Soc Psychiatry Psychiatr Epidemiol (2011) 46:521–532
DOI 10.1007/s00127-010-0219-x

ORIGINAL PAPER

# The 'DAWBA bands' as an ordered-categorical measure of child mental health: description and validation in British and Norwegian samples

**Anna Goodman · Einar Heiervang · Stephan Collishaw · Robert Goodman**

Received: 7 November 2009 / Accepted: 23 March 2010 / Published online: 9 April 2010


**Abstract**
*Purpose* To describe and validate the 'DAWBA bands'. These are novel ordered-categorical measures of child mental health, based on the structured sections of the Development and Well-Being Assessment (DAWBA).
*Methods* We developed computer algorithms to generate parent, teacher, child and multi-informant DAWBA bands for individual disorders and for groups of disorder (e.g. 'any emotional disorder'). The top two (out of 6) levels of the DAWBA bands were used as computer-generated DAWBA diagnoses. We validated these DAWBA bands in 7,912 British children (7–19 years) and 1,364 Norwegian children (11–13 years), using clinician-rated DAWBA diagnoses as a gold standard.
*Results* In general, the prevalence of clinician-rated diagnosis increased monotonically across all levels of the DAWBA bands, and also showed a dose–response association with service use and risk factors. The prevalence estimates of the computer-generated DAWBA diagnoses were of roughly comparable magnitude to the prevalence estimates from the clinician-generated diagnoses, but the estimates were not always very close. In contrast, the estimated effect sizes, significance levels and substantive conclusions regarding risk factor associations were very similar or identical. The multi-informant and parent DAWBA bands performed especially well in these regards.
*Conclusion* Computer-generated DAWBA bands avoid the cost and delay occasioned by clinical rating. They may, therefore, sometimes provide a useful alternative to clinician-rated diagnoses, when studying associations with risk factors, generating rough prevalence estimates or implementing routine mental health screening.

**Keywords** Computer-generated diagnoses · Diagnostic interview · Child mental health · Prevalence · Associations

**Electronic supplementary material** The online version of this article (doi:10.1007/s00127-010-0219-x) contains supplementary material, which is available to authorized users.

A. Goodman (✉)
Department of Epidemiology and Population Health, London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK
e-mail: anna.goodman@lshtm.ac.uk

E. Heiervang
Førde Health Authority, Førde, Norway

S. Collishaw
Cardiff University School of Medicine, Cardiff, UK

R. Goodman
King's College London Institute of Psychiatry, London, UK

## Introduction

Structured diagnostic interviews are currently the closest approximation to a 'gold standard' available in child psychiatry [1]. Some such interviews are 'respondent-based', asking pre-specified questions in a pre-set order and applying fixed algorithms to assign diagnoses. Others are 'investigator-based', pre-specifying the topics to be covered but allowing flexibility in the questioning used to cover these. Still others combine elements of the two, using fully structured sets of questions supplemented by open-ended transcripts that are reviewed by trained clinical raters before assigning diagnoses. For many clinicians, diagnoses involving some element of clinical judgement have higher face validity because they allow interpretation of informants' open-ended accounts. This can be important in

identifying and clarifying potential misunderstandings, particularly for 'out-of-the-ordinary' symptoms [2]. Clinical judgement may also be important when seeking to synthesise information across multiple informants, or for appreciating the 'bigger picture' of a child's condition and identifying children who merit a not otherwise specified diagnosis [3, 4].

It is, therefore, plausible that allowing some role for clinical judgement will improve diagnostic decision-making, provided the interviewers and raters are sufficiently trained to generate reliable results. This is, however, an issue regarding which there has been very little formal assessment [1]. Moreover, researchers and service providers often face a trade-off between maximising the validity of their measures and choosing measures that can feasibly be administered on a large scale. Employing clinical interviewers or raters may be prohibitively expensive and time-consuming for large epidemiological surveys or population screening programs. In such cases, it may be preferable to use computer-generated outcomes based on fully structured sets of closed questions. In this paper, we introduce and validate a novel outcome of this sort, the DAWBA bands.

## Description of the DAWBA interview

The Development and Well-Being Assessment (DAWBA) interview is a semi-structured interview administered to parents of children age 4–16, and to children and adolescents (henceforth 'children') over the age of 11. A briefer questionnaire is administered to teachers [5, 6]. The DAWBA consists of a mixture of open and closed questions about child mental health symptoms and their impact. Clinical raters then use responses from all informants to assign psychiatric diagnoses.

The DAWBA interview can either be administered by trained lay interviewers or else self-completed online. The main DAWBA interview is fully structured and has separate sections covering individual emotional, behavioural and hyperactivity disorders, plus sections on autistic spectrum disorders, eating disorders and tics. Teachers are not asked in detail about emotional disorders and children are not asked in detail about oppositional defiant disorder or hyperactivity; this is because of the poorer quality of information which these informants provide about these conditions [7, 8]. The questions for each disorder closely follow the diagnostic criteria operationalised in the Diagnostic and Statistical Manual of Mental Disorders, 4th edition [DSM-IV; 9] or the International Classification of Diseases [ICD-10; 10]. Each section contains around 20–25 questions, with skip-rules such that this full set of questions is only administered when children are reported to have relevant problems in initial screening questions.

If an informant completes a structured section in full, this is followed by open-ended questions. Informants are encouraged to describe difficulties in detail and give specific examples, with their answers recorded verbatim by the interviewer or typed as free text when the interview is administered online. Experienced clinicians then review the open and closed accounts of all available informants, and bring these together to rate the presence or absence of individual diagnoses [4]. We refer to these as 'clinician-rated DAWBA diagnoses'. In Britain, these clinician-rated DAWBA diagnoses have been shown to have good reliability and to be able to discriminate well between community and clinic samples and between different diagnoses [6]. Both in Britain and Norway, clinician-rated DAWBA diagnoses generated reasonable prevalence estimates for mental disorders, showed plausible comorbidity and risk factor associations, and strongly predicted contact with mental health services [11, 12].

