

Systematic Review

Assessing exposure to childhood adversity in adults: A systematic review of validated self-report childhood adversity questionnaires

Franziska Mosler^a, Aikaterini Christogianni^a, Singleton Sam^b, Tim G. Hales^b, Janine Rennie^a, Lesley A. Colvin^{a†}, Line Caes^{c†}

^a Population Health and Genomics, University of Dundee, Dundee, United Kingdom

^b Division of Neuroscience, University of Dundee, Dundee, United Kingdom

^c Division of Psychology, Faculty of Natural Sciences, University of Stirling, Stirling, United Kingdom

Evaluation of the current childhood adversity assessments in adults

Corresponding Author

Line Caes

E-mail address: Line.Caes@stir.ac.uk

†joint senior authors

Keywords: Adverse Childhood Experiences, Assessment Tools, Questionnaires

Abstract

Children exposed to adverse experiences (ACEs) are more likely to experience mental health problems in adulthood. However, ACE assessment is highly heterogeneous, hampering widespread screening and trauma-informed care provision. We aimed to systematically identify and critically evaluate all validated, self-report ACE questionnaires, working with people with lived experience (PWLE). The review followed PRISMA guidelines to systematically search databases for validated self-report measures, completed by adults, assessing at least two ACEs. Articles were excluded if they were not written in English, were not original articles, assessed poor childhood health or adverse experiences happening in adulthood, and/or only assessed one ACE. Psychometric properties were evaluated using Cohen's criteria for evidence-based assessments, the COSMIN checklist, and a content validity form co-designed with PWLE. We identified 112 eligible studies covering 31 ACE questionnaires. Cohen's criteria classified 9 questionnaires as "well-established" and 2 as "approaching well-established". No questionnaire was rated as "sufficient" quality across all the COSMIN measurement properties. The ACE Study-questionnaire, Childhood Experiences Survey (CES) and the ACE-International Questionnaire had the highest number of properties rated as sufficient. The Childhood Trauma Questionnaire-Short Form (CTQ-SF) was most frequently evaluated but received "insufficient" ratings across all measurement properties. PWLE scored content validity highest for the CES. Consequently, no ACE questionnaires received good psychometric ratings, with the most widely used questionnaire (CTQ-SF) not performing well, which has implications for selecting an appropriate instrument. With increasing emphasis on trauma-informed health care, there is an urgent need to co-develop ACE questionnaires with PWLE to balance content validity with usability.

Introduction

Adverse childhood experiences (ACEs) are the commonest and most intense childhood stressors [1]. Throughout this review, we define ACEs as events or situations likely to cause harm or distress that a child either experiences themselves or encounters within their environment, which undermine their sense of safety. The worldwide prevalence of experiencing at least one ACE is 60%, [2] with those exposed to multiple ACEs more likely to experience poor mental and physical health outcomes [3–5]. For instance, exposure to ACEs have been linked to a higher risk of developing (multiple) chronic conditions (e.g., depression, chronic pain, and cardiovascular disease) [5,6]; engaging in health-risk behaviours (e.g., smoking and the misuse of alcohol and other substances); developmental disruption; and heightened healthcare utilization [3,4]. Importantly, ACEs are a major risk factor for premature mortality in adulthood, either due to increased risk of developing a life-limiting disease or suicidal behaviours [7]. Although there is increasing recognition of the need for trauma-informed clinical care across all levels of healthcare provision, there is a lack of widespread screening, which could support more personalised health interventions. Key factors in this lack of screening are inconsistencies regarding ACE assessment [8,9] and the lack of confidence in identifying which ACE assessment is most appropriate and sensitive in assessing ACEs.

Since the seminal work by Felitti and colleagues [3], one of the most common ways to assess ACEs is their 10-item Adverse Childhood Experience Study-Questionnaire (ACE-Q) [3], involving a cumulative approach to determine a person's ACE exposure and associated risk level. However, by focusing on a limited number of events it does not capture several ACEs. Various ACE instruments were subsequently developed, which often focussed on assessing specific adverse events in more detail, such as the widely used Childhood Trauma Questionnaire (CTQ) [10]. Previous reviews [9,11,12] have highlighted a wide variety of measures available to assess ACEs, which vary in their focus, depth, length, scoring, and psychometric quality. This inconsistency highlights the lack of a unifying theoretical framework or definition of an ACE and a lack of input from people with lived experience (PWLE) guiding the development of ACE assessment. The inconsistency also makes it challenging to identify which instrument is most appropriate for a particular clinical setting, thereby limiting the impact on trauma-informed clinical and public policy approaches [8,13].

