Comparing mental health semi-structured diagnostic interviews and symptom checklists to predict poor life outcomes: an 8-year cohort study from childhood to young adulthood in Brazil

Mauricio Scopel Hoffmann, Daniel S Pine, Katholiki Georgiades, Peter Szatmari, Euripedes Constantino Miguel, Pedro Mario Pan, Ary Gadelha, Luis Augusto Rohde, Kathleen Ries Merikangas, Michael Peter Milham, Theodore Daniel Satterthwaite, Giovanni Abrahão Salum

Summary

Background Semi-structured diagnostic interviews and symptom checklists present similar internal reliability. We aim to investigate whether they differ in predicting poor life outcomes in the transition from childhood to young adulthood.

Methods For this longitudinal study, we used data from the Brazilian High Risk Cohort Study for Childhood Mental Health Conditions. Eligible participants were aged 6–14 years on the day of study enrolment (January to February, 2010) and were enrolled in public schools by a biological parent in Porto Alegre and São Paulo, Brazil. 2511 young people and their caregivers were assessed at baseline in 2010–11, and 1917 were assessed 8 years later (2018–19; 76·3% retention). Clinical thresholds were derived using semi-structured parent-report interview based on the Diagnostic and Statistical Manual of Mental Disorders, according to the Developmental and Well-being Assessment (DAWBA), and clinical scores as defined by the Child Behavior Checklist (CBCL; T-score ≥70 considered positive caseness). At 8 years, participants were assessed for a composite life-threatening outcome (a composite of death, suicide attempts, severe self-harm, psychiatric inpatient admission, or emergency department visits) and a composite poor life chances outcome (a composite of any criminal conviction, substance misuse, or school dropout). We evaluated the accuracy of DAWBA and CBCL to predict these outcomes. Logistic regression models were adjusted for age, sex, race or ethnicity, study site, and socioeconomic class.

Findings DAWBA and CBCL had similar sensitivity, specificity, predictive values, and test accuracy for both composite outcomes and their components. Any mental health problem, as classified by DAWBA and CBCL, was independently associated with the composite life-threatening outcome (DAWBA adjusted odds ratio 1·62, 95% CI 1·20–2·18; CBCL 1·66, 1·19–2·30), but only CBCL independently predicted poor life chances (1·56, 1·19–2·04). Participants classified by both approaches did not have higher odds of the life-threatening outcome when compared with participants classified by DAWBA or CBCL alone, nor for the poor life chances outcome when compared with those classified by CBCL alone.

Interpretation Classifying children and adolescents based on a semi-structured diagnostic interview was not statistically different to symptom checklist in terms of test accuracy and predictive validity for relevant life outcomes. Classification based on symptom checklist might be a valid alternative to costly and time-consuming methods to identify young people at risk for poor life outcomes.

Funding Conselho Nacional de Desenvolvimento Científico e Tecnológico; Fundação de Amparo à Pesquisa do Estado de São Paulo; and Medical Research Council, European Research Council.

Introduction

The risk for adverse outcomes associated with childhood psychopathology is continuously distributed.[1–4] However, resource limitations and the adverse effects of intervention force stakeholders to treat only some subgroups of children, who would ideally be selected using assessment tools that identify those most at risk for adverse outcomes. Globally, around 80% of young people with mental health problems do not have access to treatment.[5,6] One reason is that current classification systems require health-care professionals and services to identity those with mental health problems. This complexity might be a barrier in low-resource settings, where there can be a shortage of trained professionals and considerable stigma for those with mental health problems, and might also be a costly burden for high-income settings, which could reallocate resources from this classification step to interventions.[7,8] Because such assessments have the potential to affect care delivery, a pressing need exists to identify reliable methods that

Department of Neuropsychiatry and Mental Health Epidemiology Group (MHEG), Universidade Federal de Santa Maria, Santa Maria, Brazil
(Prof M S Hoffmann MD PhD); Graduate Program in Psychiatry and Behavioral Sciences (Prof M S Hoffmann, Prof L A Rohde MD PhD, Prof G A Salum MD PhD) and Department of Psychiatry and Legal Medicine (Prof L A Rohde, Prof G A Salum), Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil; Care Policy and Evaluation Centre, London School of Economics and Political Science, London, UK (Prof M S Hoffmann); Section on Developmental Affective Neuroscience, National Institute of Mental Health, Bethesda, MD, USA (Prof D S Pine MD); Department of Psychiatry and Behavioral Neurosciences, McMaster University, Hamilton, ON, Canada (Prof K Georgiades PhD); Cundill Centre for Child and Youth Depression, Centre for Addiction and Mental Health (CAMH), Toronto, ON, Canada (Prof P Szatmari MD); Department of Psychiatry, The Hospital for Sick Children, Toronto, ON, Canada (Prof P Szatmari); Department of Psychiatry, University of Toronto Faculty of Medicine, Toronto, ON, Canada (Prof P Szatmari); National Institute of Developmental

Psychiatry for Children and Adolescents (INCT-CNPq), São Paulo, Brazil
(Prof E C Miguel MD PhD, P M Pan MD PhD, Prof A Gadelha MD PhD, Prof L A Rohde, Prof G A Salum); **Universidade de São Paulo, São Paulo, Brazil** (Prof E C Miguel); **Laboratório Interdisciplinar de Neurociências Clínicas, Departamento de Psiquiatria, Universidade Federal de São Paulo, São Paulo, Brazil** (P M Pan); **Departamento de Psiquiatria, Escola Paulista de Medicina, Universidade Federal de São Paulo, São Paulo, Brazil** (Prof A Gadelha); **Genetic Epidemiology Research Branch, National Institute of Mental Health Intramural Research Program, Bethesda, MD, USA** (Prof K R Merikangas PhD); **Nathan S Kline Institute for Psychiatric Research, Orangeburg, NY, USA** (Prof M P Milham MD PhD); **Child Mind Institute, New York, NY, USA** (Prof M P Milham, Prof G A Salum); **Department of Psychiatry, University of Pennsylvania, Philadelphia, PA, USA** (Prof T D Satterthwaite MD); **Lifespan Informatics and Neuroimaging Center, Philadelphia, PA, USA** (T D Satterthwaite)

Correspondence to:
Prof Mauricio Scopel Hoffmann, Department of Neuropsychiatry and Mental Health Epidemiology Group (MHEG), Universidade Federal de Santa Maria, Santa Maria 97105-900, Brazil
mauricio.hoffmann@ufsm.br

Research in context

Evidence before this study

Mental health problems in young people have a major impact during their life course. The gold standard for assessing mental health problems involves trained professionals and structured interviews, which can be costly and time-consuming. Thus, it is important to understand whether simpler methods of assessing mental health problems, such as symptom checklists, could perform as well as assessments based on trained professionals and structured interviews in predicting future life events. We searched Google Scholar (peer-reviewed manuscripts, preprints, book chapters, and other grey literature) from Jan 1, 2000, to Aug 8, 2023, with the terms: comparison AND (classif* OR nosology) AND (categor* OR "structured interview" OR diagnosis OR DAWBA OR "K-SADS") AND (checklist OR dimension* OR CBCL) AND (child* OR adolesc*) AND (predict* OR utility OR validity) AND psychiatry. We searched for studies comparing the predictive validity of semi-structured interviews and symptom checklist assessments of mental health problems. We found few such studies in children and adolescents that compared predictions based on subsequent mental health diagnosis versus symptoms. Among the most common outcomes examined, namely educational problems and mental health services use, there was no clear advantage of either semi-structured diagnostic interviews or symptom checklists in predicting these outcomes. Additionally, we found no studies that compared these two assessment approaches in predicting, longitudinally, important mental health-related outcomes, including death, suicide attempt, self-harm, psychiatric inpatient admission, emergency visit due to a mental health problem, or criminal conviction.