## Rationale for creating the 'DAWBA bands'

In the task of reviewing the structured sections of the DAWBA interview, clinicians have always been assisted by computer algorithms. As initially designed, these computer algorithms assessed whether reported symptoms and impairment in the structured sections met DSM-IV and ICD-10 criteria. These automated assessments were only intended as a guide, and experienced clinical raters often overturned or modified the computer algorithm's predictions. For example, in the first large-scale epidemiological use of the DAWBA (the British Child and Adolescent Mental Health Survey of 1999 [13]), 20.8% (214/1,029) of those assessed as having a DSM-IV disorder by the computer algorithm were not given any clinician-rated diagnosis. Conversely, 1.8% (168/9,409) of those assessed as not having any disorder by the computer algorithms did receive at least one clinician-rated diagnosis.

There is, therefore, evidence that experienced raters often do disagree with the computer algorithms. When providing training to raters in everyday mental health clinics, however, RG and EH's impression was that inexperienced clinical raters were often reluctant to disagree with the yes/no computer ratings. Partly to make the computer algorithms seem less prescriptive, we created six levels of prediction of the probability of disorder, ranging from very unlikely to probable. We call these the 'DAWBA bands', and since 2005 clinicians rating DAWBAs have been presented with the DAWBA bands rather than the original binary diagnostic predictions.

In addition to being of use within the DAWBA rating process, we anticipated that the DAWBA bands would have advantages as outcomes in their own right. As

ordered-categorical measures, they could be used to examine whether there was a dose–response relationship between mental health and putative risk and protective factors, or to study children with particularly good mental health. Moreover, we are increasingly interested in how information technology can be used to conduct large-scale epidemiological studies with frequent follow-up at low cost. Online interviewing can potentially avoid the substantial costs associated with employing interviewers [14]. Computer-generated diagnostic predictions could offer additional savings by avoiding the costs of employing clinical raters. If such computer-generated measures yielded sufficiently similar substantive findings regarding prevalence or association with risk factors, then some studies might prefer to use these as outcomes.

This paper, therefore, describes the creation of the 'DAWBA bands' and validates them in two independent samples from Britain and Norway, using the clinician-rated DAWBA diagnoses as a gold standard. Specifically, we examine (1) whether the DAWBA bands are ordered-categorical measures of mental health, (2) whether the DAWBA bands show dose–response associations with known risk factors and service use, (3) whether a binary measure based on the DAWBA bands (bottom four levels vs. top two levels) provides good estimates of disorder prevalence, and (4) whether this binary measure yields correct substantive conclusions regarding associations with risk factors.

## Methods

### Sample for creating the DAWBA bands

The British Child and Adolescent Mental Health Surveys (B-CAMHS) of 1999 and 2004 were two nationally representative surveys conducted in England, Scotland and Wales [13, 15]. Children were sampled between ages 5–15 years in 1999 and 5–16 years in 2004, using the Child Benefit Register as a sampling frame and with a clustered design by postal sector. The principal caregivers ('parents') of selected children were approached to give informed consent for face-to-face interview. With the parent's permission, the child's teacher and children aged 11 or over were also approached. Between the two B-CAMHS surveys, 26,545 children were selected and 18,415 (69.4%) participated—10,438 in B-CAMHS99 and 7,977 in B-CAMHS04. Their mean age was 10.2 years and 50.7% were male. Amongst the 18,415 participating children, parent DAWBAs were available for 18,112 (98.4%); teacher DAWBAs for 14,366 (78.0%); and child DAWBAs for 7,672 (89.4% of those aged 11–16).

### Method for creating the DAWBA bands

We developed computer algorithms that used the symptoms and impact recorded in the structured sections of the DAWBA to generate ordered-categorical measures of the prevalence of clinician-based disorders. We generated both informant-specific measures (e.g. using the structured section on conduct disorder from the parent DAWBA interview to generate a parent DAWBA band for conduct disorder) and multi-informant measures (e.g. using the structured sections on conduct disorder from the parent, teacher and child DAWBA interviews). As summarised in Table 1, up to six levels were created for each DAWBA band, corresponding to the following approximate prevalences (in an epidemiological sample):

- level 0; <0.1% of children in this band have the disorder in question;
- level 1; ≈0.5% of children in this band have the disorder in question;
- level 2; ≈3% of children in this band have the disorder in question;
- level 3; ≈15% of children in this band have the disorder in question;
- level 4; ≈50% of children in this band have the disorder in question;
- level 5; >70% of children in this band have the disorder in question.

The levels were chosen to provide an approximately evenly spaced progression in terms of the log odds. Note that these prevalences for each level come from an epidemiological sample; in clinical or other high-risk samples, the observed prevalence corresponding to each level is expected to be greater, reflecting higher prior probability of disorder. We derived the corresponding 'computer-predicted DAWBA diagnoses' by combining levels 0–3 as 'absent' and levels 4–5 as 'present'.

The algorithms used to create these DAWBA bands drew partly on the symptom and impairment criteria in ICD-10 and DSM-IV, but did not always do so. For example, if the criteria suggested that four out of seven symptoms needed to be present, fewer than four would still carry some weight and more than four would add additional weight. Some symptoms within the list of seven might carry greater weight than others, as might particularly distinctive combinations of symptoms. In some instances, prediction involved symptom duration and age of onset as well as symptoms and impact. These decisions were driven by our empirical finding that using such techniques allowed us to predict the probability of disorder better than using simple 'counting' rules, and the modifications were only made if they increased predictive accuracy. At the same time, we were also guided by our

**Table 1** Summary of DAWBA bands and their available levels

| DAWBA band | DAWBA band levels available | | | |
|---|---|---|---|---|
| | Multi-informant | Parent | Teacher | Child |
| Empirically created DAWBA bands | | | | |
| Separation anxiety, DSM-IV criteria | 01235 | 1235 | | 1234 |
| Separation anxiety, ICD-IV criteria | 0124 | 024 | | 123 |
| Specific phobia | 0134 | 0134 | | 0134 |
| Social phobia | 01234 | 01234 | | 01234 |
| Post-traumatic stress disorder | 01234 | 01234 | | 01234 |
| Panic disorder | 01234 | 023 | | 0134 |
| Agoraphobia | 01234 | 023 | | 0134 |
| Obsessive compulsive disorder | 01234 | 01234 | | 01234 |
| Generalised anxiety disorder | 01234 | 1234 | | 1234 |
| Depression | 012345 | 01345 | | 01345 |
| Any emotional disorder | | | 1234 | |
| Oppositional defiant disorder | 012345 | 12345 | 12345 | |
| Conduct disorder | 012345 | 12345 | 12345 | 1234 |
| Hyperactivity, DSM-IV criteria | 012345 | 012345 | 01234 | |
| Hyperactivity, ICD-10 criteria | 012345 | 012345 | 01234 | |
| Autism | | 02345 | | |
| Eating disorder | 0124 | 0124 | | 0124 |
| Tic disorder, DSM-IV criteria | | 012345 | | |
| Tic disorder, ICD-IV criteria | | 012345 | | |
| Higher-level DAWBA (highest of constituent disorders) | | | | |
| Any emotional disorder bands | 012345 | 12345 | | 12345 |
| Any behavioural disorder | 012345 | 12345 | 12345 | |
| Any disorder | 012345 | 12345 | | |

