a major: minor status (major seizures reported at the beginning of the study that is fully controlled by medication could still leave minor seizures at the end of the study that was not evaluated) as well as other missing data (potentially underestimating the impact of the medication in minor seizures), the "most severe" score was more reliable and valid as an assessment of seizure severity. The descriptions of reliability assessments for all the tools are described in Table 5.

3.5. Contents of the tools

In general, the identified tools were presented either as a single scale or as several sections (subscales). The number of items in these tools, validity assessment, scoring range, and other features is described in Table 4. The first tool (VA scale) is divided into three subscales which are categorized by seizure types; focal-onset seizures ("simple," "complex," and "secondarily generalized"). Each of the scales consists of the components of seizure frequency, warning (aura), factors of triggering seizures, whether seizures occur as a cluster, drug levels, loss of consciousness (for focal impaired awareness seizures), diurnal patterns, and functional interference for focal aware seizures formerly known as simple partial seizures. For each type of seizure, a score is generated for a time interval. The score is then discounted by reviewing a series of altering conditions, for example, helpful warnings preceding a seizure, avoidable precipitants (e.g., lack of sleep), low drug levels, or seizures that mostly occur at night where the functionality is less likely to be impaired. Further modification of the scores is applied if focal aware seizures and focal impaired awareness seizures did not significantly disrupt functioning. Modifications are 20–80% deductions of the points assigned for frequency. Scores for all the subscales were added together to yield a total seizure score for the interval assessed in the clinical trial.

The second tool is CSSS which consists of 11 items. The items include descriptions of seizures such as loss of consciousness, aura, dropping of objects, injury, incontinence, automatism, duration of seizures and recovery time, generalized convulsion, and whether seizures occurred only in sleep. For each of the items describing seizure (except for the duration item), the absence of the symptoms is always allocated a 0 score. As described earlier, the updated version of the scale, NHS3, added a weighted score for each of the responses. One of the instances is the question on "incontinence" where the patient can choose the frequency of occurrence which is either "nearly always" or "always," "often," "occasionally" or "never." Even though this scale is simple and easy to administer, it is unable to capture fluctuations in seizure severity following changes in drug therapy.

Both the original LSSS and Revised LSSS consist of two subscales which were patients' perception of their seizure control (Percept) and the descriptions of events during and after the seizure episode (Ictal/ Postictal). The Percept subscale comprised questions on the time of seizure occurrence, the presence of aura or warning signs, and if the patient was able to expect a seizure and lessen its consequences. The Ictal/ Postictal subscale covered questions on the description of the seizure attack such as loss of consciousness, degree of postictal confusion and its duration, incontinence, fall, tongue-biting, injury, automatism, and perceived overall severity. The original version of LSSS included a total of 16 items; the Percept subscale consists of six items whereas 10 items are related to the Ictal/ postictal subscale. The inclusion of "major" and "minor" seizures into the scale, also with the addition of two items to each subscale, was established when the investigators revised the scale (Revised LSSS). The items that were included were the norms of seizure attack (particularly at night, wakening, or at any time of the day) and if the seizures come in a cluster or not (for Percept scale). As for the Ictal/ Postictal scale, the items that were added were whether the patients smacked their lips, fidgeted, or behaved in an unusual way during the attack and the presence of postictal drowsiness.

The final version of the Revised LSSS included eight items in the Percept subscale and 12 items in the Ictal/ Postictal subscale. The updated version, LSSS 2.0, eliminated the Percept subscale altogether with an inference that the Percept scale measured the impact of seizure on the quality of life and not the seizure severity per se. It came down to a 1-unit scale that comprises 12 items with individualized weighted scoring responses. Other than timeliness items (time to recover from consciousness, confusion, and return to normal activity), items on events and severity of seizures were reversely weighted (0 score for the most severe and 4 for the least severe symptoms). This is to avoid patients responding in a pattern due to recall issues. The physician will then reverse the code for the total score.

