JIDR Journal of Intellectual Disability Research
Published on behalf of mencap and in association with IASSID

Journal of Intellectual Disability Research doi: 10.1111/jir.12924

VOLUME 66 PART 4 pp 376–391 APRIL 2022

An international field study of the ICD-11 behavioural indicators for disorders of intellectual development

K. R. Lemay,[1] C. S. Kogan,[1] T. J. Rebello,[2] J. W. Keeley,[3] R. Bhargava,[4] P. Sharan,[5] M. Sharma,[6] J. V. S. Kommu,[7] M. T. Kishore,[8] J. de Jesus Mari,[9] P. Ginige,[10] S. Buono,[11] M. Recupero,[11] M. Zingale,[11] T. Zagaria,[11] S. Cooray,[12] A. Roy[13,14] & G. M. Reed[15]

1 *School of Psychology, University of Ottawa, Ottawa, Ontario, Canada*
2 *Department of Psychiatry, Vagelos College of Physicians and Surgeons, Columbia University, and Research Foundation for Mental Hygiene, New York, NY, USA*
3 *Department of Psychology, Virginia Commonwealth University, Richmond, VA, USA*
4 *Department of Psychiatry and National Drug Dependence Treatment Centre, All India Institute of Medical Sciences, New Delhi, India*
5 *Department of Psychiatry, All India Institute of Medical Sciences, New Delhi, India*
6 *School of Social Sciences, Indira Gandhi National Open University, New Delhi, India*
7 *Department of Child and Adolescent Psychiatry, National Institute of Mental Health and Neurosciences, Bengaluru, India*
8 *Department of Clinical Psychology, National Institute of Mental Health and Neuro Sciences, Bengaluru, India*
9 *Department of Psychiatry, Universidade Federal de São Paulo, São Paulo, Brazil*
10 *Faculty of Medicine, University of Peradeniya, Kandy, Sri Lanka*
11 *Department of Psychology, Oasi Research Institute-IRCCS, Troina, Italy*
12 *Faculty of Psychiatry of Learning Disability, Royal College of Psychiatrists, London, UK*
13 *Coventry and Warwickshire Partnership NHS Trust, Birmingham, UK*
14 *Centre for Educational Development, Appraisal and Research, University of Warwick, Coventry, UK*
15 *Department of Psychiatry, Columbia University Vagelos College of Physicians and Surgeons, New York, NY, USA*

Abstract

Background The World Health Organization (WHO) has approved the 11th Revision of the International Classification of Diseases (ICD-11). A version of the ICD-11 for Mental, Behavioural and Neurodevelopmental Disorders for use in clinical settings, called the Clinical Descriptions and Diagnostic Requirements (CDDR), has also been developed. The CDDR includes behavioural indicators (BIs) for assessing the severity of disorders of intellectual development (DID) as part of the section on neurodevelopmental disorders. Reliable and valid diagnostic assessment measures are needed to improve identification and treatment of individuals with DID. Although appropriately normed, standardised intellectual and adaptive behaviour assessments are considered the optimal assessment approach in this area, they are unavailable in many parts of the world. This field study tested the BIs internationally to assess the inter-rater reliability, concurrent validity, and clinical utility of the BIs for the assessment of DID.

Correspondence: Dr. Cary S. Kogan, School of Psychology, University of Ottawa, Ottawa, Ontario, Canada (e-mail: ckogan@uottawa.ca).

Methods This international study recruited a total of 206 children and adolescents (5–18 years old) with a suspected or established diagnosis of DID from four sites across three countries [Sri-Lanka (n = 57), Italy (n = 60) and two sites in India (n = 89)]. Two clinicians assessed each participant using the BIs with one conducting the clinical interview and the other observing. Diagnostic formulations using the BIs and clinical utility ratings were collected and entered independently after each assessment. At a follow-up appointment, standardised measures (Leiter-3, Vineland Adaptive Behaviour Scales-II) were used to assess intellectual and adaptive abilities.
Results The BIs had excellent inter-rater reliability (intra-class correlations ranging from 0.91 to 0.97) and good to excellent concurrent validity (intra-class correlations ranging from 0.66 to 0.82) across sites. Compared to standardised measures, the BIs had more diagnostic overlap between intellectual and adaptive functioning. The BIs were rated as quick and easy to use and applicable across severities; clear and understandable with adequate to too much level of detail and specificity to describe DID; and useful for treatment selection, prognosis assessments, communication with other health care professionals, and education efforts.
Conclusion The inclusion of newly developed BIs within the CDDR for ICD-11 Neurodevelopmental Disorders must be supported by information on their reliability, validity, and clinical utility prior to their widespread adoption for international use. BIs were found to have excellent inter-rater reliability, good to excellent concurrent validity, and good clinical utility. This supports use of the BIs within the ICD-11 CDDR to assist with the accurate identification of individuals with DID, particularly in settings where specialised services are unavailable.

Keywords Behavioural indicators, Classification, Clinical utility, Disorders of intellectual development, ICD-11, Intellectual disability, Learning disability (UK), Reliability, Validity

Background

The International Classification of Diseases (ICD) and Related Health Problems was developed by the World Health Organization (WHO) as a global diagnostic classification system for recording, reporting, and grouping health-related factors and conditions (World Health Organization 1992, 2014). The WHO finalised the 11th revision of the ICD (ICD-11) after approval by the World Health Assembly in May 2019. The WHO Mental Health and Substance Use Department also developed a version of the ICD-11 for Mental, Behavioural and Neurodevelopmental Disorders that includes more comprehensive information necessary for implementation in clinical settings, called the Clinical Descriptions and Diagnostic Requirements (CDDR) (Reed 2010; Reed *et al.* 2019). [The ICD-11 CDDR was previously referred to as the Clinical Descriptions and Diagnostic Guidelines (CDDG), but the name was changed due to the recent development of WHO policies related to documents referred to as 'guidelines', which were not considered to be applicable.] The CDDR provides information on essential (required) features, developmental presentations, differential diagnoses, boundary differentiation between normality and other disorders, culture-related features, and other features for each disorder (First *et al.* 2015). A core focus in the development of the ICD-11 chapter on Mental, Behavioural and Neurodevelopmental Disorders and the CDDR was to maximise reliability, validity, clinical utility and global applicability (International Advisory Group for the Revision of the ICD Mental and Behavioural Disorders 2011; First *et al.* 2015).