Consequently, there is an urgent need for an exhaustive overview of the psychometric qualities of the existing ACE instruments and their relevance and suitability from the perspective of PWLE. The evidence available in the existing reviews evaluating the psychometric qualities of ACE assessments

[9,11] is not sufficient. Firstly, given the rapid rise in ACE-related research, the reviewed evidence (i.e., studies up to August 2016) is outdated. Secondly, the focus of existing reviews is restricted by specifying either a particular health condition or types of ACE. Lastly, none of the reviews provide guidance for clinical practice, nor include the perspective of PWLE to comprehensively evaluate the content validity of ACE instruments. Including the perspective of PWLE in measurement evaluation and development is critical for enhancing study quality, as well as improving questionnaire comprehension, relevance and acceptability by the target group [14].

The aim of this systematic review was to identify all validated ACE questionnaires and critically evaluate and summarise 1) their psychometric quality using standardised guidelines relevant for evidence-based clinical practice and research and 2) their content validity from the perspective of PWLE.

Methods

Patient and Public involvement

This study was part of a larger project, the Consortium Against Pain Inequality (CAPE) (see <https://dundee-cape.ac.uk/>). People with lived experience of ACEs and a chronic illness (i.e., chronic pain) provided input and guidance on the aims and objectives from the grant application stage onwards and the project team included a patient partner representative. At the start of the CAPE project, an advisory group, consisting of 8 PWLE of ACEs and chronic pain (all female with an age range of 30-69), was established who offered their insights in monthly meetings to discuss all aspects of the project: project goals, approaches, results and dissemination. For this review specifically, following the guidance by Staniszewska et al., 2012 [14], the advisory group provided input on the search terms used and was actively involved in co-designing the content validity evaluation form, as well as evaluating the content validity of the identified ACE questionnaires.

Search strategy and selection criteria

The protocol was registered on PROSPERO (CRD42022299435) and PRISMA guidelines for systematic reviews were followed [15]. No ethical approval was needed for this review. Searches were carried out between 14th March 2022 (original search) and July 2023 (reference list checks). An updated search was conducted in May 2024. Databases were searched for any articles published since 1998 (based on the publication date of the seminal paper by Felitti et al. [3]) until the date of the search. Eight databases were searched, MEDLINE, PubMed, PsycINFO, Web of Science, EMBASE, Scopus, CENTRAL, and CINAHL, using the following key terms: 'adverse child experiences or negative child

experience or childhood adversity or adult survivors child abuse or child abuse survivor or childhood trauma survivor', 'adverse childhood event or ACEs', 'questionnaires or measures or survey or scale or instrument', 'adult or young adult or student or adolescent' (see Appendix 1 for complete search strategy, individualised for each database).

A two-stage search strategy, facilitated through Covidence [16], was employed with an initial title and abstract screening conducted by three independent reviewers. A 20% overlap between the reviewers was implemented to establish inter-rater reliability, using Cohen's Kappa. For practical reasons, we divided the abstract screening stage into two phases (representing a small deviation from the protocol). The initial abstract screening phase focused on identifying any relevant articles that assess ACEs using a questionnaire in participants aged 16 or older. While the typical age of maturity is often considered to be 18 years, various countries identify people aged 16 as mature, allowing them to make decisions without parental consent. To avoid excluding articles adopting this lower maturity threshold, we included articles in young people aged 16 years.

Inclusion criteria for articles were as follows: (1) participants were ≥ 16 years old; (2) ACEs were reported retrospectively by means of a self-report questionnaire; (3) the article reported on original research (irrespective of study design) and (4) the article was peer-reviewed, including conference abstracts.

