

Can standardized diagnostic assessment be a useful adjunct to clinical assessment in child mental health services?

A randomized controlled trial of disclosure of the Development and Well-Being Assessment to practitioners

Tamsin Ford · Anna Last · William Henley · Shelley Norman · Sacha Guglani · Katerina Kelesidi · Anne-Marie Martin · Pippa Moran · Harriett Latham-Cork · Robert Goodman

Received: 23 March 2012 / Accepted: 27 July 2012 / Published online: 12 August 2012
© Springer-Verlag 2012

Abstract

Purpose While research demands standardized diagnostic assessments as an indication of sufficient methodological rigour, there is debate about their application to clinical practice. The Development and Well-Being Assessment (DAWBA) provides a structured assessment of psychiatric disorder. Since it can be completed on-line, it could be used by Child and Adolescent Mental Health Services with few additional demands on staff. Access to the standardized diagnostic information as an adjunct to clinical assessment could reduce the number of appointments spent on assessment, free up practitioner time to work on engagement and improve clinical outcomes by increasing the accuracy of assessment and thus access to the appropriate evidence-based treatment.

Method Randomized controlled trial of the disclosure of the DAWBA to the assessing practitioner ($n = 117$) versus assessment at normal ($n = 118$) and analysed by “intention to disclose”.

Results Exposure to the DAWBA may increase agreement between the DAWBA and practitioners about some anxiety disorders, but detected no other statistically significant increased agreement for other disorders, nor a reduced need for further assessment, the number of difficulties recognised or influence on outcomes.

Conclusions The results may be explained by the inadequacy of the DAWBA, lack of statistical power to detect any effects that were present or a reluctance of some practitioners to use the DAWBA in their assessment. Future research might benefit from exploring the use of the DAWBA or similar assessments as a referral rather than an assessment tool, and exploring how practitioners and parents experience and use the DAWBA and what training might optimise the utility of the DAWBA to clinical practice.

Keywords DAWBA · Standardized Assessments · Diagnosis · RCT

Trial registration number: ICRCTN31394658.

T. Ford (✉) · A. Last · W. Henley · S. Norman · H. Latham-Cork
Peninsula Medical School, University of Exeter,
Veysey Building, Salmon Pool Lane, Exeter EX2 4SG, UK
e-mail: tamsin.ford@pms.ac.uk

S. Guglani
South London and Maudsley NHS Trust,
Croydon CAMHS, Lennard Lodge,
3 Lennard Road, Surrey, Croydon CR0 2UL, UK

K. Kelesidi · A.-M. Martin · P. Moran · R. Goodman
Department of Child and Adolescent Psychiatry,
Kings College London, Institute of Psychiatry,
De Crespigny Park, Box 085, London SE5 8AF, UK

Introduction

While standardized diagnostic assessments are ubiquitous in child mental health research, their absence from clinical practice is striking. Conventional secondary care assessment is costly and time-consuming, requires skilled professional input, but interventions are likely to fail without accurate and systematic assessment of the presenting problems [1]. In the UK, Child and Adolescent Mental Health Services (CAMHS) teams are multidisciplinary, and staff from different disciplines may have different conceptualizations of the same child's difficulties. It is clear from a CAMHS e-mail discussion group that there is tension between maintaining the richness of each discipline's

contribution while ensuring a minimum standard of assessment (<http://www.rcpsych.ac.uk/cru/focus>). The follow-up of a large British school-age population sample over 3 years revealed that few children attending CAMHS are seen by more than one member of the multidisciplinary team [2]. This raises the possibility that the background and orientation of the practitioner meeting the family at their initial appointment may have a major influence on the type of intervention that is subsequently offered. In addition, practitioners with generic training, such as occupational therapy, nursing and social work, may have little training in relation to working with children and are often line-managed by others in their own discipline [3–5]. A small qualitative study suggested that many CAMHS staff lack formal training in the assessment of childhood psychiatric disorder [3]. The risk of not carrying out systematic broad-based assessments is that particular types of difficulties may be missed, and may not respond to the non-specific psychological interventions that then tend to be offered [6]. Access to a standardized diagnostic assessment completed prior to the first appointment might facilitate the rational allocation of cases to practitioners with particular knowledge and skills suited to the constellation of difficulties identified, thereby increasing the chances of children receiving an appropriate evidence-based treatment. The use of a standardized assessment prior to first attendance could also free practitioners up to devote more attention to the family's concerns in the first session, leading to increased engagement, resulting in more rapid therapeutic responses to effective intervention directed towards the correct difficulties. Routine use of standardized diagnostic assessment could ultimately free up clinical capacity by reducing unnecessary assessment appointments and time spent on potentially ineffective interventions.

Weinstein et al. [7] compared diagnoses from the Diagnostic Interview Schedule for Children and Adolescents (DISC) with practitioners' admission diagnoses and reported low levels of agreement; kappas ranged from 0.03 to 0.17 ($M = 0.10$). Aronen et al. [8] later compared the same DISC diagnoses with the practitioner's discharge diagnoses and found similarly low levels of agreement; kappas ranged from -0.07 to 0.22 ($M = 0.11$). Their findings suggest that low levels of agreement were not merely due to incomplete clinical assessment at the beginning stages of treatment. This study was based on an inpatient samples, while the great majority of real-world child mental health assessment takes place in the community [2], which is a setting where diagnostic agreement may arguably be more difficult to obtain, given the briefer contact between staff and children.

Jensen and Weisz [9] compared practitioner's assessment to the diagnosis of the DISC in an outpatient setting. Agreement was poor for all individual disorders (such as

separation anxiety) and broader diagnostic clusters (such as anxiety disorders). In this American study, practitioners tended to record a diagnosis when the standardized assessment did not detect a disorder (which would have implications for funding for interventions in the USA), but also tended to record only one diagnosis and not the presence of comorbid disorders. They comment that the time pressure facing practitioners might limit clinical assessment, while the need to record a single diagnosis for reimbursement may discourage the recording of comorbidity. In an extension to this work, Hawley and Weisz [10] compared the level of agreement between parents, therapists and children about the nature of the difficulties that intervention was addressing. They reported that in more than three quarters of cases, treatment began without consensus among the triad, with the lowest level of agreement being between child and parent, and the highest between parent and therapist.

