Measurement properties of tools used to assess self-harm in autistic and general population adults

Newell, Victoria; Townsend, Ellen; Richards, Caroline; Cassidy, Sarah

Review

Measurement properties of tools used to assess self-harm in autistic and general population adults

Victoria Newell [a,*], Ellen Townsend [a], Caroline Richards [b], Sarah Cassidy [a]

[a] School of Psychology, University of Nottingham, University Park, Nottingham NG7 2RD, United Kingdom
[b] School of Psychology, University of Birmingham, 52 Pritchatts Road, Birmingham B15 2TT, United Kingdom

ARTICLE INFO

ABSTRACT

Autistic people are at increased risk of experiencing self-harm compared to the general population. However, it is unclear which tools are being used to assess self-harm in autistic people, or whether existing tools need to be adapted for this group. This two-stage systematic review aimed to identify tools used to assess self-harm in autistic and general population adults, evaluate these tools on their measurement properties, and make recommendations for their appropriate use in research and clinical practice. Four databases were systematically searched (PsycINFO, Embase, MEDLINE and Web of Science). Eight frequently used self-harm assessment tools were identified and assessed for risk of bias, criteria for good measurement properties, and quality of evidence using the COSMIN checklist. Of these, two tools had sufficient evidence of internal consistency (ISAS, QNSSI), and one had been frequently used with autistic adults (NSSI-AT). These three tools may have potential for use with autistic adults but require further investigation for content validity and measurement properties in the autistic population. More research and potential adaptations to current self-harm assessment tools are recommended in order to better conceptualise and understand self-harm and its measurement in autism.

Self-harm encompasses any act of intentional self-poisoning or self-injury irrespective of its apparent purpose (National Institute for Health and Care Excellence, 2022). Methods of self-harm can include, but are not limited to, self-cutting, self-hitting, self-burning, self-biting, and skin picking (Klonsky, 2011; Pompili et al., 2015). Non-suicidal self-injury (NSSI) is a form of self-harm where socially unsanctioned and deliberate harm is inflicted to the self without intent to end life (Nock & Favazza, 2009). The Four-function Model (Nock, 2009; Nock & Prinstein, 2004) proposes that NSSI serves either an interpersonal (e.g., to influence the behaviour of others or communicate/ express distress) or intrapersonal function (e.g., as sensory stimulation or emotion regulation), which in turn either negatively or positively reinforces one's internal affective state. Conversely, a broader definition of self-harm would include suicide attempts, where intent to end life is present (Hawton, Saunders, & O'Connor, 2012). There are debates in the literature over how best to conceptualise self-harm, such as whether self-harm with suicidal intent is distinct from self-harm without suicidal intent, or if these behaviours fall along a continuum (Hargus, Hawton, & Rodham, 2009; Kapur, Cooper, O'Connor, & Hawton, 2013). Yet

irrespective of intent, self-harm represents one of the strongest risk factors for future suicide in the general population (Favril, Yu, Uyar, Sharpe, & Fazel, 2022; Hawton et al., 2020; Hawton & Harriss, 2007; Ribeiro et al., 2016; Runeson, Haglund, Lichtenstein, & Tidemalm, 2016), along with an increased risk of other adverse outcomes such as substance misuse, subsequent mental health problems, and unemployment (Moran et al., 2015; Beckman, Lysell, Haglund, & Dahlin, 2019; Ohlis et al., 2020). It is therefore imperative these behaviours receive clinical and research attention.

Concerningly, people diagnosed with Autism Spectrum Conditions (ASC; hereafter also referred to as autistic people[1]) experience high levels of self-harm across the lifespan (Akram, Tariq, & Rafi, 2017; Licence, Oliver, Moss, & Richards, 2020), are at a greater risk of self-harming compared to non-autistic people (Cassidy, Bradley, Shaw and Baron-Cohen, 2018c; Blanchard, Chihuri, DiGuiseppi, & Li, 2021; Maddox, Trubanova, & White, 2017; Widnall et al., 2022), and adults with high autistic traits are more likely to self-harm than those with low autistic traits (Stewart et al., 2022). Moreover, approximately 50% of autistic people report engaging in at least one episode of NSSI in their

---

*   Corresponding author.
    E-mail address: victoria.newell@nottingham.ac.uk (V. Newell).
[1] Identity first language is used throughout this paper (e.g., autistic community/ person/ individual), as the preferred language of the autistic community (Kenny et al., 2016).

lifetime (Maddox et al., 2017). NSSI is also associated with suicidality (suicidal ideation, suicide plans, suicide attempts, and death by suicide) in autistic adults, and suicidality is observed more frequently in those who engage in specific methods of NSSI (i.e., cutting; Moseley, Gregory, Smith, Allison, & Baron-Cohen, 2020; Cassidy, Bradley, Shaw and Baron-Cohen, 2018c). Given the relationship of self-harm with negative outcomes, and that suicidality and mental health problems are already more prevalent in autistic people (Lai et al., 2019; Newell et al., 2023), effective identification of self-harm is crucial to inform treatment and prevention strategies. However, there is a lack of evidence regarding the availability and appropriacy of valid tools to assess self-harm in this population.

Certain characteristics of autism may present challenges to the identification of self-harm and the function it serves. Autistic people are characterised by differences in social communication and interaction, sensory processing differences, specific focused interests, and preference for routine and familiarity (American Psychiatric Association, 2013). Notably, self-injurious behaviours (SIB; e.g., head banging, biting, scratching, hair-pulling) are prevalent in 42% of autistic people (Steenfeldt-Kristensen, Jones, & Richards, 2020) but tend to be differentiated from conceptualisations of self-harm observed in non-autistic people (e.g., NSSI). Rather, SIB in autism is traditionally conceptualised as a challenging and/or repetitive behaviour (Duerden et al., 2012; South, Ozonoff, & McMahon, 2005), observed more in childhood or alongside co-occurring intellectual disability (ID; Minshawi et al., 2014). Self-harm behaviours may therefore be inappropriately attributed as an "SIB characteristic of autism" (Maddox et al., 2017). This is particularly problematic for autistic people without co-occurring ID, who are found to engage in self-harm similarly to non-autistic people in terms of age of onset, methods used, and functions (Maddox et al., 2017; Moseley, Gregory, Smith, Allison, & Baron-Cohen, 2019). There is also no research on whether this extends to autistic individuals with co-occurring ID, or whether the functions of self-harm are distinct in autistic people with versus without co-occurring ID (Steenfeldt-Kristensen et al., 2020). A lack of clear delineation in the associated terminology (Claes & Vandereycken, 2007) does not help matters, with the terms SIB and self-harm used interchangeably in the literature. Thus, research is needed to further understand and conceptualise self-harm and its function in autistic people with and without co-occurring ID. Robust measurement tools are required in order to progress this understanding, both scientifically and clinically.

In addition to how self-harm is defined, characteristics unique to autistic people may also limit the utility of existing self-harm assessment tools. Previous research adapting measurement tools for suicidality and depression have demonstrated autistic adults interpret and respond to questionnaires designed for the general population differently to the interpretation intended by the tool designers (Cassidy et al., 2020; Cassidy, Bradley, Cogger-Ward, & Rodgers, 2021). For example, challenges in identifying, understanding, and describing one's emotions (i. e., alexithymia) are common for autistic people (Kinnaird, Stewart, & Tchanturia, 2019). However, many self-harm assessment tools rely on self-report and the ability to identify internal emotional experiences (e. g., "I hurt myself… to deal with anger/ to cope with uncomfortable feelings" Non-suicidal Self-Injury – Assessment Tool [NSSI-AT]; Whitlock, Exner-Cortens, & Purington, 2014). Alexithymia is suggested to be associated with NSSI in autistic people, with NSSI functioning to regulate high-energy states (e.g., aggression, anxiety, anger; Moseley et al., 2020). Such individuals may therefore be more inclined to self-harm but could find it challenging to communicate their emotional distress and report this via a measurement tool.

Autistic people also tend to interpret information or questions literally (Happé, 1995) and may struggle with complex or abstract language and figures of speech. An example of this is the item "to get a vacation from having to try so hard" as a response to expectations connected to self-harm (Questionnaire for Non-Suicidal Self-Injury [QNSSI]; Turner, Chapman, & Layden, 2012). Moreover, autistic people can have

difficulties recalling what has happened in the past (episodic memory) or imagining what might happen in the future (episodic future thinking; Lind & Bowler, 2010; Lind, Williams, Bowler, & Peel, 2014) which may impact on their ability to answer questions about lifetime and future engagement in self-harm (e.g., "How many times in your life have you engaged in NSSI?" or "On a scale of 0 to 4, what do you think the likelihood is that you will engage in NSSI in the future?" Self-injurious Thoughts and Behaviours Interview [SITBI]; Nock, Holmberg, Photos, & Michel, 2007). Previous research on the measurement properties of the Suicidal Behaviours Questionnaire – Revised (SBQ-R; Cassidy et al., 2020) demonstrated that autistic participants found a similar question regarding future suicide intent "impossible to answer" for this reason. Therefore, the mode of assessment and language used in traditional self-harm assessment tools may be less accessible and appropriate for autistic people.

There may also be aspects of self-harm in autism which are not captured in assessment tools originally designed for other populations, such as sensory processing differences. For autistic people, sensory inputs across multiple modalities can either be experienced more intensely (hyperreactivity); not noticed at all or responded to in a delayed manner (hyporeactivity); or engaged with repeatedly and for sustained periods of time (American Psychiatric Association, 2013; Lane, 2002). Consequently, certain sensory disturbances can induce high levels of distress for an autistic person (Robledo, Donnellan, & Strandt-Conroy, 2012; Robertson & Simmons, 2015; MacLennan, O'Brien, & Tavassoli, 2021). A mixed methods study found autistic people reported using NSSI as a way to deal with overwhelm from sensory input, and that sensory differences predicted body areas targeted, lifetime engagement and frequency of NSSI (Moseley et al., 2019). Worryingly, autistic individuals with hyporeactivity might engage in self-harm at a higher frequency or severity, leading to injuries more serious than intended (Moseley et al., 2019). Self-harm measures that are used in autistic populations therefore need to be able to adequately capture functions of self-harm that are unique to autistic people.

Given the risks of failing to adequately identify and understand the function of self-harm in autism, it is extremely important to examine which self-harm assessments have been utilised in this group, whether they are validated to do so, and if not, to identify which is the most robust candidate tool to adapt for this population. Autistic adults already face significant barriers to accessing and engaging with mental health support (Brede et al., 2022) and report higher levels of unmet support needs compared to non-autistic adults (Nicolaidis et al., 2013) and autistic children (Turcotte, Mathew, Shea, Brusilovskiy, & Nonnemacher, 2016). Moreover, self-harm assessment tools developed for adults may not be valid for use with youth or vice versa. Therefore, we will focus on self-harm assessment tools used in adult samples.