Articles were excluded if they (1) focused on assessing only one particular ACE (e.g., only assessing sexual abuse), as our goal was to evaluate comprehensive assessments of ACEs; (2) focused on trauma experienced due to poor childhood health (e.g., childhood cancer), as this represents a substantially different childhood context; (3) focused on adverse experiences occurring when participants were aged 16 or older; (4) did not report on original research (i.e., reviews, books and book chapters); (5) reported on case studies; and (6) were published in a language other than English.

The second phase of the abstract screening phase aimed to identify articles that evaluated the psychometric qualities of the questionnaires. All selected abstracts from the initial abstract screening were independently screened by two reviewers. Articles were included if the abstract indicated that the study evaluated any psychometric quality of the ACE questionnaire used (e.g., internal consistency, test-retest reliability, factor structure, content validity, construct validity, cross-cultural validity, criterion validity).

Following the second round of abstract screening, full text screening and data extraction were conducted by one reviewer, with a total of 20% of records double-screened by a second reviewer. Across all stages, any discrepancies were discussed with a third reviewer. Reference lists of included articles were searched for additional relevant studies.

Data extraction

All articles were grouped by questionnaire to facilitate data extraction. Data extracted included sample characteristics (i.e., age, gender, population, sample size, ethnicity, study country), number and types of ACEs assessed, information on questionnaire design and development (i.e., name, number of items, language of questionnaire, administration details, scoring details, target population, involvement of PWLE, support provision), study design and psychometric results (i.e., data on any validity or reliability or factor analyses). All data were extracted by one reviewer, with 20% of records checked by a second reviewer. For decisions on PWLE involvement both reviewers categorised studies independently and agreed on a final rating after discussion. Studies were categorised as including PWLE if they demonstrated extensive engagement of participants beyond one-off pilot testing and participants being described as having exposure to ACEs. If available within the article, a copy of the questionnaire was retained. If no copy of the questionnaire was available within the article, authors were contacted (up to 3 times) for original versions of questionnaires. For four questionnaires, we were unable to obtain an English version of the questionnaire, and hence unable to assess their content validity.

Quality assessment of questionnaires

1. Cohen

The quality of each ACE questionnaire was assessed in three steps. In step one, all questionnaires were rated using Cohen and colleagues' [17] criteria for evidence-based assessment, which identify whether a questionnaire is well-established, approaching well-established assessment, or promising. To receive the rating "well-established", the questionnaire must have been used in at least two peer-reviewed articles by different investigators, with the articles providing sufficient detail about the questionnaire to allow critical evaluation and replication (e.g., measure and manual provided or available upon request). Additionally, detailed (e.g., statistics presented) information on validity and reliability must have been reported on, in at least one peer-reviewed article. To receive the rating "approaching well-established", the questionnaire must have been used in at least two peer-reviewed articles, which could be by the same investigators, with sufficient detail

about the questionnaire for critical evaluation and replication (e.g., measure and manual provided or available upon request). Additionally, the validity and reliability information presented was either vague or limited (e.g., no statistical analyses) or moderate values for validity and reliability were available in the literature.

The rating “promising” applied to questionnaires that have been used in just one peer-reviewed article, with sufficient detail about the questionnaire for critical evaluation and replication (e.g., measure and manual provided or available upon request). Additionally, the validity and reliability information presented was either vague or limited (e.g., no statistical analyses) or moderate values for validity and reliability were available in the literature.

2. COSMIN

In step 2, the quality of the measurement properties of the questionnaires meeting the ‘well-established’ and “approaching well-established” Cohen criteria were assessed using the COnsensus-based Standards for the selection of health Measurement InstrumeNts (COSMIN) checklist [18]. Two independent reviewers rated the questionnaire on structural validity, internal consistency, reliability, construct validity, cross-cultural validity, measurement invariance, measurement error, and criterion validity using the COSMIN checklist. The criterion “responsiveness” was not rated as this property assesses whether the questionnaire can detect changes in the measured construct over time. This is not an appropriate or relevant measurement property for an ACE questionnaire, given that a change in adult reporting of ACEs would not be expected. For each measurement property a rating of sufficient (+), insufficient (–) or indeterminate (?) was given. These ratings were made per article, with each reviewer providing a rating independently. For each assessed measurement property, the two reviewers reached a final consensus rating, based on either agreement in their independent ratings or discussions between the two reviewers to solve any conflicts in ratings. To determine the overall rating of a questionnaire, across evidence from multiple articles, the ratings for each article were summarised as sufficient (+), insufficient (–); or indeterminate (?). Following the COSMIN guidance, to rate the summarized results as sufficient or insufficient, 75% of the ratings had to meet the respective criteria. In cases of inconsistent ratings, overall ratings were based on the majority rating. In the case of an equal split between sufficient (+) and insufficient (–) ratings, the decision was made on the ratings of the most recent publication assessing that measurement property of that questionnaire.