Added value of this study

Semi-structured diagnostic interviews have been used as the gold standard in psychiatric nosology and epidemiological studies to predict the consequences of mental health problems, but little evidence supports this notion. We compared a semi-structured interview (Developmental and Well-being Assessment [DAWBA]) and a symptom checklist (Child Behavior Checklist [CBCL]) in terms of their ability to predict composite outcomes, which were defined following the UN Sustainable Development Goals. These were a composite of life-threatening indicators (death, suicide attempts, severe self-harm, psychiatric inpatient admissions, or emergency department visits due to emotional or behavioural problems) and a composite of poor life chances (any criminal conviction, substance misuse, or school dropout or expulsion) in late adolescence and early adulthood. We used the clinical threshold as defined by each assessment tool, including caseness rated by a psychiatrist in DAWBA and a T-score of at least 70 for CBCL. The outcomes were evaluated in a longitudinal Brazilian cohort who were at high risk for psychopathology and were followed up for 8 years. We found that there was no robust significant difference in predicting composite outcomes among those diagnosed with any mental health problem classified by semi-structured interview or symptoms checklist. However, semi-structured interviews independently predicted suicide attempts, whereas the symptom checklists independently predicted death by any cause, emergency visits, and criminal conviction.

Implications of all the available evidence

A semi-structured interview supervised by trained psychiatrists was not clearly superior for assessing risk of major negative life events in young people when compared with a symptom checklist. In some cases, the methods might complement each other in risk assessment. This challenges the concept that clinical diagnosis supported by structured interviews is the gold standard for assessing people with mental health problems. Moreover, symptom checklists might be a valid alternative to more resource-demanding semi-structured interviews to assess young people at risk of major negative life events due to mental health problems. Future studies need to replicate these findings with different assessment tools and settings.

can be implemented with minimal resources in different settings and also be predictive of adverse outcomes so that interventions can be delivered earlier. The current study compares two approaches to classifying children and adolescents with mental health problems in terms of ability to predict adverse outcomes longitudinally.

The first and most common type of approach identifies children using consensus-based classificatory methods such as the Diagnostic and Statistical Manual of Mental Disorders (DSM) and the International Classification of Diseases.[4,9] These approaches tally symptoms, their duration, age of onset, and other features,[4,7] and have been considered the gold standard to identify people at risk of poor life outcomes due to mental health problems.[10,11] To gather such data, researchers use semi-structured (with some open-ended questions) or fully structured (closed questions only) diagnostic interviews that apply algorithmic combinatorial rules and a rating from a trained child psychiatrist. The Development and Well-Being Assessment (DAWBA) is a semi-structured interview[12] used by many epidemiological studies around the globe for the purpose of classifying participants with mental health problems.[13]

A second approach of identifying psychopathology is by a symptom checklist.[14–16] This method relies on checklists and questionnaires answered by a respondent to sort items into one (unidimensional) or more (multidimensional) scores, which then identify children at risk of poor life outcomes based on population norms.[15,17] The Child Behavior Checklist (CBCL) provides multidimensional scores and is one of the most widely used instruments in child psychiatry, tested in different countries around the world.[16] The CBCL classifies those

with T-scores of 70 or higher as within a clinical range for dimensions of psychopathology.[17]

These two assessments might provide complimentary predictive value for risk identification.[18–20] Structured interviews and symptom checklists have similar psychometric properties (ie, internal consistency, test-retest reliability, and inter-rater reliability).[20] Dimensional models based on symptom checklists could also explain larger variance in outcomes when compared with categorical classifications based on structured interviews.[19,21] However, few studies have compared the ability of these assessments to predict serious mental health-related adverse outcomes, such as all-cause mortality, suicide attempt, and hospital admissions, or outcomes associated with poor life chances,[22] such as criminal convictions, substance misuse, and school dropout. Existing studies often operationalise categorical entities as latent dimensional factors[19,21,23,24] or simply compare continuous dimensions versus categories,[18] which does not allow for a pragmatic investigation of risk prediction when comparing group membership based on structured interviews versus symptom checklists.

To address these gaps, we compared membership assignment in childhood and early adolescence, as currently proposed by the DAWBA and CBCL, in their ability to predict important mental health-related outcomes. Based on previous studies on psychometric equivalence between structured interviews and symptom checklists[20] and their predictive validity,[18,19,21,23,24] we hypothesised that the gold-standard semi-structured interview would not be statistically different to a checklist-based assessment in predicting life-threatening and poor life chances outcomes at 8 years of follow-up.

Methods

Study design and participants

In this longitudinal study, we analysed baseline (2010–11) and 3-year (2013–14) and 8-year (2018–19) follow-up data from the Brazilian High Risk Cohort Study for Childhood Mental Health Conditions, a large, school-based cohort oversampled for high family risk for psychopathology.[25] Briefly, families were recruited from 22 state-funded schools in Porto Alegre and 35 schools in São Paulo on the day of enrolment in January to February, 2010. Attendance in schools is compulsory in Brazil for those aged 4–17 years. We interviewed 2511 young people (aged 6–14 years) and their caregivers at baseline, 2010 at first follow-up (80% retention, age 9–17 years), and 1917 at the second follow-up (76·3% retention, 14–23 years). Parent reports were collected by lay interviewers. Self-reported outcome measures were collected by trained psychologists.

