ORIGINAL CONTRIBUTION

The use of the development and well-being assessment (DAWBA) in clinical practice: a randomized trial

Marcel Aebi · Christine Kuhn ·
Christa Winkler Metzke · Argyris Stringaris ·
Robert Goodman · Hans-Christoph Steinhausen

Abstract The development and well-being assessment (DAWBA) has been used in various epidemiological studies, whereas the clinical value of the instrument needs support from further studies. In particular, it is important to document how the use of the DAWBA influences clinical decision-making. The present study employed the DAWBA in a consecutive series of 270 new referrals to a large public child and adolescent psychiatric service in Zurich, Switzerland. ICD-10 based diagnoses were obtained from clinicians for all patients and reliability of DAWBA expert raters was calculated. The DAWBA diagnoses were randomly disclosed ($n = 144$) or not disclosed ($n = 126$) before clinical decision-making. The reliability of DAWBA expert diagnoses was very satisfactory and the agreement under the disclosed versus the non-disclosed condition amounted to 77 versus 68 % for internalizing disorders and to 63 versus 71 % for externalizing disorders. The increment in agreement due to disclosure of the DAWBA diagnosis was significant for internalizing disorders. Access to DAWBA information was more likely to prompt clinicians to add an extra diagnosis. Professional background and degree of clinical experience did not affect diagnostic agreement. Overall, diagnostic agreements between DAWBA expert diagnoses and clinical diagnoses were in the fair to moderate range and comparable to previous studies with other structured diagnostic interviews. The inclusion of the DAWBA into the clinical assessment process had an impact on diagnostic decision-making regarding internalizing disorders but not regarding externalizing disorders.

Keywords Diagnosis · Standardized diagnostic interview · Clinical judgment · Child and adolescent mental health

M. Aebi and C. Kuhn contributed equally to this manuscript.

M. Aebi (✉) · C. Kuhn · C. W. Metzke · H.-C. Steinhausen
Department of Child and Adolescent Psychiatry,
University of Zurich, Zurich, Switzerland
e-mail: maebi@ppkj.uzh.ch

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

H.-C. Steinhausen
Aalborg Psychiatric Hospital, Aarhus University Hospital,
Aarhus, Denmark

H.-C. Steinhausen
Clinical Psychology and Epidemiology, Institute of Psychology,
University of Basel, Basel, Switzerland

Introduction

Research has shown several information-gathering biases of clinicians when using unstructured interviews, such as deciding on the diagnosis before collecting all relevant data, seeking information to confirm a previous diagnosis, ignoring conflicting information, combining information in ways that do not match diagnostic criteria, and various assumptions based on gender, ethnicity, and psychosocial backgrounds [8, 9]. Accordingly, the agreement between standardized diagnostic interviews (SDI) and clinical

diagnoses was rather low in a recent meta-analysis by Rettew et al. [26] with Cohen's kappa for internalizing disorders amounting to 0.28 and for externalizing disorders to 0.29. There is some evidence that treatment outcomes regarding therapy engagement and internalizing disorders were significantly worse when clinical diagnosis did not agree with diagnosis based on SDI [17]. However, others studies found diagnoses from SDI to be of limited use. For example, Duffy [5] found that SDI based on DSM-IV or ICD-10 may lead to an increase of false positive cases of bipolar disorders as a result of not considering psychiatric history. Although the reliability and validity of diagnoses based on unstructured interviews were poorer compared to SDI-based diagnoses [3, 18, 19], the clinical utility of SDI still remains a matter of concern. It is worth noting that SDI-based diagnoses underlie most clinical trials, so the evidence base for what does and does not work in child and adolescent psychiatry is particularly relevant to individuals with SDI-based diagnoses.

Many clinicians do not recognize the benefit of SDI over clinical judgments and are concerned about the practicality of SDI [16] or that SDI may damage the therapeutic alliance [21]. Furthermore, professions other than psychologists (i.e., psychiatrists, social workers) have more negative attitudes regarding the use of SDI in clinical practice [16]. To date, the effects of SDI on clinicians' diagnostic decision-making have not been systematically studied. Thus, it is unclear as to how far SDI can actually improve assessment and subsequent treatment in everyday practice.