Although existing intellectual and adaptive behaviour assessment measures are comprehensive and have robust psychometric properties, the global applicability of these measures is lacking (Robertson *et al.* 2012). For instance, underserved regions and some low-income and middle-income countries (LAMICs) may lack linguistically appropriate measures that are normed, standardised, and psychometrically validated to assess intellectual functioning and/or adaptive behaviours (Robertson *et al.* 2012; Kishore *et al.* 2019). Most measures also require specialised services and training to administer that are costly, thus rendering them unavailable or inaccessible in various contexts (Robertson *et al.* 2012; Kishore *et al.* 2019). The ICD-11 aims to address this disparity by embedding a comprehensive set of behavioural indicator (BI) tables that are affordable, accessible, and clinically useful within the CDDR for those lacking access to appropriate

resources (refer to Tassé *et al.* 2019). The use of BIs in clinical practice aims to provide further clarity, reliability, and validity by summarising behavioural observations and retrospective reports that are typically observed in the population of interest. Specific BIs characterising severity levels for disorders of intellectual development (DID) are lacking (Carulla *et al.* 2011).

It has been estimated that globally 0.62–3.00% of individuals have DID (Harris 2006; McKenzie *et al.* 2016), with mild DID being the most prevalent (85%) diagnosis (King *et al.* 2009). The ICD-11 conceptualisation of intellectual and adaptive functioning are improved by replacing the concept of arrested or incomplete development of the mind as a required feature (WHO 1992; Girimaji and Pradeep 2018). The ICD-11 also highlights the significance of impairment in adaptive behaviours together with impairment of intellectual functioning when determining severity level and diagnosing DID at various developmental stages (Schalock *et al.* 2002). Adaptive behaviour is defined according to functioning within conceptual, social, and practical skills (WHO 2018). The essential features in the CDDR for DID include significant limitations (operationalised as two or more standard deviations below the mean on standardised measures) of both intellectual functioning and adaptive behaviours with onset during the developmental period. When normed and standardised tests are available, limitations are determined separately for intellectual functioning, whereas conceptual, social, and practical domains are assessed independently and then combined to determine the overall level of adaptive behaviour. In ICD-11, there are four levels of severity (mild, moderate, severe, and profound) assigned according to intellectual functioning and adaptive behaviour functioning (WHO 2018). The BIs provide additional guidance for interpreting levels of severity for DID with the presence of co-occurring disorders. For example, when an individual with DID has co-occurring autism spectrum disorder (ASD), clinicians are advised to interpret results by placing less emphasis on the social skills domain in the assessment of adaptive behaviour functioning as this is a core feature of ASD (Ayuso-Mateos and Kogan 2020). Although the BIs for DID were developed based on expert input and factor analysis (Tassé *et al.* 2019), formal validation in different clinical settings used by diverse clinicians is necessary to assess whether the BIs function as intended (First *et al.* 2015).

Field testing of revised classification systems ensures that the proposed changes to diagnoses and diagnostic requirements function as intended and are considered an improvement over previous conceptualisations (Keeley *et al.* 2016). The present study is a multi-site international field study that evaluates reliability, validity, and clinical utility in settings where the classification will be used. The aim of this study was to assess the psychometric properties of the newly developed BIs to inform their global implementation with individuals with DID. The objectives of this study were to (1) field test the newly developed BIs embedded in the ICD-11 CDDR for classification of DID severity in three countries with different languages, cultures, and income levels; and (2) assess and evaluate the clinical utility, inter-rater reliability, and concurrent validity of the BIs within the ICD-11 CDDR.

Methods

This quantitative, observational (correlational) design is an international, collaborative, and cross-cultural psychometric study that was coordinated at the University of Ottawa in Canada. This study received ethics approval from the University of Ottawa Office of Research Ethics and Integrity (protocol no. H-03-18-371). Ethics approval was also received at each participating site [India: National Institute of Mental Health & Neuro Sciences (NIMHANS); All India Institute of Medical Sciences, New Delhi (AIIMS), Sri Lanka: Teaching Hospital Peradeniya, Italy: Oasi Research Institute-IRCCS (Oasi IRCCS)].

Participants

A total of 206 children and adolescents (aged 5 to 18 years) with a suspected or established diagnosis of DID were recruited in three countries at four study sites [India: NIMHANS (n = 35); AIIMS (n = 54), Sri Lanka: Teaching Hospital Peradeniya (n = 57), Italy: Oasi IRCCS (n = 60)]. Participating study clinicians, site coordinators, and site supervisors were required to attend an online clinician training on the ICD-11 CDDR for Neurodevelopmental Disorders and a procedural training for the BIs prior to data

collection. All clinician raters had a degree in psychology or psychiatry and were adequately trained to administer psychological tests.

The following inclusion/exclusion criteria were applied when assessing eligibility of participants: (1) aged 5 to 18 years at the time of recruitment; (2) known or suspected of having DID with or without co-occurring ASD; (3) not under the care of clinician raters; (4) not currently incapacitated due to severe physical illness or pain; (5) not currently at imminent risk of self-harm, danger to others or experiencing serious medication side effects requiring immediate medical attention. Participants under the age of five were excluded as these individuals might have received a provisional diagnosis. Adults were excluded due to feasibility given that some study sites only serve children and adolescents. Participants with severe behavioural disorders and/or severe motor impairments and/or those who appeared to not be able to sustain attention long enough for valid scores on the test of intellectual functioning were excluded at the outset of recruitment. Although data on standardised measures of intellectual functioning were unobtainable for some individuals due to their inability to complete cognitive testing, we sought to collect a sample of individuals encompassing the full range of severity.

Materials

Data collection process

All study activities were tracked by site coordinators using REDCap, a secure web platform designed to support data collection, management, and monitoring (Harris *et al.* 2009). Data collected and monitored included consent and registration, clinical interview scheduling, clinical rater assignment (clinician ID, name and role) and date of each assessment, reporting of adverse events, and confirmation of study termination. Clinicians entered their diagnostic conceptualisation and an assessment of the clinical utility of the CDDR and the BIs on an Electronic Field Study System (EFSS), a web-based data collection platform developed using Qualtrics (Provo, USA). Approximately 30–40 min were required to complete data entry. Each site, clinician, and participant were provided unique identification codes. All data were monitored, stored, and managed centrally at the University of Ottawa. Although the clinical interview could be conducted in the local languages, English was used for the standardised measures and data collection by all sites except one (Italy) where a local translation of the requirements was needed by participating clinicians.

Demographics

Both clinician and participant information were collected. Clinicians filled out an online registration form to verify their country and site, name and ID, date of birth, country of origin, language preferences, profession, and years of formal professional experience (Table 1). Participants' demographic information included date of birth, developmental stage, and gender (Table 2). Participant characteristics (i.e., previous medical diagnoses, previous psychiatric diagnoses, additional psychiatric symptoms, current medications, and previous treatment) were also collected (Table 3).

11th Revision of the International Classification of Diseases behavioural indicators and Neurodevelopmental Disorders requirements

The ICD-11 BIs are embedded in the CDDR for ICD-11 for DID as an alternative way to assess intellectual and adaptive behaviour functioning when appropriately normed and standardised tests are not available. BI are provided to distinguish severity levels (i.e., mild, moderate, severe and profound) for three age groups (0–5; 6–18; +18 years old) in order to help characterise the needs of people with DID (refer to Tassé *et al.* 2019). The BIs for adaptive behaviour functioning are identified for each adaptive skill (i.e., conceptual, social, and practical). Clinicians were provided the ICD-11 Neurodevelopmental Disorders requirements with the embedded BIs for DID.