This study was approved by the National Research Ethics Commission (Comissão Nacional de Ética em Pesquisa) under the approval number 2.448.062 (CAAE 74563817.7.1001.5327). Written informed consent was obtained from parents and participants that were able to read, write, and clearly understand the written consent. For those who were illiterate, the consent form was read aloud, doubts were explained, and verbal agreement was obtained. No participants were excluded due to parental illiteracy.[25]

Procedures

All participants underwent assessment with semi-structured diagnostic interview (DAWBA) and checklist-based classification (CBCL) at baseline. The Brazilian-Portuguese version of the DAWBA[12] was used to assess DSM-IV diagnosis. The semi-structured interview is used to generate diagnostic probabilities based on parent reports to lay interviewers. It comprises a strength and difficulties questionnaire, which is a 25-item screen for emotional and behavioural problems; algorithm-based questions containing skip-out rules; and open-questions about any problems that are reported within each section. This procedure generates DAWBA bands, which are computer-generated categories describing the probability of a positive DSM-IV diagnosis (<0·1%, ~3%, ~15%, ~50%, and >70%). Responses to DAWBA questions and open-ended responses were then evaluated by nine trained child psychiatrists (agreement ranged from 90% to 95%) who confirmed, refuted, or altered initial DAWBA bands to determine final categorical diagnostic statuses, which were used in the present analysis. DAWBA diagnoses were grouped into three broad categories: internalising (including panic, separation and social anxiety disorder, specific phobia, agoraphobia, generalised anxiety disorder, major depressive disorder and depressive disorder not other-wise specified, bipolar disorder, obsessive compulsive disorder, tic disorder, eating disorder, and post-traumatic stress disorder), externalising (including conduct and oppositional-defiant conditions), and attentional or hyperactivity (including any attention-deficit hyperactivity disorder).

CBCL is a parent-reported assessment of 120 emotional and behavioural symptoms of children and young people aged 6–18 years during the previous 6 months, answered in a three-point scale (0 represents not true; 1 represents somewhat or sometimes true; and 2 represents very true or often). It generates a total score (between 0 and 240) based on eight empirically-derived syndromes (anxious-depressed, withdrawn-depressed, somatic complaints, rule-breaking behaviour, aggressive behaviour, social problems, thought problems, and attention problems);[17] an internalising problems score (between 0 and 64, combining anxious-depressed, withdrawn-depressed, and somatic complaints); an externalising problems score (between 0 and 70, combining rule-breaking behaviour and aggressive behaviour syndromes); and an attentional or hyperactivity problems score (between 0 and 20). Scores were summed for each dimension and T-scores were generated according to the Achenbach

System of Empirically Based Assessment (ASEBA) manual.[17] Briefly, T-scores of 50 were assigned to raw scores up to the 50th percentile of each dimension (ie, total, internalising, externalising, and attention scores), as defined by the omnicultural norms.[26] The 20-point interval between T-scores of 51 and 70 was calculated by multiplying the raw score by 20 divided by two times the omnicultural SD. The 30-point interval between T-scores of 71 and 100 was calculated by multiplying the raw score by a calibration factor, calculated by dividing 30 by the result of maximum score minus the product of omnicultural mean plus two times omnicultural SD.[17] Detailed raw-to-T-score conversion can be found in appendix 2 (pp 4–9). A T-score of 70 or higher (98th percentile or beyond) defines the clinical range according to the ASEBA system and was considered a positive case (ie, any, internalising, externalising, and attentional problems are defined as scoring ≥70 in the problems dimension).

Outcomes

We evaluated a life-threatening outcome and poor life chances outcome at 8 years of follow-up. These outcomes (and their components) were selected as proxies or indicators of relevant outcomes within the UN Sustainable Developmental Goals for 2030,[5] and of core components of a child's life chances. Life chances refers to the opportunities of individuals to improve their quality of life, and is operationalised with indicators of social functioning, physical and mental health, and education or skills.[22]

The life-threatening outcome is a composite of any positive cases of: death by any cause since baseline; self-reported suicide attempt in past 4 weeks or self-harm requiring medical attention in past 6 months; parent-reported psychiatric inpatient admission since baseline; or emergency visits due to emotional or behavioural problems since baseline.

The poor life chances outcome is a composite of: parent-reported or self-reported criminal conviction, probation, or detention centre enrolment since baseline; self-reported substance misuse in the past 12 months (alcohol use disorder identified by the Alcohol Use Disorders Identification Test Consumption score of ≥5, daily tobacco smoking, or any illegal substance misuse); or school dropout (parent-reported dropout or expulsion from school) since baseline.

These components are fully described in appendix 2 (p 1). Tetrachoric correlation among these component variables is also described in appendix 2 (p 10).

Statistical analysis

Our analysis was conducted to answer three questions. First, do semi-structured interview and symptom checklist assessments of psychopathology predict life-threatening and poor life chances outcomes? Second, do these approaches independently predict the outcomes when adjusting for one another? Third, are those individuals classified by both approaches at increased risk for poor outcomes when compared with either approach alone (only CBCL or DAWBA)? We conducted all analyses for any type of mental health disorder and subgroups of mental health disorder (internalising, externalising, and attention or hyperactivity problems).

The primary objective of this research is to examine the hypothesis that there is no statistical difference between the semi-structured interview (DAWBA) and the checklist-based assessment (CBCL) in predicting relevant outcomes. To evaluate this, we computed 95% CIs for all analyses and overlapping 95% CIs were considered to indicate that estimates were not statistically different between the DAWBA and CBCL approaches.

For the first question, we calculated sensitivity, specificity, positive predictive value, negative predictive value (NPV), positive likelihood ratio, negative likelihood ratio, and test accuracy of assessment approaches to detect the life-threatening and poor life chance composite outcomes at 8-year follow-up, supplemented by area under the receiver operating characteristic curve (AUC) analysis. Furthermore, we used logistic regression models to calculate adjusted odds ratios (ORs). Regressions were weighted to account for missing data. We estimated separate models for each assessment approach. All regression models were adjusted for age, sex (male *vs* female), race or ethnicity (White *vs* non-White [Asian, Black, Mixed race or ethnicity (phrased as "between white and black" in the questionnaire), and Indigenous]), study site (São Paulo *vs* Porto Alegre), and socioeconomic class as defined by the Brazilian Association of Research Companies (ABEP; class A or B *vs* other classes). The ABEP is a parent-reported, asset-based assessment that also includes parental education and sanitation to classify individuals from socioeconomic class A (highest) to E (lowest).

For the second question, the same regression structure was used but outcomes were regressed on the DAWBA and CBCL approaches simultaneously. This shows independent associations of the approaches with external validators (direct comparisons on their independent or additive associations).

For the third question, the predictor assessments were combined to generate a nominal classification variable (not classified by any approach, DAWBA only, CBCL only, or both). Then, outcomes were regressed on the nominal classification. Being classified by both approaches was used as a comparator for other strategies. This analysis was designed to understand whether using both approaches was statistically different to using only one (or none) to predict those at risk for poor outcomes. Supplementary regression analyses were also made for each outcome's component variables.