To address the research aim; the current review will utilise COSMIN (the Consensus-based Standards for the Selection of Health Measurement Instruments; Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018), which is a validated research tool developed to assess the measurement properties of patient reported outcome measures. A growing number of previous COSMIN reviews have highlighted a distinct lack of evidence for the measurement properties of outcome measures in autistic people for constructs such as depression, suicidality, and stress (Cassidy, Bradley, Bowen, et al., 2018a; Cassidy, Bradley, Bowen, Wigham, & Rodgers, 2018b; Thoen, Steyaert, Alaerts, Evers, & Van Damme, 2021). The COSMIN method firstly involves a systematic search of the literature to identify which tools have been used to assess a pre-specified health outcome in a defined population. The most frequently used tools (where frequency is defined as being used at least twice), which have some evidence of validity (i.e., reference to a previously published study assessing its measurement properties), are then subjected to a second stage of review. In this second stage, the selected measures are searched for using a comprehensive filter validated to find studies assessing their measurement properties (Terwee, Jansma, Riphagen, & de Vet, 2009), and are subsequently rated using the

COSMIN checklist. COSMIN therefore allows for a systematic assessment of risk of bias, measurement properties and quality of evidence in frequently used assessment tools, in order to make evidence-based recommendations on the appropriateness of these for particular research and clinical contexts (Prinsen et al., 2018; Mokkink et al., 2018; Terwee et al., 2018).

Employing this robust method, the current review is split into two stages. The aims of the first stage are to: 1) identify what tools are frequently used to assess self-harm in a) autistic adults, and b) general population adults, in research and clinical practice. The aim of the second stage is to: 2) evaluate the evidence regarding the appropriateness and measurement properties of the previously identified tools used to assess self-harm in autistic and general population adults using the COSMIN checklist. This review will enable a comparison of the most frequently used self-harm assessment tools and their measurement properties within each group. Should none, or few tools be available which have been used with autistic adults, we can instead identify the most robust tool used in general population adults that has the potential to be adapted. Finally, from the synthesis of evidence for self-harm assessment tools and their measurement properties, we aim to 3) make recommendations for the effective assessment of self-harm in these groups.

1. Review methods: Stage 1

The protocol for this review is registered with the International Prospective Register of Systematic Reviews (Registration number: CRD42022352501) and can be accessed online (https://www.crd.york. ac.uk/prospero/display_record.php?RecordID=352501). This systematic review follows the guidelines for Preferred Reporting Items for Systematic Reviews and Meta-Analyses standards (PRISMA; Page et al., 2021).

1.1. Search strategy

Four electronic databases were systematically examined (Embase, PsycINFO, MEDLINE and Web of Science) using two search engines (PubMed and OVID) from inception to August 9, 2023. The Cochrane Database of Systematic Reviews was also searched to confirm that no other systematic reviews of the current study topic already existed. Two separate searches were carried out in stage 1 for measures of self-harm used in: a) autistic adults with or without co-occurring ID, and then for b) general population adults without any co-occurring conditions or ID. We chose to include autistic adults with co-occurring ID in the search, due to the prevalence of SIB in this group (Minshawi et al., 2014), which may highlight additional relevant measures. Search terms (see Table 1) were derived from similar recent COSMIN reviews (Cassidy, Bradley, Bowen, et al., 2018a; Cassidy, Bradley, Bowen, et al., 2018b) and were adapted to fit the specific search criteria of each database. Searches were restricted to human research, with articles published in the English Language.

1.2. Selection criteria

A standardized approach using a well-defined group and outcome was employed for study selection, similar to that used in previous COSMIN reviews (Cassidy, Bradley, Bowen, et al., 2018a; Cassidy, Bradley, Bowen, et al., 2018b). Studies had to focus on a tool specifically used to assess self-harm, defined by the NICE guideline [NG225] for self-harm: assessment, management and preventing recurrence (National Institute for Health and Care Excellence, 2022). As a result, tools could use either a specific (i.e., NSSI, SIB) or broader conceptualisation of self-harm (irrespective of intent, e.g., including suicide attempts). However, those using a broader conceptualisation were required to cover self-harm both with and without suicidal intent, as a suicidality assessment tool has already been adapted for autistic adults (Cassidy et al., 2020). Studies utilising a single self-harm related question, item(s) or subscale contained within a more general measure (e.g., Mini-International Neuropsychiatric Interview (MINI; Sheehan et al., 1998) were not eligible for inclusion. This was because single items or subscales were less likely to capture the construct of self-harm compared to a standalone measure, given its complex nature (Townsend et al., 2016). Additionally, we required that tools had been used in at least two studies for consideration in the next stage, so those that were created ad hoc (i. e., for the purpose of the current study), or without evidence of validity were excluded. This was necessary to maximise the probability that identified tools would have evidence of the measurement properties necessary for stage 2.

Eligible studies were required to utilise tools assessing either the prevalence of self-harm (e.g., epidemiological/ population studies) or assessing self-harm as an outcome (e.g., treatment/ intervention and longitudinal/ cohort studies). Studies also needed to focus on adults (aged 18 years and over). If the age range was partly outside this, studies were only included if 50% or more of the total sample was over 18 years old or the mean age of the sample was 18 years and above. In line with previous similar COSMIN reviews, this was to ensure that the tools identified were likely to be appropriate for adults (Cassidy et al., 2018b; Cassidy et al., 2018c). Where a study used had a tool adapted for a specific population (e.g., older adults, a particular gender, or a specific culture), it was excluded. This was so the tool would be more widely applicable to assess self-harm in autistic or general populations, rather than just a narrow subgroup.

1.3. General population adult search criteria

Studies were included if focusing on general population adults (i.e., non-clinical, without any co-occurring conditions or ID). Data for the general population were required to be presented separately from any other population(s), and to comprise of at least 50% or more of the sample. Studies with an autistic comparison group were excluded and instead considered for inclusion in the ASC search.

Table 1
Stage 1 review search terms.

1. ("General population" or "general public" or "population sampl*"or "community sampl*"or "national sampl*"or "national survey" or "household survey" or "non referred" or nonreferred or "non clinical" or nonclinical or "population screen*")
2. ("Autis* spectrum*"or ASC or ASD or asperg* or autis* or "pervasive developmental disorder*"or PDD or "unspecified PDD" or PDD-NOS)
3. (adult* or "young adult*"or "middle-aged" or "old* adult*"or elder*)
4. (Assess* or measur* or test* or tool* or "treatment outcome*"or scale* or survey or screen* or questionnaire* or quotient* or inventor* or instrument* or interview* or checklist* or index* or indices)
5. ("Self harm*"or selfharm* or "self injur*"or selfinjur* or "non suicid*" self injur*"or
6. "nonsuicid* self injur*"or NSSI or "self mutilat*"or "deliberate harm*"or "deliberate selfharm*"or "deliberate self harm*"or DSH or "self inflict*"or selfinflict* or "self cut*"or selfcut* or "self poison*"or selfpoison* or "self destruct*"or selfdestruct* or parasuicid* or "para suicid*")
7. General Population Search (1 AND 3 AND 4 AND 5)
8. Autism Spectrum Condition Search (2 AND 4 AND 5)
9. Limit 6 and 7 to English Language

* Denotes wildcard search terms.

1.4. Autistic adults search criteria

Studies were included if focusing on autistic adults with or without co-occurring ID (IQ below 70), where data for the autistic group were presented separately, and 50% or more of the sample was comprised of individuals with a diagnosis of ASC.

1.5. Screening and data extraction

As independent searches were conducted for studies of general population adults and autistic adults, the respective results were screened separately. Duplicates were removed prior to screening. The first author (VN) screened the titles and abstracts of articles for inclusion. Where there was insufficient information at screening on whether an article should be read in full, it was included. VN then conducted a full text screen of the remaining articles. To reduce the risk of researcher bias, 25% of papers at both stages were also screened by an independent reviewer (NH) and inter-rater reliability was calculated. Inter-rater agreement was almost perfect for the title and abstract screen (PABAK = 0.83, 91.4%) of articles with autistic adults, and substantial (PABAK = 0.72, 86.1%) of articles with general population adults. Inter-rater agreement was substantial for the full-text screen of both articles with autistic adults (PABAK = 0.78, 88.9%) and general population adults (PABAK = 0.61, 80.4%). All discrepancies were discussed to reach a consensus, but where this could not be resolved, the opinion of a third reviewer was sought (SC, CR). Data extraction of assessment tool characteristics was performed by VN (i.e., original authors, year published, assessment aim, mode of administration, number of items, subscales, and response options).

2. Results: Stage 1

2.1. Autistic adults

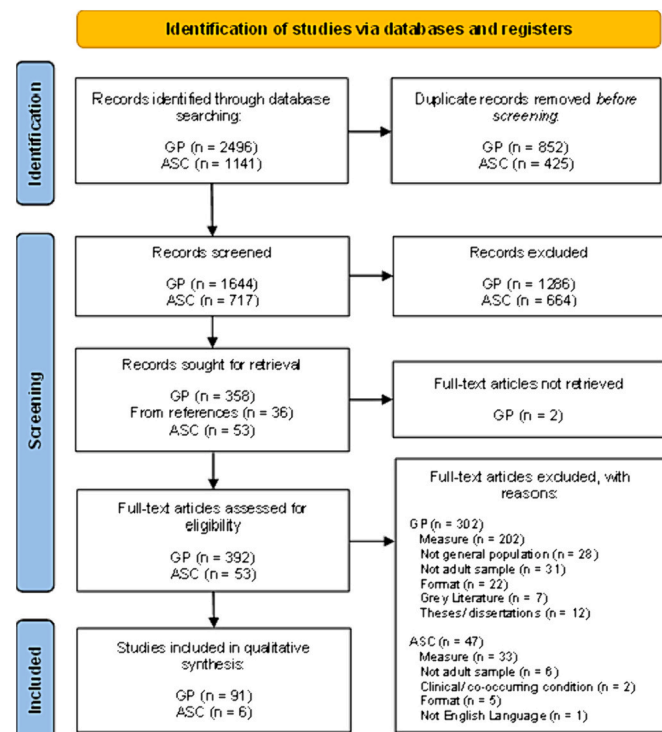The search for studies using tools to assess self-harm in autistic adults

Fig. 1. PRISMA flow diagram for the selection of articles for general and autistic population searches.

with or without co-occurring ID identified 717 articles which were screened, where six of these articles were retained for analysis (Fig. 1). These six studies all utilised a cross-sectional design to explore forms of self-harm (NSSI or SIB) in autistic adults and included between 42 and 314 autistic participants. Two tools were used to assess self-harm across the six studies: the NSSI-AT (Whitlock et al., 2014) and the Inventory for the functional assessment of self-injurious behaviour (IfES; Bienstein & Nußbeck, 2010). However, the IfES was not considered for stage 2 as it had only been used in one study of autistic adults with co-occurring ID and the tool was not available in English.

2.2. General population adults

The search for studies using tools to assess self-harm in general population adults identified 1644 articles which were screened, 91 of which were retained for analysis (Fig. 1). The majority of studies utilised a cross-sectional design ($n = 74$, 81.32%), and sample sizes ranged from 97 to 11,529 general population adults. Within these, fourteen different tools were used to assess self-harm. Self-report questionnaires included the: Body-focused Self-damaging Behaviour Expectancies Questionnaire (BSBEQ; Forbes, Tull, Lavender, Dixon-Gordon, & Gratz, 2022); Cardiff Self-Injury Inventory (CSII; Snowden, Tiley, & Gray, 2023) Deliberate Self-Harm Inventory (DSHI; Gratz, 2001); Functional Assessment of Self-Mutilation (FASM; Lloyd, 1997); Inventory of Statements about Self-Injury (ISAS; Klonsky & Glenn, 2009); NSSI-AT (Whitlock et al., 2014); QNSSI (Turner et al., 2012); Self-Harm Inventory (SHI; Sansone, Wiederman, & Sansone, 1998); Self-Injury Questionnaire – Treatment Related (SIQ-TR; Claes & Vandereycken, 2007b); The Ottawa Self-Injury Inventory (OSI; Martin et al., 2013); The Repetitive Non-Suicidal Self-Injury Questionnaire (R-NSSI-Q; Manca, 2009); and The Self-Harm Behaviour Questionnaire (SHBQ; Gutierrez, 1998). The search also identified two clinician interviews, the: Clinician-Administered Non-suicidal Self-injury Disorder Index (CANDI; Gratz, Dixon-Gordon, Chapman, & Tull, 2015); and SITBI (Nock et al., 2007).