The SSQ, which was first commenced in 1999, consisted of a total of 24 items and has three main subscales addressing warning, ictal, and postictal events. For the warning scale, the developer addressed whether the warning symptoms could help the patient in terms of readiness of having the attack. Both Ictal and Postictal subscales tackled the degree of severity and bothersomeness of the events with an additional query on the frequency for the postictal subscale. Items on the Postictal subscale cover the cognitive effects, emotional effects, recovery, and physical effects (sleepy, tired, weak, sore muscles, headache) after the seizure episode. Three items addressed the overall intensity (severity), bothersomeness (interference with patients' livelihood), and which phases of seizures bother the most (warnings, events during a seizure, and

Table 4
Extractions from studies included (subjects, content and validity assessment).

Tool, year (Ref. no.)	Average Age of subjects (years)	Scale Rater	No. of items	Subscale (domain)	Scoring and interpretation	Validity assessment used	Usability
VA, 1983 [20]	Adults (Age range not mentioned)	Physician	6–8 items per subscale	Subscale based on seizure types (divided into 3 sections: Generalised clonic-tonic, complex partial, and simple partial)	Total score unclear	Not specified in the article but was undertaken by other investigators (Wijsman, 1991)	Clinical Trials for ASMs Clinical practice
Chalfont, 1991 [23]	Adults (Age range not mentioned)	Interview with Observer/Carer/ Witness	11	No subscale	1–178 – Scored as per type of seizure, higher score = more severe	Face & Content Validity	Longitudinal Studies
NHS3, 1996 [24]	31 (median) (IQR range: 27–44)	Observer/Carer/ Witness	8	No subscale	1–27 – Higher score = more severe seizure	Construct Validity	Clinical Trials for ASMs
Original LSSS, 1991 [25]	31 (mean) (range: 15–80)	Patient	16	i. Percept subscale ii. Ictal/- postictal subscale	Not mentioned in the article	Not specified in the article	Clinical Trials for ASMs
Revised LSSS, 1998 [26]	32 (mean) (range: 15–70)	Patient	20	i. Minor Percept subscale ii. Major Percept Subscale iii. Minor Ictal/postictal subscale iv. Major Ictal/postictal subscale	Percept scale: 7–32 Ictal/post ictal scale: 10–48 – Higher score = more severe seizure	Criterion Validity	Clinical Trials for ASMs
LSSS 2.0, 2001 [27]	39.4 (mean)	Patient	12	No subscale (Ictal scale only)	0–100 – Higher score = more severe seizure	Construct Validity	Clinical Trials for ASMs
SSQ, 2002 [28]	39 (mean) (range: 17–77)	Physician (facilitated interview)	24	i. Warning ii. Activity movement (ictal) iii. Recovery (post ictal)	7 (maximum score) – Higher score = more severe seizure	Face & Content Validity (pilot Study) Construct Validity	– Clinical Trials for ASMs – Clinical setting
PIES, 2015 [29]	42.7 (mean) (range: 21–71)	Patient	9 (seizure domain)	No subscale (seizure domain)	0–36 (seizure domain) – Higher score = more severe seizure	Construct Validity	– Clinical Trials for ASMs – Healthcare setting

Table 5
Findings of reliability assessments of the tools.

Tool, year (Ref. no.)	Scale Tested	Reliability Assessment Utilized			Statistical Method		
		Test-retest	Inter-rater or inter-observer	Internal consistency of the scale	Pearson Correlation (r-value)	Cronbach's alpha	Others (or unclear statistical tests)
VA, 1983 [20]	Overall	-	✓	-	0.9	-	-
Chalfont Scale, 1991 [23]	Overall	✓	✓	-	-	-	Coefficient of reliability: Inter-rater reliability = 13.4 Test-retest = 15.9
NHS3, 1996 [24]	Overall	✓	✓	✓	Interclass correlation: test-retest = 0.9 Inter observer = 0.9	0.77	-
Original LSSS, 1991 [25]	Subscale: i. Ictal/ postictal ii. Percept	✓ (by subscale; ictal/post ictal and Percept subscale)	✓ (only 4 of the items drawn from the scale (not specified)	✓	Test-retest result: ictal/ postictal = 0.8 Percept = 0.79	ictal/post ictal = 0.85 percept = 0.69	-
Revised LSSS, 1998 [26]	Subscale: Major seizures i. Percept ii. Ictal/ postictal Minor seizures i. Percept ii. Ictal/ postictal	✓	-	✓	Subscale: Major seizures i. Percept = 0.96 ii. ictal/ postictal = 0.93 Minor seizures i. Percept = 0.72 ii. ictal/post ictal = 0.78	Subscale: Major seizures i. Percept = 0.62 ii. ictal/postictal = 0.85 Minor seizures i. Percept = 0.68 ii. ictal/postictal = 0.86	-
LSSS 2.0, 2001 [27]	Subscale: Major seizures Minor seizures Most severe	-	-	✓	-	From ALERT study: Most severe = 0.87 From LAM30/31 study: Most severe = 0.73	-
SSQ, 2002 [28]	Each of the subscales were tested for reliability (22 items)	✓	✓	-	✓ (Correlations among summary score and derived variables)	-	Coefficient of reliability(Kappa coefficient) : Test-retest: Summary score = 0.74 Inter-rater: Summary score = 0.76
PIES, 2015 [29]	Overall	✓	-	✓	Internal consistency coefficient (seizure domain): Range: 0.644–0.837	0.87 (overall)	-