Vineland Adaptive Behaviour Scale – Second Edition

The Vineland Adaptive Behaviour Scale – Second Edition (VABS-II) is a widely used, individually administered interview designed to measure adaptive behaviour (Sparrow *et al.* 2005). The structured interview version was selected as a measure to assess the three domains of adaptive behaviour that reflect areas assessed by the BIs. The domains

Table 1 Clinician demographics by site

Variable	Total (*N* = 51)	Italy (Oasi IRCCS; *N* = 5)	Sri Lanka (Teaching Hospital Peradeniya; *N* = 5)	India (AIIMS; *N* = 18)	India (NIMHANS; *N* = 23)
Age, years (Mean ± SD)	35.2 ± 8.7	47.8 ± 10.5	44.4 ± 13.1	28.9 ± 2.6	35.4 ± 5.4
Gender, *N* (%)					
Men	26 (51.0)	0	0	13 (72.2)	13 (56.5)
Women	25 (49.0)	5 (100)	5 (100)	5 (27.8)	10 (43.5)
Level of proficiency in English, *N* (%)					
Low	1 (2.0)	0	0	0	1 (4.3)
Intermediate	8 (15.7)	5 (100)	0	0	3 (13.0)
Advanced	16 (31.4)	0	4 (80.0)	8 (44.4)	4 (17.4)
Completely Fluent	26 (51.0)	0	1 (20.0)	10 (55.6)	15 (65.2)
Clinical Profession, *N* (%)					
Psychiatry	39 (76.5)	0	3 (60.0)	16 (88.9)	20 (87.0)
Psychology	9 (17.6)	5 (100)	0	2 (11.1)	2 (8.7)
Other	3 (6.0)	0	2 (40.0)	0	1 (4.3)
Years of experience, (Mean ± SD)	6.6 ± 7.7	22.2 ± 11.1	9.0 ± 8.4	1.9 ± 1.3	6.4 ± 4.8

Table 2 Participant demographics by site

Variable	Total (*N* = 206)	Italy (Oasi IRCCS; *N* = 60)	Sri Lanka (Teaching Hospital Peradeniya; *N* = 57)	India (AIIMS; *N* = 54)	India (NIMHANS; *N* = 35)
Age, years (Mean ± SD)		9.55 ± 3.68	10.3 ± 3.59	12.6 ± 3.66	12.21 ± 3.834
Gender, *N* (%)					
Male	142 (69)	44 (73)	39 (68)	37 (69)	22 (63)
Female	64 (31)	16 (27)	18 (32)	17 (31)	13 (37)

and subdomains used in this study were communication (receptive, expressive, and written), daily living skills (personal, domestic, and community) and socialisation (interpersonal relationships, play and leisure time, and coping skills). Scores for each subdomain are summed to yield the domain composite scores, which are then summed to form the adaptive behaviour composite (ABC). Raw scores were converted to standard scores, V-scale scores, and percentile ranks. The VABS communication domain was compared with BIs conceptual skills, the VABS socialisation domain was compared with the BIs social skills, and the VABS daily living skills was compared with the BIs practical skills.

The inter-rater reliability reported in the literature across the three domains was between 0.71–0.83 (Sparrow *et al.* 2005). The validity for the VABS-II was deemed robust based on test content, response process, test structure, clinical groups, and relationships to other measures (Sparrow *et al.* 2005). The internal consistency reported in the literature for the three domains and ABC were between 0.84–0.93 and 0.93–0.97, respectively (Sparrow *et al.* 2005). The internal consistency in our study across all sites were found to be acceptable for communication (n = 215; α = 0.74), good for daily living skills (n = 213; α = 0.83), excellent for socialisation (n = 215; α = 0.90), and excellent for the ABC (n = 213; α = 0.93).

Table 3 Participant characteristics by site

Variable	Total (*N* = 206)	Italy (Oasi IRCCS; *N* = 60)	Sri Lanka (Teaching Hospital Peradeniya; *N* = 57)	India (AIIMS; *N* = 54)	India (NIMHANS; *N* = 35)
Previous medical diagnoses, *N* (%)					
Epilepsy/Seizure disorder	27 (13)	9 (15)	5 (9)	9 (17)	4 (11)
Down syndrome	5 (2)	0	3 (5)	0	2 (6)
Dysmorphism	3 (1)	3 (5)	0	0	0
Previous psychiatric diagnoses, *N* (%)					
Disorders of intellectual development	20 (10)	13 (22)	0	4 (7)	3 (9)
Attention-deficit hyperactivity disorder	17 (8)	5 (8)	7 (12)	0	5 (14)
Autism spectrum disorders	10 (5)	10 (17)	0	0	0
Disorders of speech and language	9 (4)	6 (10)	0	0	3 (9)
Mixed specific developmental disorders	5 (2)	5 (8)	0	0	0
Global developmental delay	5 (2)	5 (8)	0	0	0
Behavioural disorders	3 (1)	3 (5)	0	0	0
Emotional disorders	3 (1)	3 (5)	0	0	0
Learning disorder	3 (1)	3 (5)	0	0	0
Obsessive compulsive disorder	3 (1)	3 (5)	0	0	0
Additional psychiatric symptoms, *N* (%)					
Attention related difficulties	18 (9)	5 (8)	8 (14)	2 (4)	3 (9)
Aggressiveness	6 (3)	3 (5)	0	3 (6)	0
Irritability	5 (2)	0	0	3 (6)	2 (6)
Emotional disturbances	3 (1)	3 (5)	0	0	0
Oppositional defiant features	3 (1)	0	3 (5)	0	0
Anxiety in social situations	3 (1)	0	0	0	3 (9)
Difficulty with academics	2 (1)	0	0	0	2 (6)
Current medications in use, *N* (%)					
Atypical antipsychotic	30 (28)	7 (12)	0	21 (39)	2 (6)
Anticonvulsant	28 (26)	8 (13)	9 (16)	7 (13)	4 (11)
Stimulant	10 (9)	0	8 (14)	0	2 (6)
Hormone	4 (4)	2 (3)	2 (4)	0	0
Benzodiazepines	5 (2)	0	0	5 (9)	0
Antihypertensive	2 (1)	0	0	0	2 (6)
Selective serotonin reuptake inhibitor	2 (1)	0	0	0	2 (6)
Previous treatment, *N* (%)					
Psychomotor therapy	38 (36)	38 (63)	0	0	0
Speech therapy	15 (15)	15 (25)	0	0	0
Psychoeducational intervention	6 (3)	6 (10)	0	0	0

Note: Only variables that affected >5% of the sample at any site are reported.