The development and well-being assessment (DAWBA) [13] is a potentially promising SDI for use in routine clinical work for several reasons. Firstly, its mixture of structured and open-ended questions means that clinicians can review descriptions of problems in the respondent's own words. Secondly, the DAWBA can be administered via a secure internet connection, with advantages for respondents, clinicians and service managers. Online completion is often convenient for respondents. In addition, the DAWBA covers all major diagnoses, including co-morbidities which might be missed in clinical interviewing due to focusing on the referral complaints only. Finally, DAWBA items refer to ICD-10 and DSM-IV diagnostic criteria and inbuilt diagnostic algorithms provide useful guides to the likelihood of different diagnoses. Despite these possible advantages, the effects of the DAWBA and other SDI on clinicians' diagnostic decision-making have not been systematically addressed.

The initial validation of the DAWBA involved a study of both a community and a clinic sample [13] and the instrument has been used ever since in various epidemiological studies in Britain [7], Brazil [6], Bangladesh [24], Norway [15] and Russia [14]. In addition, two clinical studies based on the DAWBA [1, 24] found higher agreements between DAWBA expert-rated diagnoses and clinicians' diagnoses (Cohen's kappa of 0.63–0.94) than expected from previous research based on other SDI [26]. However, both of these studies were performed in low-income countries and were based only on small teams of clinicians. To date, there are no published studies of the agreement between DAWBA and clinical diagnoses in large samples drawn from sizeable psychiatric clinic in high-income countries.

The present study addressed the issue of how DAWBA information influences routine clinical practice. The DAWBA diagnoses generated by expert raters were compared with independent diagnoses generated by clinicians in a consecutive series of new referrals of children and adolescents to a large public service. To test whether or not the disclosure of DAWBA diagnoses has an effect on clinical decision-making, the DAWBA diagnoses were randomly either disclosed or not disclosed to clinicians prior to their routine assessment. It was assumed that the disclosure of DAWBA information before final decision-making increases the total diagnostic agreement as well as the positive agreements regarding internalizing and externalizing disorders. This part of the study was modeled after the design of the study by Ford et al. (submitted) so that the findings may serve as cross-validation.

Given the heterogeneity in clinical expertise within a large clinical service and in line with previous findings [27], the present study analyzed whether or not the degree of clinical experience (more or less than 4 years clinical practice) and professional background (psychiatrists vs. psychologists) affected the agreement of DAWBA and clinical diagnoses with and without previous disclosure of DAWBA.

Method

Sample

A total of 875 children and adolescents older than 5 years admitted to the outpatient department of the Child and Adolescent Psychiatric Service, University of Zurich, Switzerland between September 2007 and June 2009 were eligible for the present study (see Fig. 1). After exclusion of 521 subjects due to a lack of parent motivation or insufficient knowledge of the German language, ratings were available for 354 children and adolescents. Furthermore, data on patients with adjustment disorders as primary diagnosis had to be excluded from analyses because there is no equivalent in the DAWBA. Lack of parent information led to further exclusions. The final sample consisted of a total of 270 children and adolescents aged 5–18 years (mean 10.46 years, SD = 3.56 years) including 184 (68.1 %) boys and 86 (31.9 %) girls. The sample where DAWBA information was
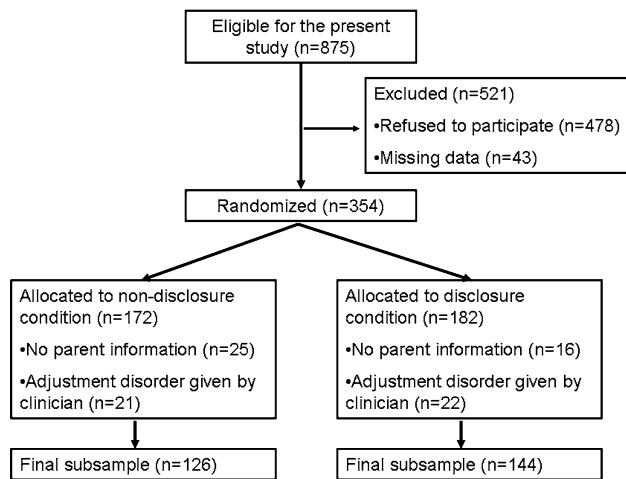
Fig. 1 Flow chart describing progress of participants trough randomized trial