Given the focus of the review, we neither assessed the quality of the questionnaire development (as the included studies assessed the psychometric quality of ACE questionnaires rather than describing the questionnaire development) nor applied the risk of bias checklist to score the quality of the

evidence (which was integrated as part of the Cohen's evaluation). We also did not use the COSMIN content validity checklist in this context, but used the checklist to inform the content validity evaluation tool co-created with PWLE.

3. Content validity evaluation tool

In step 3, content validity was rated by PWLE. The evaluation form to rate the content validity was co-designed with PWLE (i.e., the advisory group members) and Scottish clinicians (identified through Scottish professional networks). This co-development process ensured that the evaluation matched the aspects that PWLE find most relevant in an ACE questionnaire and are relevant in a clinical setting where ACEs are prevalent [18]. The co-designed evaluation form consisted of 10 items rating acceptability and the scope of each questionnaire using Likert scales and leaving space for score explanations (see Appendix 2). Six PWLE contributed to this evaluation, with each person rating 8-12 questionnaires. To account for differences between raters, each available questionnaire, regardless of their Cohen criteria ratings, was independently rated by two PWLE. For each item, a mean score across the two raters was calculated. Subsequently, these item mean scores were labelled, representing good (i.e., scores 4&5 for section 1; scores 2&3 for section 2), intermediate (i.e., scores 2&3; scores 1 & 4 for section 2), or poor (i.e., scores 0&1; scores 0&5 for section 2) content validity. Lastly, for each questionnaire, the percentage of good ratings across all content validity items was calculated (i.e., number of items labelled as good divided by 10 (the total number of items)*100).

Results

Figure 1 shows the PRISMA diagram. A total of 17,522 records were identified across the databases, with a final 112 included in the review. After removing 6,768 duplicates, 5,978 articles were excluded using the first title and abstract screening; 4,620 articles were excluded during the second title and abstract screening, targeting psychometric evaluations only. A further 79 records were removed during the full-text screening, which led to 77 included records. Their reference lists were screened, which added another 35 records. The Cohen's Kappa coefficient across the three raters for the abstracts screening ranged from 0.62 to 0.89, with the Cohen's Kappa coefficient between the two raters for the full text screening being 0.76.

- Insert Figure 1 about here -

Across the 112 included articles (see Appendix 3 for all references), 31 different ACE questionnaires were identified. Table 1 shows a summary of the study characteristics for each included questionnaire. The CTQ-Short Form (CTQ-SF) had the highest number ($n=37$) of publications. The psychometric qualities of the CTQ-SF were also investigated across the longest time period (i.e., 22

years) and across the largest number of countries (i.e., 19 different countries). Across all the questionnaires, the USA was the most common country for testing the psychometric quality, with less than half of all questionnaires tested in more than one country. A quarter of the questionnaires had psychometric evaluations in more than one language, with 13 languages being the most languages in which a questionnaire was evaluated (CTQ-SF). Populations evaluated fell within three broad categories of general, clinical, and special interest groups (such as the National Collegiate Athletic Association (NCAA) Division 1 college athletes), with more than a third of questionnaires tested in more than one population. Data collection methods included online forms, paper and pen as well as phone interviews, with only the Childhood Attachment and Relational Trauma Screen (CARTS) requiring purpose-built software. Sample sizes varied widely from 30 participants to 85,248.