Leiter International Performance Scale – Third Edition (Leiter-3)

The Leiter-3 (Roid *et al.* 2013) is a norm-referenced individually administered non-verbal measure of intelligence. The four subtests required to derive the nonverbal intellectual functioning include figure ground, form completion, classification and analogies, and sequential order. Raw scores on each test were converted into normalised/scaled scores (mean = 10, standard deviation = 3), percentiles, and age equivalence in accordance with the manual (Roid *et al.* 2013). The sum of scaled scores is used to establish nonverbal intellectual functioning scores for each individual (Roid *et al.* 2013). The nonverbal intellectual functioning domain was compared with the BIs intellectual functioning domain.

Evidence of adequate to good content, criterion, and construct validity has been reported in the literature (Roid *et al.* 2013). The internal consistency reported in the literature for the four subtests and nonverbal intellectual functioning across all age levels were between 0.78–0.95 and 0.94–0.98, respectively (Roid *et al.* 2013). The internal consistency in our study across all sites was found to be excellent for non-verbal intellectual functioning (n = 211; α = 0.93).

Clinical utility questions

Clinical utility of the CDDR including the BIs were analysed quantitatively [i.e., eight questions scored on a 4-point Likert scale from 0 (*not useful/clear at all*) to 4 (*extremely useful/clear*)] on interpretability, ease of application, usefulness for treatment selection, prognosis assessments, patient communication, and education efforts.

Procedure

After obtaining the ethical approval from their respective institutional ethics body, each site recruited participants to the study by referral, and child assent and parent/caregiver consent were obtained. At each site, participants were assessed by a clinician rater registered in their country to assess and manage neurodevelopmental disorders using the ICD-11 CDDR for DID, including the BIs, to ascertain their intellectual and adaptive behaviour functioning. A second clinician observed the assessment conducted by the primary interviewer (i.e., the first clinician), and was instructed to only ask follow-up questions at the end of the assessment. Clinicians were prevented from accessing prior diagnoses (when applicable) and were instructed to collect sufficient clinical information to make a diagnosis regarding the index condition as well as additional mental disorders. Clinicians then independently entered their diagnostic conceptualisation and completed an assessment of the clinical utility of the CDDR including the BIs on the EFSS data collection platform within 48 h of assessment. This was performed to reduce recall bias. These data were used to establish inter-rater reliability of the BIs. The child or adolescent was also assessed by a separate clinician rater using standardised, normed instruments to determine their nonverbal intellectual (Leiter-3) and adaptive behaviour (VABS-II) functioning to establish the concurrent validity of the BIs with standardised tests.

Data analysis

Data were scanned for extreme outliers (z score $> \pm$ 3.29), which were adjusted in the dataset through winsorisation (Kwak and Kim 2017). Data normality for the Leiter-3 and VABS-II were assessed using tests of skewness and kurtosis, the Shapiro–Wilk test of normality, and visually using histograms. Subsequently, all data were analysed using SPSS for Windows (version 24; IBM Corp 2016). The data were analysed based on the final diagnostic formulations for DID. Inter-rater reliability and concurrent validity were assessed using a variant of the intra-class correlation (ICC; Hallgren 2012). Data were combined across sites to increase the power to detect differences across severity levels. In the present study, clinicians were not able to select 'not applicable' when designating severity levels for domains in adaptive functioning. Although this was not an issue for intellectual functioning, as this is an essential feature for DID, this may have been an issue for selecting severity levels for adaptive behaviour as an individual could have normal functioning in one domain and be found to be in the clinical range (e.g., 'mild') in the other two domains, likely resulting in an overall rating of adaptive behaviour in the clinical range. To account for this issue, the concurrent validity was analysed two ways: first using the raw data provided (concurrent validity – raw); second by collapsing functioning that was considered higher than meeting criteria for 'mild' on the Vineland-2 (i.e., 'not applicable' and 'mild' were all considered 'mild' cases; concurrent validity – adjusted). Therefore, the concurrent validity for the BIs in this study are likely found between the raw and adjusted values; both are presented. The ICC is a one-way, random effects, single-measures model. Internal consistency of the BIs (both primary and observer clinician raters) and standardised measures were evaluated using nonparametric Kendall's tau-b correlation coefficients to examine the correlations between the final diagnostic formulation variables (i.e., intellectual functioning, adaptive behaviours, and overall severity) for DID. The validity, reliability and internal consistency of the BIs were compared to

previously established criteria (Hunsley and Mash 2008). Descriptive statistics were performed to assess the clinical utility of the BIs for DID embedded within the ICD-11 Neurodevelopmental Disorders requirements.

Results

Clinician demographics are reported in Table 1. On average, clinicians were English-proficient psychiatrists and clinical psychologists between their late twenties to late forties with approximately 7 years of experience. Participant demographics are reported in Table 2. On average, participants were 11 years old and most were male. Participant characteristics (combined and individual countries/sites) are reported in Table 3. Most participants had not been previously diagnosed with medical or psychiatric conditions, did not present with other psychiatric symptoms, had not received treatment, and were not taking medication.

Psychometric analyses

The ICCs for inter-rater reliability and concurrent validity for the BIs collapsed across sites are shown in Table 4. The inter-rater reliability across domains and overall functioning were excellent (ICCs > 0.90). The concurrent validity (raw and adjusted) was good for intellectual functioning (0.71–0.73), excellent for conceptual skills (0.75–0.77), good for social skills (0.66–0.68), good for practical skills (0.68–0.73), and excellent for overall functioning (0.76–0.82). All ICCs were statistically significant ($P < 0.05$).

Table 4 Intra-class correlations (ICCs) for ICD-11 BIs for individuals receiving a DID diagnosis collapsed across sites

Variables	Inter-rater reliability (*n*)	Concurrent validity – raw (*n*)	Concurrent validity – adjusted (*n*)
Intellectual	0.97 (175)	0.71 (167)	0.73 (167)
Conceptual	0.91 (175)	0.75 (172)	0.77 (172)
Social	0.93 (175)	0.66 (172)	0.68 (172)
Practical	0.94 (175)	0.68 (172)	0.73 (172)
Overall	0.97 (176)	0.76 (166)	0.82 (166)

Note: All results were statistically significant at $p < 0.05$ (two-tailed).

Internal consistency of the BIs are presented in Tables 5 and 6 for diagnostic formulations conducted by the primary clinician and observer clinician, respectively. Correlation coefficients were comparable between clinicians for the BIs, ranging from 0.84 to 0.97. All correlations were statistically significant ($P < 0.001$, two-tailed). Internal consistencies of the standardised measures (Leiter-3 and VABS-II) are presented in Table 7. Correlation coefficients for the standardised measures ranged from 0.53–0.83.