not disclosed ($n = 126$) consisted of 90 boys (71.4 %) and 36 girls (28.6 %) with a mean age of 10.2 years (SD = 3.5 years). In contrast, the sample where DAWBA was disclosed ($n = 144$) consisted of 94 boys (65.3 %) and 50 girls (34.7 %) with a mean age of 10.7 years (SD = 3.6 years). The two samples did not differ regarding to sex ($\chi^2 = 1.17$, df = 1, $p > 0.05$) and age ($t = 0.97$, df = 268, $p > 0.05$) of the participants. In addition to a parent DAWBA in all instances, there was a youth self-report DAWBA in 79 instances and a teacher DAWBA in 117 instances. Informed consent for participation in the study was given by all participating parents and teachers. In addition, the study was approved by the local Ethics Committee of the Canton of Zurich and is registered as a randomized clinical trial (ISRCTN 19935149).

Measures

Development and well-being assessment

The development and well-being assessment (DAWBA) interview consists of questionnaires, interviews, and rating approaches designed to generate ICD-10 and DSM-IV diagnoses on children aged 5–16 years [13]. There are parallel interviews for parents and 11–16 year olds; and there is a brief questionnaire for teachers. Respondents initially complete the Strength and Difficulties Questionnaire [11, 12] before moving on to detailed interview sections covering a wide range of specific diagnoses. The DAWBA can be completed online via the internet. The computer program of the DAWBA brings together the different sorts of information and proposes likely diagnoses [10]. Experienced clinical raters have to decide then whether to accept or overturn the diagnoses in the light of all the data, including transcripts. In a recent comparison

between the DAWBA and two other SDIs, i.e., the Diagnostic Interview Schedule for Children (DISC) and the Child and Adolescent Psychiatric Assessment (CAPA), the DAWBA was a relatively conservative measure, generating fewer diagnoses than the other two measures [2]. The German version of the DAWBA was used in the present study. Further information on the DAWBA including translations in many languages and online demonstrations of the clinical rating procedure are available via http://www.dawba.info.

Procedure

Respondents with internet access completed the online interview at a time and place of their choosing. Respondents without internet access were able to complete the online interview immediately prior to their clinic appointment using a dedicated computer in the clinic. The parents had to have a sufficient knowledge of the German language to understand the DAWBA. After completion of the interview by the parents (and by youth or teachers in some instances), the three expert raters (CK, MA, CWM; all senior board-certified clinicians) generated diagnostic ratings on the basis of the DAWBA information, blind to information independently collected by the clinic. All the three raters were initially trained in DAWBA rating by the author of the measure (RG). Subsequently, the raters met regularly to discuss difficult cases to maintain consistency between raters. Random assignment for the disclosure or non-disclosure condition of the study was made by coin toss of the DAWBA raters after their diagnostic rating. There were no additional restrictions for the randomization process. In case of disclosure, the corresponding clinicians received all available DAWBA information including expert rated diagnoses, SDQ results, and all information from parents, youths and teachers within 4 weeks of the DAWBA being filled out. In case of non-disclosure, the clinicians were blind to DAWBA information until the assessment was finished. There were no adverse events or side effects in each group to report.

The clinical diagnoses were based on ICD-10 and came from a large group of clinicians ($n = 65$), comprising 23 child and adolescent psychiatrists and 42 clinical psychologists, collaborating in teams guided by a senior board-certified child psychiatrist. A total of 30 (46.2 %) of the clinicians had more than 4 years of clinical expertise. To maintain diagnostic standards, junior clinicians were usually closely supervised by a senior team member.