Level 0 created to correspond to <0.1% prevalence disorder in question in an epidemiological sample; level 1 ≈0.5% prevalence; level 2 ≈3% prevalence; level 3 ≈15% prevalence; level 4 ≈50% prevalence; level 5 > 70% prevalence

Some DAWBA bands do not have all six levels because no set of responses corresponded empirically to that nominal prevalence of diagnosis

Some DAWBA bands do not exist at all because the DAWBA interview for that informant does not contain a section on the disorder in question

understanding of the items, such that we only made modifications that made theoretical and clinical sense. In this way, we aimed to create algorithms that were more likely to be valid in other study samples.

We initially created the DAWBA algorithms using the B-CAMHS04 baseline data, using these to generate bands with the prevalences set out above. In most instances, the observed prevalences were very similar when the algorithm was then applied to B-CAMHS99 baseline data; where this was not the case, further adjustments were made to ensure the algorithm worked well for both samples. For separation anxiety, hyperactivity disorder and tic disorder, separate bands were created for the ICD-10 and DSM-IV diagnostic criteria as the two international classifications have substantially different criteria in these instances. For all other disorders, the close similarity of the ICD-10 and DSM-IV diagnostic criteria meant that we created only a single set of bands. As Table 1 shows, not all of the DAWBA bands had all levels. This corresponds to cases in which no set of responses corresponded empirically to one of the approximate nominal prevalences of diagnosis. For example a '<0.1% prevalence (level 0)' category could not be created for separation anxiety (DSM-IV criteria) by parent report because even when the parent DAWBA provided no indication of a separation anxiety disorder, the observed prevalence of diagnosis was still closer to 0.5% than 0.1%. In such cases, diagnosis would be based on a convincing account of separation anxiety in the parent open-ended transcript or in the child DAWBA.

These empirically created DAWBA bands were based on individual sections of the DAWBA. As shown in Table 1, we also created several higher-level ordered-categorical DAWBA bands. These were any emotional disorder (multi-level, parent and child DAWBA bands), any behavioural disorder (multi-level, parent and teacher DAWBA bands) and any mental disorder (multi-level and parent DAWBA bands). These were created as being equal to the value of the highest DAWBA band level observed

for any constituent empirically created DAWBA band; for example, if the parent DAWBA band for oppositional defiant disorder was level 3 and the parent DAWBA band for conduct disorder was level 1, then the parent DAWBA band for any behavioural disorder would be 3. We did not create teacher or child 'any disorder' bands because the teacher DAWBA provides only limited coverage of emotional disorders and the child DAWBA provides only limited coverage of behavioural/hyperactivity disorders. As such, we felt it would be misleading to claim that interviews with these informants really can generate a DAWBA band that provides proper coverage across all common child mental disorders.

## Samples for validating the DAWBA bands

### *British sample*

Both B-CAMHS surveys included a 3-year follow-up. B-CAMHS99 oversampled participants with disorders at baseline [11], while B-CAMHS04 attempted to follow-up all participants [16]. In total, 11,969 children were selected for follow-up, and 7,912 (66.1%) participated (2,586 from B-CAMHS99 and 5,326 from B-CAMHS04). Their mean age was 13.3 years (range 7–19 years) and 51.7% were male.

### *Norwegian sample*

We also evaluated the parent DAWBA band using a Norwegian sample from the second wave of the Bergen Child Study. The Bergen Child Study is a series of cross-sectional surveys of children living in the medium-sized Norwegian city of Bergen and born between 1993 and 1995 [12, 17]. The second cross-sectional study was carried out in 2006/2007, when the children were in the fifth to seventh grade (aged 11–13 years). All state-funded schools and most private schools in Bergen agreed to send parents an invitation to participate, giving a total eligible population of 9,218 children. Parents of these children were initially invited to complete a paper questionnaire survey, including information on demographic characteristics and some putative risk factors. Parents who completed the questionnaire ($N = 5{,}128$) were then given a unique ID number and password, and invited to log onto a special website and complete the DAWBA interview online. In total, 1,364 parents (26.6% of those invited, 14.8% of total eligible population) completed the online DAWBA in full, and these form the study population for the present paper. These parents were unrepresentative in a number of ways, leading to an underestimation of prevalence in some instances [14]. For the purposes of this paper, this selection bias does not matter as our primary interest is in comparing computer-generated and clinician-rated DAWBA diagnoses. The Norwegian sample was 47.9% male with a mean age of 12.1 years.

## Mental health measures

We applied the DAWBA band computer algorithms to our validation samples, generating both the DSM-IV and the ICD-10 versions where applicable. We used the DSM-IV versions when creating the DAWBA bands for broad diagnostic groups, e.g. emotional disorders. The only DAWBA band we were unable to calculate was the parent autism DAWBA band for the British sample, because the B-CAMHS follow-up survey interviews did not contain the DAWBA autism section. In addition, the B-CAMHS99 follow-up did not include the DAWBA eating and tic disorder sections.

In both Britain and Norway, clinician-rated DAWBA diagnoses were assigned by experienced child psychiatrists who were trained by and subsequently supervised by the author (RG) who developed the instrument. High inter-rater reliabilities have previously been reported for the British and Norwegian studies [12, 15]. Diagnoses were made according to both DSM-IV and ICD-10 criteria. In this paper, we always use the DSM-IV disorders, except for comparisons with the ICD-10 versions of the separation anxiety, hyperactivity and tic disorder DAWBA bands. In DSM-IV, children who meet the criteria for conduct disorder cannot also receive a diagnosis of oppositional defiant disorder, even if they otherwise meet the diagnostic criteria. For the purposes of this paper, however, we counted such children as having oppositional defiant disorder. We did this because knowing that a child meets the criteria for both oppositional defiant as well as conduct disorder adds useful information, e.g. for understanding patterns of comorbidity [5].