- Insert Table 1 about here -

The details of the 31 questionnaires are described in detail in Table 2. Numbers of items range widely from 3 to 75 items. Three measures were not stand-alone questionnaires but amalgamated from items within wider population health surveys (HRS, MIDUS) or subscales of a more extensive questionnaire (Life Stressor Checklist Revised (LSC-R)). The focus of the questionnaires ranged widely, with abuse and neglect being the most assessed domains (shown in Appendix 4). A handful of questionnaires (e.g., the Questionnaire of Unpredictability in Childhood (QUIC), History of Social Punishment (HoSP)) had a very specific focus, leading to some domains, e.g., parental predictability or aversive control, only being assessed by one questionnaire. While three of the questionnaires (i.e., Trauma and Distress Scale (TADS), QUIC, and Childhood survey) included some positive items, the Benevolent Childhood Experiences (BCE) questionnaire was the only instrument solely focused on assessing the absence or presence of positive childhood experiences to infer ACE exposure. PWLE were explicitly involved in the development of only one questionnaire, the LSC-R.

- Insert Table 2 about here -

Nine questionnaires met Cohen's criteria for being "well-established" and two for "approaching well-established". Therefore, eleven questionnaires (evaluated across 91 papers) were assessed using the COSMIN checklist (see Table 3). No questionnaire had assessments of all measurement properties; with measurement error not addressed by any of them. The ACE Study questionnaire, the ACE International Questionnaire (ACE-IQ), and the CTQ-SF were most comprehensively assessed, with five properties being covered across the available articles. None of the questionnaires achieved a "sufficient rating" in all their assessed properties. The ACE questionnaire, the ACE-IQ and the Childhood Experiences Survey (CES) received the most positive ratings, with each having three

properties rated positively, whereas the CTQ-SF received ratings of “insufficient” across all five assessed properties.

With respect to the content validity as rated by PWLE, the average positive rating for the 27 questionnaires was 38% (see Table 3). The CES scored highest with 63% of items rated positively and the ICAST-R scored the lowest with only 11% positive ratings. The most well-rated aspects of content validity across all the questionnaires were providing participants the chance to reflect on their experience (64% of questionnaires received a positive rating) and the use of accessible language (59% of questionnaires received a positive rating).

- Insert Table 3 about here -

Discussion

Our comprehensive review identified a wide variety of questionnaires available to assess ACEs, but their psychometric quality was mixed and, overall, quite limited. Of the 11 questionnaires for which the psychometric qualities could be evaluated, not one was rated sufficient from a psychometric perspective. With over half of the included articles (60/112; 54%) published after August 2016, the cut-off date for the most recent available review [9], our findings provide an important update on outdated reviews evaluating the psychometric properties of ACE questionnaires. The ACE Study, ACE-IQ and CES questionnaire received the strongest positive psychometric evaluation, while the most frequently evaluated scale, the CTQ-SF, received the lowest psychometric quality rating.

Comparing the review’s findings with the most recent and comparable review, conducted by Saini and colleagues [9], reveals some important similarities and differences. With respect to similarities, both reviews identified a wide variability in the psychometrical quality of the identified instruments and a lack in comprehensive assessment of the COSMIN measurement properties across all instruments. Consequently, Saini and colleagues [9], concluded that there was no single instrument that is superior, and the choice of the instrument largely depends on the context and target population. Despite the lack of a superior instrument, Saini and colleagues [9] found moderate to strong evidence for many measurement properties of the CTQ (55%) and CTQ-SF (44%) and highlighted these instruments as the most thoroughly investigated, with the strongest level of evidence. While our review also identifies the CTQ-SF as the most frequently evaluated questionnaire, the evidence reveals insufficient support for all assessed measurement qualities (i.e., structural validity, internal consistency, cross-cultural validity, reliability, and criterion validity), which is in contrast with Saini and colleagues [9]. This discrepancy is likely due to newer evidence (with 14 articles on the CTQ-SF published since August

2016) downgrading the ratings in this review. Content validity ratings by the PWLE further supported concerns, with 33% negative and 20% positive ratings for the CTQ-SF. These inconsistencies are important given the CTQ-SF's widespread use in assessing ACE.