In the present sample, baseline maternal education (none or incomplete primary, complete primary, complete secondary, or complete tertiary) and study site

	Baseline, 2010–11 (n=2511)	Follow-up, 2018–19 (n=1917)
Age, years, mean (SD)	10·2 (1·9)	18·2 (2·0)
Sex		
Female	1136 (45·2%)	876 (45·7%)
Male	1375 (54·8%)	1041 (54·3%)
Race or ethnicity		
Asian	5 (0·2%)	4 (0·2%)
Black	264 (10·5%)	210 (11·0%)
Mixed race or ethnicity	706 (28·1%)	519 (27·1%)
Indigenous	11 (0·4%)	9 (0·5%)
White	1519 (60·5%)	1170 (61·0%)
Missing	6 (0·2%)	5 (0·3%)
Study site (cities)		
Porto Alegre	1255 (50·0%)	1010 (52·7%)
São Paulo	1256 (50·0%)	907 (47·3%)
Socioeconomic group*		
A or B	428 (17·0%)	392 (20·4%)
C	1788 (71·2%)	1118 (58·3%)
D or E	295 (11·7%)	285 (14·9%)
Missing	0	122 (6·4%)
IQ classification		
Below average (IQ <90)	587 (23·4%)	472 (24·6%)
Average (IQ ≥90 and <110)	1112 (44·3%)	831 (43·3%)
Above average (IQ ≥110)	542 (21·6%)	422 (22·0%)
Missing	270 (10·8%)	192 (10·0%)
Maternal educational level		
None or incomplete primary	607 (24·2%)	522 (27·2%)
Complete primary	857 (34·1%)	412 (21·5%)
Complete secondary	934 (37·2%)	694 (36·2%)
Complete tertiary	85 (3·4%)	113 (5·9%)
Missing	28 (1·1%)	176 (9·2%)
Group membership according to assessment approaches		
Any DAWBA-based diagnosis	652 (26·0%)	..
Any CBCL-based caseness (T-score ≥70)	448 (17·8%)	..
Any internalising DAWBA-based diagnosis	343 (13·7%)	..
CBCL-based internalising caseness	319 (12·7%)	..
Any externalising DAWBA-based diagnosis	171 (6·8%)	..
CBCL-based externalising caseness	414 (16·5%)	..
Any attention or hyperactivity DAWBA-based diagnosis	274 (10·9%)	..
CBCL-based attention or hyperactivity caseness	797 (31·7%)	..
Nominal classifications based on DAWBA and CBCL (any diagnosis)		
DAWBA only	338 (13·5%)	..
CBCL only	134 (5·3%)	..
Both DAWBA and CBCL	314 (12·5%)	..
None	1725 (68·7%)	..
		(Table 1 continues in next column)

	Baseline, 2010–11 (n=2511)	Follow-up, 2018–19 (n=1917)
(Continued from previous column)		
Nominal classifications DAWBA and CBCL (internalising)		
DAWBA only	178 (7·1%)	..
CBCL only	154 (6·1%)	..
Both DAWBA and CBCL	165 (6·6%)	..
Nominal classifications DAWBA and CBCL (externalising)		
DAWBA only	40 (1·6%)	..
CBCL only	283 (11·3%)	..
Both DAWBA and CBCL	131 (5·2%)	..
Nominal classifications DAWBA and CBCL (attention or hyperactivity)		
DAWBA only	34 (1·4%)	..
CBCL only	557 (22·2%)	..
Both DAWBA and CBCL	240 (9·6%)	..
Outcomes		
Life-threatening outcome		
Any life-threatening outcome	..	232 (12·1%)
Death	..	18 (0·9%)
Suicide attempt or severe self-harm in past 4 weeks (self-reported)	..	164 (8·6%)
Any psychiatric inpatient admission since baseline (parent-reported)	..	16 (0·8%)
Any visits to emergency department due to emotional or behavioural problems since baseline (parent-reported)	..	58 (3·0%)
Missing	..	295 (15·4%)
Poor life chances outcome		
Any poor life chances outcome	..	788 (41·1%)
Any criminal conviction (parent-reported or self-reported)	..	81 (4·2%)
Substance misuse in past 12 months (self-reported)	..	485 (25·3%)
School dropout (parent-reported dropout or expulsion)	..	475 (24·8%)
AUDIT-C high-risk category (score ≥5 based on past 12 months)	..	264 (13·8%)
AUDIT-C score, median (range)	..	1·00 (0, 12·0)
Smoking frequency in past 12 months		
Never	..	1066 (55·6%)
Once or twice	..	165 (8·6%)
Monthly	..	71 (3·7%)
Weekly	..	62 (3·2%)
Daily or almost daily	..	115 (6·0%)
Any substance use in past 12 months		
Cannabis	..	306 (16·0%)
Cocaine	..	32 (1·7%)
Crack cocaine	..	3 (0·2%)
Shoe glue	..	3 (0·2%)
Ecstasy	..	52 (2·7%)
		(Table 1 continues in next column)

	Baseline, 2010–11 (n=2511)	Follow-up, 2018–19 (n=1917)
(Continued from previous column)		
LSD	..	24 (1·3%)
Solvent	..	47 (2·5%)
Smoked amphetamine	..	5 (0·3%)
Inhaled amphetamine	..	1 (0·1%)
Missing	..	311 (16·2%)

Data are n (%) unless otherwise indicated. IQ=intelligence quotient. DAWBA=Developmental and Well-being Assessment. CBCL=Child Behavior Checklist. AUDIT-C=Alcohol Use Disorders Identification Test for Consumption. *A and B represent the high or wealthy social classes; C is considered middle class; and D and E represent the low or poor social classes.

Table 1: Patient characteristics at baseline and 8-year follow-up, mental health assessments at baseline, and outcomes at 8-year follow-up

predicted response at follow-up.[27] Therefore, we used these variables to compute inverse probability weight to address sample attrition in all regression analysis. We contrasted the explained variance (pseudo R^2) from each regression model and compared them with each other to understand the incremental value of combining assessment approaches to predict the main outcomes. We also tested whether the associations of DAWBA and CBCL with composite outcomes were confounded by parental psychiatric diagnosis (appendix 2 p 2).

All analyses were done between May, 2022, and January, 2023, using R version 4.2.1 and RStudio version 2022.07.1. Test characteristics were calculated using the *epiR* package.[28] Regression models were carried out using the *glm* function for binary outcomes. Multicollinearity between classification approaches was tested using the *vif* function. Statistical significance was indicated by p values of less than 0·05 and by overlap of 95% CIs. Post-hoc power analysis was calculated as described in appendix 2 (p 2).

Role of the funding source
The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results
Baseline (n=2511) and second follow-up sample (mean follow-up 8·04 years [SD 0·45], n=1917) characteristics and outcomes are described in table 1. The frequencies of baseline DAWBA classification within the categories of any disorder, internalising disorders, externalising disorders, and attention or hyperactivity disorders are described in appendix 2 (p 11). Mean CBCL T-scores for each checklist-based classification and overall mean CBCL T-scores are also described in appendix 2 (pp 12–13).