When clinicians were rating DAWBAs from the B-CAMHS99 follow-up, they were presented with the original binary computer-generated assessments as to whether the child met ICD-10/DSM-IV diagnostic criteria. They therefore assigned clinical diagnoses blind to the DAWBA bands, which were only calculated retrospectively. In contrast, in the B-CAMHS04 follow-up and the Norwegian surveys, clinicians were presented with the DAWBA bands when making diagnoses. To address this possible source of circularity, we therefore conducted sensitivity analyses restricted to the B-CAMHS99 follow-up.

## Mental health service contact

Both B-CAMHS follow-up surveys contained information from parents about whether their child had used a child mental health specialist service over the past three years.

This was ascertained from a single question at 3-year follow-up in B-CAMHS04 [16] and by combining information between 2- and 3-year follow-up in B-CAMHS99 [18]. The Norwegian study asked parents if their child had ever used a mental health specialist service [12].

### Child and family covariates

Two of our aims involve examining relationships with child mental health risk factors. To do this, we selected and defined a priori a small number of established correlates of child mental disorder covering a range of domains (child characteristics, family composition, and family socioeconomic position).

In the British sample, these covariates were gender; age at baseline (5–8, 9–12, 13–16 years); survey year (1999 vs. 2004); child's general health by parent report (5-point scale with response options very good, good, fair, bad and very bad); parent report of whether the child had a learning difficulty or dyslexia; family type ('traditional' two-parent family, lone parent family, stepfamily); parent's mental health (measured using the 12-item General Health Questionnaire [19]); parent's highest educational level (no qualifications, GCSE's, A-level/diploma, degree); and housing tenure (owner occupied vs. rented).

In the Norwegian sample, the covariates were gender; parent report of whether the child had a specific learning difficulty or mental retardation; family type ('traditional' two-parent family vs. 'non-traditional family'); parent rating of household income (very poor, poor, fair, good, very good); and whether one or both parents were immigrants from low or middle income countries.

### Statistical methods

All analyses were conducted in Stata 10.1. We adjusted for the clustered sampling design of the British surveys when comparing prevalence estimates and risk factor associations between the computer-generated and clinician-rated DAWBA diagnoses. These analyses also used appropriate weights to correct for the oversampling of children with disorder in the follow-up to B-CAMHS99 [11], and to allow for the ways in which the Norwegian full-responders were not representative of those invited to complete the DAWBA [14].

## Results

### The DAWBA bands as an ordered-categorical measure

The DAWBA bands worked well as ordered-categorical measures of the prevalence of disorder in both the British and Norwegian samples. As shown in Table 2, the proportion of children receiving a clinician-rated DAWBA diagnosis for common child mental disorders generally increased across the full range of the DAWBA bands. These results were very similar in sensitivity analyses restricted to the B-CAMHS99 follow-up, when clinical raters were blind to the DAWBA bands (see Supplementary material). A monotonic increase in disorder prevalence was also observed for the DAWBA bands for individual diagnoses (see Supplementary material), although the number of children was insufficient to provide a meaningful evaluation for the higher levels of some of the rarer disorders.

There were only a few notable violations of the monotonic increase in the prevalence of diagnosis across the DAWBA bands. These were a few cases in which the multi-informant level 1 band contained zero children with a disorder, despite containing large numbers of children. In Table 2 this applied to behavioural disorder, with none of the 2,114 level 1 children receiving a clinician-rated diagnosis. The same was also true of the DAWBA bands for separation anxiety, social phobia, generalised anxiety disorder, oppositional defiant disorder and eating disorder (see Supplementary material).

### Assessing dose–response relationships using the DAWBA bands

As Figs. 1 and 2 show both mental health service use and the prevalence of most risk factors increased across the full range of the multi-informant and parent DAWBA bands. This dose–response relationship included levels 0–3 on the DAWBA bands, i.e. those children who make up most of the population, but who would typically be grouped together as 'non-disordered' children. This highlights the ways in which the DAWBA bands may allow researchers to increase statistical power by making distinction amongst children in the normal range.

### Comparison of prevalence estimates between the clinician-rated and computer-generated DAWBA diagnoses

Having examined the DAWBA bands as ordered-categorical measures, the remaining analyses focus instead on a derived binary measure: the computer-generated DAWBA diagnoses (levels 4–5 vs. 0–3). Figure 3 presents the estimated prevalence of diagnosis for each disorder using the multi-informant and parent DAWBA bands, and gives chance-corrected kappas for their agreement with the clinician-rated diagnoses. Supplementary material provides full results for all DAWBA bands and all informants, plus