Statistical analyses

Before entering into the main analyses, inter-observer reliability of the three DAWBA expert raters with an

experienced DAWBA rater (AS) was tested by calculation of Cohen's kappa coefficients. The main study consisted of various comparisons of rates of diagnoses. First, the distributions of diagnoses made by DAWBA experts and independent diagnoses in clinicians were compared using McNemar Chi-Square or binomial tests in the sample without DAWBA disclosure (i.e., where the two diagnoses were independent). Second, the overlap of DAWBA- and clinician-generated diagnoses was examined separately for those with and without disclosure of DAWBA ratings to clinicians; indices of agreement included the total and the positive agreements as well as the kappa coefficients. Third, the total and the positive agreements between the two conditions (disclosure, non-disclosure) were compared using Pearsons's Chi-square statistics for any, internalizing, externalizing, and other disorders. Fourth, more versus less-experienced clinicians and psychiatrists versus psychologists were compared regarding the total agreements of internalizing and externalizing disorders using Pearsons's Chi-square statistics. To avoid alpha-error accumulation by multiple comparisons of diagnoses, the Benjamini–Hochberg method was used for adjusting the significance level of 0.05 [4].

Results

Reliability of DAWBA expert diagnoses

A random series of 60 DAWBA expert ratings were blindly re-rated by a senior rater who had been using the DAWBA before and had been involved in various studies of it. Kappa coefficients were 0.83 (95 %CI = 0.68–0.97) for any disorder, 0.84 (95 %CI = 0.69–0.99) for any internalizing disorder, 0.89 (95 %CI = 0.77–1.00) for externalizing disorder, and 0.79 (95 %CI = 0.39–1.00) for any other disorder. According to Landis and Koch [20] all coefficients have to be regarded as almost perfect.

Frequencies of DAWBA expert diagnoses and diagnoses from clinician

Among the 381 DAWBA expert-rated diagnoses and the 277 diagnoses given by clinicians to the present sample, 161 (42.3 %) and 58 (20.9 %) were rated as "unsure", respectively. For the following analyses "unsure" diagnoses were scored as present. However, it has to be noticed that significantly more DAWBA expert diagnosis were rated as unsure compared to diagnoses from clinicians ($m = 0.59$, SD = 0.71 versus $m = 0.21$, SD = 0.47, $t = 7.59$, $p < 0.001$).

Table 1 shows the frequency with which different psychiatric diagnoses were made by DAWBA expert raters and by clinicians (including only those clinician ratings made blind to DAWBA information). The percentages did not sum to 100 % because some children and adolescents had more than one diagnosis. According to the DAWBA expert ratings, 98 (77.8 %) of the children and adolescents had at least one ICD-10 psychiatric disorder. Similarly, 97 (77.0 %) of the children and adolescents were considered to have at least one ICD-10 diagnosis by the corresponding clinicians after full psychiatric assessment. The total number of disorders did not differ significantly. However, DAWBA expert raters more frequently diagnosed multiple disorders, whereas clinicians more frequently diagnosed a single disorder only. Furthermore, significantly more diagnoses were given by DAWBA expert raters for internalizing disorders, in particular for specific phobias and affective disorders. Although DAWBA raters and clinicians did not differ significantly in the number of externalizing disorders, ODD as a specific externalizing disorder was diagnosed significantly more frequently by DAWBA expert raters. In contrast, the clinicians identified significantly more frequently "other non-specified diagnoses". Further analysis of these 29 cases of "other non-specified diagnoses" showed that clinicians most frequently diagnosed "other behavioural and emotional disorders with onset in childhood and adolescence" (10 cases, 34.5 %) and enuresis (8 cases, 27.6 %). Enuresis is not covered by the DAWBA.