Label: '✓': tests done, '-': test not done/not applicable.

recovering from a seizure). The final item asked about any changes in terms of severity or bothersomeness after changing seizure therapy. Following this tool development, two updated versions were accessible for use. SSQ V3 (2003) did not have much change in the scale except an elaboration on the recovery responses whereas the latest version (SSQ V2.2 by Cramer et al., 2014) has an additional subitem which addresses any change of seizure severity since baseline following changes in seizure treatment (baseline and follow-up versions).

As mentioned earlier, the SSQ items were used as a reference in developing the construct for seizure assessment in the PIES. Ultimately, the domain comprises nine items. These items include the duration from the last seizure attack (any type and most severe), with a response category ranging from 1 day to > 1 year. Other aspects of seizures such as the overall intensity, loss of awareness, warning, bothersomeness, and post-seizure bothersomeness within the last 3 months are covered in items number 3 to 7. The last two items gauged the response for injuries suffered as a consequence of an attack, and lastly the duration and presence of a seizure cluster. These items were fitted with a 5-point scale ranging from 0 to 4 which indicates from "never" or "not at all" to "very often" or "very much".

4. Discussion

This systematic review compared various assessment tools that were developed to measure seizure severity among adults with epilepsy. One of the most crucial parts of developing a tool or a questionnaire is the initial selection of items to be included. Many researchers have acknowledged that the process of scale development entails complicated procedures that necessitate theoretical and methodological disciplines [32]. This includes the design and development of factors/items, preliminary questionnaire testing (pilot testing), validity and reliability testing, and subsequent validation. The researcher can adopt or follow a pre-existing questionnaire with a standard structure from existing literature while devising a new questionnaire [33].

The qualitative approach to the initial development varies from one study to another. The main developers must first identify the theorized structures that are relevant to patients, and evaluate the employment of the assessment tool, the reports of trials that addressed the constructs, and the features of the patients [34]. The selection of the approach is dependent on the aim of the study. Since these studies focused on obtaining the patients' experience of seizure events, interviews would help the researchers to gain in-depth information on persons' subjective experiences, feelings, and motivations. Interviews also have the advantage of being interactive, allowing unexpected topics to arise and be identified by researchers [35]. Different methods such as face-to-face interviews, self-completed questionnaires, diary reviews, and various tool applications such as electronic devices or computers are also able to gauge patients' perspectives.

Unlike the individual interview approach, focus group discussions (FGD) are beneficial for bringing together similar participants with relevant experience and sharing detailed information about their experience [36]. One of the drawbacks of this approach includes the participants' reservations about sharing sensitive topics in a group. In addition, the moderators and the data analysts required a certain level of experience [35]. It is worth noting that all the publications in this review complied with the standard format of initial tool development using one or more of the standard qualitative research approaches.

The validity assessment of an instrument is one of the key requirements in its development [37]. Most of the studies in this review have carried out validity tests. Face validity was the frequently used test in these studies. It is defined as a subjective assessment of the arrangement of items in a tool as well as whether the items appear to be appropriate, sensible, and straightforward in measuring the intended construct [38]. It is the easiest and least precise approach to establishing validity, as it is entirely dependent on the assessor's competence and acquaintance with the topic area [39]. The external validity of the tool can be improved by obtaining population representation through approaches e.g., random selection, using diversified groups, utilizing non-reactive measures, and applying precise description to allow study repetition or replication across different populations and settings [37]. This was well executed by developers of the updated versions of the seizure severity assessment tools.