**Table 2** Prevalence of clinician-rated diagnosis for each disorder by DAWBA band level

| Disorder | British sample (B-CAMHS follow-up) | | | | | | | | | Norwegian sample | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Level | Multi-informant ($N = 7,816$) | | Parent ($N = 7,777$) | | Teacher ($N = 4,775$) | | Child ($N = 4,974$) | | Parent ($N = 1,364$) | |
| | | *N* | *%D* | *N* | *%D* | *N* | *%D* | *N* | *%D* | *N* | *%D* |
| Any disorder | 0 | 334 | 0.0 | – | – | – | – | – | – | – | – |
| | 1 | 3,342 | 0.2 | 2,991 | 1.5 | – | – | – | – | 1,039 | 0.1 |
| | 2 | 2,491 | 1.6 | 3,430 | 3.8 | – | – | – | – | 195 | 4.6 |
| | 3 | 947 | 24.2 | 767 | 24.5 | – | – | – | – | 81 | 29.6 |
| | 4 | 454 | 74.7 | 309 | 70.6 | – | – | – | – | 31 | 58.1 |
| | 5 | 248 | 92.7 | 280 | 92.9 | – | – | – | – | 18 | 83.3 |
| Any emotional disorder | 0 | 1,184 | 0.1 | – | – | – | – | – | – | – | – |
| | 1 | 4,964 | 0.3 | 5,902 | 1.1 | 4,244 | 2.9 | 3,727 | 1.2 | 1,152 | 0.1 |
| | 2 | 811 | 2.8 | 1,189 | 4.1 | 192 | 8.3 | 620 | 3.5 | 135 | 4.4 |
| | 3 | 565 | 21.6 | 447 | 21.3 | 278 | 12.6 | 460 | 16.7 | 55 | 27.3 |
| | 4 | 224 | 73.2 | 149 | 67.8 | 61 | 36.1 | 142 | 63.4 | 18 | 61.1 |
| | 5 | 68 | 85.3 | 90 | 83.3 | – | – | 25 | 88.0 | 4 | 75.0 |
| Any behavioural disorder | 0 | 3,629 | 0.0 | – | – | – | – | – | – | – | – |
| | 1 | 2,114 | 0.0 | 3,690 | 0.3 | 3,102 | 1.1 | – | – | 1,226 | 0.0 |
| | 2 | 1,133 | 0.1 | 3,295 | 1.9 | 1,316 | 5.7 | – | – | 75 | 0.0 |
| | 3 | 536 | 22.0 | 383 | 16.7 | 140 | 21.4 | – | – | 35 | 5.7 |
| | 4 | 241 | 73.4 | 206 | 58.7 | 133 | 61.7 | – | – | 14 | 14.3 |
| | 5 | 163 | 87.1 | 203 | 89.2 | 84 | 96.4 | – | – | 14 | 71.4 |
| Hyperactivity DSM-IV criteria | 0 | 5,623 | 0.0 | 6,278 | 0.0 | 2,760 | 0.3 | – | – | 1,238 | 0.0 |
| | 1 | 1,210 | 0.4 | 611 | 1.3 | 1,196 | 1.1 | – | – | 50 | 0.0 |
| | 2 | 511 | 3.3 | 326 | 2.1 | 429 | 3.5 | – | – | 26 | 3.8 |
| | 3 | 279 | 11.8 | 389 | 11.1 | 314 | 15.0 | – | – | 38 | 26.3 |
| | 4 | 145 | 39.3 | 123 | 41.5 | 76 | 35.5 | – | – | 7 | 14.3 |
| | 5 | 48 | 72.9 | 50 | 70.0 | – | – | – | – | 5 | 60.0 |

*%D* percent with clinician-rated disorder for the disorder in question, using DSM-IV criteria

Some DAWBA bands do not have all six levels because no set of responses corresponded empirically to that nominal prevalence of diagnosis

Some DAWBA bands do not exist at all because the DAWBA interview for that informant does not contain a section on the disorder in question; for full results for all individual diagnoses, see Supplementary material

the sensitivities, specificities, positive predictive values and negative predictive values.

In most cases, the computer-generated DAWBA bands gave prevalence estimates that were of broadly comparable magnitude to the clinician-rated diagnoses. The prevalence estimates generated from the teacher and child DAWBA bands were, however, typically further from the clinician-rated prevalences than those generated from multi-informant and parent DAWBA bands. Moreover, even for multi-informant and parent informant bands, there were several instances where the confidence intervals for the prevalence estimates did not overlap. In particular, the total prevalence of any mental disorder was underestimated by the computer-generated DAWBA diagnoses: 9.5% of children in the British sample had a clinician-rated diagnosis, but the computer-rated prevalence was only 7.8% by the multi-informant DAWBA band and 6.3% by the parent DAWBA band. Similarly, 6.0% of the Norwegian sample had a clinician-rated diagnosis but only 4.7% had a computer-generated diagnosis from the parent DAWBA band.

These prevalence comparisons address the extent to which the clinician-rated and computer-generated diagnoses produced similar findings at the *group* level. As for their *individual-level* agreement, in most cases the kappas were 0.4–0.7, the sensitivity was 0.4–0.8 and the specificity was 0.98–0.99. The corresponding positive predictive values were usually 0.5–0.8 and the negative predictive values 0.96–0.99 (see Supplementary material). Agreement was poorer, however, for the British teacher DAWBA bands for any emotional or hyperactivity disorder and the Norwegian parent band for hyperactivity disorder; in these cases, the kappas were under 0.30, the sensitivity values under 0.3

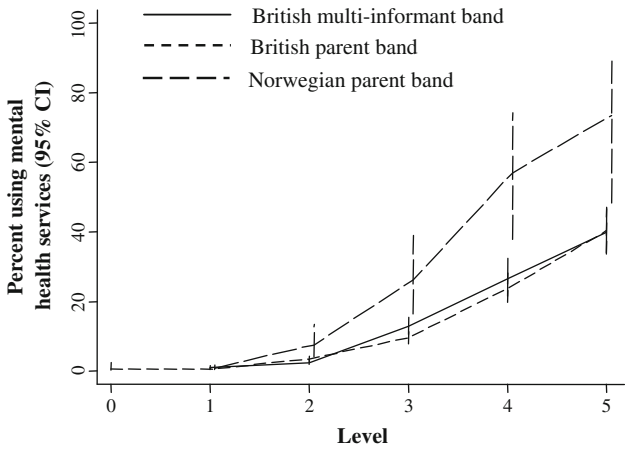


**Fig. 1** Proportion of children with mental health services use, by the multi-informant and parent DAWBA bands for any disorder. The British sample asked about mental health service use over the past 3 years, while the Norwegian sample asked about lifetimes service use

and the positive predictive values under 0.4. Agreement was also somewhat lower in the Norwegian sample than in the British sample. For example, the average value of the four Norwegian kappas presented in Fig. 3 was 0.47, as compared with 0.61 for the corresponding four kappas from the British sample.

### Comparison of risk factor associations using the clinician-rated and computer-generated DAWBA diagnoses

We compared the association with risk factors in multivariable models using two sorts of diagnostic outcome: the computer-generated DAWBA diagnoses and the clinician-rated DAWBA diagnoses (Table 3). In both the British and the Norwegian samples, there was close similarity in both the estimated effect sizes and significance levels, resulting in substantive conclusions which were in most cases identical. As shown in Supplementary material, this similarity extended to models predicting separately to emotional, behavioural and hyperactivity disorders (although in the smaller Norwegian sample, these analyses were underpowered). These analyses generated comparable substantive conclusions about the relative importance of different risk factors for different types of disorder. For example, in the clinician-rated and computer-generated diagnoses alike, child's general health was generally strongly associated with emotional disorders; non-traditional family types with behavioural disorders; and learning difficulties/dyslexia with hyperactivity. As was the case for prevalence estimates, the agreement between the clinician-rated and computer-generated DAWBA diagnoses was particularly good when based on the multi-informant and parent DAWBA bands.