Clinical utility of the ICD-11 CDDR and BIs are presented for single diagnoses of DID in Table 8, and for co-occurring diagnoses such as DID with ASD and DID with other neurodevelopmental (e.g., attention deficit hyperactivity disorder), mental (e.g., major depressive disorder) and/or behavioural disorders (e.g., oppositional defiant disorder) in Table 9. An independent samples *t* test grouping single diagnosis with co-occurring diagnoses for intellectual, adaptive functioning and overall functioning across sites had *P* values above 0.05, indicating that there were no significant differences between single and co-occurring presentations. In general, most clinicians indicated that the BIs within the ICD-11 CDDR were quite to extremely clear and understandable, that the BIs had about the right to too much amount of detail and specificity; were quite easy to apply across severities; were shorter or took about the same amount of time to apply as their usual clinical practice; and were quite useful for treatment selection, prognosis assessments, communication with other health care professionals, and education efforts.

Discussion

The ICD-11 includes BIs for DID within the CDDR for Neurodevelopmental Disorders to improve assessment and diagnostic practices for individuals, particularly in settings where appropriately normed and standardised testing is not available or not feasible (Tassé *et al.* 2019). This international field study examined the inter-rater reliability, concurrent validity, and clinical utility of the BIs in three countries to inform the global implementation of the ICD-11 and the CDDR. The ability to recognise, detect, and assign severity levels for intellectual, adaptive, and overall levels of functioning for DID in

Table 5 Internal consistency (Kendall's tau-b) and 95% confidence intervals [CIs] of the primary clinician ratings using the BIs

	1	2	3	4	5
1. Intellectual (*n* = 177)	–	–	–	–	–
2. Conceptual (*n* = 177)	0.89 [0.86–0.91]	–	–	–	–
3. Social (*n* = 177)	0.88 [0.86–0.90]	0.88 [0.86–0.90]	–	–	–
4. Practical (*n* = 177)	0.89 [0.87–0.91]	0.84 [0.80–0.86]	0.88 [0.85–0.90]	–	–
5. Overall (*n* = 177)	0.97 [0.96–0.97]	0.91 [0.89–0.93]	0.92 [0.90–0.93]	0.91 [0.89–0.92]	–

Note: All results were statistically significant at $p < 0.001$ (two-tailed).

Table 6 Internal consistency (Kendall's tau-b) of the observer clinician ratings using the BIs

	1	2	3	4	5
1. Intellectual (*n* = 179)	–	–	–	–	–
2. Conceptual (*n* = 179)	0.90 [0.88–0.92]	–	–	–	–
3. Social (*n* = 179)	0.83 [0.80–0.86]	0.84 [0.81–0.87]	–	–	–
4. Practical (*n* = 179)	0.88 [0.85–0.90]	0.86 [0.83–0.89]	0.85 [0.82–0.88]	–	–
5. Overall (*n* = 179)	0.96 [0.95–0.96]	0.93 [0.92–0.94]	0.88 [0.85–0.90]	0.92 [0.90–0.93]	–

Note: All results were statistically significant at $p < 0.001$ (two-tailed).

Table 7 Internal Consistency (Kendall's tau-b) of the standardised measures clinician ratings using the Leiter-3 (nonverbal intellectual functioning) and the VABS-II (communication, socialisation and daily living skills)

	1	2	3	4	5
1. Nonverbal Intellectual Functioning (*n* = 113)	–	–	–	–	–
2. Communication (*n* = 112)	0.60 [0.53–0.66]	–	–	–	–
3. Socialisation (*n* = 112)	0.53 [0.46–0.60]	0.76 [0.71–0.80]	–	–	–
4. Daily living skills (*n* = 112)	0.59 [0.52–0.65]	0.75 [0.70–0.79]	0.75 [0.71–0.79]	–	–
5. Overall (*n* = 112)	0.72 [0.67–0.77]	0.83 [0.80–0.86]	0.81 [0.77–0.84]	0.83 [0.80–0.86]	–

Note: All variables were $P < 0.001$ (two-tailed).

a valid and reliable way through direct behavioural observations and informant reported behaviours would be a major innovation with the potential to impact the identification and treatment of people with DID around the world, particularly in low- and middle-income countries (LAMICs) where most of the world's population lives (The World Bank Group 2022).

Inter-rater reliability of the BIs was established by comparing the severity levels of each of the domains (i.e., intellectual, conceptual, social, and practical) and the overall severity level assigned using the BIs by independent clinicians presented with the same clinical information. The results of the ICCs for inter-rater reliability across sites was found to be excellent (Table 4; Hunsley and Mash 2008), indicating that clinicians derive similar diagnoses with the BIs when presented with the same information. Overall, compared to other standardised, normed, and validated measures, the ICD-11 BIs were generally consistent with other studies of measures of intellectual functioning [e.g., Wechsler Intelligence Scale for Children – Fourth and Fifth Edition (WISC-I*V*/V; Wechsler 2014)], and adaptive behaviour functioning [e.g., Adaptive Behaviour Assessment System – Third Edition (ABAS-3;

Table 8 Clinical utility of the ICD-11 CDDR and BIs across DID severity from all clinician raters

Clinical utility variables	Response options	DID severity (*n* = 267)			
		Mild (*n* = 135)	Moderate (*n* = 61)	Severe (*n* = 37)	Profound (*n* = 32)
Extent to which the diagnostic requirements were clear and understandable overall as applied to this patient	Not at all clear and understandable	0	0	0	0
	Somewhat clear and understandable	5 (4)	8 (13)	3 (8)	0
	Quite clear and understandable	114 (84)	45 (74)	33 (89)	29 (91)
	Extremely clear and understandable	16 (12)	8 (13)	1 (3)	3 (9)
Level of detail and specificity of the disagnostic requirements/BIs for the diagnosis/diagnoses applied to this patient	Insufficient detail and specificity	4 (3)	2 (3)	1 (3)	0
	About the right amount of detail and specificity	77 (57)	47 (77)	31 (84)	32 (100)
	Too much detail and specificity	54 (40)	12 (20)	5 (13)	0
Extent to which the ICD-11 CDDR imposed requirements that were difficult to assess/apply to this patient	Very difficult to apply	0	0	0	0
	Somewhat difficult to apply	15 (11)	5 (8)	3 (8)	2 (6)
	Quite easy to apply	112 (83)	51 (84)	31 (84)	29 (91)
	Extremely easy to apply	8 (6)	5 (8)	3 (8)	1 (3)
Amount of time needed to apply all of the diagnostic requirements/BIs to this patient compared to usual clinical practice	Much longer than my usual clinical practice	0 (0)	1 (2)	0 (0)	0 (0)
	Somewhat longer than my usual clinical practice	41 (30)	14 (23)	10 (27)	4 (13)
	About the same as my usual clinical practice	86 (64)	41 (67)	23 (62)	27 (84)
	Shorter than my usual clinical practice	8 (6)	5 (8)	4 (11)	1 (3)
How useful would the diagnostic requirements be in helping to select a treatment for this patient?	Not at all useful	0 (0)	0 (0)	0 (0)	1 (3)
	Somewhat useful	20 (15)	9 (15)	4 (11)	3 (9)
	Quite useful	112 (83)	48 (79)	33 (89)	28 (88)
	Extremely useful	3 (2)	4 (7)	0 (0)	0 (0)
How useful would the diagnostic requirements be in helping you to assess this patient's prognosis	Not at all useful	3 (2)	0 (0)	0 (0)	1 (3)
	Somewhat useful	26 (19)	9 (15)	4 (11)	4 (13)
	Quite useful	99 (73)	50 (82)	32 (86)	27 (84)
	Extremely useful	7 (5)	2 (3)	1 (3)	0 (0)
Communication with other health care professionals	Not at all useful	0 (0)	0 (0)	0 (0)	0 (0)
	Somewhat useful	10 (7)	5 (8)	2 (5)	5 (16)
	Quite useful	112 (83)	51 (84)	33 (89)	27 (84)
	Extremely useful	13 (10)	5 (8)	2 (5)	0 (0)
Educating patients/family about condition	Not at all useful	0 (0)	0 (0)	0 (0)	0 (0)
	Somewhat useful	26 (19)	9 (15)	4 (11)	3 (9)
	Quite useful	97 (72)	45 (74)	32 (86)	29 (91)
	Extremely useful	12 (9.0)	7 (11)	1 (3)	0 (0)