The lack of agreement in the studies above does not necessarily mean that the research standardized assessments are correct and the practitioners are incorrect. However, when compared to other indicators of validity, standardized assessments have greater validity and reliability than clinician-generated diagnoses. For example, Basco et al. [11] found that standardized assessments were more likely than clinician diagnoses to agree with "gold standard diagnoses" that were generated by experts who reviewed all available information including medical records, the clinician diagnoses and the results of the standardized assessment.

There are several mechanisms by which poor or incomplete assessment could have a negative impact on a child's course of treatment. Agreement on therapy goals is arguably an essential component of a strong working alliance [12]. Given the significant association between therapy alliance and treatment engagement [10], inaccurate assessment may be associated with decreased engagement in therapy. Inaccurate assessment could also have a negative impact on the outcomes of treatment. Many psychotherapies and medications are developed and tested in research studies in which structured diagnostic interviews are used to select children into the study. Indeed, Jensen-Doss and Weisz [13] report as above that diagnostic agreement between a practitioner-generated diagnosis and a structured diagnostic assessment predicted better engagement with therapy, with the "agree group" having fewer therapy "no-shows" and cancellations and a decreased likelihood of therapy dropout. The "agree group" also obtained larger reductions in parent-reported internalizing problems during treatment.

The current study aimed to explore the application of a standardized diagnostic assessment as an adjunct to clinical practice with children in the UK. It was nested within a

study in which the Development and Well-Being Assessment (DAWBA, see <http://www.dawba.info>) was used with all participants and involved a randomized controlled trial in which the results of the DAWBA were provided to the assessing practitioner in a random half of participants. Quantitative and qualitative information was collected from parents and practitioners about the process. We predicted that access to the DAWBA would increase the level of agreement between the clinical and the standardized diagnosis, reduce the need for further assessment and improve clinical outcomes by leading to an increase in the number of children receiving evidence-based treatments as a result of rational allocation of cases to practitioners with particular skills or background.

Method

The study received ethical approval from the joint South London and Maudsley NHS Trust and Institute of Psychiatry research ethics committee. It involved the Children's Team (Tier 3) and Early Intervention Team (Tier 2) from one CAMHS and the Tier 3 team from another CAMHS according to the Health Advisory Service's classification (1995). In this model, Tier 1 comprises all professionals working with children who are expected to be able to support minor, transient difficulties and detect more significant problems. Tier 2 comprises early intervention and occasionally uni-disciplinary mental health specialists whose work supports Tier 1. Tier 3 describes multidisciplinary mental health specialist teams of professionals. Both CAMHS were in greater London and serve catchment areas that are broadly representative of the British population [3].

Participants

The trial included children aged 5–10 years 9 months, who were consecutive referrals to the two participating CAMHS, once they had been accepted onto the waiting list. Children were excluded if they were looked after by the Local Authority, presented as an emergency; so that the DAWBA assessment could not take place prior to their clinical assessment, or if their parent did not speak sufficient English to complete the measures.

Measures

Parent completed measures

The DAWBA (<http://www.dawba.info>) combines highly structured questions closely based on the DSM-IV and ICD-10 diagnostic criteria with qualitative descriptions of

any areas of difficulty. A computer algorithm provides a probability of a child having any given disorder, based on answers to the structured questions and national survey data [14], but practitioners can review the symptoms, impairment and qualitative information in addition to these probabilities to make their own assessment. The validation study demonstrated clear differentiation between clinical and community samples, and the inter-rater reliability for two clinical raters who independently rated 500 cases was 0.86 for any psychiatric disorder, 0.57 for internalising disorders and 0.98 for externalising disorders [15]. All parents completed the DAWBA prior to the child's first appointment at the clinic, and the details of the DAWBA was shared with the assessing practitioners in a random half of cases. If the parents consented, the child's teacher was mailed a paper version of the DAWBA. The Strengths and Difficulties Questionnaire (SDQ) is a widely used, valid and reliable dimensional measure of childhood psychopathology [16]. Parents completed the SDQ as part of the DAWBA assessment prior to attending CAMHS and then 6 months later.

Practitioner completed measures

After the child's initial assessment, practitioners completed a pro forma (available on request) listing the disorders covered by the DAWBA plus some additional disorders suggested by participating practitioners. The options were "definite", "possible" and "no" for each disorder. Practitioners also reported whether the child required additional assessment or not. At 6 months follow-up, practitioners completed a second pro forma (available on request), which included the same list of disorders, plus the number of appointments attended and offered, as well as the interventions that had been offered to the child or his/her family since assessment. Possible interventions were listed and practitioners were asked to tick all that had been offered. The list was generated from the list of evidence-based interventions compiled from *Drawing on the Evidence* [17] and supplemented with common procedures such as liaison with other involved agencies, plus any other interventions requested by the participating practitioners at the outset of the study (available on request). *Drawing on the Evidence* is a brief pamphlet that summarised the evidence-based interventions used with childhood psychiatric disorders, which was distributed to all English CAMHS free of charge in 2006, so it should have been available to the participating practitioners during the current study.

The Children's Global Assessment Scale (CGAS) provides an estimate of the level of functioning for children aged 4–16, and has demonstrated high levels of inter-rater reliability (intraclass correlation coefficient 0.84), retest stability (intraclass correlation coefficients 0.69–0.95), and

discriminant validity ($p = 0.001$) for difference between mean scores for outpatients and inpatients; it can detect clinically significant change [18, 19]. Practitioners completed the CGAS after the child's first attended appointment and 6 months later.