Generally, all the investigators have conducted one or more reliability assessment methods to verify the scales' precision, consistency, repeatability, and trustworthiness. The reliability of the test–retest and other forms is commonly determined using the statistical test of correlation [40]. For a low-stakes setting or less important situations, values of 0.7 or 0.8 are considered sufficient. Most of the studies managed to achieve a coefficient of reliability of more than 0.7. As a general rule, reliability values greater than 0.8 are considered high [41]. To prevent threats of reliability, it is important to provide clear standardized instruction, minimize ambiguity in the items, arrange the items in proper order, and provide a simple comprehensive questionnaire that is easy to read and does not take a long time to respond to [42]. These are seen in updated versions and more recently developed seizure severity assessment tools (NHS3, LSSS 2.0, SSQ, PIES).

An established guideline issued by the FDA; Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labelling Claims outlined the requirement and guidance for the development of patient-reported outcome instruments [43]. The guidelines comprehensively described the proper and systematic way of tools' development, from hypothesizing and adjusting the conceptual framework to modifying the instrument which includes culturally adapting and translating to other languages.

The difference in the number of items and contents of the seizure severity tools is notable, even for the revised version. The most notable similarity across all the tools is the ability of the tool to capture a certain degree of disruptive automatisms and functional impairments experienced by the patients in one way or another. This is an important feature as functional impairment (disability) is probably a good marker for the effectiveness of treatment which is especially appropriate for use in drug trials [1]. One of the major differences between these tools is the number of scale sets, either multiple subscales (VA, Original, Revised LSSS, and SSQ) or just one-segmented tool. These subscales may help the rater to recall the events and distinguish between seizure phases. The LSSS 2.0 removed the Percept scale that represents the events before seizures and overall patients' perception toward seizure control, due to the lack of function for severity assessment. This may give rise to an argument on whether the limited number of questions may or may not cover an individual's seizure severity [1]. Earlier tools (VA, CSSS, NHS3, LSSS) may also bear this drawback, coupled with subjective and ambiguous scoring systems, cumbersomeness, an undetermined degree of score improvement/worsening with drug therapy changes, and a lack of sensitivity of the scale to gauge an individual's specific response. This issue was addressed with the development of the second most recent seizure severity tool [3].

In the 90s when new ASMs were being discovered, drug trials utilized seizure severity tools as one of the instruments to measure drug effectiveness [34,35]. The Original LSSS was used and was adequately able to assess the change in seizure severity, however, it did not correlate with the change in seizure frequency. Thus, it was considered an independent measure. The ictal subscale of

the tool is capable of detecting changes in seizure severity hence suggesting a true additional treatment effect. Thereafter, when the SSQ has been published, the newer ASMs trials utilized the tool as a part of effectiveness measurement, adding the advantage of its capability of detecting changes in clinical status. A recently published ASM trial that utilized the SSQ found significant improvements in seizure severity with eslicarbazepine treatment, with improvement in health-related quality of life as well [44].

Throughout the years of researching and developing severity assessments, researchers have acknowledged the need to assess the overall impact of epilepsy and not just seizure events. This led to the development of the most current scale, the PIES. As mentioned earlier, this scale provides an assessment of the overall impact of epilepsy which are aspects of seizure, adverse drug reactions, comorbidities as well as the quality of life as rated by patients, in a single tool. Its development complied with the standard guideline mentioned above with additional analysis on the minimal clinically important difference (MCID), which was carried out recently [28,43]. Prior to this, the SSQ research team demonstrated that the tool plays an important role in assessing the change of seizure severity as an outcome measure in future ASM trials [19]. Therefore, it is worth considering this analysis as an ideal prerequisite element for the establishment of future patient-reported outcome instruments.