Five of these tools had only been used in one study of general population adults without co-occurring conditions (BSBEQ; CANDI; OSI; R-NSSI-Q; SHBQ) and were not considered further.

3. Summary

Overall, nine tools were identified that had been used frequently in either general population or autistic adults with some evidence of validity (CSII; DSHI; ISAS; FASM; NSSI-AT; QNSSI; SHI; SITBI; SIQ-TR). From the six studies with autistic adults, a single self-harm assessment tool (NSSI-AT) had been used frequently. However, the NSSI-AT had not been developed or validated for autistic people.

4. Review methods: Stage 2

This stage of the review searched specifically for evidence of the measurement properties of the tools identified in stage 1. Studies on measurement properties are recognised as difficult to find due to poor indexing, variations in terminology, and poor reporting (Terwee et al., 2009). PubMed is the only database with a filter designed and validated specifically to identify studies assessing the measurement properties of a health outcome assessment tool (Terwee et al., 2009).

Consequently, a second comprehensive search was conducted from inception to August 26, 2023 using the PubMed database and supplemented with handsearching references of analogous reviews. Similarly, to stage 1, eligible studies needed to explore the measurement properties of these tools in adults (autistic or general population) and were only included if 50% or more of the total sample was over 18 years old or the mean age of the sample was 18 years and above. However, studies of adults with co-occurring conditions were also considered for stage 2. This was because self-harm assessment tools used frequently in the general population may also be validated in clinical samples and could

highlight additional contexts where the application of such tools would be useful. As in stage 1, 25% of papers at both title and abstract and full-text screen were checked by an independent reviewer (SC). Inter-rater reliability was almost perfect for the title and abstract screen (PABAK = 0.83, 91.7%) and substantial for the full-text screen (PABAK = 0.78, 88.9%). Any discrepancies were resolved through discussion and the opinion of a third reviewer was sought if necessary (CR).

4.1. Data extraction method

The COSMIN (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018) risk of bias checklist was used to assess the methodological quality of each study identified from the search. COSMIN rates evidence against each of the following 9 measurement properties: 1) content validity – the degree to which the content of a tool is an adequate reflection of the construct to be measured; 2) structural validity – the degree to which the scores of a tool are an adequate reflection of a constructs' dimensionality; 3) internal consistency – the degree of interrelatedness among the items; 4) reliability – the proportion of total variance in a measurements' 'true' differences between participants; 5) measurement error – the systematic and random error of a participant's score not attributed to true changes in the construct; 6) hypothesis testing for construct validity[2] – the degree to which the scores of a tool are consistent with hypotheses[3]; 7) cross-cultural validity/ measurement invariance – the degree to which items on a translated or adapted (e.g., culturally, different groups) tool adequately reflect the performance of items from the original version; (8) criterion validity – the degree to which the scores of a tool are an adequate reflection of a 'gold standard'; and (9) responsiveness – the ability of a tool to detect change over time in the construct measured (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018). Each study was rated on these measurement properties using a 4-point system of 'very good', 'adequate', 'doubtful' or 'inadequate' quality. The overall quality for every measurement property was then determined from a "worst score counts" principle, taking the lowest score provided. For example, if any criteria for reliability was scored as 'inadequate', the overall methodological quality of reliability for that specific study would then be 'inadequate'. Structural validity and internal consistency were not rated for behaviour-specific measures (CSII, ISASI I subscale, DSHI, SHI); as the presence or forms of self-harm do not reflect an underlying latent construct. These measurement properties would hold little theoretical or empirical meaning.

The quantitative findings of each single study on a measurement property was also rated against the updated criteria for good measurement properties (see Table A1), and scored as sufficient (+; in support of the measurement property), insufficient (−; evidence against the measurement property) or indeterminate (?; not possible to deduce whether the evidence is for or against the measurement property).

Checklists were completed by VN, and 20% of the articles were rated for risk of bias by an independent reviewer (SC) trained and experienced in using COSMIN (Cassidy et al., 2018b; Cassidy et al., 2018c). Inter-rater agreement was substantial for the risk of bias checklist (PABAK = 0.71, 85.7%), similar to that of previous COSMIN reviews (Cassidy, Bradley, Bowen, et al., 2018a; Cassidy, Bradley, Bowen, et al., 2018b). Any disagreements were resolved through discussion, and agreed ratings were utilised in the subsequent evidence synthesis. As recommended by COSMIN, data were extracted on the characteristics of eligible self-harm assessment tools, included sample(s), and the results for measurement properties. Information about interpretability and feasibility of the score (s) of the eligible tools was also qualitatively summarised, where

interpretability refers to how easily meaning can be derived from the tools score(s), and feasibility, meaning how easily the tool can be applied in its intended setting (Mokkink et al., 2018).

4.2. Evidence synthesis

This step focused on the quality of each tool as a whole by qualitatively summarising the results of all available studies per measurement property (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018). The summarised results were again compared against the criteria for good measurement properties to determine whether the overall measurement property of each tool was sufficient (+), insufficient (−), inconsistent (±) or indeterminate (?). To be classed as sufficient or insufficient, 75% of the results should meet the corresponding criteria, e. g., at least 75% of results should be in accordance with hypotheses to rate the overall result for hypothesis testing as 'sufficient', versus 75% not in accordance with hypotheses for 'insufficient'. If results of single studies were inconsistent and the inconsistency was unexplained, the overall result would be 'inconsistent', likewise, if results per study were all indeterminate, the overall rating would also be 'indeterminate' (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018).

A modified version of the Grading of Recommendations Assessment, Development and Evaluation (GRADE; Schunemann, Brożek, Guyatt, & Oxman, 2013) system was then used to evaluate the quality of the summarised evidence per measurement property per tool (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018). This was assessed by four factors: 1) risk of bias – the methodological quality of studies; 2) inconsistency – unexplained inconsistency of results across studies; 3) imprecision – total sample size of the available studies; and 4) indirectness – evidence from different populations other than the defined population of interest. However, due to the stringent inclusion criteria, indirectness was not considered applicable to the current review and therefore not evaluated. Factors 1 to 3 were graded as high, moderate, low, or very low evidence to reflect the confidence in the true measurement property being reflected by that of the summarised result. As recommended by COSMIN, grading was completed both by VN and independently by a second reviewer (SC), where inter-rater agreement was substantial (PABAK = 0.61, 81.0%).

The final step of the COSMIN method was to formulate recommendations on the most suitable assessment tool for the construct of interest and study population (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018). These were categorised as: A) assessment tools with evidence for sufficient content validity (any level) AND at least low-quality evidence for sufficient internal consistency, B) assessment tools categorised not in A) or C), or C) assessment tools with high quality evidence for an insufficient measurement property. Those categorised as A can be recommended for use (or adaptation), B have the potential to be recommended for use but require further research to assess quality, and C are not recommended for use (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018).

5. Results: stage 2

The PubMed search for studies assessing the measurement properties of self-harm assessment tools identified 394 articles that were eligible for title and abstract screening. Of these, 44 were screened in full, and 19 were retained for analysis (Fig. 2). No studies were identified which assessed measurement properties of any of the self-harm assessment tools in autistic samples, nor the measurement properties of the FASM in adult samples. Table A3 shows the characteristics of the study populations included in the analysis, and Table A4 describes the characteristics of each of the self-harm assessment tools that were evaluated.

The methodological quality of the included studies were rated using the risk of bias checklist and criteria for good measurement properties. However, none had explored content validity, measurement error, cross cultural validity/ measurement invariance, or responsiveness, and so

---

[2] Hypotheses to evaluate construct validity were formulated a-priori by the review team based on those in Prinsen et al., (2018), see Table A2.

[3] For the hypothesis testing measurement property, (a) refers to convergent or divergent relationships with other instruments, and (b) to discriminant differences between relevant groups.
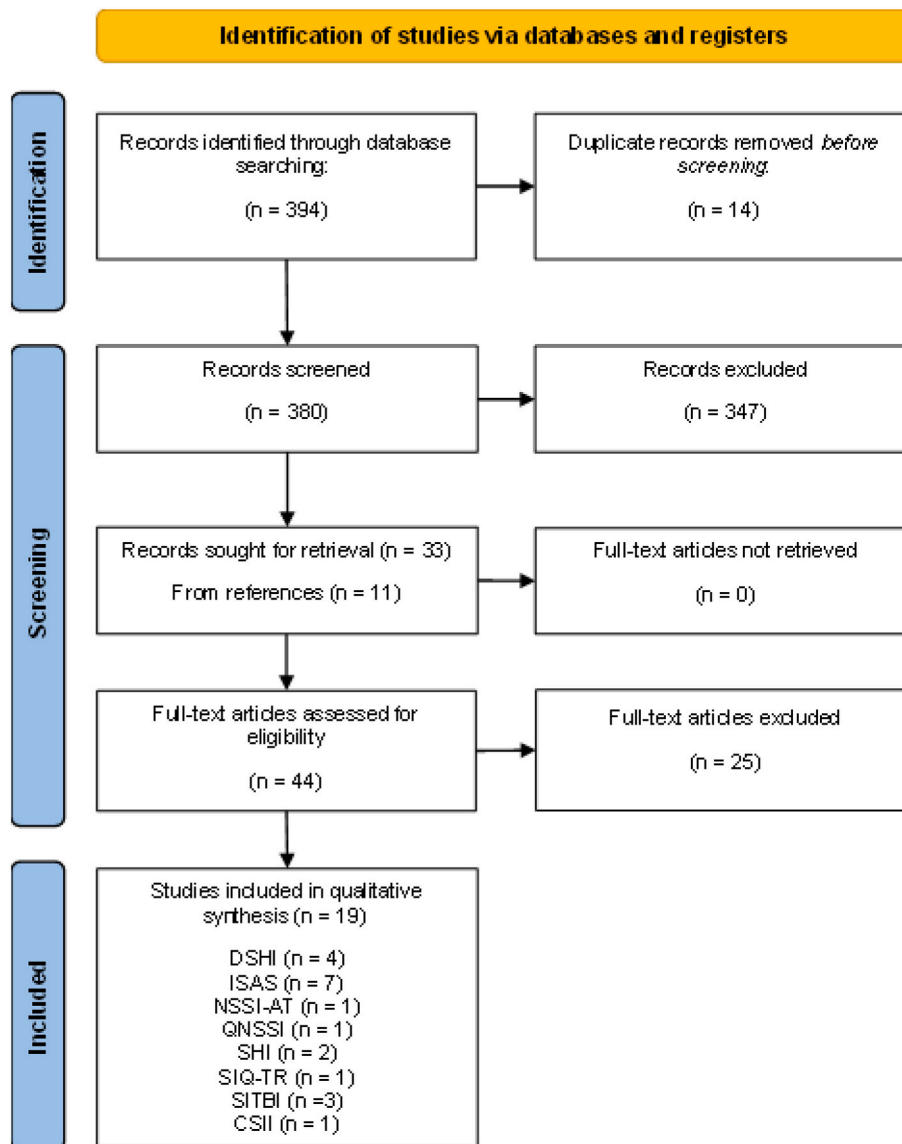
Fig. 2. PRISMA flow diagram for the selection of articles with evidence for measurement properties.

these properties could not be rated. The risk of bias for each measurement property of the included studies is presented in Table 2, and the collated evidence for the criteria for good measurement properties of each tool is presented in Table 3. Table 4 provides a qualitative summary of results, overall rating, and quality of evidence for each tool.