Assessing whether interventions were evidence-based

Interventions reported by practitioners were scored as to the extent of the supporting evidence base in relation to the practitioner reported disorder(s) using the weighting given by *Drawing on the Evidence* [17], which ranks the strength of the evidence behind practice recommended [1] for each disorder. We added a "0" category for therapeutic interventions that were not in the booklet, and a "9" category for procedures that are accepted as good practice, such as liaison with the network of professionals around the child, but which are not supported by an evidence-base. Where doubt existed, we assumed that practitioners applied evidence-based practice unless we had concrete evidence to the contrary.

Procedure

Eligible families were approached after being accepted onto the clinic waiting list. Parents completed the DAWBA over the internet or as an interview with a research assistant. Randomisation was carried out by an independent clinical trials unit; it used permuted block randomization [20], stratifying by clinical team to ensure balance for the different length of waiting list among the three clinical teams. It was obviously impossible to blind practitioners and researchers to allocation, but there was little scope for researchers to influence the completion of postal questionnaires.

A printed report of the DAWBA was either emailed to the practitioner if they were identified before the first appointment with the child, and if they could not be identified, then a copy was printed and placed at the front of the child's medical notes. The report contained a summary of the symptoms and impact upon the child and their family as reported by parents and teacher for each disorder with the computer predictions of how likely a diagnosis was. It also contained a complete breakdown of the response of each information to each question (both structured and qualitative) about each area of difficulty plus background characteristics such as physical health, life events, parental psychopathology, family structure and function, and whether the child had additional support at school, contact with other services and was taking any psychotropic medication (examples are available on request). Practitioners were invited to a single training session about the use of the DAWBA, during which

example reports were shown to illustrate the types of information available and to generate discussion about what it might mean.

Referral source and presenting problem were extracted by the researchers from the referral letter to the clinic. KK generated the following categories from examination of the letters: emotional, behavioural, autism spectrum, attention deficit hyperactivity disorder (ADHD), eating and feeding disorders, tic disorders, and attachment issues. Multiple problems could be endorsed for a single child. Two experienced practitioners (SG and TF) independently classified the presenting problems according to these categories across the whole sample, with 88–99 % agreement and good or better than chance corrected agreement using the kappa statistic in all but one out of seven categories [21]. There was poor agreement about the presence of attachment issues ($\kappa = 0.37$, agreement 96 %); otherwise kappa's ranged from 0.68 to 0.93. Instances of moderate or poorer agreement might have been depressed by the extremely high rates of agreement, since the kappa statistic functions poorly at extremes of agreement or disagreement [22].

The Index of Multiple Deprivation (IMD) provides an estimate of the neighbourhood deprivation in which a family is living. It is made up of seven area level indices, which include income, employment, health/disability, education/skills/training, barriers to social housing/services, living environment and crime to provide a summary measure of neighbourhood deprivation [23]. The IMD for each child was extracted using his/her parent's postcode to link to data at lower super output area level; these contain a minimum of 1,000 people and 400 households.

The Paddington Complexity Scale [24] was designed to use objective information that is easily available to clinics about characteristics of a child or family that might influence the outcome of clinical intervention. It consists of 16 items to be rated by the practitioner that can be combined to give a total complexity score.

Analysis

A descriptive analysis assessed how representative the children who participated were compared to those who were eligible but did not participate and those who were referred during the same time period but were not eligible for the study. Similarly, we assessed the balance between the two arms of the trial.

The level of agreement between practitioner-reported diagnoses and standardized DAWBA computer diagnoses was assessed separately for the two arms of the trial using weighted kappa statistics with linear weights using three by three tables. We selected the computer diagnosis in preference to the expert clinical rater, as the computer

information was what was provided to the practitioner. Clinical options were “definite”, “possible” or “no” and were matched to probabilities of diagnoses against data from the British Child Mental Health Surveys [25, 26], which provide probabilities of a clinically rated diagnosis. Computer predictions that fewer than 0.5 % would have a disorder was coded as “no”, 50 % or more was coded as “definite” and those in between 0.5 and 50 % were coded as “possible” [14]. The 50 % cut point was chosen for “definite” disorder as it provided the best match between DAWBA computer algorithm-generated and clinician-generated estimates of prevalence in a large population survey [14]. These predictions were generated from data from all available informants. A sensitivity analysis collapsed the “no” and “possible” groups for both DAWBA and practitioner. Agreement statistics were calculated using the kap procedure in Stata for each diagnostic category where the sample prevalence was at least 10 %, for grouped and individual diagnoses such as emotional disorders (anxiety and depressive disorder were analysed separately where appropriate), disruptive behaviour, ADHD and autism spectrum disorder (ASD). The reliability of estimates based on individual disorder categories will be lower than that based on grouped disorders with higher numbers, but are provided for reference. Statistical significance of the intervention effect on level of agreement was assessed by considering whether or not there was overlap between the 95 % confidence intervals (CIs) for the relevant kappa statistics.

Given the indication from previous studies that standardized diagnoses detect unreported comorbidity, we compared the number of practitioner-reported diagnoses between the two trial arms in two ways: first, we compared reports of “definite” disorder, and second, “definite” and “possible” combined.

The effect of access to the DAWBA on receipt of evidence-based treatment was tested using separate logistic regression models for the four main disorder groups and an overall model for diagnosis with any disorder. The relatively small number of children allocated to the two trial arms for some of the disorder groups made it unlikely that randomization alone would control entirely for confounding. This was addressed by adjusting the models for the following potential confounders: age, gender and intellectual or learning disability. The models were extended to reflect the effect of clinician-level clustering by adding a random effect for clinician. The resulting hierarchical models assume that observations are independent across clusters (children seen by different clinicians), but not necessarily within clusters (children seen by the same clinician). The effect of access to the DAWBA on the need for further assessment was explored using the same approach.