Agreement of DAWBA expert diagnoses and clinical diagnoses with and without previous DAWBA disclosure

The left column of Table 2 shows the agreement of independent diagnoses based on DAWBA experts versus clinicians. Total agreement amounted to 78 % ($\kappa = 0.30$) for any diagnosis, between 67 and 83 % for diagnostic categories ($\kappa = 0.22$–0.38), and between 74 and 99 % for specific diagnoses ($\kappa = 0.15$–0.66). In the left column, agreements are shown for the sample with DAWBA information disclosed before clinical decision-making. In this instance, total agreement amounted to 76 % ($\kappa = 0.15$) for any diagnosis, between 63 and 87 % for diagnostic categories ($\kappa = 0.25$–0.46), and between 73 and 100 % for specific diagnoses ($\kappa = 0.24$–1.00). Under this condition, total agreement for internalizing disorders was higher when the DAWBA diagnoses was disclosed to clinicians before final decision-making ($\chi^2 = 3.13$ df = 1, $p < 0.05$). In contrast, disclosure did not significantly influence the rate at which clinicians rated any diagnosis ($\chi^2 = 0.36$, df = 1, $p > 0.05$), externalizing disorders ($\chi^2 = 1.99$, df = 1, $p > 0.05$), or other disorders ($\chi^2 = 0.00$, df = 1, $p > 0.05$). In addition, there were no significant differences between the two conditions for positive agreements between DAWBA and

Table 1 Frequencies of ICD-10 diagnoses by DAWBA expert raters and by clinicians in the sample without DAWBA information (n = 126)

	Diagnoses given by DAWBA raters (n = 126)	Diagnoses given by clinicians (n = 126)	Statistical test [a,b]
Any disorder	98 (77.8 %)	97 (77.0 %)	0.00 ns
One diagnosis	47 (37.3 %)	71 (56.3 %)	9.45**
Two diagnoses	36 (28.8 %)	19 (15.1 %)	6.24*
Three or more diagnoses	15 (11.9 %)	7 (5.6 %)	ns
Internalizing disorders	45 (35.7 %)	26 (20.6 %)	7.90*
Anxiety disorders	30 (23.8 %)	19 (15.1 %)	3.45 ns
Separation anxiety disorders	5 (4.0 %)	1 (0.8 %)	ns
Specific phobias	14 (11.1 %)	2 (1.6 %)	**
Social phobias	7 (5.6 %)	4 (3.2 %)	ns
PTSD	1 (0.8 %)	1 (0.8 %)	ns
OCD	4 (3.2 %)	4 (3.2 %)	ns
Generalized anxiety disorders	7 (5.6)	7 (5.6 %)	ns
Affective disorders	20 (15.9 %)	7 (5.6 %)	*
Depression	13 (10.3 %)	7 (5.6 %)	ns
Externalizing disorders	50 (39.7 %)	47 (37.3 %)	0.11 ns
Hyperactivity disorders	43 (34.1 %)	42 (33.3 %)	0.00 ns
ODD	29 (23.0 %)	7 (5.6 %)	15.75***
CD	7 (5.6 %)	6 (4.8 %)	ns
Other disorders	22 (17.5 %)	43 (34.1 %)	12.25***
PDD/Autism	8 (6.3 %)	8 (6.3 %)	ns
Eating disorder	3 (2.4 %)	6 (4.8 %)	ns
Selective mutism	1 (0.8 %)	2 (1.6 %)	ns
Other non specified disorder	10 (7.9 %)	29 (23.0 %)	10.45**

PTSD posttraumatic stress disorders, *OCD* obsessive–compulsive disorders, *ODD* oppositional defiant disorders, *CD* conduct disorders

* Significance (two sided), $p < 0.05$, ** Significance (two sided), $p < 0.01$, *** Significance (two sided), $p < 0.001$

[a] McNemar χ^2

[b] Binomial distribution (If fewer than 25 cases change values binominal distribution was used instead of χ^2 statistics)

clinician diagnoses regarding any diagnosis ($\chi^2 = 1.30$, df $= 1$, $p > 0.05$), internalizing disorders ($\chi^2 = 1.23$, df $= 1$, $p > 0.05$), externalizing disorders ($\chi^2 = 1.25$, df $= 1$, $p > 0.05$) or other disorders ($\chi^2 = 0.84$, df $= 1$, $p > 0.05$).