5.1. Cardiff self-injury inventory (CSII)

One study assessed the measurement properties of the CSII in general population adults (Snowden et al., 2023). Reliability was indeterminate, reported using r values (not the intraclass correlation coefficient (ICC) as required by COSMIN), and of very low quality due to a doubtful rating and small sample size ($n = 48$). On the other hand, evidence for criterion validity was sufficient and of high quality, where the CSII demonstrated an acceptable correlation with the DSHI ($r = 0.81$). Hypothesis testing (a) was also sufficient and of high quality, also showing an acceptable correlation with an instrument measuring a similar construct (SBQ-R).

5.2. Deliberate self-harm inventory (DSHI)

Four studies (see Table A3) assessed the measurement properties of

the DSHI in general population adults (Fliege et al., 2006; Gratz, 2001; Ohira, Munesue, Oi, Suzuki, & Saito, 2018; Vigfusdottir et al., 2020). Three studies used versions translated into German (Fliege et al., 2006), Japanese (Ohira et al., 2018) or Norwegian (Vigfusdottir et al., 2020). For the populations of included studies, one consisted of a clinical sample of psychosomatic inpatients (Fliege et al., 2006) and three of general population adults. The DSHI demonstrated indeterminate evidence for reliability which was of low quality due to multiple inadequate ratings and inappropriate reporting of statistics (r and omega values, instead of ICC and weighted Kappa). Evidence for criterion validity was inconsistent and also of low quality, the DSHI demonstrated sufficient coefficients with the ISAS ($r > 0.70$) but not with clinician rating (ICC/ k < 0.70). Evidence for hypothesis testing (a) was sufficient and of high quality, with acceptable correlations for similar measures (e.g., SHBQ), related but dissimilar measures (e.g., DERS), and unrelated measures (e. g., MCSDS). Evidence for hypothesis testing (b) was also sufficient but of moderate quality, with expected differences between groups (e.g., rates of self-harm in women vs men).

Table 2

COSMIN Risk of Bias checklist for studies included in the qualitative synthesis.

Measure	Study	Structural Validity	Internal Consistency	Reliability	Criterion Validity	Hypothesis testing (a)	(b)
CSII	Snowden et al. (2023)			Doubtful	Very good	Very good	
DSHI	Gratz (2001)			Inadequate		Adequate	Doubtful
	Fliege et al. (2006)			Inadequate	Inadequate	Very good	Doubtful
	Ohira et al. (2018)			Inadequate		Adequate	Adequate
	Vigfusdottir et al. (2020)				Very good	Adequate	Doubtful
ISAS	Glenn and Klonsky (2011)			Inadequate		Adequate	
	Kim et al. (2019)	Very good	Very good	Doubtful	Very good	Adequate	
	Klonsky and Glenn (2009)	Adequate	Very good			Adequate	
	Klonsky and Olino (2008)			Doubtful		Adequate	Adequate
	Kortge et al. (2013)	Very good	Very good				Doubtful
	Pérez et al. (2020)	Very good	Very good	Inadequate		Very good	
	Vigfusdottir et al. (2020)	Inadequate	Very good		Very good	Adequate	Doubtful
NSSI-AT	Whitlock et al. (2014)	Adequate	Very good	Doubtful	Inadequate	Doubtful	
QNSSI	Turner et al. (2012)	Doubtful	Very good			Adequate	
SHI	Latimer et al. (2009)						Adequate
	Müller et al. (2016)					Adequate	Adequate
SIQ-TR	Claes and Vandereycken (2007)	Inadequate	Doubtful			Adequate	
SITBI	Fox et al. (2020)			Adequate		Adequate	
	García-Nieto et al. (2013)			Inadequate		Doubtful	
	Lee et al. (2021)			Inadequate	Inadequate	Adequate	

CSII = Cardiff Self-Injury Inventory; DSHI = Deliberate Self-Harm Inventory (DSHI); ISAS = Inventory of Statements about Self-Injury; NSSI-AT = Non-Suicidal Self-Injury – Assessment Tool; QNSSI = Questionnaire for Non-Suicidal Self-Injury; SHI = Self-Harm Inventory; SIQ-TR = Self-Injury Questionnaire – Treatment Related; SITBI = Self-injurious Thoughts and Behaviours Interview.

Table 3

Collated evidence of the criteria for good measurement properties for each tool.

Measure	Structural Validity	Internal Consistency	Reliability	Criterion Validity	Hypothesis testing (a)	(b)
CSII			?	+	+	
DSHI			???	+?	+ + + + − −	+ + + +
ISAS	+ + − − −	+ + + + + −	???	+ +	+ + + + +	+ + +
NSSI-AT	?	−	+	?	+	
QNSSI	+	+			+	
SHI					+	+ +
SIQ-TR	+	−			+ −	
SITBI			−	?	+ + +	

"+" = sufficient," –"= insufficient, "?" = indeterminate.

CSII = Cardiff Self-Injury Inventory; DSHI = Deliberate Self-Harm Inventory (DSHI); ISAS = Inventory of Statements about Self-Injury; NSSI-AT = Non-Suicidal Self-Injury – Assessment Tool; QNSSI = Questionnaire for Non-Suicidal Self-Injury; SHI = Self-Harm Inventory; SIQ-TR = Self-Injury Questionnaire – Treatment Related; SITBI = Self-injurious Thoughts and Behaviours Interview.

5.3. Inventory of statements about self-injury (ISAS)

Seven studies (see Table A3) assessed the measurement properties of the ISAS in general population adults (Glenn & Klonsky, 2011; Klonsky & Glenn, 2009; Klonsky & Olino, 2008; Kortge, Meade, & Tennant, 2013; Kim, Kim, & Hur, 2019; Pérez, García-Alandete, Cañabate, & Marco, 2020; Vigfusdottir et al., 2020). Three studies used versions translated into Korean (Kim et al., 2019), Spanish (Pérez et al., 2020) or Norwegian (Vigfusdottir et al., 2020). The ISAS consists of two subscales for frequency (ISAS I) and functions (ISAS II). Two studies examined the measurement properties of ISAS I only (Klonsky & Olino, 2008; Vigfusdottir et al., 2020), three ISAS II only (Klonsky & Glenn, 2009; Kortge et al., 2013; Pérez et al., 2020), and two both ISAS I and II (Glenn & Klonksy, 2011). One of the seven studies included a clinical sample of participants diagnosed with an eating disorder or cluster B personality disorder (Pérez et al., 2020).

For the ISAS II subscale, evidence for structural validity was inconsistent and the subsequent quality was low. Four studies conducted exploratory factor analysis only (EFA), two conducted confirmatory factor analysis (CFA) and one used a Rasch model. EFA indicated a two-factor solution (interpersonal and intrapersonal) across respective studies, with 3 or more coefficients' of ≥ 0.4 loading on each factor (although there was evidence of minor cross-loading in two studies, Kortge et al., 2013; Vigfusdottir et al., 2020). Only one CFA supported the two-factor solution with acceptable model fit, and this was in the clinical sample (Pérez et al., 2020). The Rasch model met the assumptions for dimensionality and independence but did not report the appropriate fit statistics specified for a sufficient rating by COSMIN (Kortge et al., 2013). Conversely, evidence for internal consistency of the ISAS II subscale was sufficient and of high quality, where all studies reporting on this had very good ratings. Three of the four studies reported the acceptable level for Cronbach's alpha ($\alpha \geq 0.70$), and the one study that had an insufficient rating was only below acceptable on the intrapersonal factor (Kortge et al., 2013; $\alpha = 0.52$).

There was indeterminate evidence for reliability of both ISAS I and ISAS II subscales, which were of low quality. Similarly to the CSII and DSHI, the four studies reporting on this only did so using r values rather than ICC, along with either doubtful or inadequate ratings. Evidence for criterion validity was sufficient and of high quality, where the ISAS demonstrated acceptable correlations ($r > 0.70$) with both the FASM and the DSHI. Evidence for hypothesis testing (a) was sufficient and of high quality, where correlations were demonstrated with related but dissimilar measures (e.g., suicidality items on the YRBS) and unrelated measures (e.g., MCSDS). Hypothesis testing (b) was also sufficient but of moderate quality, showing expected differences between groups (e.g., higher scores on ISAS II corresponding to BPD diagnosis).

5.4. Non-suicidal self-injury-assessment tool (NSSI-AT)

One study (see Table A3) assessed the measurement properties of the NSSI-AT in general population adults (Whitlock et al., 2014). Evidence in support of structural validity was indeterminate but of moderate quality. EFA was conducted on suitable modules of the measure (NSSI habituation, NSSI functions, NSSI practice patterns and NSSI personal

Table 4
Summary of findings.

Measure		Structural Validity	Internal Consistency	Reliability	Criterion Validity	Hypothesis testing	
						(a)	(b)
CSII	Summary result			$r = 0.82$; total sample size: 58	DSHI ($r = 0.82$)	1 out of 1 hypotheses confirmed (100%)	
	Overall rating			Indeterminate (?)	Sufficient (+)	Sufficient (+)	
	Quality of evidence			Very low	High	High	
DSHI	Summary result			r range = 84–92; ø range = 0.49–0.73; total sample size 176	Clinician rating (ICC/ k < 0.70); ISAS (r > 0.70)	10 out of 13 hypotheses confirmed (76%)	4 out of 4 hypotheses confirmed (100%)
	Overall rating			Indeterminate (?)	Inconsistent (±)	Sufficient (+)	Sufficient (+)
	Quality of evidence			Low	Low	High	Moderate
ISAS	Summary result	**EFA:** 3+ item coefficients >0.4 loading on each factor with some minor cross-loading	ISAS II: α range =. 52–92 (one below >0.70); total sample size: 973	ISAS I: r range = 0.52–0.94; ISAS II: r range = 0.35–0.92 total sample size: 268	DSHI (r range = 0.81–0.96); FASM (r range = 0.74–0.77)	8 out of 8 hypotheses confirmed (100%)	3 out of 3 hypotheses confirmed (100%)
	Overall rating	Inconsistent (±)	Sufficient (+)	Indeterminate (?)	Sufficient (+)	Sufficient (+)	Sufficient (+)
	Quality of evidence	Low	High	Low	High	High	Moderate
NSSI-AT	Summary result	**EFA:** Unable to determine	α all <0.70; total sample size: 1773	k = 0.74–0.85; ICC = 0.63–0.91; total sample size: 25	FASM (k > 0.77)	3 out of 3 hypotheses confirmed (100%)	
	Overall rating	Indeterminate (?)	Insufficient (−)	Sufficient (+)	Indeterminate (?)	Sufficient (+)	
	Quality of evidence	Moderate	High	Very low	Very low	Low	
QNSSI	Summary result	**EFA:** 3+ item coefficients >0.4 on each factor and no cross-loadings	α = 0.66–0.85 (one below >0.70); total sample size: 162			3 out of 3 hypotheses confirmed (100%)	
	Overall rating	Sufficient (+)	Sufficient (+)			Sufficient (+)	
	Quality of evidence	Low	High			Moderate	
SHI	Summary result					1 out of 1 hypotheses confirmed (100%)	2 out of 2 hypotheses confirmed (100%)
	Overall rating					Sufficient (+)	Sufficient (+)
	Quality of evidence					Moderate	High
SIQ-TR	Summary result	**EFA:** 3+ co-efficients >0.4 on each factor and no cross-loadings	α = 0.51–0.89; total sample size: 83			1 out of 3 hypotheses confirmed (100%)	
	Overall rating	Sufficient (+)	Insufficient (−)			Insufficient (−)	
	Quality of evidence	Very low	Very low			Moderate	
SITBI	Summary result			k = 0.33–1; ICC = 0.15–0.91; total sample size: 212	DSM-5 diagnoses: SBD (k = 1.0); NSSI (k = 0.94)	5 out of 5 hypotheses confirmed (100%)	
	Overall rating			Insufficient (−)	Indeterminate (?)	Sufficient (+)	
	Quality of evidence			Low	Very low	High	