Unlike previous reviews, our review not only provides a rigorous assessment of measurement properties relevant for research purposes but is the first to provide crucial guidance to inform trauma-informed clinical practice by including ratings from PWLE on content validity and the Cohen's criteria aimed to guide the selection of evidence-based assessments. Table 2 provides a comprehensive overview of other relevant information to help guide clinicians in their decision of choosing an appropriate questionnaire for trauma-informed care. Strikingly, only one questionnaire (LSC-R) met our criterion for PWLE involvement in its development, which was reflected in the generally low ratings for content validity, as rated by PWLE, across all questionnaires. Based on all three evaluations, the ACE-IQ stood out as the most psychometrically sound, evidence-based and acceptable (as rated by PWLE) assessment. While the CES and Traumatic Experiences Checklist (TEC) also received high ratings for PWLE-rated content validity, the CES was less widely evaluated and there was mixed evidence with respect to the quality of the TEC measurement properties.

A unique aspect of this review is the content validity evaluation by PWLE, highlighting the lack of ACE questionnaires in reflecting the true diversity of ACEs and the need for scoring methods that better reflect the real-life impact of ACEs. The most common strategy to calculate the scores remains an unweighted summary approach, which assumes that all ACEs have an equal impact and confer equal risk. Previous reviews [8,11] and discussions with PWLE as part of this review process, identify that this approach does not reflect the personalised impact of various ACEs on peoples' lives.

Implementation of a more appropriate scoring mechanism to reflect the diverse impact ACEs can have on a person's life depending on the wider context in which the ACEs took place (i.e., other vulnerability and resilience factors within the child and their environment) is needed. Alternative approaches included weighted ACE models that account for differential impact as rated by the person, as well as theoretically or empirically driven approaches (e.g., machine learning-based predictive models or latent class analysis approaches), which aim to identify potentially different combinations of adversities and how these combinations impact the outcome under investigation [8]. However, none of these approaches have been rigorously and systematically evaluated and hence need to be comprehensively evaluated and compared to PWLE perspectives before implementation. Taken together, the evidence identifies an urgent need for standardised approaches to assessing ACEs, which demonstrate strong psychometric qualities and are co-developed with PWLE. Such a co-development approach would ensure that assessments 1) use sensitive, trauma-

informed and clear language, 2) reflect the diversity of ACEs (e.g., including cyberbullying, which was not specifically or directly covered by any of the included questionnaires) and 3) assesses all the relevant ACE domains, using accepted and agreed-upon definitions [12,19].

This review needs to be considered in the light of various limitations. Firstly, only articles published in English were included, and only those questionnaires with an available English version were evaluated on content validity. Furthermore, most included studies were conducted in North America. Indeed, only two of the included questionnaires have been developed in Non-Western countries (i.e., RACE Q (developed in Iran) and ICAST-R (developed across 7 different countries, including non-Western countries such as Russia and India (see Table 1 for details) and three have been validated in Non-Western countries (i.e., ACE-IQ, CTQ-SF and ETISR (See Table 1 for more details). Consequently, most of the included questionnaires had a strong Western perspective on ACEs, thereby limiting their generalizability to adverse experiences in non-Western cultures. Further exploration and validation of ACE questionnaires across different cultures is essential as cultural differences may influence how childhood adversity is defined, reported, and perceived. Indeed, ACE questionnaires developed in Western countries may not fully capture culturally specific adversities that might affect children in non-Western settings [20–22]. Furthermore, social and structural constructs, cultural taboos, stereotypes and stigma around ACEs may influence how and to what extent people are able to report ACEs [20,21]. Therefore, identifying culturally specific factors and ensuring ACE questionnaires cover all relevant events, using sensitive language, may support more comprehensive and authentic reporting of ACEs [21,23]. The involvement of PWLE from a diverse cultural background will be critical in developing or adapting ACE questionnaires with attention to culturally appropriate language and traditions [22,24]. Relatedly, while many studies evaluated the psychometric qualities in a sample spanning the entire adult lifespan, most participants were females, thereby raising concerns about the generalizability of psychometric evaluation and applicability of the ACE questionnaires in the male population. Secondly, the COSMIN checklist was modified to suit ACE questionnaires, as some of the COSMIN criteria were not applicable (i.e., responsiveness), or the analytical approaches used in the included articles differed from those recommended in the COSMIN criteria. Consequently, despite the detailed COSMIN manual, subjectivity of the decisions made in the rating process cannot be completely excluded.