Clinical experience and professional background

Under both conditions of disclosure and non-disclosure of DAWBA diagnoses, the total agreements between DAWBA expert and clinician diagnoses was independent of the degree of professional expertise (Table 3) as well as of specialist group (psychiatrists vs. psychologists) (Table 4) for any disorder, internalizing disorders, externalizing disorders, and other disorders.

Discussion

The present study addressed the agreement of diagnoses based on the DAWBA versus the ordinary clinical process of diagnostic assessment. Furthermore, the impact of DAWBA on clinical decision-making in a clinical sample from a large child and adolescent mental health service was

analyzed. To the best of our knowledge, so far, no published study has tested the effects on SDI measures on clinical decision-making by use of a randomized design like the parallel study by Ford et al. (submitted) and the present study. Due to well-known limitations, neither DAWBA expert-rated diagnoses nor clinicians' diagnoses were taken as the "gold standard". In consequence, the present study did not focus on the validation of DAWBA diagnoses in a strict sense but, rather, addressed the clinical utility and practicality of the DAWBA. In comparison to the parallel study by Ford et al. (submitted) there were a few minor differences in the design. First, the sample in the UK study was younger than in the present study and did not contain adolescent patients. Secondly, Ford et al. (submitted) disclosed the computer diagnosis of the DAWBA to clinicians, whereas this study disclosed the diagnosis generated by expert clinical raters. Thirdly, the UK study used the clinical options of "definite", "possible", and "no" diagnosis and matched these options to probabilities of diagnoses against data from the British Child Mental Health Survey [22, 23]. Given the lack of similar data from Switzerland, the present study used only the two categories of "definite" and "no" by collapsing the "possible" option into the "definite" option.

Table 2 Level of agreement between DAWBA expert diagnoses and clinical diagnoses according to whether DAWBA was or was not disclosed to clinicians