At baseline, almost a third of the sample (786 [31·3%] of 2511) was classified with a mental health problem according to at least one of the two assessment approaches (DAWBA, 652 [26·0%]; CBCL, 448 [17·8%]). 314 (12·5%) participants were classified as positive by both approaches (40% agreement; table 1). The numbers of participants at baseline classified by DAWBA and CBCL alone and by both, by the composite life-threatening outcome and poor life chances outcome and the components of each, are described in appendix 2 (pp 31–34).

Test characteristics for the life-threatening outcome and the poor life chances outcome are described in table 2, and AUCs are depicted in appendix 2 (pp 17–18). For both outcomes, both assessment approaches had high specificity and NPV for life-threatening outcomes, with moderate test accuracy and low AUC. For the poor life chances outcome, specificity was high but accuracy was low. Positive and negative likelihood ratios were low for both assessment approaches for both outcomes. Test characteristics and AUCs for outcome components are presented in appendix 2 (pp 14–16, 19–25).

wParticipants classified as having a mental health problem with DAWBA or CBCL approaches were also at increased odds of experiencing poor life chances, except when classified specifically as having internalising problems. There was no statistical difference in the association with poor life chances when using DAWBA (adjusted OR 1·45, 95% CI 1·18–1·80, p=0·0005) or CBCL (1·72, 1·36–2·19, p<0·0001). The association with poor life chances was not statistically different between diagnosis by DAWBA or CBCL when considering externalising (DAWBA 2·37, 1·55–3·64, p<0·0001; CBCL 1·93, 1·43–2·61, p<0·0001) and attention or hyperactivity (DAWBA 1·68, 1·18–2·41, p=0·0085; CBCL 1·58, 1·23–2·04, p<0·0001) classifications (figure 1A, appendix 2 p 26).

Having a mental health problem as classified with DAWBA and CBCL approaches independently predicted the life-threatening composite outcome (DAWBA adjusted OR 1·62, 95% CI 1·20–2·18, p=0·0017; CBCL 1·66, 1·19–2·30, p=0·0027). Associations were significant only for CBCL for internalising (1·78, 1·23–2·57, p=0·0020) and attention or hyperactivity (1·46, 1·10–1·93, p=0·0080) classifications, and for both assessment approaches for the externalising classification (DAWBA 1·73, 1·07–2·79, p=0·025; CBCL 1·66, 1·18–2·32, p=0·0032; figure 1B, appendix 2 p 27). Associations for life-threatening components are described in appendix 2 (p 27).

Classification using CBCL (adjusted OR 1·56, 95% CI 1·19–2·04, p=0·0013) but not DAWBA (1·22, 0·96–1·54, p=0·11) independently predicted the composite poor life chances outcome. Associations were significant when externalising problems were assessed with DAWBA (2·74, 1·77–4·24, p<0·0001) and CBCL (1·68, 1·28–2·21, p=0·0002). Attention or hyperactivity problems were associated with poor life chances only when assessed with CBCL (1·77, 1·43–2·19, p<0·0001) but not with DAWBA (1·06, 0·77–1·47, p=0·73; figure 1B, appendix 2

	Sensitivity (95% CI)	Specificity (95% CI)	Positive predictive value (95% CI)	Negative predictive value (95% CI)	Positive likelihood ratio (95% CI)	Negative likelihood ratio (95% CI)	Accuracy (95% CI)
Life-threatening outcome							
Any diagnosis							
DAWBA	0·37 (0·31–0·44)	0·74 (0·72–0·77)	0·19 (0·16–0·23)	0·88 (0·86–0·89)	1·45 (1·20–1·75)	0·85 (0·76–0·94)	0·69 (0·67–0·71)
CBCL	0·26 (0·21–0·32)	0·83 (0·81–0·85)	0·21 (0·16–0·26)	0·87 (0·85–0·89)	1·55 (1·21–1·98)	0·89 (0·82–0·96)	0·75 (0·73–0·77)
Internalising							
DAWBA	0·20 (0·15–0·26)	0·86 (0·84–0·88)	0·19 (0·15–0·25)	0·87 (0·85–0·88)	1·45 (1·09–1·92)	0·93 (0·87–0·99)	0·77 (0·74–0·79)
CBCL	0·19 (0·15–0·25)	0·88 (0·86–0·90)	0·21 (0·16–0·27)	0·87 (0·85–0·88)	1·61 (1·19–2·16)	0·92 (0·86–0·98)	0·78 (0·76–0·80)
Externalising							
DAWBA	0·12 (0·08–0·16)	0·94 (0·93–0·95)	0·25 (0·18–0·35)	0·86 (0·85–0·88)	2·05 (1·35–3·10)	0·94 (0·89–0·98)	0·83 (0·81–0·84)
CBCL	0·25 (0·19–0·31)	0·84 (0·82–0·86)	0·2 (0·16–0·25)	0·87 (0·85–0·89)	1·52 (1·18–1·96)	0·90 (0·83–0·97)	0·75 (0·73–0·77)
Attention or hyperactivity							
DAWBA	0·14 (0·10–0·19)	0·89 (0·87–0·91)	0·17 (0·12–0·24)	0·86 (0·84–0·88)	1·26 (0·89–1·80)	0·97 (0·92–1·02)	0·78 (0·76–0·80)
CBCL	0·37 (0·31–0·44)	0·69 (0·67–0·71)	0·17 (0·14–0·20)	0·87 (0·85–0·89)	1·2 (0·99–1·44)	0·91 (0·82–1·01)	0·64 (0·62–0·67)
Poor life chances outcome							
Any diagnosis							
DAWBA	0·33 (0·30–0·37)	0·77 (0·74–0·80)	0·56 (0·51–0·61)	0·56 (0·53–0·59)	1·43 (1·22–1·68)	0·87 (0·82–0·93)	0·56 (0·54–0·59)
CBCL	0·25 (0·22–0·28)	0·86 (0·83–0·88)	0·61 (0·55–0·66)	0·56 (0·53–0·59)	1·75 (1·42–2·15)	0·88 (0·83–0·92)	0·57 (0·55–0·59)
Internalising							
DAWBA	0·18 (0·15–0·20)	0·87 (0·84–0·89)	0·54 (0·47–0·60)	0·54 (0·51–0·57)	1·30 (1·04–1·64)	0·95 (0·91–0·99)	0·54 (0·51–0·56)
CBCL	0·15 (0·13–0·18)	0·88 (0·86–0·90)	0·53 (0·47–0·60)	0·54 (0·51–0·56)	1·28 (1·00–1·64)	0·96 (0·92–1·00)	0·54 (0·51–0·56)
Externalising							
DAWBA	0·11 (0·09–0·14)	0·96 (0·95–0·97)	0·73 (0·64–0·80)	0·55 (0·52–0·57)	2·97 (2·00–4·40)	0·92 (0·90–0·95)	0·56 (0·54–0·59)
CBCL	0·25 (0·22–0·29)	0·88 (0·86–0·90)	0·65 (0·60–0·71)	0·57 (0·54–0·60)	2·11 (1·70–2·63)	0·85 (0·81–0·89)	0·58 (0·56–0·61)
Attention or hyperactivity							
DAWBA	0·14 (0·12–0·17)	0·91 (0·89–0·93)	0·58 (0·51–0·65)	0·54 (0·52–0·57)	1·56 (1·18–2·06)	0·94 (0·91–0·98)	0·55 (0·52–0·57)
CBCL	0·41 (0·37–0·44)	0·75 (0·72–0·78)	0·59 (0·55–0·63)	0·59 (0·56–0·61)	1·61 (1·40–1·86)	0·79 (0·74–0·85)	0·59 (0·56–0·61)

Any CBCL diagnosis is defined as a participant having a T-score ≥70, according to the CBCL system. DAWBA=Developmental and Well-being Assessment. CBCL=Child Behavior Checklist.