Conclusion

Notwithstanding these limitations, our comprehensive approach identified that, despite a wide variety (31) of validated ACE self-report questionnaires being available, psychometric quality and content validity is lacking. There is an urgent need for a standardised, high-quality approach towards ACE assessment, co-developed with PWLE. The findings of this review represent a critical first step in this co-development process by identifying which aspects of existing questionnaires are appropriate and which aspects need more careful consideration. See Figure 2 for an infographic summarising the key aspects and conclusions of this review.

- Insert Figure 2 about here –

Statements

Acknowledgment

The authors of this study are members of the Advanced Pain Discovery Platform and the Consortium Against Pain Inequality (CAPE) studying the impact of adverse childhood experiences on pain and responses to treatment. Members of the CAPE advisory group provided input on the search strategy based on their insight through their lived experiences.

Statement of Ethics

Statement of Ethics is not applicable because this study is based exclusively on published literature.

Conflict of Interest Statement

We declare no competing interests.

Funding Sources

This study was supported by a UKRI, Lily and Versus Arthritis Grant: MR/W002566/1

Author Contributions

FM contributed to the screening process, conducted the data extraction, conducted the analyses for the COSMIN, Cohen and content validity ratings, completed Table 3 and took the lead on writing the Abstract, Methods and Results section.

AC conducted the updated searches, contributed to the screening process and data extraction of the updated searches, completed the Tables 1-2 and appendices and contributed to the writing of the article.

SS contributed to the screening process and the writing of the article.

TGH is the principal investigator of CAPE and provided input on all aspects of the review and contributed to the writing of the article.

JR took the lead on engaging PWLE in the entire process of the review, ensuring the appropriateness of the process to engage PWLE in the content validity evaluation and providing feedback on drafts of the article.

LAC co-supervised all aspects of the review, assisted in resolving conflicts during the screening process and contributed to the writing of the article.

LC co-supervised all aspects of the review, contributed to the development of the search strategy, the screening process, the data extraction, and COSMIN ratings and took the lead on gathering the content validity ratings of the PWLE and writing up the Introduction and Discussion.

Data Availability Statement

There was no new data collected for this study, and all the gathered data used in the narrative review is made directly available Tables 1 – 3 and appendices. Upon request from the corresponding/senior author (line.caes@stir.ac.uk) the complete data extraction file can be obtained.

References [Numerical]

- 408 1 World Health Organization (WHO). Child Maltreatment [Internet]. 2022 Available from:
409 <https://apps.who.int/violence-info/child-maltreatment>
- 410 2 Madigan S, Deneault A, Racine N, Park J, Thiemann R, Zhu J, et al. Adverse childhood
411 experiences: a meta-analysis of prevalence and moderators among half a million adults in 206
412 studies. *World psychiatry*. 2023;22(3):463–71.
- 413 3 Felitti VJ, Anda RF, Nordenberg D, Williamson DF, Spitz AM, Edwards V, et al. Relationship of
414 childhood abuse and household dysfunction to many of the leading causes of death in adults:
415 The Adverse Childhood Experiences (ACE) Study. *Am J Prev Med*. 1998;14(4):245–58.
- 416 4 Hughes K, Bellis MA, Hardcastle KA, Sethi D, Butchart A, Mikton C, et al. The effect of multiple
417 adverse childhood experiences on health: a systematic review and meta-analysis. *Lancet*
418 *public Heal*. 2017;2(8):e356–66.
- 419 5 Senaratne DNS, Thakkar B, Smith BH, Hales TG, Marryat L, Colvin LA. The impact of adverse
420 childhood experiences on multimorbidity: a systematic review and meta-analysis. *BMC Med*.
421 2024;22(1):315.
- 422 6 Nelson CA, Bhutta ZA, Harris NB, Danese A, Samara M. Adversity in childhood is linked to
423 mental and physical health throughout life. *bmj*. 2020;371.
- 424 7 Pournaghash-Tehrani SS, Zamanian H, Amini-Tehrani M. The impact of relational adverse
425 childhood experiences on suicide outcomes during early and young adulthood. *J Interpers*
426 *Violence*. 2021;36(17–18):8627–51.
- 427 8 Lacey RE, Minnis H. Practitioner review: twenty years of research with adverse childhood
428 experience scores—advantages, disadvantages and applications to practice. *J Child Psychol*
429 *Psychiatry*. 2020;61(2):116–30.
- 430 9 Saini SM, Hoffmann CR, Pantelis C, Everall IP, Bousman CA. Systematic review and critical
431 appraisal of child abuse measurement instruments. *Psychiatry Res*. 2019;272:106–13.
- 432 10 Bernstein DP, Fink L, Handelsman L, Foote J, Lovejoy M, Wenzel K, et al. Initial reliability and
433 validity of a new retrospective measure of child abuse and neglect. *Am J Psychiatry*.
434 1994;151(8):1132–6.
- 435 11 Appleton AA, Holdsworth E, Ryan M, Tracy M. Measuring childhood adversity in life course
436 cardiovascular research: a systematic review. *Psychosom Med*. 2017;79(4):434–40.