## Discussion

This paper introduces the DAWBA bands, a novel six-level ordered-categorical measure of child mental health based on the structured sections of the DAWBA. In our validation of this measure in samples of British and Norwegian children, we have shown that the DAWBA bands function well as ordered-categorical measures, show a dose–response association with mental health service contact and can also be used to examine dose–response associations with risk factors. We compared the computer-generated DAWBA diagnoses (corresponding to children with levels 4 or 5 of the DAWBA bands) with the 'gold standard' clinician-rated diagnoses. The computer-generated diagnoses yielded very similar results regarding associations with risk factors, but only approximately similar prevalence estimates.

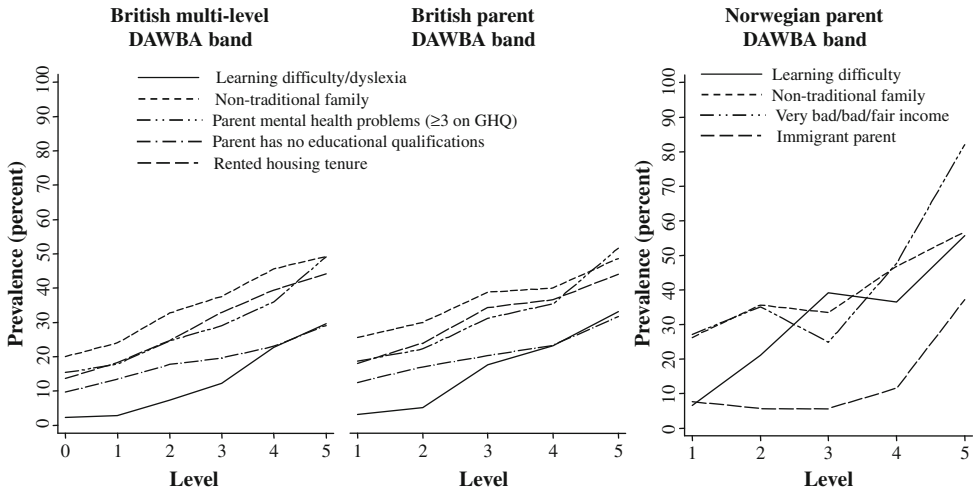


**Fig. 2** Prevalence of child and family risk factors according to the multi-informant and parent DAWBA bands for any disorder (with dichotomisation of continuous variables)

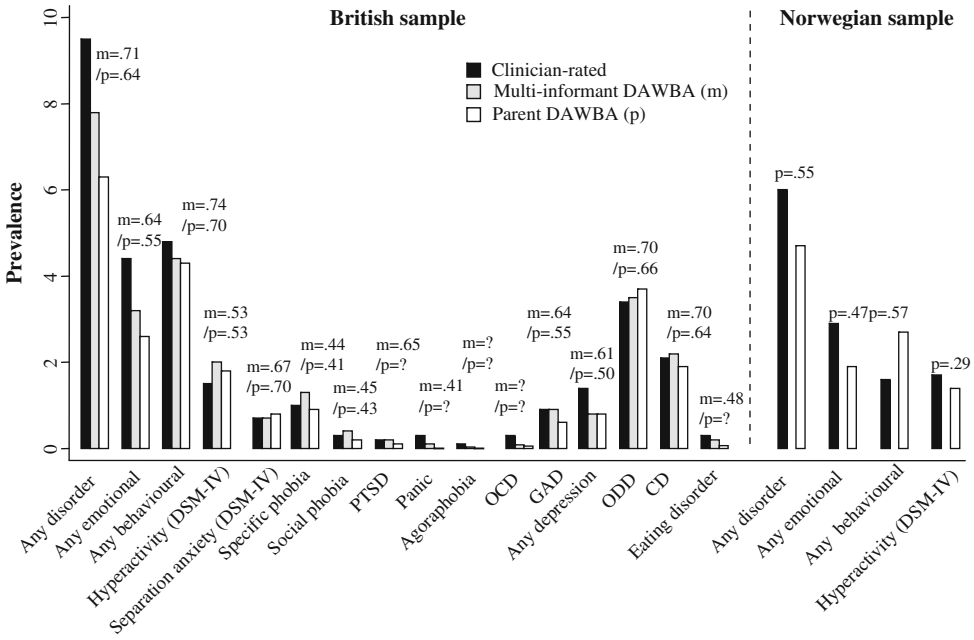


**Fig. 3** Prevalence of disorder estimates for the clinician-rated and computer-generated DAWBA diagnoses. *PTSD* post-traumatic stress disorder, *GAD* generalised anxiety disorder, *OCD* obsessive compulsive disorder, *ODD* oppositional defiant disorder, *CD* conduct disorder, *m* kappa for agreement between the clinician-rated and the multi-informant computer-generated diagnoses, *p* kappa for agreement between the clinician rated and the parent computer-generated diagnoses. Kappas were not calculated when there were fewer than ten children with computer-generated diagnoses, and are indicated by '?'. For full results, see Supplementary material

Before discussing these findings in detail, it is worth reviewing some of the limitations of this study. In the B-CAMHS04 and Norwegian samples, clinical raters were presented with a child's DAWBA bands when assigning diagnoses and might, therefore, have been influenced by these. Nevertheless, the results were very similar in sensitivity analyses restricted to the B-CAMHS99 follow-up sample in which raters were blind to the DAWBA bands. Moreover, as outlined in the "Introduction", experienced clinical raters frequently did overturn the yes/no computer-generated diagnoses in the original version of the DAWBA. Therefore, we believe they will not have been unduly influenced by the even less prescriptive DAWBA bands.