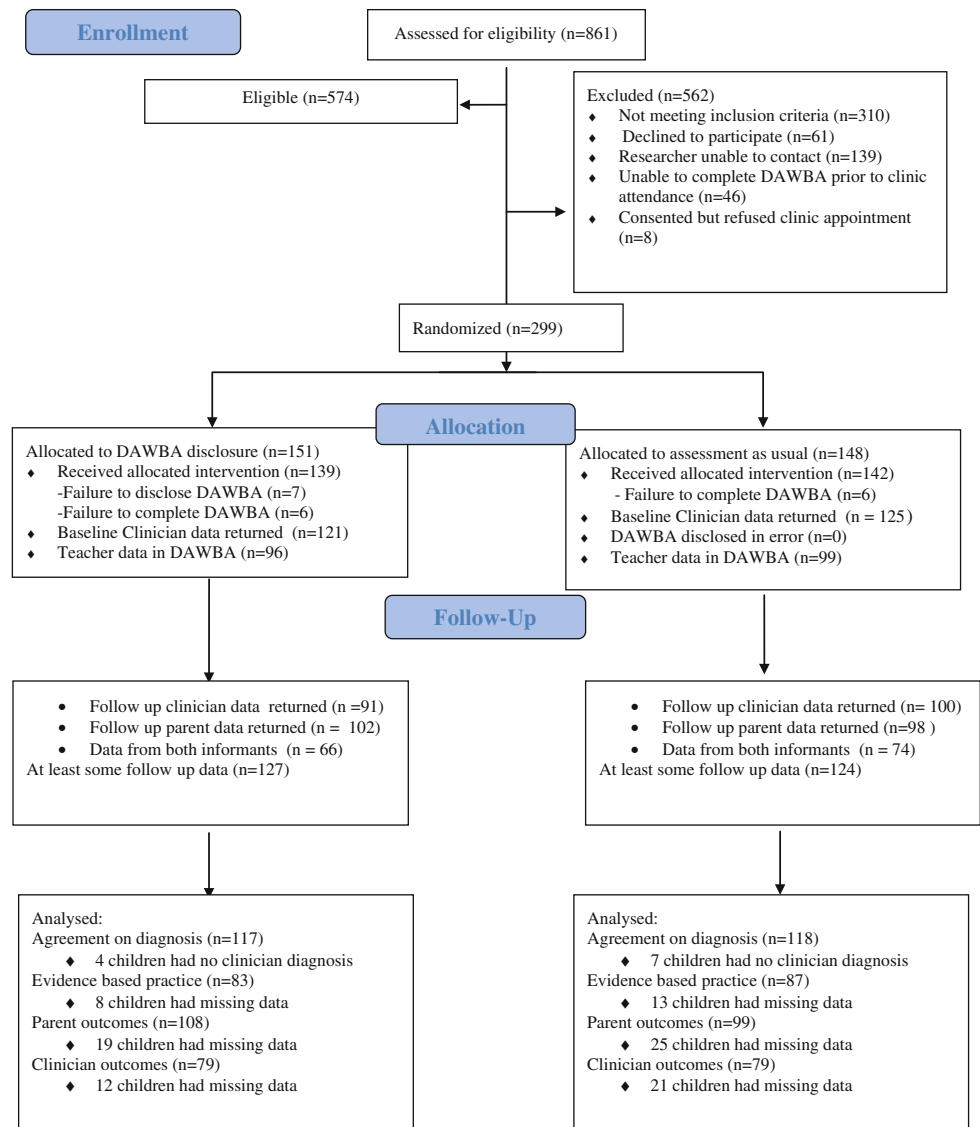
Differences in parent-reported and clinical outcomes between trial arms were analysed on an intention-to-treat basis using analysis of covariance (ANCOVA). The ANCOVA model was specified with change in score from baseline to 6 months post-intervention as the dependent variable, and baseline outcome level included as a covariate. Separate models were fitted for CGAS score, SDQ total difficulties score and SDQ impact score. The models were extended to include the following potential confounders, which were identified a priori as likely to influence clinical outcomes: age, gender, global or specific learning disability, Paddington Complexity Scale and IMD. The effect of clustering due to clinician on trial outcomes was explored by fitting hierarchical linear models with clinician as a random effect [27].

All analyses were based on intention to treat, where “treat” means “disclose DAWBA diagnosis”. However, a full application of the intention to treat approach is possible only when complete outcome data are available for all randomized subjects. We adopted a complete-case approach, in which children with missing responses were excluded from the analysis, under the assumption that the mechanism for generating missing data was “missing at random” [28]. Under this assumption, dropping cases with missing data does not introduce bias into model inferences, provided the regression models control for all variables that influence the probability of missingness. All analyses were carried out using Stata 11.1

Results

The Consort diagram shows the response rate and follow-up of the parents and practitioners who were approached to participate in the trial. There were 861 children referred to the clinics during the recruitment period; of these, 562 did not participate in the trial, mostly due to ineligibility ($n = 310$) (see Fig. 1). Only 61 families declined to participate, research staff were unable to contact 139, a further 46 did not complete the DAWBA prior to attending the clinic and eight families consented to participate but did not attend the clinic. Most families consented to a teacher being contacted, and teacher data were received for 195 children (65 %) prior to the first assessment, although additional teacher data were received after the assessment in some cases [29]. The commonest reason for the failure of allocation was that the DAWBA was not completed in time, usually because of a sudden crisis after the initial referral or new practitioners taking multiple cases off the waiting list. A high proportion (94 % of cases) received the allocation as randomized. Where no DAWBA was completed at all, we were unable to analyse as “intention to treat”; however, where a DAWBA was completed, we

Fig. 1 Consort diagram illustrating the flow of participants



have analysed according to randomized allocation regardless of whether disclosure happened as planned. Thus, we analysed 118 controls (77 %) and 117 disclosures (79 %). Clinician data were present for 88 % of children in both arms at baseline and at least some follow-up data were available for 89 % of those in the intervention arm and 87 % of controls. There were data from both informants for 66 children in the intervention arm and 74 in the control arm.