Thus far, COSMIN Checklist is the only validated tool appropriate for assessing the methodological quality of PROM studies. One of the concerns of its use in these studies is the tendency to underestimate a study that uses terminology other than that used by COSMIN [45]. This was handled by assigning at least two independent assessors in which the final score is based on consensus. From the assessment, the SSQ and the PIES met all the key requirements and have the advantage of having a number of earlier tools to refer to and were able to tackle the limitations of the latter. Understanding the lack of sensitivity and responsiveness of the tool, the updated versions of the SSQ were developed [18,19]. It has the baseline and follow-up versions to capture any changes that occur along the course of ASM therapy. Thus, the SSQ and its later versions, are more appropriate for use in clinical practice as compared to the earlier tools. In contrast to the SSQ, the PIES provides a more concise assessment of seizure severity which would make it more favorable for daily clinical practice.

Researchers have investigated the possibilities of using other methods such as incorporating diagnostic measures with the existing seizure severity tools to measure seizure severity. A recent study conducted by Pattnaik et. al [10] was reported to have a positive finding on the association between preictal iEEG recordings with severe seizure activity. In the study, they combined the three components of seizure; seizure semiology (as measured by NHS3), spread, and duration of seizure as measured by iEEG network recordings into a single scoring tool. An important point to note with regard to this approach is, it is unable to detect changes in severity following changes in ASM therapy. Besides, there may be discrepancies in seizure severity scores in an ambulatory EEG setting, where more seizure triggers may be inflicted. Nevertheless, this approach is a valuable breakthrough and is anticipated to be further expanded to assess long-term changes in seizure severity.

5. Conclusion

The development of seizure severity assessment tools has come a long way since first initiated four decades ago. Despite the variety and vast differences in the process of developing the tools, these existing tools were able to serve their function which is mainly as an instrument to assess the effectiveness of ASM therapy. Improvements and lessons learned from the limitations of the earlier tools would heighten the usability and functionality of the new ones, supported by additional tests to assess their sensitivity and responsiveness. In recent years, methods other than patient-reported outcomes have been investigated for measuring seizure severity. It would be interesting to see the integration of the objective measures and PROM in a single module to be established and further utilized in clinical trials, other research, and daily clinical practice.

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.

Acknowledgments

The authors would like to thank the Ministry of Health Malaysia for the studentship support for this research.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.yebeh.2023.109154.