A further limitation is that the DAWBA bands for rare individual disorders had very few children in the higher levels. There was therefore insufficient powered to provide a very rigorous evaluation of these DAWBA bands. Nevertheless, what analyses were possible for the individual DAWBA bands provided no grounds for concern. In addition, the fact that these bands were created using the same successful method that was used for more common bands provides some grounds for reassurance. It is also reassuring that the British and Norwegian surveys generate fairly similar findings, despite having different strengths and limitations. Nevertheless, the kappa for computer/clinician agreement were lower in Norway than in Britain, raising questions as to whether the validity of the bands would fall further in populations with even greater social and cultural differences from Britain. Further validation of the DAWBA bands is, therefore, required in a broader range of settings.

## Potential uses of the DAWBA bands

These findings suggest several potential uses for the DAWBA bands. As an ordered-categorical measure, they can be used to examine associations with risk factors of interest across the full range of child mental health—for example, examining whether there is a dose–response relationship. Ordered-categorical measures may also: increase statistical power; be more sensitive when seeking to detect change over time (including treatment efficacy); and reduce residual confounding when adjusting for mental health as a covariate. The fact that separate parent, teacher and child DAWBA bands exist provides greater scope for triangulating findings across informants, and thereby examining issues such as inter-informant agreement or situation specificity. We have previously shown that the clinician-rated DAWBA diagnoses can be used as a 'gold standard' reference point when examining potential reporting biases between groups on brief questionnaire measures [17]. The ordered-categorical DAWBA bands can increase power to perform such analyses, particularly when examining groups which have a mental health advantage and unusually low prevalences of disorder. When using the DAWBA bands as ordered-categorical variables, however, future researchers may wish to combine levels 1 with 0 for those multi-informant DAWBA bands where we found a zero prevalence of disorder in both levels (any behavioural disorder, separation anxiety, social phobia, generalised anxiety disorder, oppositional defiant disorder and eating disorder).

At a minimum, therefore, DAWBA bands represent a useful complement to the clinician-rated DAWBA diagnoses, since they are ordered-categorical measures and can

**Table 3** Risk factor associations with any mental disorder in multi-variable logistic regression models: comparison of clinician-rated and computer-generated DAWBA diagnoses

| | Clinician-rated DAWBA diagnosis (OR and 95% CI) | Computer-generated multi-informant DAWBA diagnosis (OR and 95% CI) | Computer-generated parent DAWBA diagnosis (OR and 95% CI) |
|---|---|---|---|
| British sample | | | |
| *N* | 7,837 | 7,747 | 7,713 |
| Gender | | | |
| Male | 1** | 1*** | 1** |
| Female | 0.80 (0.68, 0.94) | 0.70 (0.59, 0.84) | 0.74 (0.61, 0.91) |
| Age at baseline | | | |
| 5–8 years | 1 | 1 | 1 |
| 9–12 years | 1.12 (0.93, 1.34) | 1.09 (0.88, 1.34) | 1.00 (0.79, 1.26) |
| 13–16 years | 1.15 (0.94, 1.40) | 1.06 (0.87, 1.29) | 0.88 (0.69, 1.11) |
| Survey year | | | |
| B-CAMHS99 | 1 | 1 | 1 |
| B-CAMHS04 | 0.94 (0.80, 1.11) | 1.19 (0.98, 1.43) | 1.07 (0.87, 1.32) |
| Poor general health | | | |
| Change per level | 1.54 (1.38, 1.72)*** | 1.54 (1.37, 1.73)*** | 1.54 (1.36, 1.75)*** |
| Learning difficulty or dyslexia | | | |
| No | 1*** | 1*** | 1*** |
| Yes | 4.65 (3.78, 5.71) | 3.91 (3.15, 4.85) | 4.46 (3.54, 5.63) |
| Parent mental health | | | |
| Change per point | 1.11 (1.08, 1.13)*** | 1.11 (1.08, 1.14)*** | 1.13 (1.10, 1.16)*** |
| Family type | | | |
| Traditional | 1*** | 1*** | 1** |
| Lone parent | 1.78 (1.42, 2.21) | 1.86 (1.45, 2.38) | 1.60 (1.21, 2.11) |
| Step family | 1.19 (0.96, 1.48) | 1.32 (1.04, 1.68) | 1.14 (0.89, 1.45) |
| Parent's highest educational level | | | |
| No qualifications | 1 | 1 | 1 |
| GCSEs | 0.90 (0.72, 1.14) | 0.88 (0.69, 1.13) | 0.75 (0.58, 0.98) |
| A-levels/diploma | 0.89 (0.69, 1.13) | 0.81 (0.61, 1.06) | 0.73 (0.55, 0.98) |
| Degree | 0.89 (0.66, 1.20) | 0.80 (0.56, 1.13) | 0.82 (0.59, 1.16) |
| Housing tenure | | | |
| Owner occupied | 1*** | 1*** | 1** |
| Rented | 1.58 (1.30, 1.91) | 1.57 (1.27, 1.93) | 1.47 (1.17, 1.85) |
| Norwegian sample | | | |
| *N* | 1,315 | – | 1,315 |
| Gender | | | |
| Male | 1 | – | 1 |
| Female | 0.68 (0.39, 1.19) | – | 0.62 (0.34, 1.15) |
| Learning difficulty or mental disorder | | | |
| No | 1*** | – | 1*** |
| Yes | 5.18 (2.85, 9.42) | – | 5.07 (2.64, 9.76) |
| Family type | | | |
| Traditional | 1 | – | 1 |
| Non-traditional | 1.34 (0.70, 2.58) | – | 1.34 (0.66, 2.72) |
| Household affluence | | | |
| Change per point | 0.58 (0.38, 0.89)*** | – | 0.48 (0.29, 0.80)*** |
| Parent from low and middle income country | | | |
| No | 1 | – | 1 |
| Yes | 1.29 (0.49, 3.39) | – | 2.10 (0.84, 5.22) |

*OR* odds ration, *CI* confidence interval

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

be informant specific. Our analyses also provide some indication of when computer-generated DAWBA diagnoses may be an adequate substitute for the clinician-rated diagnoses. In the case of prevalence estimates, the rough approximations provided by the DAWBA bands may be adequate for some purposes, such as service planning or tracking change over time. Our analyses indicate that they do not necessarily provide more than 'ballpark' estimates, however, and that in particular they may underestimate total rates of child mental disorder by 20–30%. Considerable caution would, therefore, be needed before comparing prevalences based on the computer-generated DAWBA diagnoses with prevalences in other populations based on different methods. On the other hand, when two or more samples are assessed using the computer-generated DAWBA diagnoses, this may increase the comparability of prevalence estimates by eliminating variability between clinical raters.