There were no differences between participants and non-participants that could not be explained by the inclusion/exclusion criteria. Children who participated ($M = 7.91$, $SD = 1.64$) and those who were eligible but did not participate ($M = 7.81$, $SD = 1.87$) were significantly younger than the children who were ineligible ($M = 8.69$, $SD = 2.66$, $F = 14.2$, $p < 0.001$), because of the inclusion criteria that excluded children aged 11–16 who were still

referred to both services. Similarly, there were more children referred from Social Services in the ineligible group than those children who were eligible (14 vs. 6 %, $\chi^2 = 42.9$, $p < 0.001$), which relates to the exclusion of children who were looked after from the trial. The arms of the trial were balanced with regard to most socio-demographic characteristics, but a higher proportion of the children in the intervention arm had intellectual or specific learning disabilities (27 vs. 18 %, $\chi^2 = 52.0$, $p < 0.001$), which might be expected to complicate their response to treatment.

Table 1 shows weighted kappa coefficients quantifying the level of agreement between the clinical and standardized diagnoses for disorder groups, separately for the intervention and control arms of the trial. After excluding children for whom the practitioner had either not returned the form ($n = 53$) or returned the form without a diagnosis

Table 1 Agreement between computer diagnosis and practitioner diagnosis separately for each trial arm (individual diagnoses and disorder groups, with computer and practitioner diagnoses classified as ‘no diagnosis’, ‘possible diagnosis’ or ‘probable diagnosis’)

Diagnosis	Intervention (<i>n</i> = 117)				Control (<i>n</i> = 118)			
	DAWBA + clinician ^a	Total agreement (%)	Kappa	95 % CI	DAWBA + clinician ^a	Total agreement	Kappa	95 % CI
Any emotional disorder	34	72.2	0.35	0.22–0.48	37	70.8	0.19	0.06–0.32
Anxiety disorders								
Separation anxiety	11	79.9	0.22	0.08–0.36	7	83.1	0.18	0.04–0.31
Specific phobia	6	89.8	0.40	0.25–0.55	0	90.7	–0.03	–0.13 to 0.07
Social phobia	1	86.2	–0.04	–0.19 to 0.12	2	85.6	–0.04	–0.20 to 0.13
Generalized anxiety	12 ^b	77.6	0.12	–0.02 to 0.26	12	80.5	0.21	0.09–0.33
Any anxiety disorder	30	74.8	0.37	0.23–0.50	31	71.6	0.20	0.08–0.33
Depressive disorders	5	88.9	0.21	0.06–0.35	3	87.7	0.07	–0.08 to 0.23
Any conduct disorders	36	50.9	0.12	0.05–0.19	37	60.0	0.14	0.07–0.22
Conduct disorder	17	76.5	0.25	0.14–0.36	15	71.6	0.13	0.02–0.24
Oppositional defiant disorder	26	46.6	0.09	0.02–0.15	31	51.3	0.13	0.06–0.21
Attention disorders (ADHD)	36	49.6	0.10	0.04–0.16	40	55.5	0.16	0.08–0.25
Autistic spectrum disorders	26	85.0	0.54	0.39–0.68	24	83.9	0.43	0.28–0.59

Weighted kappa statistics (with 95 % confidence intervals) are included to quantify the level of agreement beyond that expected due to chance for all diagnoses with a prevalence of at least 10 % (average of DAWBA and clinician prevalence)

^a Number classified as possible or probable diagnosis by both practitioner and DAWBA

^b One additional missing clinician assessment (*n* = 11)

(*n* = 11), the sample sizes available for assessing levels of agreement were 117 and 118 in the intervention and control arms, respectively. We provide data on individual disorders of interest but consider that the very small number of children contributing to each individual diagnosis means that the analysis lacks power; therefore, greater attention should be paid to the analysis of broader diagnostic groups since these will have greater precision. Absolute levels of agreement in the intervention arm ranged from negative kappa statistics for social phobia to a kappa statistic of 0.54 for autistic spectrum disorders. Higher levels of agreement were observed in the intervention arm compared to the control arm for several disorders including autistic spectrum disorders, depressive disorders and separation anxiety, but the only statistically significant difference was for specific phobia, as the 95 % confidence intervals for the kappa coefficients in the control and intervention arms overlapped for all other individual and grouped disorders. Table 2 shows the same analysis at group level, with the “possible” or “no” groups of diagnosis combined for both practitioners and the computer DAWBA diagnosis. The results are very similar, except that there is statistically significant improved agreement about anxiety disorders in the intervention arm using this method.

The number of “definite” diagnoses reported by practitioners ranged from 0 to 12; the mean in the intervention group was 0.38 and in the control arm was 0.34. After adjusting for age, gender, learning disability and clustering

by practitioner, the mean increase in the number of practitioner reported diagnoses per child compared to the control group was 0.07 (95 % CI –0.10 to 0.25; *p* = 0.41). The slight increase in the number of disorders reported by practitioners was diminished if both “possible” and “definite” diagnoses were examined. In this sensitivity analysis, the mean increase in the number of practitioner diagnoses per child in the intervention arm was 0.03 (95 % CI –0.35 to 0.41; *p* = 0.89).

There was no evidence that access to evidence-based practice was increased in the intervention arm compared to the control arm for any of the disorder groups, before or after adjusting for the clustering of cases within clinicians (Table 3). Similarly, there was no evidence that access to the DAWBA influenced the reported need for additional assessment (further information available on request) before or after adjusting for clustering by clinician.

Practitioner access to the DAWBA was not associated with improvements in outcomes, before or after accounting for clustering (Table 4). The extent of clustering by clinician varied considerably by outcome measure: intra-class correlations ranged from 87 % for the CGAS score (95 % CI 79–92) to 17 % and 9 % for the SDQ impact and total difficulties scores, respectively, and probably relates to individual practitioners scoring the CGAS within idiosyncratic ranges, despite training for reliability against the freely available CGAS vignettes (Table 4).