Table 2: Sensitivity, specificity, predictive values, likelihood ratios, and accuracy of a mental health diagnosis, assessed by DAWBA or CBCL, for predicting composite life-threatening and poor life chances outcomes

p 27). Associations for poor life chances components are described in appendix 2 (p 27).

We investigated whether independent associations of DAWBA and CBCL with composite outcomes were confounded by parental psychiatric diagnosis. We found no association between parental diagnosis and the outcomes, nor change in regression coefficients for the independent variables compared with models without the parental psychiatric diagnosis variable (appendix 2 pp 2, 28).

Participants classified with either DAWBA or CBCL alone did not have a statistically different odds for life-threatening outcomes when compared with people classified with both approaches, for any type of mental health problem (figure 1C, appendix 2 pp 29–30). Participants classified with any problems (adjusted OR 0·55, 95% CI 0·38–0·79, p=0·0013) and attention or hyperactivity problems (0·31, 0·13–0·73, p=0·0075) with DAWBA alone had significantly lower odds of poor life chances when compared with those classified by both DAWBA and CBCL. For the same outcome, participants

with externalising problems classified only with CBCL had lower odds when compared with those classified by both DAWBA and CBCL (0·36, 0·21–0·63, p=0·0003). Therefore, people classified with both approaches were at higher risk for poor life chances compared with those classified only with DAWBA (for any problems and for attention or hyperactivity problems) or only with CBCL (for externalising problems). Results from the regression models using the components of the life-threatening and poor life chances outcomes are presented in appendix 2 (p 28).

Figure 2 shows the explained variance of the aforementioned regression models compared with a regression model containing covariates only (appendix 2 p 35). The model containing only covariates explained 4% of the composite life-threatening outcome and 24% of the composite poor life chances outcome (figure 2A, absolute explained variance). Overall, combining classification approaches had little additive value, either as independent predictors or combined classification approaches. For the life-threatening outcome, there was
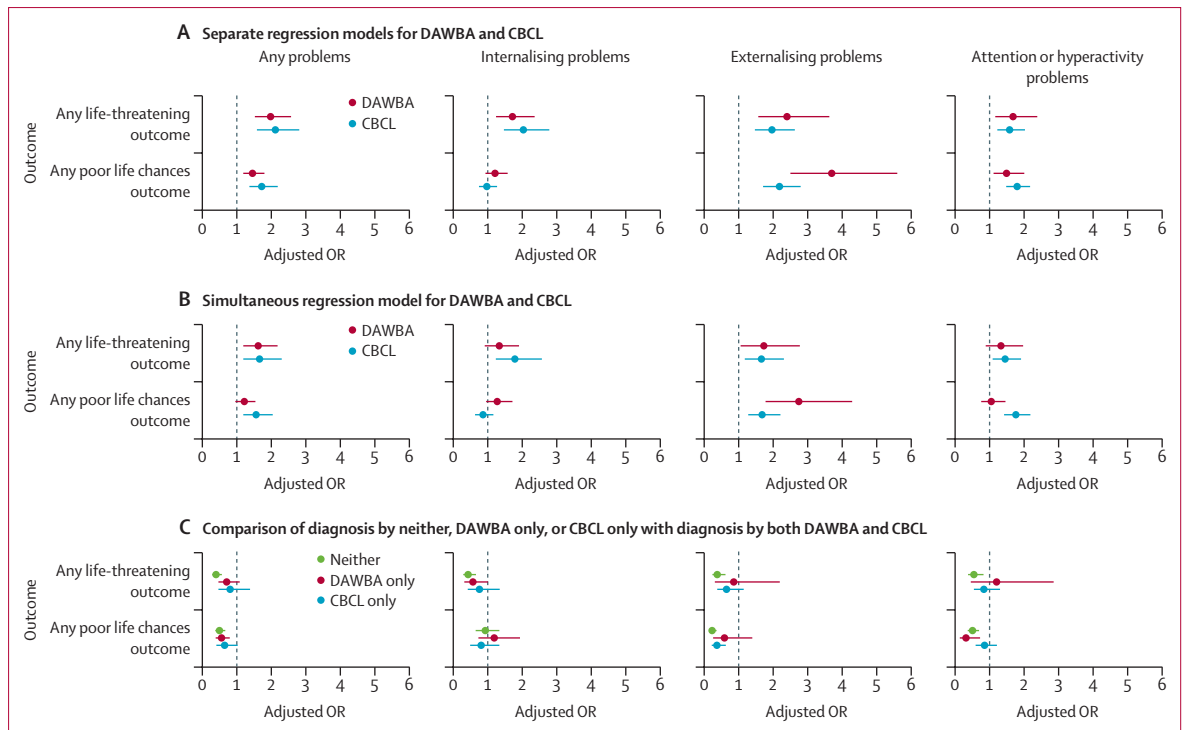
Figure 1: Associations of any mental health diagnosis, as assessed by DAWBA and CBCL, with composite life-threatening and poor life chances outcomes
The analysis was performed in separated regression models (A); using both classification approaches in the same regression models (B); and by comparing not being classified by either approach, by DAWBA only, or by CBCL only, with being classified by both DAWBA and CBCL (C). All regression models were adjusted for covariates (age, sex, ethnicity, socioeconomic class, and study site) and weighted for attrition. CBCL=Child Behavior Checklist. DAWBA=Developmental and Well-being Assessment. OR=odds ratio.

a 75% increase in explained variance (from 4% to 7%) for predicting it using DAWBA, CBCL, both as independent predictors, and by a model in which participants were detected using a combination of approaches (figure 2B, increased explained variance relative to covariate-only regression model). For the poor life chances outcome, there was an 8% increase (from 24% to 26%) in explained variance for all models except that using DAWBA, which presented a 4% increase (from 24% to 25%).