When estimating associations with risk factors, the estimated effect sizes, significance levels and substantive conclusions were very similar or identical for the clinician-rated and computer-generated DAWBA diagnoses. This mirrors our previous findings from the Norwegian sample regarding the effect of using non-representative samples based on low response rates; again, risk factor associations were more robust than prevalence estimates [14]. This suggests that studies whose primary research aim is to explore the correlates of mental disorder may be able to dispense with clinician rating of disorders. The resultant decrease in diagnostic accuracy may be more than counterbalanced using the cost savings to achieve considerably larger sample sizes or to improve other aspects of study design. Dispensing with the open-ended sections of the DAWBA (which are not necessary for the DAWBA bands) also reduces participant burden. This may increase compliance, particularly in web-based designs and for surveys with repeated assessments.

If seeking to use the DAWBA bands in this way, our results indicate that the multi-informant results produced prevalence estimates and associations with risk factors that most closely resembled the clinician-rated DAWBA diagnoses. If collecting information from parents, teachers and children is not possible, then the best single informant was the parent. The parent and multi-informant DAWBA bands also have the advantage of detailed coverage of emotional, behavioural and hyperactivity disorders. By contrast, coverage is less comprehensive using teacher or child DAWBAs.

The above discussion focuses on applications of the DAWBA bands in epidemiological studies. In clinical practice, we expect that most clinicians and families would be dissatisfied with positive predictive values in the range 40–80%. Most will feel that clinical judgement remains essential when assessing and treating individual children. In a clinic, therefore the computer-generated DAWBA bands and diagnoses should not be the end point, but rather one of the starting points for clinical assessment. The DAWBA bands might also be useful when carrying out screening for child mental health problems in schools or other community settings, with children whose DAWBA band identified them as being at high risk then receiving more detailed assessments by mental health specialists.

## Conclusion

The DAWBA bands represent a useful complement to the clinician-rated DAWBA diagnoses. Moreover, given limited resources, epidemiological studies often have to choose between expensive 'gold standard' measures and cheaper measures which can be administered to larger samples. In such circumstances, the computer-generated DAWBA diagnoses may sometimes be preferable to clinician-rated diagnoses. This may apply particularly to studies examining associations with risk factor, monitoring changes in child mental health over time or seeking to generate 'ballpark' prevalence figures to inform service use. Especially in combination with web-based methods of survey data collection, we believe these DAWBA bands offer novel possibilities for conducting larger surveys with more frequent follow-up and thereby, increasing the opportunity for testing causal hypotheses regarding the aetiology of child mental disorders.

**Conflict of interest statement** AG and RG are directors and RG is the owner of Youthinmind, which provides no-cost and low-cost software and web sites related to the DAWBA. SC consulted to Youthinmind in the development of the DAWBA bands. EH is the director and owner of Careahead, which provides teaching and supervision to clinics on the use of the DAWBA. EH and RG's opportunities to charge for clinical rating of the DAWBA may be reduced by the findings of this paper.

## References

1. Angold A (2002) Diagnostic interviews with parents and children. In: Rutter M, Taylor EA (eds) Child and adolescent psychiatry. Blackwell, Oxford
2. Breslau N (1987) Inquiring about the bizarre: false positives in Diagnostic Interview Schedule for Children (DISC) ascertainment of obsessions, compulsions, and psychotic symptoms. J Am Acad Child Adolesc Psychiatry 26(5):639–644
3. Goodman R, Scott S (1997) Child psychiatry. Blackwell Scientific Publications, Oxford
4. Goodman R et al (1996) Rating child psychiatric caseness from detailed case histories. J Child Psychol Psychiatry 37(4):369–379
5. Ford T, Goodman R, Meltzer H (2003) The British Child and Adolescent Mental Health Survey 1999: the prevalence of

DSM-IV disorders. J Am Acad Child Adolesc Psychiatry 42(10):1203–1211
6. Goodman R et al (2000) The development and well-being assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. J Child Psychol Psychiatry 41(5):645–655
7. Angold A (1989) Structured assessments of psychopathology in children and adolescents. In: Thompson C (ed) The instruments of psychiatric research. Wiley, New York
8. Loeber R et al (1991) Differences and similarities between children, mothers, and teachers as informants on disruptive child behavior. J Abnorm Child Psychol 19(1):75–95
9. American Psychiatric Association (1994) Diagnostic and statistical manual of mental disorders, 4th edition (DSM-IV). American Psychiatric Association, Washington, DC
10. WHO (1993) The ICD-10 classification of mental and behavioural disorders; diagnostic criteria for research. World Health Organisation, Geneva
11. Meltzer H et al (2003) Persistence, onset, risk factors and outcomes of childhood mental disorders. The Stationery Office, London
12. Heiervang E et al (2007) Psychiatric disorders in Norwegian 8- to 10-year-olds: an epidemiological survey of prevalence, risk factors, and service use. J Am Acad Child Adolesc Psychiatry 46(4):438–447
13. Meltzer H et al (2000) Mental health of children and adolescents in Great Britain. The Stationery Office, London
14. Heiervang E, Goodman R (2010) Advantages and limitations of web-based surveys: evidence from a child mental health survey. Soc Psychiatry Psychiatr Epidemiol
15. Green H et al (2005) Mental health of children and young people in Great Britain, 2004. MacMillan, Palgrave
16. Parry-Langdon N (2008) Three years on: survey of the development and emotional well-being of children and young people. Office for National Statistics, Cardiff
17. Heiervang E, Goodman A, Goodman R (2008) The Nordic advantage in child mental health: separating health differences from reporting style in a cross-cultural comparison of psychopathology. J Child Psychol Psychiatry 49(6):678–685
18. Ford T et al (2008) Predictors of service use for mental health problems among British schoolchildren. Child Adolesc Ment Health 13(1):32–40
19. Goldberg DP, Williams P (1998) A user's guide to the General Health Questionnaire. NFER-Nelson, Windsor