Table 2 Agreement between computer diagnosis and practitioner diagnosis separately for each trial arm (disorder groups only with computer and practitioner diagnoses classified as ‘no/possible diagnosis’ or ‘probable diagnosis’)

Diagnosis	Intervention (<i>n</i> = 117)				Control (<i>n</i> = 118)			
	DAWBA + clinician ^a	Total agreement (%)	Kappa	95 % CI	DAWBA + clinician ^a	Total agreement (%)	Kappa	95 % CI
Anxiety disorders	10	83.8	0.43	0.26 to 0.59	1	83.1	0.01	−0.15 to 0.18
Depressive disorders	0	93.2	−0.03	−1.81 to 0.13	0	94.9	−0.02	−0.19 to 0.15
Conduct disorders	4	52.1	0.05	−0.02 to 0.12	5	53.4	0.08	0.01 to 0.15
Attention disorders (ADHD)	2	50.9	0.03	−0.01 to 0.08	6	55.9	0.08	−0.02 to 0.17
Autistic spectrum disorders	8	90.4	0.54	0.36 to 0.72	4	92.7	0.46	0.28 to 0.65

Weighted kappa statistics (with 95 % confidence intervals) are included to quantify the level of agreement beyond that expected due to chance

^a Number classified as probable diagnosis by both practitioner and DAWBA

Table 3 Logistic regression analysis of access to evidence-based practice

Diagnosis	Base model					Adjustment for clustering by clinician		
	<i>N</i>	% EBP	OR	95 % CI	<i>p</i> value	OR	95 % CI	<i>p</i> value
Any disorder								
Intervention	83	42	1.17	0.61–2.25	0.63	1.32	0.64–2.72	0.45
Control	87	34	1.00			1.00		
Emotional disorders								
Intervention	36	28	0.73	0.24–2.16	0.57	0.79	0.26–2.39	0.68
Control	40	30	1.00			1.00		
Conduct disorders								
Intervention	28	61	1.77	0.50–6.22	0.38	1.94	0.54–6.88	0.31
Control	26	38	1.00			1.00		
Attention disorders (ADHD)								
Intervention	23	39	0.66	0.18–2.42	0.53	0.68	0.16–2.84	0.60
Control	26	38	1.00			1.00		
Autistic spectrum disorders								
Intervention	28	7	0.28	0.05–1.72	0.17	0.27	0.04–1.68	0.16
Control	33	18	1.00			1.00		

Discussion

We predicted that access to the DAWBA would increase the level of agreement between the clinical and the standardized diagnosis, and improve clinical outcomes by leading to an increase in the number of children receiving evidence-based treatments as a result of rational allocation of cases to practitioners with particular skills or background. Higher levels of agreement between the clinical and the standardized diagnosis were observed when using the DAWBA for several disorders, but differences did not reach significance, except for specific phobias in the three-category analysis and anxiety disorders in general in the two-category analysis. There was no evidence that access to evidence-based treatments across all disorders was

higher when using the DAWBA, or that access to the DAWBA improved outcomes.

There are a few potential explanations for the mostly negative findings. One possibility is that standardized assessments in general or DAWBAs in particular are not helpful even in the best of circumstances, although previous reliability and validity studies suggest otherwise [15]. Before accepting this, it is worth noting circumstances that may have undermined the difference between the two arms of the trial, thereby reducing the likelihood of significant differences emerging. Disclosure involved providing a copy of the DAWBA assessment to the clinician, but anecdotal evidence suggests that some clinicians never looked at it; Last et al. [29] found qualitative evidence that certain clinicians were reluctant to use a standardized

Table 4 Intention-to-treat analysis of trial outcomes as reported by practitioners and parents

Diagnosis	Base model					Adjustment for clustering by clinician			ICC (%)
	<i>N</i>	Mean score at time 2	Coefficient	95 % CI	<i>p</i> value	Coefficient	95 % CI	<i>p</i> value	
SDQ total difficulties score									
Intervention	108	19.92	−0.07	−8.71 to 8.57	0.99	−0.09	−8.01 to 7.82	0.98	12.7
Control	99	19.55	0.00			0.00			
SDQ impact score									
Intervention	108	3.81	0.50	−3.98 to 4.99	0.82	0.56	−3.62 to 4.73	0.79	0.0
Control	99	3.72	0.00			0.00			
CGAS score									
Intervention	79	44.7	7.82	−44.64 to 60.28	0.77	−4.40	−25.06 to 16.26	0.68	90.9
Control	81	48.9	0.00			0.00			

SDQ Strengths and Difficulties Questionnaire, CGAS Children's Global Assessment Scale

diagnostic measures. The method of disclosing the DAWBA to practitioners may also have impacted upon the findings as it meant that there was no way to ensure that practitioners actually read the DAWBA. In England, gaps have been identified in the training of practitioners with generic training, such as primary mental health workers, nurses, family therapists and social workers, in relation work with children, with suggestions for curricula and of joint training between health, education and social work to ensure adequate pre-registration training and continuing professional development [3–5]. Accurate assessment should form a key part of such training. Practitioners who are not trained to undertake systematic diagnostic assessments, and/or who are actively hostile to the use of diagnoses, may not make appropriate use of the findings of structured diagnostic assessments and are plausibly less likely to perceive them as useful. Sadly, we lacked statistical power to empirically test whether the practitioners orientation influenced the impact of the DAWBA, but related work in the same study and with many of the same practitioners in relation to routine outcome monitoring (ROM) suggests that there might be. A mixed methods study explored the attitudes of the same group of practitioners that participated in the current study towards ROM [30] and standardized assessments [3]; both suggest that the practitioners were neither overwhelmingly positive nor negative towards standardized diagnostic assessment or ROM, but many expressed concerns about the restraints of standardized diagnoses [3]. There was a marked difference in opinion dependent upon the background of the practitioner, with practitioners with a psychiatric/psychological background identifying more positive aspects of ROM than those from a nursing/counselling/social work background [30]. Further research could explore these issues in more detail.