Harrison and Oakland 2015)] reported in the literature on inter-rater reliability.

Concurrent validity of the BIs was established by comparing the severity levels of each of the domains (i.e., intellectual, conceptual, social, and practical) and the overall severity level that clinicians' assigned using the BIs with those determined by the standardised measures (i.e., VABS-II and Leiter-3). The results of the ICCs using both methods of analysis ranged from good to excellent levels of reliability (Table 4; Hunsley and Mash 2008). Lower concurrent validity data may be due, in part, to the nature of combining two separate standardised measures with different administration methods and domains and comparing them to a single measure (i.e., the BIs). Specifically, the Leiter-3 is a test of

Table 9 Clinical utility of the ICD-11 CDDR and BIs for co-occurring diagnoses

Clinical utility variables	Response options	DID and ASD (*n* = 29)				DID and other neurodevelopmental/ mental/behavioural disorders (*n* = 91)			
		Mild (*n* = 11)	Moderate (*n* = 10)	Severe (*n* = 8)	Profound (*n* = 0)	Mild (*n* = 57)	Moderate (*n* = 23)	Severe (*n* = 8)	Profound (*n* = 3)
Extent to which the diagnostic requirements were clear and understandable overall as applied to this patient	Not at all clear and understandable	0	0	0	0	0	0	0	0
	Somewhat clear and understandable	0	1 (10)	0	0	1 (2)	1 (4)	2 (25)	0 (0)
	Quite clear and understandable	6 (55)	7 (70)	4 (50)	0 (0)	50 (88)	14 (61)	6 (75)	3 (100)
	Extremely clear and understandable	5 (45)	2 (20)	4 (50)	0 (0)	6 (10)	8 (35)	0 (0)	0 (0)
Level of detail and specificity of the disagnostic requirements/BIs for the diagnosis/diagnoses applied to this patient	Insufficient detail and specificity	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	About the right amount of detail and specificity	7 (64)	4 (40)	6 (75)	0 (0)	29 (51)	12 (52)	5 (63)	2 (67)
	Too much detail and specificity	4 (36)	6 (60)	2 (25)	0 (0)	28 (49)	11 (48)	3 (37)	1 (33)
Extent to which the ICD-11 CDDR imposed requirements that were difficult to assess/ apply to this patient	Very difficult to apply	0 (0)	1 (10)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	Somewhat difficult to apply	2 (18)	0 (0)	0 (0)	0 (0)	2 (3)	1 (4)	0 (0)	0 (0)
	Quite easy to apply	6 (55)	6 (60)	5 (63)	0 (0)	50 (88)	18 (78)	7 (88)	3 (100)
	Extremely easy to apply	3 (27)	3 (30)	3 (37)	0 (0)	5 (9)	4 (17)	1 (12)	0 (0)
Amount of time needed to apply all of the diagnostic requirements/BIs to this patient compared to usual clinical practice	Much longer than my usual clinical practice	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (4)	0 (0)	0 (0)
	Somewhat longer than my usual clinical practice	1 (9)	0 (0)	0 (0)	0 (0)	14 (25)	2 (9)	2 (25)	1 (33)
	About the same as my usual clinical practice	5 (46)	9 (90)	6 (75)	0 (0)	36 (63)	14 (61)	6 (75)	2 (67)
	Shorter than my usual clinical practice	5 (45)	1 (10)	2 (25)	0 (0)	7 (12)	6 (26)	0 (0)	0 (0)
	Not at all useful	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)

Table 9. *(Continued)*

Clinical utility variables	Response options	DID and ASD (n = 29)				DID and other neurodevelopmental/ mental/behavioural disorders (n = 91)			
		Mild (n = 11)	Moderate (n = 10)	Severe (n = 8)	Profound (n = 0)	Mild (n = 57)	Moderate (n = 23)	Severe (n = 8)	Profound (n = 3)
How useful would the diagnostic requirements be in helping to select a treatment for this patient?	Somewhat useful	0 (0)	1 (10)	2 (25)	0 (0)	9 (16)	4 (17)	3 (38)	1 (33)
	Quite useful	10 (91)	9 (90)	6 (75)	0 (0)	47 (82)	17 (74)	5 (62)	2 (67)
	Extremely useful	1 (9)	0 (0)	0 (0)	0 (0)	1 (2)	2 (9)	0 (0)	0 (0)
How useful would the diagnostic requirements be in helping you to assess this patient's Prognosis	Not at all useful	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	Somewhat useful	0 (0)	1 (10)	3 (37)	0 (0)	9 (16)	3 (13.0)	2 (25)	0 (0)
	Quite useful	8 (73)	9 (90)	5 (63)	0 (0)	48 (84)	19 (83)	6 (75)	3 (100)
	Extremely useful	3 (27)	0 (0)	0 (0)	0 (0)	0 (0)	1 (4)	0 (0)	0 (0)
Communication with other health care professionals	Not at all useful	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	Somewhat useful	0 (0)	0 (0)	1 (12)	0 (0)	1 (2)	1 (4.3)	2 (25)	1 (33)
	Quite useful	9 (82)	10 (100)	7 (88)	0 (0)	51 (89)	21 (91)	6 (75)	2 (67)
	Extremely useful	2 (18)	0 (0)	0 (0)	0 (0)	5 (9)	1 (4)	0 (0)	0 (0)
Educating patients/family about condition	Not at all useful	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
	Somewhat useful	0 (0)	2 (20)	3 (38)	0 (0)	7 (12)	2 (9)	3 (38)	1 (33)
	Quite useful	9 (82)	8 (80)	5 (62)	0 (0)	45 (79)	20 (87)	5 (62)	2 (67)
	Extremely useful	2 (18)	0 (0)	0 (0)	0 (0)	5 (9)	1 (4)	0 (0)	0 (0)

nonverbal intellectual functioning, and the VABS-II is a structured interview that assesses communication, socialisation, and daily living skills whereas the ICD-11 CDDR uses BIs to assess intellectual and adaptive behaviour functioning. The concurrent validity of the BIs were also generally consistent with the reported concurrent validity among other measures of intellectual and adaptive functioning (refer to Sparrow *et al.* 2005; Banks and Franzen 2010; Roid *et al.* 2013; Wechsler 2014).