CSII = Cardiff Self-Injury Inventory; DSHI = Deliberate Self-Harm Inventory (DSHI); ISAS = Inventory of Statements about Self-Injury; NSSI-AT = Non-Suicidal Self-Injury – Assessment Tool; QNSSI = Questionnaire for Non-Suicidal Self-Injury; SHI = Self-Harm Inventory; SIQ-TR = Self-Injury Questionnaire – Treatment Related; SITBI = Self-injurious Thoughts and Behaviours Interview.

reflections and advice), however no coefficients were provided for any of these. Evidence for internal consistency was insufficient, yet of high quality. Cronbach's alpha (α = ≥ 0.70) was below the acceptable level for each of the aforementioned modules but rated as very good for risk of bias. Conversely, reliability was sufficient, with acceptable kappa (> 0.70) and ICC reported for all but one of the corresponding variables (number of wound locations), but we did not downgrade on this occasion due to the possibility of real change occurring. Criterion validity

was indeterminate in establishing NSSI-AT scores as an adequate reflection of the FASM as no correlation or AUC were reported. However, evidence for both reliability and criterion validity were of very low quality due to risk of bias from missing necessary information or small sample size ($n = 25$). There was sufficient evidence for hypothesis testing (a) with acceptable correlations for related but dissimilar constructs (e.g., suicidal thoughts and behaviours), and unrelated constructs (e.g., number of sexual partners in the last year). Although this was of low quality due to a lack of evidence for validity of how these constructs were measured.

5.5. Questionnaire for non-suicidal self-injury (QNSSI)

One study (see Table A3) assessed the measurement properties of the QNSSI in general population adults (Turner et al., 2012). Evidence in support of structural validity was sufficient but of low quality. EFA indicated a five-factor solution (emotion relief, feeling generation, interpersonal influence, interpersonal communication, and self-punishment) with 3 or more co-efficients of ≥ 0.4 loading on each factor, although sample size was inadequate with < 5 participants per item. On the other hand, internal consistency was sufficient and of high quality. Cronbach's alpha was above acceptable for four of the five factors ($\alpha \geq 0.70$), but self-punishment ($\alpha = 0.66$) was considered close enough to be of an acceptable level. Evidence for hypothesis testing (a) was sufficient and of moderate quality. Acceptable correlations were demonstrated between factors with related but dissimilar items on other measures (e.g., emotion relief with affective intensity, expressive suppression, and difficulties in emotion regulation), and unrelated items (e. g., feeling generation with affective reactivity or expressive suppression).

5.6. Self-harm inventory (SHI)

Two studies (see Table A3) assessed the measurement properties of the SHI in general population adults (Latimer, Covic, Cumming, & Tennant, 2009; Müller, Claes, Smits, Brähler, & de Zwaan, 2016). One study used a version translated into German (Müller et al., 2016). The SHI is a behaviour specific measure, where the only measurement property of relevance was hypothesis testing. The evidence in support of hypothesis testing (a) was sufficient and of moderate quality, where the SHI demonstrated acceptable correlations with related but dissimilar measures (e.g., PHQ-4). Similarly, hypothesis testing (b) was sufficient and of high quality, with expected differences found between groups (e. g., DASS-21 scores differed across severity levels of the SHI).

5.7. Self-injury questionnaire – treatment related (SIQ-TR)

One study (see Table A3) assessed the measurement properties of the SIQ-TR in a clinical sample of inpatients with an eating disorder (Claes & Vandereycken, 2007). Evidence for structural validity was sufficient, where EFA indicated a two-factor solution on the emotion subscale (preceding feelings and consequent feelings) and a three-factor solution on the function subscale (social positive reinforcement, automatic positive reinforcement and automatic negative reinforcement). Both scales had 3 or more co-efficients of ≥ 0.4 loading on each factor. Conversely, the evidence for internal consistency was insufficient; there was no overall Cronbach's alpha for any of subscale's factors, and individual alphas ranged below the acceptable level on each subscale ($\alpha < 0.70$). Quality of evidence was very low for both structural validity and internal consistency due to unclear and flawed reporting on these measurement properties which increased risk of bias. There was also insufficient evidence for hypothesis testing (a), although this was of moderate quality, where the SIQ-TR demonstrated inconsistent relationships across subscales with related but dissimilar measures (e.g., Self-Expression and Control Scale).

5.8. Self-injurious thoughts and behaviours interview (SITBI)

Three studies (see Table A3) assessed the measurement properties of the SITBI in general population adults. One study used a revised version of the SITBI (SITBI – R; Fox et al., 2020), and two studies used versions translated into Spanish (García-Nieto, Blasco-Fontecilla, Yepes, & Baca-García, 2013) or Korean (Lee, Cho, & Hyun, 2021). One study included a clinical sample of psychiatric inpatients admitted for self-harmful (with or without suicidal intent) ideation and behaviour (García-Nieto et al., 2013). Despite not being a wholly behaviour-specific tool, the authors suggest that structural validity and internal consistency are not meaningful to the SITBI which uses various formats of items to measure a wide range of constructs and was therefore not reported by any of the studies. Evidence in support of reliability was insufficient and of low quality. ICC and k values reported within all three of the studies ranged below the acceptable level (< 0.70) and were at risk of bias due to inappropriate time intervals, inconsistent types of administration or small sample sizes. Similarly, criterion validity was indeterminate and of very low quality due to missing necessary information. Nonetheless, there was sufficient evidence for hypothesis testing (a) that was of high quality, with acceptable correlations of the SITBI with similar (e.g., SSI) and related but dissimilar (e.g., SIS, BDI-II) measures and items.

5.9. Interpretability and feasibility

For interpretability, the distribution of the overall or subscale score of each tool (i.e., mean and standard deviation or equivalent) was provided in all studies. This was less relevant in behaviour specific measures (CSII, DSHI, SHI), where the proportion of self-harm behaviours endorsed, or frequency of behaviours were given instead. Eight studies reported exclusion of participants due to missing data or addressed how missing data was/ would be handled, although it was not necessarily clear whether this was specific to data from the self-harm assessment tool. Only one study reported the number of responses for each item and subscale on the tool (NSSI-AT; Whitlock et al., 2014). Two studies explicitly reported a floor effect, which was present at the more severe end of self-harm on the DSHI (Latimer et al., 2009), and for the interpersonal factor on the ISAS (Kortge et al., 2013). Scores and change scores for relevant (sub)groups, minimal important change and information on response shift were not reported in any of the studies.

In terms of feasibility, there was no evidence of content validity for any of the self-harm assessment tools in the population of interest, and therefore neither patient nor clinician comprehensibility could be evaluated. Length of the tools varied from 8 items (CSII) to 169 items (SITBI), with a duration of 1 to 30 minutes to complete depending on the measure. It was assumed that respondents would require at least average IQ (> 70) to complete the tools, as there was no evidence to suggest otherwise. In general, the availability of guidance or instructions for the administration and scoring of the self-harm assessment tools was inconsistent, although somewhat self-explanatory. Copyright information was provided for the CSII and SHI, and only one tool was not freely available online but was provided at no cost with permission from the authors (QNSSI; Turner et al., 2012).

5.10. Formulated recommendations

Based on the COSMIN checklist for formulating recommendations, seven of the eight self-harm assessment tools were classified as B (having potential to be recommended for use but require further research to assess quality), however, the NSSI-AT was classified as C (not recommended for use) due to high quality evidence for insufficient internal consistency. The ISAS and QNSSI were the only assessment tools which had sufficient internal consistency, in line with classification A (recommended for use), but did not assess content validity which was necessary to meet the full requirements. Moreover, the ISAS had the most evidence overall for its measurement properties in general and

clinical populations ($n = 7$), compared to the single study for the QNSSI. Thus, as no evidence was available for autistic adults, either the ISAS or QNSSI would likely be the most suitable tool to recommend for use with general population adults, but both need further research to assess content validity.

6. Discussion

Previous research has shown high rates of self-harm in autistic people (Akram et al., 2017; Licence et al., 2020). However, prior to this review, it was unknown which tools (if any) had been frequently used to assess self-harm in autistic versus general population adults. Results revealed a small number of studies with autistic adults that had used a validated tool to assess self-harm ($n = 6$) compared to general population adults ($n = 91$). Moreover, only one frequently used self-harm assessment tool (NSSI-AT) had been utilised with autistic adults. The NSSI-AT had also been used in the general population with some evidence of validity, however it has not been developed or validated for autistic people. This corresponds with previous COSMIN reviews which have highlighted a distinct lack of robust health outcomes measures for autistic people (Cassidy et al., 2018b; Cassidy et al., 2018c; Thoen et al., 2021). The current findings therefore emphasise the need for assessment tools that are developed and validated for autistic people to improve the conceptualisation of self-harm and its measurement in research and clinical practice.

Returned search results from stage 1 of the review demonstrated that studies with autistic people commonly employed singular items or subscales contained within a more general tool to determine the presence of self-harm. Many also assessed SIB in this manner as a challenging behaviour (e.g., Behaviour Problems Inventory; Rojahn, Matson, Lott, Esbensen, & Smalls, 2001), which did not meet the inclusion criteria. This was mostly the case in samples of autistic children or autistic people with co-occurring ID, although a small number of studies did use such items/ subscales for autistic adults without co-occurring ID (e.g., Challenging Behaviour Questionnaire, Hyman, Oliver, & Hall, 2002; Licence, Oliver, Moss, & Richards, 2020). Only one study was found in the current review which had utilised a tool specifically for the SIB conceptualisation of self-harm (IfES; Bienstein & Nußbeck, 2010), but was not considered for stage 2. In addition, some studies used items that did not distinguish between self-harm and suicidality (e.g., "Thoughts you would be better off dead or hurting yourself" Patient Health Questionnaire-9; Kroenke, Spitzer, & Williams, 2001). However, we do not know enough about self-harm in autistic people to assess its function as commensurate to suicidal ideation (Cassidy, 2020). There is a clear need for future research of self-harm in autism to better understand these conceptualisations and functions, using self-harm assessment tools that have high quality evidence in support of their measurement properties (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018). Based on a synthesis of the available evidence, and the outcomes of the COSMIN checklist, we therefore make recommendations for the use of self-harm assessment tools in autism research.