References

[1] Cramer JA, French J. Quantitative assessment of seizure severity for clinical trials: A review of approaches to seizure components. Epilepsia 2001;42 (1):119–29. https://doi.org/10.1046/j.1528-1157.2001.19400.x.
[2] Schmidt D. Evaluation of clinical efficacy of antiepileptic drug trials. Epilepsy Res 1991;3:69–77. PMID: 1777084.
[3] Cramer JA, Baker GA, Jacoby A. Development of a new seizure severity questionnaire: initial reliability and validity testing. Epilepsy Res 2002;48 (3):187–97. https://doi.org/10.1016/s0920-1211(02)00003-7.
[4] Baker GA, Smith DF, Dewey M, Morrow J, Crawford PM, Chadwick DW. The development of a seizure severity scale as an outcome measure in epilepsy. Epilepsy Res 1991;8:245–51. https://doi.org/10.1016/0920-1211(91)90071-m.
[5] DeVellis RF. Scale development: theory and applications. 2nd ed., vol. 26. Thousand Oaks, CA: SAGE Publications; 2003.
[6] DeVellis RF. Scale development. 4th ed. Thousand Oaks, CA: SAGE Publication; 2016.
[7] Dimitrov D. Statistical methods for validation of assessment scale data in counseling and related fields. Article in Applied Psychological Measurement. Virginia: American Counseling Association; 2012.
[8] Delgado-García G, Wiebe S, Josephson CB. The use of patient-reported measures in epilepsy care: the Calgary Comprehensive Epilepsy Program experience. J Patient Rep Outcomes 2021;5(Suppl 2):83. https://doi.org/10.1186/s41687-021-00356-4.
[9] Nelson EC, Eftimovska E, Lind C, Hager A, Wasson JH, Lindblad S. Patient reported outcome measures in practice. BMJ 2015;350:. https://doi.org/10.1136/bmj.g7818. PMID: 25670183g7818.
[10] Pattnaik AR, Ghosn NJ, Ong IZ, Revell AY, Ojemann WKS, Scheid BH, et al. A quantitative tool for seizure severity: diagnostic and therapeutic applications. BMJ 1 November 2022. medRxiv Preprint.10.26.22281569; doi:https://doi.org/10.1101/2022.10.26.22281569.
[11] Borghs S, Tomaszewski EL, Halling K, de la Loge C. Understanding the Patient Perspective of Seizure Severity in Epilepsy: Development of a Conceptual Model. Patient 2016;9(5):419–31. https://doi.org/10.1007/s40271-016-0165-0. PMID: 27002318.
[12] Cramer JA. Seizure Measurement in Clinical Trial. J Epilepsy 1998;11:256–60. https://doi.org/10.1016/S0896-6974(98)00027-9.
[13] Wiebe S. Seizure control as an indicator of therapeutic usefulness. Can J Neurol Sci 2000;27(1):S97–S105. https://doi.org/10.1017/s0317167100000743.
[14] Chan CJ, Zou G, Wiebe S, Speechley KN. Global assessment of the severity of epilepsy (GASE) scale in children: Validity, reliability, responsiveness. Epilepsia 2015;56(12):1950–6. https://doi.org/10.1111/epi.13216.
[15] Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, et al. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. Qual Life Res 2018;27(5):1171–9. https://doi.org/10.1007/s11136-017-1765-4.
[16] 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.
[17] Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:. https://doi.org/10.1136/bmj.n71. PMID: 33782057; PMCID: PMC8005924n71.
[18] Borghs S, de la Loge C, Brabant Y, Cramer JA. Sensitivity testing of the Seizure Severity Questionnaire (SSQ). Epilepsy Behav 2014;31:281–5. https://doi.org/10.1016/j.yebeh.2013.10.010.
[19] Cramer JA, de La Loge C, Brabant Y, Borghs S. Determining minimally important change thresholds for the Seizure Severity Questionnaire (SSQ). Epilepsy Behav 2014;31:286–90. https://doi.org/10.1016/j.yebeh.2013.09.006.
[20] Cramer JA, Smith DB, Mattson RH, Delgado-Escueta AV, Collins JF and the VA Epilepsy Cooperative Study Group. A method of quantification for the evaluation of antiepileptic drug therapy. Neurology 1983;33(1):26–37. doi:10.1212/wnl.33.3_suppl_1.26.
[21] Mattson RH, Cramer JA, Collins JF and the Dept. of Veterans Affairs Epilepsy Cooperative Study No. 264 Group. A comparison of valproate with carbamazepine for the treatment of partial seizures and secondarily generalized tonic-clonic seizures in adults. N Engl J Med 1992;327:765–771. doi:10.1056/NEJM199209103271104.
[22] Mattson RH, Cramer JA, Collins JF, Smith DB, Delgado-Escueta AV, Browne TR, et al. Comparison of carbamazepine, phenobarbital, phenytoin and primidone in partial and secondarily generalized tonic clonic seizures. N Engl J Med 1985;313:145–51. https://doi.org/10.1056/NEJM_198507183130303.
[23] Duncan JS, Sander JWAS. The Chalfont Seizure Severity Scale. J Neurol Neurosurg Psychiatry 1991;54:873–6. https://doi.org/10.1136/jnnp.54.10.873.