- 437 12 Karatekin C, Mason SM, Riegelman A, Bakker C, Hunt S, Gresham B, et al. Adverse childhood
438 experiences: A scoping review of measures and methods. *Child Youth Serv Rev.*
439 2022;136:106425.
- 440 13 Nelson SM, Cunningham NR, Kashikar-Zuck S. A conceptual framework for understanding the
441 role of adverse childhood experiences in pediatric chronic pain. *Clin J Pain.* 2017;33(3):264–
442 70.
- 443 14 Staniszewska S, Haywood KL, Brett J, Tutton L. Patient and public involvement in patient-
444 reported outcome measures: evolution not revolution. *Patient-Patient-Centered Outcomes*
445 *Res.* 2012;5:79–87.
- 446 15 Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA
447 2020 statement: an updated guideline for reporting systematic reviews. *bmj.* 2021;372.
- 448 16 Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia
449 [Internet]Available from: www.covidence.org
- 450 17 Cohen LL, La Greca AM, Blount RL, Kazak AE, Holmbeck GN, Lemanek KL. Introduction to
451 Special Issue: Evidence-based Assessment in Pediatric Psychology. *J Pediatr Psychol.* 2008
452 Oct;33(9):911–5.
- 453 18 Mokkink LB, Prinsen C, Patrick DL, Alonso J, Bouter L, De Vet HC, et al. COSMIN methodology
454 for systematic reviews of patient-reported outcome measures (PROMs). *User Man.*
455 2018;78(1):3–6.
- 456 19 Krinner LM, Warren-Findlow J, Bowling J, Issel LM, Reeve CL. The dimensionality of adverse
457 childhood experiences: a scoping review of ACE dimensions measurement. *Child Abuse Negl.*
458 2021;121:105270.
- 459 20 Baumont M, Wandasari W, Agastya NLPM, Findley S, Kusumaningrum S. Understanding
460 childhood adversity in West Sulawesi, Indonesia. *Child Abuse Negl.* 2020;107:104533.
- 461 21 Cluver L, Orkin M, Boyes ME, Sherr L. Child and adolescent suicide attempts, suicidal behavior,
462 and adverse childhood experiences in South Africa: A prospective study. *J Adolesc Heal.*
463 2015;57(1):52–9.
- 464 22 Mendel WE, Sperlich M, Fava NM. “Is there anything else you would like me to know?”:
465 Applying a trauma-informed approach to the administration of the adverse childhood
466 experiences questionnaire. *J Community Psychol.* 2021;49(5):1079–99.
- 467 23 Behr D, Shishido K. The translation of measurement instruments for cross-cultural surveys.

- 468 SAGE Handb Surv Methodol. 2016;55:269–87.
- 469 24 Vu TH, Bishop J, McGill L, Valmadrid L, Golden S, Emmerling D, et al. Using systems-mapping
470 to address Adverse Childhood Experiences (ACEs) and trauma: A qualitative study of
471 stakeholder experiences. PLoS One. 2022;17(8):e0273361.
- 472

Figure Legends

Fig 1. PRISMA diagram

Fig 2. Infographic