Nineteen studies had explored the measurement properties of validated self-harm assessment tools in general population adults with or without co-occurring conditions: the CSII, DSHI, ISAS, NSSI-AT, QNSSI, SHI, SITBI and SIQ-TR. The DSHI, ISAS, SIQTR and SITBI had also been validated in clinical samples. However, no studies were identified assessing the measurement properties of these tools in autistic adults, nor the FASM in any adult samples (despite a number of studies in stage 1 utilising it within this age group, e.g., Morales, Buser, & Farag, 2018). Of the included measures, three were wholly behaviour specific (DSHI, SHI) measuring the type, frequency, or severity of self-harm, and one (CSII) only measured frequency. Generally, a key barrier to providing effective support is poor understanding of the function that self-harm serves for an individual (Edmondson, Brennan, & House, 2016). In this sense, the CSII, DSHI and SHI are valuable for epidemiological research but have limited utility for understanding personal experiences. Given the paucity of research on self-harm in autistic people, tools which also measure the underlying latent construct (e.g., functions) are more likely to be beneficial within clinical and research contexts. Consequently, the CSII, DSHI and SHI are not considered for further recommendation within this review.

Only one study had explored the measurement properties of each of the following assessment tools in general and clinical population adults: NSSI-AT (Whitlock et al., 2014), QNSSI (Turner et al., 2012) and the SIQTR (Claes & Vandereycken, 2007). For the NSSI-AT, structural validity was rated indeterminate as EFA coefficients were not provided, and internal consistency was insufficient despite the quality of evidence being high. However, the authors recognise several possible reasons for low alphas, such as the small number of items per subscale (3–4), that the items were not originally designed as scales, and the scoring as mostly dichotomous (Whitlock et al., 2014). Reliability, criterion validity and hypothesis testing (a) all had a sufficient overall rating, although quality of evidence was low to very low. The NSSI-AT was the only measure to demonstrate a sufficient rating for reliability despite a small test-retest subsample ($n = 25$). Likewise, the QNSSI had promising ratings of sufficient for each of the measurement properties that were explored (structural validity, internal consistency, and hypothesis testing), although the quality of evidence was inconsistent, ranging from low to high. On the other hand, the SIQTR only had sufficient evidence for structural validity, whereas internal consistency and hypothesis testing (a) were insufficient, and quality of evidence was very low in all but hypothesis testing (where it was moderate). The NSSI-AT was the only tool to have been utilised in autism research, and the QNSSI demonstrated some initial strengths, but there was not enough evidence to confidently recommend these tools for use in research or clinical practice.

Three studies explored the measurement properties of the SITBI, each adapted into a different language (Spanish, García-Nieto et al., 2013; Korean, Lee et al., 2021), version (e.g., revised, Fox et al., 2020; short, García-Nieto et al., 2013), or format (e.g., self-report, interview; Fox et al., 2020). Despite this, overall ratings were consistent across studies; reliability was insufficient with low quality of evidence, and hypothesis testing (a) was sufficient with high quality of evidence. It is also important to note that this tool is the only one that assessed self-harm via a module within a wider measure of suicidality. Whilst broader definitions include self-harm both with and without suicidal intent (Hargus et al., 2009; Kapur et al., 2013), the SITBI focuses more on suicidality, and the SBQ-ASC has already been successfully adapted to measure this construct in autistic people (Cassidy, Bradley, Cogger-Ward and Rodgers, 2021). Combined with this, and that the SITBI has limited evidence in support of just two measurement properties, it is unlikely to be the most optimal self-harm assessment tool to recommend for use in future research.

The ISAS had the most studies ($n = 7$) of all the tools exploring its measurement properties in general and clinical population adults. Notably, the evidence for structural validity was inconsistent with low quality of evidence, but this may be partly explained by the samples used. The only study which determined adequate model fit for CFA was in a clinical group of adults (Pérez et al., 2020). Therefore, it could be that the general population samples in the other studies may engage in self-harm less frequently or at a subclinical level, limiting the generalisability of the factor structure to specific presentations (Klonsky, Glenn, Styer, Olino, & Washburn, 2015). Internal consistency, criterion validity and hypothesis testing (a) and (b) were sufficient across studies, where all had high quality of evidence with the only exception being hypothesis testing (b) with a moderate rating. Reliability was indeterminate as the statistics provided for this did not meet COSMIN criteria (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018), and quality of evidence was low. Based on its measurement properties and outcomes from COSMIN, the ISAS could potentially be recommended for use in future research, but evidence is needed for its content validity and measurement properties in autistic adults.

Overall, none of the studies explored measurement error, cross cultural validity/ measurement invariance, or responsiveness in any of the self-harm assessment tools identified. In particular, there was a wide variation in languages which the tools were translated into using the forward-backward method, yet no evidence that these adequately reflected the performance of items from the original tool (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018). A consistent issue across assessment tools was also that of test-retest reliability, whereby the statistics used to report on this were not optimal or were below 0.70 and time intervals were inadequate (i.e., longer than 4 weeks). However, antecedent factors and the temporal proximity of these to an episode of self-harm are complex (Townsend et al., 2016), and the predictive utility of self-harm and suicide assessments for future behaviour is already known to be poor (Quinlivan et al., 2016; Quinlivan et al., 2017). It is therefore unsurprising that reports of self-harm frequency and function do not appear to be stable between repeated measurements. Studies should aim to also include tests of responsiveness to change in order to address the nature or assessment of changes in self-harm over time.

In summary, the current review revealed mixed evidence for measurement properties of eight frequently used self-harm assessment tools (CSII, DSHI, ISAS, NSSI-AT, QNSSI, SHI, SITBI and SIQ-TR). The ISAS had the most evidence overall, and this was generally favourable on all measurement properties aside from structural validity and reliability. Certain tools demonstrated stronger evidence for their internal structure (ISAS, QNSSI) and therefore a better reflection of self-harm as a construct (Mokkink et al., 2010), or had both adequate criterion and construct validity (ISAS, NSSI-AT). Moreover, using the COSMIN categories to formulate recommendations, only the ISAS and the QNSSI had at least low-quality evidence for sufficient internal consistency, meeting one of the two criteria for tools that can be recommended for use (or adaptation). To our knowledge, this review is the first to use a validated research tool (COSMIN) to identify which self-harm assessment tools have been frequently used in autistic and general population adults, and to synthesise the available evidence for these on a range of measurement properties across a number of studies.

6.1. Future research

Only one self-harm assessment tool (NSSI-AT) was found to be frequently used in studies of autistic adults. However, this was developed for general population adults and there was no evidence for its measurement properties in autistic samples. Even in general population adults, the NSSI-AT was the only tool with high quality evidence for an insufficient measurement property (internal consistency), meaning it would not be recommended for future use according to COSMIN. In addition, the validity of this tool in autistic samples may be affected by previously highlighted characteristics such as alexithymia (Kinnaird et al., 2019), literal interpretations (Happé, 1995), difficulties with episodic memory and future thinking (Lind et al., 2014; Lind & Bowler, 2010), along with differing presentations of self-harm (e.g., SIB; Steenfeldt-Kristensen et al., 2020) and its function (e.g., to modulate sensory overwhelm; Moseley et al., 2019). Future studies should aim to explore the content validity for existing self-harm assessment tools in autistic people to determine whether items are relevant, comprehensive, and clear in regard to the construct of self-harm. Performance of any adapted measure should then be compared with its original in autistic and general population groups, as recommended good practice by COSMIN (Mokkink et al., 2018; Prinsen et al., 2018; Terwee et al., 2018). In theory, an adapted tool would be superior in capturing the construct of self-harm in autistic people, potentially demonstrated by a different factor structure, along with better internal consistency and criterion validity (i.e., ability to predict future behaviour). Assessment tools for suicidality and depression have already been adapted with and for autistic adults, and demonstrate that improving the clarity and relevance of items also increases sensitivity to detect associated constructs (Cassidy, Bradley, Cogger-Ward, Graham, & Rodgers, 2021; Cassidy,

Bradley, Cogger-Ward, & Rodgers, 2021). This shows that not only is there a need to develop or adapt robust self-harm assessment tools for autistic adults, but also that this can be done successfully.

Based on their measurement properties, this review suggests that the ISAS or QNSSI may be the most promising candidate tools to begin exploring self-harm in autism, with the potential to be adapted for autistic adults. The ISAS consists of 59 items and is split into two sections; the first section assesses lifetime frequency of 12 NSSI behaviours along with 5 questions about descriptive and contextual factors, whereas the second section assesses 13 potential intra- and interpersonal functions of NSSI with 3 items per function (Klonksy & Glenn, 2009). The QNSSI consists of 39 items across 5 subscales of emotion relief, feeling generation, interpersonal communication, interpersonal influence and self-punishment. Despite this, neither measure appears to be superior in terms of feasibility and both are free to access. The QNSSI is shorter in length than the ISAS but is more complicated in its formatting and uses a wider variety of response options.

6.2. Strengths and limitations

The current review had a number of strengths and limitations. The main strength lies in the rigorous COSMIN method used to identify and evaluate studies. However, this rigour may have conferred disadvantages through the exclusion of studies from the evidence synthesis. For example, larger measures that included a self-harm subscale were not eligible for inclusion, so there was the possibility of missing potentially relevant data, particularly as SIB was mainly assessed this way within tools for challenging behaviour. Nevertheless, subscales frequently have narrower definitions and fewer items to aid holistic understanding of the construct (Cassidy, Bradley, Bowen, et al., 2018b), therefore we felt it important to focus on assessment tools specific to self-harm. As results were consistent with previous COSMIN of other patient reported outcomes showing a paucity of evidence for autistic people (Cassidy, Bradley, Bowen, et al., 2018a; Cassidy, Bradley, Bowen, et al., 2018b; Thoen et al., 2021), we believe this is unlikely to be a cause for concern.

In addition, the strict nature of the COSMIN checklist meant our results may have trended more towards negative ratings (i.e., inadequate, insufficient or no rating), whereby very specific or excellent statistics were necessary to meet adequate or sufficient criteria (Modini, Abbott, & Hunt, 2015). For example, we adapted the criteria for good measurement properties so that it would take EFA into account when evaluating structural validity. Under original circumstances, these studies would have received a negative rating despite having a well-adjusted EFA according to best practice (Costello & Osborne, 2005). Searches were also limited by only including studies which were published and available in the English language (due to not having the resources to allow for the translation of articles). Specifically, the QNSSI was originally developed in German (Kleindienst et al., 2008), meaning that relevant studies with evidence in support of this tools measurement properties may have also been missed within the searches.

7. Conclusion

This is the first systematic review to use COSMIN as a robust research tool to identify, assess, and synthesise the evidence for the assessment of self-harm in autistic and general population adults. Eight validated self-harm assessment tools were identified which had been frequently used in general population adults with or without co-occurring conditions, but only one of these had been used with autistic adults. No studies had investigated the measurement properties of any of the tools in autistic adults or had been developed or validated to assess self-harm in this group. Future research should explore the content validity of tools designed for the general population in autistic people, to determine whether they also adequately capture self-harm in this group. If not, more work is needed to adapt available measures using focus groups and cognitive interviews, and to explore whether measurement properties

improve as a result.

Role of funding sources

Declaration of competing interest

All authors declare that they have no conflicts of interest.

Acknowledgement

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cpr.2024.102412.

References

Akram, B., Tariq, A., & Rafi, Z. (2017). Non-suicidal self-injury among children with hearing loss and intellectual disability. *JPMA. Journal of the Pakistan Medical Association, 67*(10), 1506–1511.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Publishing.

Beckman, K., Lysell, H., Haglund, A., & Dahlin, M. (2019). Prognoses after self-harm in youth: Exploring the gender factor. *Social Psychiatry and Psychiatric Epidemiology, 54,* 437–444.