[24] O'Donoghue MF, Duncan JS, Sander JW. The National Hospital Seizure Severity scale: A further development of the Chalfont Seizure Severity scale. Epilepsia 1996;37(6):563–71. https://doi.org/10.1111/j.1528-1157.1996.tb00610.x.
[25] Baker GA, Smith DF, Jacoby A, Hayes JA, Chadwick DW. The Liverpool Seizure Severity Scale revisited. Seizure 1998;8:201–5. https://doi.org/10.1016/s1059-1311(98)80036-8.
[26] Scott-Lennox J, Bryant-Comstock L, Lennox R, Baker GA. Reliability, validity and responsiveness of a revised scoring system for the Liverpool Seizure Severity Scale. Epilepsy Res 2001;44(1):53–63. https://doi.org/10.1016/s0920-1211(01)00186-3.
[27] Fisher RS, Nune G, Roberts SE, Cramer JA. The Personal Impact of Epilepsy Scale (PIES). Epilepsy Behav 2015;42:140–6. https://doi.org/10.1016/j.yebeh.2014.09.060. Epub 2014 Nov 20 PMID: 25450530.
[28] Martz G, Fisher RS, Folley B, Panza GA, Ando F, McEachern C, et al. Minimum Clinically Important Difference (MCID) of the Personal Impact of Epilepsy Scale (PIES). Epilepsy Behav 2022;130:. https://doi.org/10.1016/j.yebeh.2022.108691. Epub 2022 Apr 19 PMID: 3545304108691.
[29] Cramer J. Seizure Severity Questionnaire used in LCM clinical trials. Available from: https://sites.google.com/site/seizureseverity/; 2007.
[30] Smith D, Baker GA, Davies G, Dewey M, Chadwick DW. Outcomes of add-on treatment with lamotrigine in partial epilepsy. Epilepsia 1993;34:312–22. https://doi.org/10.1111/j.1528-1157.1993.tb02417.x.
[31] Smith T, Shumaker S, Rapp S, McFarlane M, Tennis P. Adjunctive lamictal in epilepsy: Response to treatment (ALERT)-Baseline sample characteristics. Epilepsia 1996;37(5):165.
[32] Morgado FFR, Meireles JFF, Neves CM. Scale development: Ten main limitations and recommendations to improve future research practices. Psicol Refl Crit 2017;30:3. https://doi.org/10.1186/s41155-016-0057-1.
[33] Aithal A, Aithal PS. Development and validation of survey questionnaire & experimental data – A systematical review-based statistical approach. IJMTS 2020;5(2):233–51. https://doi.org/10.2139/ssrn.3724105.
[34] Calvert M, Blazeby J, Altman DG, Revicki DA, Moher D, Brundage MD. Reporting of patient-reported outcomes in randomized trials: The CONSORT PRO extension. J Am Med Assoc 2013;309:814–22. https://doi.org/10.1001/jama.2013.879.
[35] Busetto L, Wick W, Gumbinger C. How to use and assess qualitative research methods. Neurol Res Pract 2020;2, 14(1-10). https://doi.org/10.1186/s42466-020-00059-z.
[36] Van Royen P, Peremans L. Exploring with focus group conversations: the "voice" of the group under the magnifying glass. Houten: Bohn Stafleu van Loghum. Netherlands; 2007. p. 53–64.
[37] Mohajan H. Two criteria for good measurement in research: Validity and reliability. Annals of Spiru Haret University 2017;17(3):58–82. https://doi.org/10.26458/1746.
[38] Taherdoost H. Validity and Reliability of the Research Instrument; How to Test the Validation of a Questionnaire/Survey in a Research. Int J Acad Res Manag 2016;5:28–36. https://doi.org/10.2139/ssrn.3205040.
[39] Nwana OC. Textbook on educational measurement and evaluation. Owerri: Bomaway Publishers; 2007.
[40] Traub RE, Rowley GL. An NCME instructional module on understanding reliability. Educ Meas Issues Pract 1991;10(1):37–45. https://doi.org/10.1111/j.1745-3992.1991.tb00183.x.
[41] Downing SM. Reliability: On the reproducibility of assessment data. Med Education 2004;38:1006–12. https://doi.org/10.1111/j.1365-2929.2004.01932.x.
[42] Church J. Reliability, threats to reliability and the assessment of reliability. School of Educational Studies and Human Development University of Canterbury, Christchurch, New Zealand; 2021. http://www.tecks.co.nz [accessed 27 October 2021].
[43] U.S. Department of Health and Human Services. Food and Drug Administration (FDA). Guidance for Industry Patient Reported Outcome Measure: Use in Medical Product Development to Support Labeling Claims 2009.
[44] Cramer JA, Colman S, Anastassopoulos K, Grinnell T, Mehta D, Williams G. Associations between seizure severity change and patient characteristics, changes in seizure frequency, and health-related quality of life in patients with focal seizures treated with adjunctive eslicarbazepine acetate: Post hoc analyses of clinical trial results. Epilepsy Behav 2020;112:107312. 10.1016/j.yebeh.2020.107312.
[45] Mason SJ, Catto JWF, Downing A, Bottomley SE, Glaser AW, Wright P. Evaluating patient-reported outcome measures (PROMs) for bladder cancer: A systematic review using the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist. BJU Int 2018;122:760–73.