It might, therefore, have been more useful to test the value of DAWBA during the referral process, for example

carrying out assessments in primary care and schools, with the findings being screened by senior CAMHS practitioners. This could inform the basis for allocation to the practitioner with the appropriate skills and training to complete a clinical assessment by the pattern of difficulties suggested by the DAWBA. The clinician who then assessed the child might have been more likely to use the information provided by the DAWBA if it matched their expertise and interest. Such a referral process could then free up practitioners to devote more attention to the family's concerns in the first session, perhaps leading to increased engagement and more rapid therapeutic responses to intervention. In the current study, and anecdotally in CAMHS practice, referral letters and referral forms contained very limited information, and often seemed skewed by “key terms”, such as “possible ADHD” that could be employed by referrers in order maximize access to CAMHS, regardless of the level of need. Systematic diagnostic assessment prior to entry to CAMHS might allow CAMHS practitioners to more accurately triage referrals, and therefore to better target the limited CAMHS resource to the children who are most likely to benefit from the interventions offered. Alternatively, DAWBA ratings by senior clinicians, rather than the computer predictions of the likelihood of diagnoses, might be expected to agree more closely with the practitioners' reports. Our study is one of only two trials of this nature that we are aware of. Aebi and colleagues [31] used the same measure, but with a wider age group, in a different service context and feed back the expert clinician rated DAWBA assessments rather than the computer DAWBAs as we chose to. It is interesting, therefore, that their results were similar to ours, in that they suggest a statistically significant increase in the recognition of emotional disorders, which adds strength to our tentative finding. In addition, their results echo the DAWBA validation study by suggesting that the DAWBA prompted practitioners to report, and perhaps therefore

consider comorbidity [17]. That the current study failed to replicate a significant difference in the reports of comorbidity could indicate that we lacked statistical power and/or that practitioners with access to the DAWBA were not applying the results provided to them to their assessment. Clinical rating of the DAWBA allows the assignment of “not otherwise” specified diagnoses to children with significant symptoms and impairment who do not quite fit diagnostic criteria, can detect where informants have misunderstood the questions and allows clinical judgment to be applied to inconsistent reporting, or inter-informant disagreement about the child’s difficulties. However, practitioners in the current study had access to the details of individual difficulties and their impacts as well as background risk factors, from each informant as well as the overall diagnostic predictions, and we hoped that they would use all this information to support their assessment. The decision about whether to clinical rate or use the computer prediction in the DAWBA provided to practitioners, therefore can only be guided by the diagnostic time available for a senior clinician to review DAWBAs balanced against the potential to save time if the initial accuracy of assessments were increased, and if this translated into better or most quickly attained therapeutic change.

Like the Swiss trial of the DAWBA, our study has considerable strength. First, the consideration of the impact of a diagnostic assessment and how decisions about the results are used in clinical practice is an important but neglected area of research [32]. Other strengths include independent randomization and the test of a measure that would not add greatly to the burden of assessment in CAMHS as it can be completed independently via the internet. However, the randomized sample size was smaller than intended and inevitably missing data further diminished the statistical power available to our analysis: another alternative explanation for our findings. The judgment about what was and was not an evidence-based intervention is inevitably somewhat subjective and prone to socially desirability responses from practitioners. However, a more rigid application of the coding requiring a higher standard of evidence did not reveal improved outcomes [33], and there may have been an insufficient range of interventions and outcomes to detect any impact of the level of evidence base applied.

At a time when health services in many countries are expected to reduce the costs of services while still maintaining a high quality of care [34], there is a pressing need for studies into ways of increasing the appropriateness of referrals for specialist services, reducing the number of missed or incorrect diagnose, reducing missed appointments, and increasing the speed at which children access effective interventions. Since the DAWBA was very well

accepted by parents in the current study [29], it may also be a useful bridge between family and clinic since computer use is an everyday experience for many people, while visiting a psychiatric clinic and being interviewed is not.

Certain considerations would need to be addressed if the DAWBA were to be used in such a way. For example, what if the child does not meet the level of need for specialist involvement or is on the borderline? How would the family and referrer be best supported at this stage? It would be necessary for triaging CAMHS practitioner to have a wide knowledge of voluntary services that match the description of the child’s difficulties, for example, bereavement charities, counseling services or behavior supports service in schools. Alternatively, the referrers’ would need to be prepared to undertake this role. There would need to be a protocol to manage the situation where the DAWBA did not meet the threshold for referral, but where the family or referrers levels of anxiety was not contained by reassurance or direction to self-help or the voluntary sector, or the child subsequently deteriorated. Another important consideration is the potential damage to engagement that could be done, had a parent spent an hour completing the DAWBA, or any other measure, and for it to be ignored by the practitioner. If we place demands on parents and practitioners to complete measures, it is essential that we explain how the data provided will be used and demonstrate that these have been used in this way.

Future research should be conducted in two areas: first, the views and experiences of parents, young people and both educational and health professionals should be elicited to ascertain how the DAWBA or similar diagnostic assessments could best fit into clinical practice and what support could or should be provided to facilitate its completion and interpretation. Second, the possibility of a cluster randomized trial of systematic assessment during the referral process should be explored, and in particular, whether the orientation of the practitioner influences how it is used in order to develop tailored training packages to optimize any benefits of its use. While we could not demonstrate a convincing benefit for the use of the DAWBA, we believe that it is important to further explore the use of standardized assessment as an adjunct to clinical practice further before designating it as a practice without evidence.

Acknowledgments We are extremely grateful to the two child and adolescent mental health services that hosted the study, and to the parents and practitioners who gave up their time to participate, and to the Medical Research Council in England who funded the original project as part of a Clinician Scientist Fellowship for Tamsin Ford.

Conflicts of interest RG is the owner of Youthinmind, which provides no-cost and low-cost software and web sites related to the SDQ and DAWBA. TF was supported by a Clinician Scientist award from the Medical Research Council in England to complete this work.

Anna Last, Shelley Norman and William Henley completed this work while supported by the Collaboration for Applied Health Services Research and Care (CLAHRC) for the South West of England. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. None of these funders had any influence over the design, analysis or writing up of this work.