Discussion
This study compares the ability of different parent-reported mental health classification approaches to detect children and early adolescents in the community who are at risk for life-threatening and poor life chances outcomes 8 years later. Overall, classification based on semi-structured diagnostic interview (DAWBA) was not superior to symptom checklist (CBCL) in the ability to detect individuals at risk. Moreover, they are not interchangeable. Of those classified with any mental health problems, 43% would not have been diagnosed if DAWBA were not used, and 17% would not if CBCL were not used. Considering any mental health problems, these approaches also presented independent associations with life-threatening outcomes but CBCL, and not DAWBA, was a significant predictor of the composite poor life chances outcome. Children and adolescents

classified by both approaches as having a mental health problem were not at higher risk for life-threatening outcome when compared with being classified by either approach alone. However, the results indicate that DAWBA alone is insufficient to detect those at higher risk for poor life chances when considering any mental health problems or attention or hyperactivity problems, and CBCL alone is insufficient when considering externalising problems.

A notable finding is the absence of a clear advantage for interview over the checklist approach in detecting participants at risk. Similarities between these approaches might explain these findings. Overall, DAWBA and CBCL capture signs and symptoms of mental health problems that are related to several outcomes during the life course.[5,10] Hence, the focus of inquiring about these signs and symptoms might be enough to predict negative outcomes, and differences between these approaches might not be relevant at the primary outcome level. However, some approaches might be better suited than others to detect those at risk for specific outcomes. For example, all types of mental health problems as classified with CBCL presented significant independent associations with future emergency department visits due to emotional or behavioural problems. CBCL-based attention or hyperactivity problems independently predicted all

components of the poor life chances outcome, whereas DAWBA did not predict any of them. Conversely, DAWBA-based classifications of any, internalising, and externalising problems independently predicted suicide attempts and severe self-harm, whereas CBCL did not (specific independent associations are shown in appendix 2 p 27).

In terms of implementation, the use of a checklist has some advantages over a diagnosis given by a clinician that is aided by a semi-structured algorithm. Most children and adolescents with mental health problems live in low-income and middle-income countries, where 75–90% of those in need do not receive a mental health diagnosis and professionals are scarce.[5,6,29] Moreover, despite high-income countries having higher mental health expenditure and more health-care professionals than other countries,[6] bureaucracy (eg, the UK National Health Service requires a referral from a general practitioner to obtain a psychiatrist appointment, where a clinical diagnosis can be given) and costs involved in a clinician-based diagnosis could impose barriers to treatment and also relocate resources.[7] Thus, regardless of socioeconomic context, classifications based on symptom checklists could reduce the treatment gap[5,6,8,29] by providing a low-cost and easy-to-implement alternative to detect young people at risk for poor life outcomes based on their mental health assessment.

Our study does not aim to replace diagnostic approaches but to highlight their utility. Checklists are unlikely to replace the clinician-based diagnosis because diagnosis is also a process rather than the mere result of classifying. However, if we consider classification pluralism,[30] different approaches can be incorporated in measurement-based care for different purposes, such as diagnosis and follow-ups, as is common when diagnosing diabetes or hypertension.[8] After the diagnosis is made (eg, with a structured interview-based approach), follow-up could be carried out by monitoring a specific characteristic (using a symptom checklist), which also informs on risk of future events.

Although this cohort had multiple strengths, including a large, community-based sample, participants with normally distributed cognition, and a low attrition rate in long-term follow-up, several limitations should be noted. These include reliance on self-reporting and parent reports; rare outcomes leading to low power; no exploration of different taxonomies; no evaluation of different T-score threshold; external validity concerns; potential cultural variations; focus on predictive validity; limitations in assessing informants and age groups; the need for replication with different tools; and variability in prediction properties across outcomes. These limitations are discussed in more detail in appendix 2 (p 3).

In summary, we found no evidence of a clear advantage of semi-structured diagnostic interviews over symptom checklists for risk assessment of important outcomes. Comparing classification approaches analysed separately
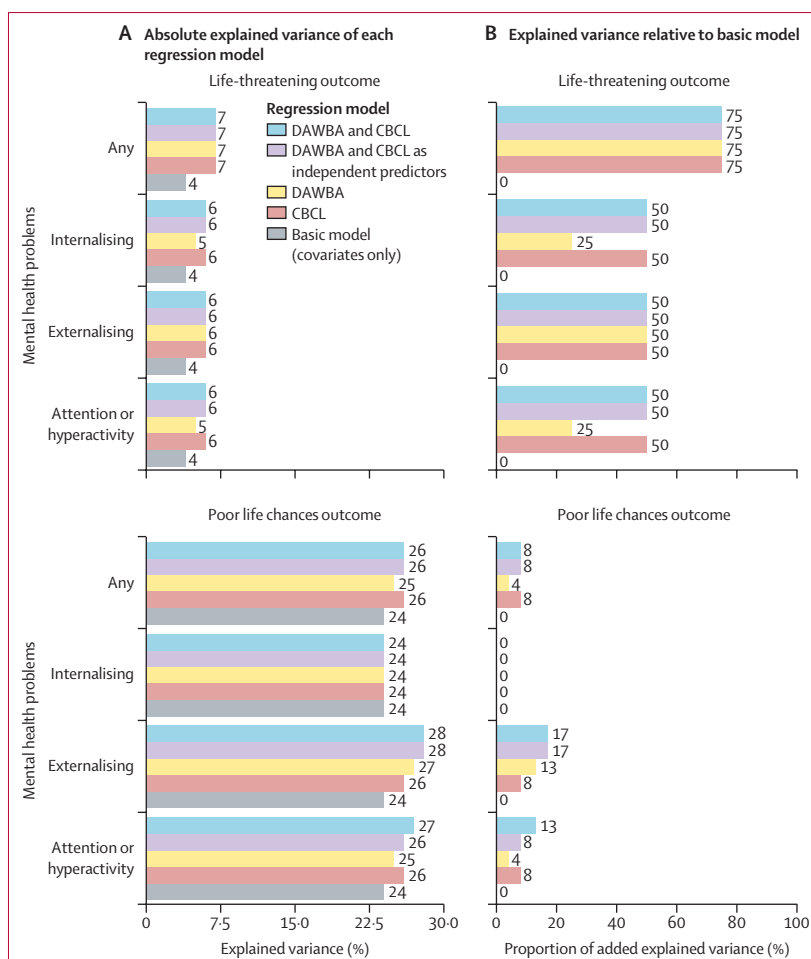
Figure 2: Absolute explained variance (A) and explained variance relative to regression models using covariates only (B)

The x-axis shows explained variance in A and the proportion of added explained variance in B, calculated by dividing models with classification approaches as predictors by the explained variance of regression models including covariates only. Regression models used the composite life-threatening or poor life chances outcomes as dependent variables. The independent variables were the basic model (age, sex, ethnicity, and study site, and socioeconomic group), DAWBA, CBCL (as separated models), DAWBA and CBCL in the multiple regression model, and a nominal classification variable containing participants classified by both DAWBA and CBCL (comparator group), DAWBA only, CBCL only, or neither. All regression models were adjusted for covariates (age, sex, ethnicity, study site, and socioeconomic group) and weighted for attrition. CBCL=Child Behavior Checklist. DAWBA=Developmental and Well-being Assessment.

showed that structured interview and symptom checklists have a considerable amount of overlap when informing future outcomes. Moreover, they also present independent associations, with no clear addition in explained variance when compared with separated regression models. Multiple classification approaches could capture those at risk for specific poor life outcomes due to specific mental health problems and the approaches might have different roles and purposes in a patient's care. Moving forward, other aspects should be considered when comparing the utility of classification approaches, such as the capacity to capture clinical improvement from interventions, and their feasibility in different clinical settings.