As a naturalistic study, clinicians were provided with minimal training on the ICD-11 CDDR and associated BIs, prevented from accessing prior diagnoses, provided with minimal guidance on how to conduct their clinical interviews or how to make a diagnosis, and instructed to enter their diagnostic formulation independently without consultations. Nonetheless, the ICCs for the BIs were comparable, and in some cases greater than standardised diagnostic instruments. An explanation for these findings may be that the BIs are aligned with the routine history taking and mental state examination and the clinicians were trained and experienced to gather details outlined in the BIs and CDDR. These findings were also consistent across sites and countries, supporting the use of BIs internationally given their good reliability and validity.

In general, results from the internal consistency analyses suggest a high degree of interrelatedness between intellectual functioning, adaptive behaviours and the overall severity of DID using the BIs (Tables 5 and 6), whereas the standardised measures (Leiter-3 and VABS-II) appear to tap more distinct constructs (Table 7). As discussed, this may be due to the nature of combining two separate standardised measures that use different administration methods and domains. Based on the correlation coefficient confidence intervals, the internal consistencies of the standardised measures (ranging from 0.53 to 0.83) were significantly lower than the BIs (ranging from 0.83 to 0.97). These findings suggest that there is more diagnostic overlap between intellectual and adaptive functioning using the BIs, likely due to the inherently less rigid diagnostic requirements and more weight being placed on the decision-making abilities of the clinician (First *et al.* 2015).

The clinical utility of the BIs were assessed across levels of severity (i.e., mild to profound) separately for DID without co-occurring disorders (Table 8), and DID with co-occurring disorders (Table 9) to account for the complexities and large degree of clinical heterogeneity in DID with co-occurring disorders (especially ASD; Casanova *et al.* 2020). The BIs may be clinically useful for individuals with DID regardless of whether co-occurring disorders are present, given that there was no significant difference in ratings of clinical utility or in severity levels for intellectual functioning, adaptive behaviours, and overall functioning between single and co-occurring presentations. Despite smaller sample sizes in the present study for people diagnosed with DID and co-occurring disorders, it appears that the ICD-11 CDDR and BIs are generally clear and understandable, quick and easy to apply across severities and useful in facilitating treatment selection, prognosis assessments, communication with other health care professionals, and education efforts, although according to clinicians they include perhaps too much detail and specificity. On the other hand, providing less detail and specificity could affect consistency across classification users and whether the CDDR are used as intended. The results of this study suggest that the ICD-11 BIs may be used as an alternative resource to establish the severity of DID under conditions when psychometrically sound, appropriately normed and standardised measures for intellectual functioning and adaptive behaviour are not available or not feasible.

Limitations and future directions

There are several potential limitations to our study. First, findings from our study would benefit from replication with larger samples with greater diversity across sites and countries to improve generalisability and global applicability of the ICD-11 CDDR and BIs for DID. Although clinicians varied widely by age (range: 29 to 48 years old) and years of experience (range: 2–22 years), most were psychiatrists. A sample of clinicians that vary in profession (e.g., psychologists, primary care physicians, and paediatricians) and setting may improve generalisability among healthcare and system users, especially in LAMICs where resources and access to healthcare professionals may be limited. A larger sample size is recommended to further assess for clinically meaningful differences between single diagnoses of DID from those with co-occurring

diagnoses in which clinicians are instructed to adjust their use of the CDDR and BIs (e.g., placing less weight on the social domain for adaptive behaviours in individuals with co-occurring ASD).

Second, our findings would benefit from replication studies that include clinics in various countries that service more diverse clinical presentations to ensure clinicians are applying the CDDR and BIs across the full spectrum of severity levels for DID. Within our study, most participants were diagnosed with mild to moderate DID at the Oasi IRCCS site in Italy and the NIMHANS site in India, mild DID at the Peradeniya site in Sri Lanka, and severity levels spanning from mild to profound at the AIIMS site in India. These variations in clinical presentation help explain the large discrepancy in concurrent validity among sites with individuals presenting with subthreshold to mild DID. Furthermore, if clinicians typically see certain levels of severity for DID at their site, they may be primed to designate similar diagnostic formulations within a smaller range of options without needing to use all the information provided by the BIs. A narrowed range of participant clinical presentations may increase inter-rater reliability due to fewer severity levels considered when diagnosing.

Finally, the current study's eligibility criteria required participants to be aged 5 to 18 years. Although some clinicians used the BIs for early childhood (i.e., participants aged 5 years), most used the BIs for childhood and adolescence (i.e., participants aged 6–18 years). These age restrictions were put in place due to the limited sample size per site, to accommodate participants typically seen across sites, and to prioritise the facilitation of early DID diagnoses. Given these restrictions, the results from this study may not generalise to populations in early childhood (<5 years) or adulthood (>18 years). Future studies should conduct similar investigations focusing on these populations to assess the reliability, validity, and clinical utility of the ICD-11 CDDR and BIs for individuals with DID across the lifespan.

Conclusion

This international field study was conducted in clinical settings in three countries of varying language, culture, and income level to assess the performance of the BIs embedded in the ICD-11 CDDR for DID. The BIs were assessed in several ways including whether clinicians interpret and apply the BIs as intended by the developers of the requirements, derive similar diagnostic formulations using the BIs when presented with the same case presentation information (inter-rater reliability), and compared to the use of validated, normed, and standardised assessment measures of intellectual and adaptive functioning (concurrent validity). The BIs were also assessed for their clinical utility. Findings suggest that the BIs have excellent inter-rater reliability, good to excellent concurrent validity, and good clinical utility across sites, suggesting good to excellent global applicability of the BIs in various clinical settings, irrespective of language, culture, economic status, and assessment/diagnostic practices. These findings will be used to inform the global implementation of the ICD-11 and the CDDR. Implemented in the context of the CDDR, BIs provide a much needed tool to identify and evaluate the severity of limitations in intellectual and adaptive functioning for DID in a valid and reliable way through direct behavioural observations and informant-reported behaviours. Their use in settings where appropriately normed, standardised measures and specialised services are unavailable has the potential to impact the identification and treatment of people with DID around the world, particularly in lower-resource settings. By facilitating early diagnosis and intervention in these settings, the BIs can contribute to improved individual and community outcomes and reduce health disparities for people living with DID.

Acknowledgements

The authors wish to thank Ms Shivani Purnima, Ms Feba Philip, Dr Grazia Trubia, Dr Giovanna Sciacca, Dr Santina Città, and Dr Vitalba Mongelli for their participation as clinicians in the study. We also thank Dr Graccielle Rodrigues da Cunha for her assistance with the design of the study.