Bienstein, P., & Nußbeck, S. (2010). *IfES: Inventar zur funktionellen Erfassung selbstverletztenden Verhaltens bei Menschen mit intellektueller Beeinträchtigung.* Hogrefe.

Blanchard, A., Chihuri, S., DiGuiseppi, C. G., & Li, G. (2021). Risk of self-harm in children and adults with autism spectrum disorder: A systematic review and meta-analysis. *JAMA Network Open, 4*(10), e2130272.

Brede, J., Cage, E., Trott, J., Palmer, L., Smith, A., Serpell, L., & Russell, A. (2022). "We have to try to find a way, a clinical bridge"-autistic adults' experience of accessing and receiving support for mental health difficulties: A systematic review and thematic meta-synthesis. *Clinical Psychology Review, 93,* Article 102131.

Cassidy, S. (2020). *Suicidality and self-harm in autism spectrum conditions* (In *The Oxford Handbook of Autism and Co-Occurring Psychiatric Conditions).*

Cassidy, S., Bradley, L., Cogger-Ward, H., Graham, J., & Rodgers, J. (2021). *Development and Validation of the Autistic Depression Assessment Tool–Adult (ADAT-A) in Autistic Adults.*

Cassidy, S. A., Bradley, L., Bowen, E., Wigham, S., & Rodgers, J. (2018a). Measurement properties of tools used to assess suicidality in autistic and general population adults: A systematic review. *Clinical Psychology Review, 62,* 56–70.

Cassidy, S. A., Bradley, L., Bowen, E., Wigham, S., & Rodgers, J. (2018b). Measurement properties of tools used to assess depression in adults with and without autism spectrum conditions: A systematic review. *Autism Research, 11*(5), 738–754.

Cassidy, S., Bradley, L., Shaw, R., & Baron-Cohen, S. (2018c). Risk markers for suicidality in autistic adults. *Molecular Autism, 9*(1), 1–14.

Cassidy, S. A., Bradley, L., Cogger-Ward, H., & Rodgers, J. (2021). Development and validation of the suicidal behaviours questionnaire-autism spectrum conditions in a community sample of autistic, possibly autistic and non-autistic adults. *Molecular Autism, 12*(1), 46.

Cassidy, S. A., Bradley, L., Cogger-Ward, H., Shaw, R., Bowen, E., Glod, M., & Rodgers, J. (2020). Measurement properties of the suicidal behaviour questionnaire-revised in autistic adults. *Journal of Autism and Developmental Disorders, 50,* 3477–3488.

Claes, L., & Vandereycken, W. (2007). Self-injurious behavior: Differential diagnosis and functional differentiation. *Comprehensive Psychiatry, 48*(2), 137–144.

Claes, L., & Vandereycken, W. (2007b). *The Self-Injury Questionnaire—Treatment Related (SIQ-TR): Construction, Reliability, and Validity in a Sample of Female Eating Disorder Patients.*

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation, 10*(1), 7.

Duerden, E. G., Oatley, H. K., Mak-Fan, K. M., McGrath, P. A., Taylor, M. J., Szatmari, P., & Roberts, S. W. (2012). Risk factors associated with self-injurious behaviors in children and adolescents with autism spectrum disorders. *Journal of Autism and Developmental Disorders, 42*(11), 2460–2470.

Edmondson, A. J., Brennan, C. A., & House, A. O. (2016). Non-suicidal reasons for self-harm: A systematic review of self-reported accounts. *Journal of Affective Disorders, 191,* 109–117.

Favril, L., Yu, R., Uyar, A., Sharpe, M., & Fazel, S. (2022). Risk factors for suicide in adults: Systematic review and meta-analysis of psychological autopsy studies. *Evidence-Based Mental Health, 25*(4), 148.

Fliege, H., Kocalevent, R. D., Walter, O. B., Beck, S., Gratz, K. L., Gutierrez, P. M., & Klapp, B. F. (2006). Three assessment tools for deliberate self-harm and suicide behavior: Evaluation and psychopathological correlates. *Journal of Psychosomatic Research, 61*(1), 113–121.

Forbes, C. N., Tull, M. T., Lavender, J. M., Dixon-Gordon, K. L., & Gratz, K. L. (2022). Development and initial validation of the body-focused self-damaging behavior expectancies questionnaire. *Journal of Psychopathology and Behavioral Assessment, 44* (3), 875–897.

Fox, K. R., Harris, J. A., Wang, S. B., Millner, A. J., Deming, C. A., & Nock, M. K. (2020). Self-injurious thoughts and behaviors interview—Revised: Development, reliability, and validity. *Psychological Assessment, 32*(7), 677.

García-Nieto, R., Blasco-Fontecilla, H., Yepes, M. P., & Baca-García, E. (2013). Translation and validation of the "self-injurious thoughts and Behaviours interview" in a Spanish population with suicidal behaviour. *Revista de Psiquiatría y Salud Mental (English Edition), 6*(3), 101–108.

Glenn, C. R., & Klonsky, E. D. (2011). One-year test-retest reliability of the inventory of statements about self-injury (ISAS). *Assessment, 18*(3), 375–378.

Gratz, K. L. (2001). Measurement of deliberate self-harm: Preliminary data on the deliberate self-harm inventory. *Journal of Psychopathology and Behavioral Assessment, 23*(4), 253–263.

Gratz, K. L., Dixon-Gordon, K. L., Chapman, A. L., & Tull, M. T. (2015). Diagnosis and characterization of DSM-5 nonsuicidal self-injury disorder using the clinician-administered nonsuicidal self-injury disorder index. *Assessment, 22*(5), 527–539.

Gutierrez, P. M. (1998). *Self-Harm Behavior Questionnaire (SHBQ).* Unpublished manuscript. DeKalb: Northern Illinois University.

Happé, F. G. (1995). Understanding minds and metaphors: Insights from the study of figurative language in autism. *Metaphor and Symbol, 10*(4), 275–295.

Hargus, E., Hawton, K., & Rodham, K. (2009). Distinguishing between subgroups of adolescents who self-harm. *Suicide and Life-threatening Behavior, 39*(5), 518–537.

Hawton, K., Bale, L., Brand, F., Townsend, E., Ness, J., Waters, K., & Geulayov, G. (2020). Mortality in children and adolescents following presentation to hospital after non-fatal self-harm in the multicentre study of self-harm: A prospective observational cohort study. *The Lancet Child & Adolescent Health, 4*(2), 111–120.

Hawton, K., & Harriss, L. (2007). Deliberate self-harm in young people: Characteristics and subsequent mortality in a 20-year cohort of patients presenting to hospital. [CME]. *Journal of Clinical Psychiatry, 68*(10), 1574.

Hawton, K., Saunders, K. E., & O'Connor, R. C. (2012). Self-harm and suicide in adolescents. *The Lancet, 379*(9834), 2373–2382.

Hyman, P., Oliver, C., & Hall, S. (2002). Self-injurious behavior, self-restraint, and compulsive behaviors in Cornelia de Lange syndrome. *American Journal on Mental Retardation, 107*(2), 146–154.

Kapur, N., Cooper, J., O'Connor, R. C., & Hawton, K. (2013). Non-suicidal self-injury v. attempted suicide: New diagnosis or false dichotomy? *The British Journal of Psychiatry, 202*(5), 326–328.

Kenny, L., Hattersley, C., Molins, B., Buckley, C., Povey, C., & Pellicano, E. (2016). Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism, 20*(4), 442–462.

Kim, S., Kim, Y., & Hur, J. W. (2019). Nonsuicidal self-injury among Korean young adults: A validation of the Korean version of the inventory of statements about self-injury. *Psychiatry Investigation, 16*(4), 270.

Kinnaird, E., Stewart, C., & Tchanturia, K. (2019). Investigating alexithymia in autism: A systematic review and meta-analysis. *European Psychiatry, 55,* 80–89.

Kleindienst, N., Bohus, M., Ludäscher, P., Limberger, M. F., Kuenkele, K., Ebner-Priemer, U. W., … Schmahl, C. (2008). Motives for nonsuicidal self-injury among women with borderline personality disorder. *The Journal of Nervous and Mental Disease, 196*(3), 230–236.

Klonsky, E. D. (2011). Non-suicidal self-injury in United States adults: Prevalence, sociodemographics, topography and functions. *Psychological Medicine, 41*(9), 1981–1986.

Klonsky, E. D., & Glenn, C. R. (2009). Assessing the functions of non-suicidal self-injury: Psychometric properties of the inventory of statements about self-injury (ISAS). *Journal of Psychopathology and Behavioral Assessment, 31*(3), 215–219.

Klonsky, E. D., Glenn, C. R., Styer, D. M., Olino, T. M., & Washburn, J. J. (2015). The functions of nonsuicidal self-injury: Converging evidence for a two-factor structure. *Child and Adolescent Psychiatry and Mental Health, 9*(1), 1–9.

Klonsky, E. D., & Olino, T. M. (2008). Identifying clinically distinct subgroups of self-injurers among young adults: A latent class analysis. *Journal of Consulting and Clinical Psychology, 76*(1), 22.

Kortge, R., Meade, T., & Tennant, A. (2013). Interpersonal and intrapersonal functions of deliberate self-harm (DSH): A psychometric examination of the inventory of statements about self-injury (ISAS) scale. *Behaviour Change, 30*(1), 24–35.

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16*(9), 606–613.

Lai, M. C., Kassee, C., Besney, R., Bonato, S., Hull, L., Mandy, W., & Ameis, S. H. (2019). Prevalence of co-occurring mental health diagnoses in the autism population: A systematic review and meta-analysis. *The Lancet Psychiatry, 6*(10), 819–829.

Lane, S. J. (2002). Sensory modulation. In A. C. Bundy, S. J. Lane, & E. A. Murray (Eds.), *Sensory Integration: Theory and Practice* (pp. 101–122). F A Davis.

Latimer, S., Covic, T., Cumming, S. R., & Tennant, A. (2009). Psychometric analysis of the self-harm inventory using Rasch modelling. *BMC Psychiatry, 9*(1), 1–9.

Lee, S. J., Cho, Y. J., & Hyun, M. H. (2021). Self-injurious thoughts and behaviors interview-Korean version: Psychometric properties. *Psychiatry Investigation, 18*(2), 157.

Licence, L., Oliver, C., Moss, J., & Richards, C. (2020). Prevalence and risk-markers of self-harm in autistic children and adults. *Journal of Autism and Developmental Disorders, 50*(10), 3561–3574.

Lind, S. E., & Bowler, D. M. (2010). Episodic memory and episodic future thinking in adults with autism. *Journal of Abnormal Psychology, 119*(4), 896.

Lind, S. E., Williams, D. M., Bowler, D. M., & Peel, A. (2014). Episodic memory and episodic future thinking impairments in high-functioning autism spectrum disorder: An underlying difficulty with scene construction or self-projection? *Neuropsychology, 28*(1), 55.

Lloyd, E. E. (1997). *Self-mutilation in a community sample of adolescents*. Louisiana State University and Agricultural & Mechanical College.

MacLennan, K., O'Brien, S., & Tavassoli, T. (2021). In our own words: The complex sensory experiences of autistic adults. *Journal of Autism and Developmental Disorders*, 1–15.

Maddox, B. B., Trubanova, A., & White, S. W. (2017). Untended wounds: Non-suicidal self-injury in adults with autism spectrum disorder. *Autism, 21*(4), 412–422.

Manca, M. (2009). *Condotte autolesive in adolescenza* (Doctoral dissertation,. Roma e Firenze: Tesi di dottorato.