References

1 Salum GA, Mogg K, Bradley BP, et al. Threat bias in attention orienting: evidence of specificity in a large community-based study. *Psychol Med* 2013; **43:** 733–45.

2 Shaw P, Gilliam M, Liverpool M, et al. Cortical development in typically developing children with symptoms of hyperactivity and impulsivity: support for a dimensional view of attention deficit hyperactivity disorder. *Am J Psychiatry* 2011; **168:** 143–51.

3 Markon KE, Chmielewski M, Miller CJ. The reliability and validity of discrete and continuous measures of psychopathology: a quantitative review. *Psychol Bull* 2011; **137:** 856–79.

4 Zachar P, Kendler KS. The philosophy of nosology. *Annu Rev Clin Psychol* 2017; **13:** 49–71.

5 Patel V, Saxena S, Lund C, et al. The *Lancet* Commission on global mental health and sustainable development. *Lancet* 2018; **392:** 1553–98.

6 Ridley M, Rao G, Schilbach F, Patel V. Poverty, depression, and anxiety: causal evidence and mechanisms. *Science* 2020; **370:** eaay0214.

7 Lahey BB, Tiemeier H, Krueger RF. Seven reasons why binary diagnostic categories should be replaced with empirically sounder and less stigmatizing dimensions. *JCPP Adv* 2022; **2:** e12108.

8 Bhugra D, Tasman A, Pathare S, et al. The WPA–*Lancet Psychiatry* Commission on the future of psychiatry. *Lancet Psychiatry* 2017; **4:** 775–818.

9 Kendler KS. DSM disorders and their criteria: how should they inter-relate? *Psychol Med* 2017; **47:** 2054–60.

10 Momen NC, Plana-Ripoll O, Agerbo E, et al. Association between mental disorders and subsequent medical conditions. *N Engl J Med* 2020; **382:** 1721–31.

11 Dalsgaard S, McGrath J, Østergaard SD, et al. Association of mental disorder in childhood and adolescence with subsequent educational achievement. *JAMA Psychiatry* 2020; **77:** 797–805.

12 Goodman R, Ford T, Richards H, Gatward R, Meltzer H. The Development and Well-Being Assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. *J Child Psychol Psychiatry* 2000; **41:** 645–55.

13 Aebi M, Kuhn C, Metzke CW, Stringaris A, Goodman R, Steinhausen HC. The use of the Development and Well-Being Assessment (DAWBA) in clinical practice: a randomized trial. *Eur Child Adolesc Psychiatry* 2012; **21:** 559–67.

14 Cuthbert BN, Insel TR. Toward the future of psychiatric diagnosis: the seven pillars of RDoC. *BMC Med* 2013; **11:** 126.

15 Kotov R, Krueger RF, Watson D, et al. The Hierarchical Taxonomy of Psychopathology (HiTOP): a quantitative nosology based on consensus of evidence. *Annu Rev Clin Psychol* 2021; **17:** 83–108.

16 Rescorla L, Ivanova MY, Achenbach TM, et al. International epidemiology of child and adolescent psychopathology ii: integration and applications of dimensional findings from 44 societies. *J Am Acad Child Adolesc Psychiatry* 2012; **51:** 1273–83.e8.

17 Achenbach TM, Rescorla LA. Manual for the ASEBA school-age forms & profiles. Burlington, VT: University of Vermont Research Center for Children, Youth & Families, 2001.

18 Rosenman S, Korten A, Medway J, Evans M. Dimensional vs. categorical diagnosis in psychosis. *Acta Psychiatr Scand* 2003; **107:** 378–84.

19 Kim H, Turiano NA, Forbes MK, Kotov R, Krueger RF, Eaton NR. Internalizing psychopathology and all-cause mortality: a comparison of transdiagnostic vs. diagnosis-based risk prediction. *World Psychiatry* 2021; **20:** 276–82.

20 Boyle MH, Duncan L, Wang L, Georgiades K. Problem checklists and standardized diagnostic interviews: evidence of psychometric equivalence for classifying psychiatric disorder among children and youth in epidemiological studies. *J Child Psychol Psychiatry* 2023; **64:** 779–86.

21 Kotov R, Cicero DC, Conway CC, et al. The Hierarchical Taxonomy of Psychopathology (HiTOP) in psychiatric practice and research. *Psychol Med* 2022; **52:** 1666–78.

22 Tucker J, Milburn A, Marmot M, Lister R, Stewart K. Improving children's life chances. Sept 29, 2016. https://cpag.org.uk/shop/ policy-book/improving-childrens-life-chances (accessed Oct 25, 2022).

23 Martin EA, Jonas KG, Lian W, et al. Predicting long-term outcomes in first-admission psychosis: does the Hierarchical Taxonomy of Psychopathology aid DSM in prognostication? *Schizophr Bull* 2021; **47:** 1331–41.

24 Krueger RF, Hobbs KA, Conway CC, et al. Validity and utility of Hierarchical Taxonomy of Psychopathology (HiTOP): II. Externalizing superspectrum. *World Psychiatry* 2021; **20:** 171–93.

25 Salum GA, Gadelha A, Pan PM, et al. High risk cohort study for psychiatric disorders in childhood: rationale, design, methods and preliminary results. *Int J Methods Psychiatr Res* 2015; **24:** 58–73.

26   Rescorla LA, Ginzburg S, Achenbach TM, et al. Cross-informant agreement between parent-reported and adolescent self-reported problems in 25 societies. *J Clin Child Adolesc Psychol* 2013; **42:** 262–73.

27   Ziebold C, Evans-Lacko S, Andrade MCR, et al. Childhood poverty and mental health disorders in early adulthood: evidence from a Brazilian cohort study. *Eur Child Adolesc Psychiatry* 2023; **32:** 903–14.

28   Stevenson M, Sergeant E. epiR: tools for the analysis of epidemiological data. 2022. https://CRAN.R-project.org/package=epiR (accessed Nov 1, 2022).

29   Pan American Health Organization. A new agenda for mental health in the Americas: report of the Pan American Health Organization high-level commission on mental health and COVID-19. May 17, 2023. https://iris.paho.org/handle/10665.2/57508 (accessed June 13, 2023).

30   Markon KE. Epistemological pluralism and scientific development: an argument against authoritative nosologies. *J Pers Disord* 2013; **27:** 554–79.