Source of funding

This research was supported in part by an Azrieli Foundation grant to CK.

Conflict of interest

The authors certify that they have no affiliations with or involvement in any organisation or entity with any financial or non-financial interest in the subject matter or materials discussed in this manuscript. Unless specifically stated, the views expressed in this paper are those of the authors and do not represent the official policies or positions of the WHO. There was no source of funding for the conduct of this study.

No conflicts of interest have been declared.

Data availability statement

Author elects to not share data.

References

Ayuso-Mateos J. & Kogan C. (2020) Management and treatment of intellectual disability. In: *New Oxford Textbook of Psychiatry* (eds J. Geddes, N. Andreasen & G. Goodwin), pp. 231–6. Oxford University Press, Oxford.

Banks S. H. & Franzen M. D. (2010) Concurrent validity of the TONI-3. *Journal of Psychoeducational Assessment* **28**, 70–9.

Carulla L. S., Reed G. M., Vaez-Azizi L. M., Cooper S. A., Leal R. M., Bertelli M. *et al.* (2011) Intellectual developmental disorders: towards a new name, definition and framework for "mental retardation/intellectual disability" in ICD-11. *World Psychiatry* **10**, 175–80.

Casanova M. F., Frye R. E., Gillberg C. & Casanova E. L. (2020) Editorial: Comorbidity and autism spectrum disorder. *Frontiers in Psychiatry* **11**(617395), 1–7.

First M. B., Reed G. M., Hyman S. E. & Saxena S. (2015) The development of the ICD-11 clinical descriptions and diagnostic guidelines for mental and behavioural disorders. *World Psychiatry* **14**, 82–90.

Girimaji S. C. & Pradeep A. J. V. (2018) Intellectual disability in international classification of Diseases-11: a developmental perspective. *Indian Journal of Social Psychiatry* **34**, 68–74.

Hallgren K. A. (2012) Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in Quantitative Methods for Psychology* **8**, 23–34.

Harris J. C. (2006) *Intellectual Disability: Understanding its Development, Causes, Classification, Evaluation, and Treatment*. Oxford University Press, Oxford.

Harris P. A., Taylor R., Thielke R., Payne J., Gonzalez N. & Conde J. G. (2009) Research electronic data capture (REDCap) – a metadata-driven methodology and workflow process for providing translational research informatics support. *Journal of Biomedical Informatics* **42**, 377–81.

Harrison P. L. & Oakland T. (2015) *ABAS-3: Adaptive Behavior Assessment System*. Western Psychological Services, Los Angeles, CA.

Hunsley, J., & Mash, E. J. (2008) Developing criteria for evidence-based assessment: an introduction to assessments that work. In: *A Guide to Assessments that Work* (eds J. Hunsley & E. J. Mash), pp. 3–14. Oxford University Press, Oxford.

IBM Corp. (2016) *IBM SPSS Statistics for Windows, Version 24.0*. IBM Corp, Armonk, NY, USA.

International Advisory Group for the Revision of the ICD Mental and Behavioural Disorders. (2011) A conceptual framework for the revision of the ICD-10 classification of mental and behavioural disorders. *World Psychiatry* **10**, 86–92.

Keeley J. W., Reed G. M., Roberts M. C., Evans S. C., Medina-Mora M. E., Robles R. *et al.* (2016) Developing a science of clinical utility in diagnostic classification systems: field study strategies for ICD-11 mental and behavioral disorders. *American Psychologist* **71**, 3–16.

King B. H., Toth K. E., Hodapp R. M. & Dykens E. M. (2009) Intellectual disability. In: *Comprehensive Textbook of Psychiatry* (eds B. J. Sadock, V. A. Sadock & P. Ruiz), 9th edn, pp. 3444–74. Lippincott Williams &Wilkins, Philadelphia.

Kishore M. T., Udipi G. A. & Seshadri S. P. (2019) Clinical practice guidelines for assessment and management of intellectual disability. *Indian Journal of Psychiatry* **61**, 194–210.

Kwak S. K. & Kim J. H. (2017) Statistical data preparation: management of missing values and outliers. *Korean Journal of Anesthesiology* **70**, 407–11.

McKenzie K., Milton M., Smith G. & Ouellette-Kuntz H. (2016) Systematic review of the prevalence and incidence of intellectual disabilities: current trends and issues. *Current Developmental Disorders Reports* **3**, 104–15.

Reed G. M. (2010) Toward ICD-11: improving the clinical utility of WHO's International Classification of mental disorders. *Professional Psychology: Research and Practice* **41**, 457–64.

Reed G. M., First M. B., Kogan C. S., Hyman S. E., Gureje O., Gaebel W. *et al.* (2019) Innovations and changes in the ICD-11 classification of mental, behavioural and neurodevelopmental disorders. *World Psychiatry* **18**, 3–19.

Robertson J., Hatton C., Emerson E. & Yasamy M. T. (2012) The identification of children with, or at significant risk of, intellectual disabilities in low- and middle-income countries: a review. *Journal of Applied Research in Intellectual Disabilities* **25**, 99–118.

Roid G. H., Miller L. J., Pomplun M. & Koch C. (2013) *Leiter International Performance Scale, (Leiter-3)*. Western Psychological Services, Los Angeles.

Schalock R. L., Brown I., Brown R., Cummins R. A., Felce D., Matikka L. *et al.* (2002) Conceptualization, measurement, and application of quality of life for persons with intellectual disabilities: report of an international panel of experts. *Mental Retardation* **40**, 457–70.

Sparrow S. S., Cicchetti D. V. & Balla D. A. (2005) *Vineland-II: Vineland Adaptive Behaviour Scales*, 2nd edn. Pearson Assessments, Minneapolis.

Tassé M. J., Balboni G., Navas P., Luckasson R., Nygren M. A., Belacchi C. *et al.* (2019) Developing behavioural indicators for intellectual functioning and adaptive behaviour for ICD-11 disorders of intellectual development. *Journal of Intellectual Disability Research* **63**, 386–407.

The World Bank Group. (2022) World Bank Country and Lending Groups. Available at: https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups (Accessed: 29 January 2022).

Wechsler D. (2014) *WISC-V: Technical and Interpretive Manual*. Pearson, Bloomington, MN.

World Health Organization. (1992) *The ICD-10 Classification of Mental and Behavioural Disorders: Clinical Descriptions and Diagnostic Guidelines*, vol. **1**. World Health Organization, Geneva.

World Health Organization. (2014) *Basic documents*. 48th ed. World Health Organization, Geneva. http://apps.who.int/gb/bd/PDF/bd48/basic-documents-48th-edition-en.pdf

World Health Organization. (2018) ICD-11 for mortality and morbidity statistics. November 11, 2018. https://icd.who.int/browse11/l-m/en

Accepted 2 February 2022