Martin, J., Cloutier, P. F., Levesque, C., Bureau, J. F., Lafontaine, M. F., & Nixon, M. K. (2013). Psychometric properties of the functions and addictive features scales of the Ottawa self-injury inventory: A preliminary investigation using a university sample. *Psychological Assessment, 25*(3), 1013.

Minshawi, N. F., Hurwitz, S., Fodstad, J. C., Biebl, S., Morriss, D. H., & McDougle, C. J. (2014). The association between self-injurious behaviors and autism spectrum disorders. *Psychology Research and Behavior Management, 7*, 125.

Modini, M., Abbott, M. J., & Hunt, C. (2015). A systematic review of the psychometric properties of trait social anxiety self-report measures. *Journal of Psychopathology and Behavioral Assessment, 37*, 645–662.

Mokkink, L. B., De Vet, H. C., Prinsen, C. A., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). COSMIN risk of bias checklist for systematic reviews of patient-reported outcome measures. *Quality of Life Research, 27*, 1171–1179.

Mokkink, L. B., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., … de Vet, H. C. (2010). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology, 63*(7), 737–745.

Morales, J. J., Buser, T. J., & Farag, M. S. (2018). Comparison of brief measures of the prevalence of nonsuicidal self-injury in a nonclinical sample of young adults. *Journal of Mental Health Counseling, 40*(2), 156–171.

Moran, P., Coffey, C., Romaniuk, H., Degenhardt, L., Borschmann, R., & Patton, G. C. (2015). Substance use in adulthood following adolescent self-harm: A population-based cohort study. *Acta Psychiatrica Scandinavica, 131*(1), 61–68.

Moseley, R. L., Gregory, N. J., Smith, P., Allison, C., & Baron-Cohen, S. (2020). Links between self-injury and suicidality in autism. *Molecular Autism, 11*(1), 1–15.

Moseley, R. L., Gregory, N. J., Smith, P., Allison, C., & Baron-Cohen, S. J. M. A. (2019). A 'choice', an 'addiction', a way 'out of the lost': Exploring self-injury in autistic people without intellectual disability. *Molecular Autism, 10*(1), 1–23.

Müller, A., Claes, L., Smits, D., Brähler, E., & de Zwaan, M. (2016). Prevalence and correlates of self-harm in the German general population. *PLoS One, 11*(6), Article e0157928.

National Institute for Health and Care Excellence. (2022). Self-Harm: Assessment, Management and Preventing Recurrence. NG225. September. Available from: https://www.nice.org.uk/guidance/ng225.

Newell, V., Phillips, L., Jones, C., Townsend, E., Richards, C., & Cassidy, S. (2023). A systematic review and meta-analysis of suicidality in autistic and possibly autistic people without co-occurring intellectual disability. *Molecular Autism, 14*(1), 1–37.

Nicolaidis, C., Raymaker, D., McDonald, K., Dern, S., Boisclair, W. C., Ashkenazy, E., & Baggs, A. (2013). Comparison of healthcare experiences in autistic and non-autistic adults: A cross-sectional online survey facilitated by an academic-community partnership. *Journal of General Internal Medicine, 28*, 761–769.

Nock, M. K. (2009). Why do people hurt themselves? New insights into the nature and functions of self-injury. *Current Directions in Psychological Science, 18*(2), 78–83.

Nock, M. K., & Favazza, A. R. (2009). Nonsuicidal self-injury: Definition and classification. In M. K. Nock (Ed.), *Understanding nonsuicidal self-injury: Origins, assessment, and treatment* (pp. 9–18). American Psychological Association.

Nock, M. K., Holmberg, E. B., Photos, V. I., & Michel, B. D. (2007). Self-injurious thoughts and behaviors interview: Development, reliability, and validity in an adolescent sample. *Psychological Assessment, 19*(3), 309–317.

Nock, M. K., & Prinstein, M. J. (2004). A functional approach to the assessment of self-mutilative behavior. *Journal of Consulting and Clinical Psychology, 72*(5), 885.

Ohira, T., Munesue, T., Oi, M., Suzuki, K., & Saito, D. (2018). Investigation of the reliability and validity of the Japanese deliberate self-harm inventory. *Journal of Brain Science, 48*, 14–42.

Ohlis, A., Bjureberg, J., Lichtenstein, P., D'Onofrio, B. M., Fruzzetti, A. E., Cederlöf, M., & Hellner, C. (2020). Comparison of suicide risk and other outcomes among boys and girls who self-harm. *European Child & Adolescent Psychiatry, 29*, 1741–1746.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., & Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *International Journal of Surgery, 88*, Article 105906.

Pérez, S., García-Alandete, J., Cañabate, M., & Marco, J. H. (2020). Confirmatory factor analysis of the inventory of statements about self-injury in a Spanish clinical sample. *Journal of Clinical Psychology, 76*(1), 102–117.

Pompili, M., Goracci, A., Giordano, G., Erbuto, D., Girardi, P., Klonsky, E. D., et al. (2015). Relationship of non-suicidal self-injury and suicide attempt: A psychopathological perspective. *Journal of Psychopathology, 21*(4), 348–353.

Prinsen, C. A., Mokkink, L. B., Bouter, L. M., Alonso, J., Patrick, D. L., De Vet, H. C., & Terwee, C. B. (2018). COSMIN guideline for systematic reviews of patient-reported outcome measures. *Quality of Life Research, 27*, 1147–1157.

Prinsen, C. A., Vohra, S., Rose, M. R., Boers, M., Tugwell, P., Clarke, M., & Terwee, C. B. (2016). How to select outcome measurement instruments for outcomes included in a "Core outcome set"–a practical guideline. *Trials, 17*(1), 1–10.

Quinlivan, L., Cooper, J., Davies, L., Hawton, K., Gunnell, D., & Kapur, N. (2016). Which are the most useful scales for predicting repeat self-harm? A systematic review evaluating risk scales using measures of diagnostic accuracy. *BMJ Open, 6*(2), Article e009297.

Quinlivan, L., Cooper, J., Meehan, D., Longson, D., Potokar, J., Hulme, T., & Kapur, N. (2017). Predictive accuracy of risk scales following self-harm: Multicentre, prospective cohort study. *The British Journal of Psychiatry, 210*(6), 429–436.

Ribeiro, J. D., Franklin, J. C., Fox, K. R., Bentley, K. H., Kleiman, E. M., Chang, B. P., & Nock, M. K. (2016). Self-injurious thoughts and behaviors as risk factors for future suicide ideation, attempts, and death: A meta-analysis of longitudinal studies. *Psychological Medicine, 46*(2), 225–236.

Robertson, A. E., & Simmons, D. R. (2015). The sensory experiences of adults with autism spectrum disorder: A qualitative analysis. *Perception, 44*(5), 569–586.

Robledo, J., Donnellan, A. M., & Strandt-Conroy, K. (2012). An exploration of sensory and movement differences from the perspective of individuals with autism. *Frontiers in Integrative Neuroscience, 6*, 107.

Rojahn, J., Matson, J. L., Lott, D., Esbensen, A. J., & Smalls, Y. (2001). The behavior problems inventory: An instrument for the assessment of self-injury, stereotyped behavior, and aggression/destruction in individuals with developmental disabilities. *Journal of Autism and Developmental Disorders, 31*, 577–588.

Runeson, B., Haglund, A., Lichtenstein, P., & Tidemalm, D. (2016). Suicide risk after nonfatal self-harm: A national cohort study, 2000-2008. *The Journal of Clinical Psychiatry, 77*(2), 20475.

Sansone, R. A., Wiederman, M. W., & Sansone, L. A. (1998). The self-harm inventory (SHI): Development of a scale for identifying self-destructive behaviors and borderline personality disorder. *Journal of Clinical Psychology, 54*(7), 973–983.

Schunemann, H., Brożek, J., Guyatt, G., & Oxman, A. (2013). GRADE Handbook for Grading Quality of Evidence and Strength of Recommendation. Version 3.2 http://www.cc-ims.net/gradepro.

Sheehan, D. V., Lecrubier, Y., Sheehan, K. H., Amorim, P., Janavs, J., Weiller, E., … Dunbar, G. C. (1998). The Mini-international neuropsychiatric interview (MINI): The development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *Journal of Clinical Psychiatry, 59*(20), 22–33.

Snowden, R. J., Tiley, O., & Gray, N. S. (2023). The Cardiff self-injury inventory (English version): Convergent validity and psychometric properties. *Health Science Reports, 6*(1), Article e1028.

South, M., Ozonoff, S., & McMahon, W. M. (2005). Repetitive behavior profiles in Asperger syndrome and high-functioning autism. *Journal of Autism and Developmental Disorders, 35*(2), 145–158.

Steenfeldt-Kristensen, C., Jones, C. A., & Richards, C. (2020). The prevalence of self-injurious behaviour in autism: A meta-analytic study. *Journal of Autism and Developmental Disorders, 50*(11), 3857–3873.

Stewart, G. R., Corbett, A., Ballard, C., Creese, B., Aarsland, D., Hampshire, A., … Happé, F. (2022). Self-harm and suicidality experiences of middle-age and older adults with vs. without high autistic traits. *Journal of Autism and Developmental Disorders*, 1–13.

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., … de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology, 60*(1), 34–42.

Terwee, C. B., Jansma, E. P., Riphagen, I. I., & de Vet, H. C. (2009). Development of a methodological PubMed search filter for finding studies on measurement properties of measurement instruments. *Quality of Life Research, 18*, 1115–1123.

Terwee, C. B., Prinsen, C. A., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., … Mokkink, L. B. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: A Delphi study. *Quality of Life Research, 27*, 1159–1170.

Thoen, A., Steyaert, J., Alaerts, K., Evers, K., & Van Damme, T. (2021). A systematic review of self-reported stress questionnaires in people on the autism spectrum. *Review Journal of Autism and Developmental Disorders*, 1–24.

Townsend, E., Wadman, R., Sayal, K., Armstrong, M., Harroe, C., Majumder, P., … Clarke, D. (2016). Uncovering key patterns in self-harm in adolescents: Sequence analysis using the card Sort task for self-harm (CaTS). *Journal of Affective Disorders, 206*, 161–168.

Turcotte, P., Mathew, M., Shea, L. L., Brusilovskiy, E., & Nonnemacher, S. L. (2016). Service needs across the lifespan for individuals with autism. *Journal of Autism and Developmental Disorders, 46*, 2480–2489.

Turner, B. J., Chapman, A. L., & Layden, B. K. (2012). Intrapersonal and interpersonal functions of non suicidal self-injury: Associations with emotional and social functioning. *Suicide and Life-threatening Behavior, 42*(1), 36–55.

Vigfusdottir, J., Dale, K. Y., Gratz, K. L., Klonsky, E. D., Jonsbu, E., & Høidal, R. (2020). The psychometric properties and clinical utility of the Norwegian versions of the deliberate self-harm inventory and the inventory of statements about self-injury. *Current Psychology*, 1–11.

Whitlock, J., Exner-Cortens, D., & Purington, A. (2014). Assessment of nonsuicidal self-injury: Development and initial validation of the non-suicidal self-injury–assessment tool (NSSI-AT). *Psychological Assessment, 26*(3), 935.

Widnall, E., Epstein, S., Polling, C., Velupillai, S., Jewell, A., Dutta, R., … Downs, J. (2022). Autism spectrum disorders as a risk factor for adolescent self-harm: A retrospective cohort study of 113,286 young people in the UK. *BMC Medicine, 20*(1), 1–14.