References

- Foreman D, Foreman D, Prendergast M, Minty B (2001) Is clinic prevalence of ICD-10 hyperkinesia underestimated? Impact of increasing awareness by a questionnaire. *Eur J Child Adolesc Psychiatry* 10:130–134
- Ford T, Hamilton H, Goodman R, Meltzer H (2005) Service contacts among the children participating in the British child and adolescent mental health. *Child Adolesc Mental Health* 10:2–9
- Martin A, Fishman R, Baxter L, Ford T (2011) Practitioners' attitudes to standardized diagnostic assessment in routine practice: a qualitative study in two Child and Adolescent Mental Health Services. *Clin Child Psychol Psychiatry* 16(3):407–420
- Jones J (2004) The post-registration education and training for nurses working with children and young people with mental health problem. RCN Institute, London
- Lindsay C, Griffiths P (2004) Developing a comprehensive CAMHS. *YoungMinds Magazine* 73:31–33
- Shah P, Mountain D (2007) The medical model is dead: long live the medical model. *Br J Psychiatry* 19:375–377
- Weinstein SR, Stone K, Noam GG, Grimes K, Schwab-Stone M (1989) Comparison of DISC with practitioner's DSM-III diagnoses in psychiatric inpatients. *J Am Acad Child Adolesc Psychiatry* 28:53–60
- Aronen ET, Noam GG, Weinstein SR (1993) Structured diagnostic interviews and practitioner's discharge diagnoses in hospitalized adolescents. *J Am Acad Child Adolesc Psychiatry* 32: 674–681
- Jensen AL, Weisz JR (2002) Assessing match and mismatch between practitioner-generated and standardized interview-generated diagnoses for clinic-referred children and adolescents. *J Consult Clin Psychol* 70:158–168
- Hawley KM, Weisz JR (2005) Youth versus parent working alliance in usual clinical care: distinctive associations with retention, satisfaction and treatment outcome. *J Clin Child Adolesc Psychol* 34:117–128
- Basco MR, Bostic JQ, Davies D, Rush AJ, Witte B, Hendrickse W et al (2000) Methods to improve diagnostic accuracy in a community mental health setting. *Am J Psychiatry* 157(10):1599–1605
- Horvath AO, Luborsky L (1993) The role of the therapeutic alliance in psychotherapy. *J Consult Clin Psychol* 61:561–573
- Jensen Doss A, Weisz JR (2008) Diagnostic agreement predicts treatment process and outcomes in youth mental health clinics. *J Consult Clin Psychol* 76:711–722
- Goodman A, Heiervang E, Collishaw S, Goodman R (2011) The 'DAWBA bands' as an ordered categorical measure of child mental health: description and validation in British and Norwegian samples. *Soc Psychiatry Psychiatr Epidemiol* 46:521–532
- Goodman R, Ford T, Simmons H, Gatward R, Meltzer H (2000) Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *Br J Psychiatry* 177:534–539
- Goodman R (2001) Psychometric properties of the Strengths and Difficulties Questionnaire (SDQ). *J Am Acad Child Adolesc Psychiatry* 40:1337–1345
- Wolpert M, Fuggle P, Cottrell D, Fonagy P, Phillips J, Pilling S, Stein S, Target M (2006) Drawing on the evidence, vol 2. CAMHS Publications, London, p 17
- Shaffer D, Gould MS, Brasic J, Ambrosini P, Fisher P, Bird H, Aluwahlia S (1983) A children's global assessment scale (CGAS). *Arc Gen Psychiatry* 40(11):1228–1231
- Sourander A, Helenius H, Piha J (1996) Outcome of short-term child psychiatric hospitalization (teacher evaluation at 5-month and 12-month follow up). *Eur J Child Adolesc Psychiatry* 5: 204–211
- Matts JP, Lachin JM (1998) Properties of permuted-block randomization in clinical trials. *Control Clin Trials* 9:327–344
- Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics* 33:159–174
- Dewey ME (1983) Coefficients of agreement. *Br J Psychiatry* 143:487–489
- Department of Communities and Local Government (2007) The English indices of multiple deprivation. HMSO, London
- Hunter J, Higginson I, Garralda E (1996) Systematic literature review: outcome measures for child and adolescent mental health services. *J Public Health Med* 18(2):197–206
- Meltzer M, Gatward R, Goodman R, Ford TJ (2000) The Mental Health of children and adolescents in Great Britain. The Stationery Office, London
- Meltzer H, Lader D, Corbin T, Goodman R, Ford T (2004) The mental health of young people looked after by local authorities in Scotland. The Stationery Office, Edinburgh
- Gelman A, Hill J (2006) Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, Cambridge
- Schenker N, Gentleman JF (2001) On judging the significance of differences by examining the overlap between confidence intervals. *Am Statist* 55(3):182–186
- Last A, Ford T, Henley W, Norman S, Goodman R. Feasibility of the Development and Well Being Assessment as an adjunct to clinical assessment in child and adolescent mental health services. *Child Adolesc Mental Health* (submitted 16 July 2012)
- Norman SM, Dean S, Hansford L, Ford T. Clinical practitioner's attitudes towards the use of routine outcome monitoring within Child and Adolescent Mental Health Services: a qualitative study of two Child and Adolescent Mental Health Services (in submission)
- Aebi M, Kuhn C, Winkler Metke C, Stringaris A, Goodman R, Steinhausen HC. The use of the Development and Well-Being Assessment in clinical practice: a clinical trial. *Eur Child Adolesc Psychiatry* (submitted 20 February 2012)
- Ferrante di Ruffano L, Hyde CJ, McCaffery KM, Bossuyt PMM, Deeks JJ (2012) Assessing the value of diagnostic tests: a framework for designing and evaluating trials. *BMJ* 344:18–22
- Norman SM (2011) Factors influencing the outcomes of children utilizing CAMHS. PhD Thesis, Peninsula College of Medicine and Dentistry, University of Exeter, Exeter
- Martin AM, Fishman R, Baxter L, Ford TF (2011) Practitioner Attitudes to Standardised Diagnostic Assessment: a qualitative study in two CAMHS. *Clin Child Psychol Psychiatry* 16:407–420