	DAWBA diagnosis disclosed to clinicians (n = 144)					DAWBA diagnosis not disclosed to clinicians (n = 126)				
	DAWBA + clinician	DAWBA only	Clinician only	Total agreement	κ	DAWBA + clinician	DAWBA only	Clinician only	Total agreement	κ
Any disorders	103 (71.5 %)	20 (13.9 %)	14 (9.7 %)	110 (76.4 %)	0.15 ns	82 (65.1 %)	16 (12.7 %)	15 (11.9 %)	95 (75.4 %)	0.30**
Internalizing disorders	24 (16.7 %)	30 (20.8 %)	3 (2.1 %)	111 (77.1 %)	0.46***	15 (11.9 %)	30 (23.8 %)	11 (8.7 %)	85 (67.5 %)	0.22*
Anxiety disorders	17 (11.8 %)	22 (15.3 %)	5 (3.5 %)	117 (81.3 %)	0.45***	10 (7.9 %)	20 (15.9 %)	9 (7.1 %)	97 (77.0 %)	0.27**
Separation anxiety disorder	2 (1.4 %)	7 (4.9 %)	1 (0.7 %)	136 (94.4 %)	0.31***	0 (0.0 %)	5 (4.0 %)	1 (0.8 %)	120 (95.2 %)	−0.01 ns
Specific phobias	2 (1.4 %)	10 (6.9 %)	2 (1.4 %)	132 (91.7 %)	0.22**	1 (0.8 %)	13 (10.3 %)	1 (0.8 %)	112 (88.9 %)	0.10 ns
Social phobias	2 (1.4 %)	6 (4.2 %)	6 (4.2 %)	132 (91.7 %)	0.21*	2 (1.6 %)	5 (4.1 %)	2 (1.6 %)	119 (94.4 %)	0.34***
PTSD	1 (0.7 %)	0 (0.0 %)	0 (0.0 %)	144 (100.0 %)	1.00***	0 (0.0 %)	1 (0.8 %)	1 (0.8 %)	124 (98.4 %)	−0.01 ns
OCD	2 (1.4 %)	3 (2.1 %)	0 (0.0 %)	141 (98 %)	0.56***	2 (1.6 %)	2 (1.6 %)	2 (1.6 %)	122 (96.8 %)	0.48***
Generalized anxiety disorders	3 (2.1 %)	7 (4.9 %)	3 (2.1 %)	134 (93.1 %)	0.34***	1 (0.7 %)	6 (4.8 %)	6 (4.8 %)	114 (90.5 %)	0.09 ns
Affective disorders	8 (5.6 %)	15 (11.4 %)	4 (2.8 %)	125 (86.8 %)	0.39***	3 (2.4 %)	4 (3.2 %)	17 (13.5 %)	105 (83.3 %)	0.15*
Depression	7 (4.9 %)	9 (6.3 %)	3 (2.1 %)	132 (91.7 %)	0.50***	3 (2.4 %)	9 (7.1 %)	3 (2.4 %)	114 (90.5 %)	0.29**
Externalizing disorders	43 (29.9 %)	23 (16.0 %)	31 (21.5 %)	90 (62.5 %)	0.25**	30 (23.8 %)	20 (15.9 %)	17 (13.5 %)	89 (70.6 %)	0.38***
Hyperactivity disorders	35 (24.3 %)	23 (16.0 %)	16 (11.1 %)	105 (72.9 %)	0.43***	26 (20.6 %)	17 (13.5 %)	16 (12.7 %)	93 (73.8 %)	0.41***
ODD	15 (10.4 %)	36 (25.0 %)	2 (1.4 %)	106 (73.6 %)	0.32**	4 (3.2 %)	25 (19.8 %)	3 (2.4 %)	98 (77.8 %)	0.15*
CD	4 (2.8 %)	12 (8.3 %)	6 (4.2 %)	126 (87.5 %)	0.24**	2 (1.6 %)	5 (4.0 %)	4 (3.2 %)	117 (92.9 %)	0.27**
Other disorders	24 (16.7 %)	15 (10.4)	26 (18.1 %)	103 (71.5 %)	0.34***	16 (12.7 %)	7 (5.6 %)	29 (23.0 %)	90 (71.4 %)	0.30***
PDD/Autism	8 (5.6 %)	3 (2.1 %)	8 (5.6 %)	133 (92.4 %)	0.55***	4 (3.2 %)	4 (3.2 %)	4 (3.2 %)	118 (93.7 %)	0.47***
Selective mutism	1 (0.7 %)	1 (0.7 %)	0 (0.0 %)	143 (99.3 %)	0.66***	1 (0.8 %)	1 (0.8 %)	0 (0.0 %)	125 (99.2 %)	0.66***
Eating disorders	2 (1.4 %)	2 (1.4 %)	2 (1.4 %)	140 (97.2 %)	0.49***	3 (2.4 %)	0 (0.0 %)	3 (2.4 %)	123 (97.6 %)	0.66***

PTSD posttraumatic stress disorders, *OCD* obsessive–compulsive disorders, *ODD* oppositional defiant disorders, *CD* conduct disorders, κ = Cohen's kappa

* Significance (two sided), $p < 0.05$, ** Significance (two sided), $p < 0.01$, *** Significance (two sided), $p < 0.001$

Table 3 Frequencies of total agreements of more-experienced and less-experienced clinicians with and without DAWBA information

	Frequency of total diagnostic agreement by clinicians with DAWBA information ($n = 144$)			Frequency of total diagnostic agreement by clinicians without DAWBA information ($n = 126$)		
	More experienced clinicians ($n = 64$)	Less experienced clinicians ($n = 80$)	χ^2	More experienced clinicians ($n = 53$)	Less experienced clinicians ($n = 73$)	χ^2
Any disorders	51 (79.7 %)	59 (73.8 %)	0.70 ns	40 (75.5 %	55 (75.3 %)	0.00 ns
Internalizing disorders	51 (79.7 %)	60 (75.0 %)	0.44 ns	36 (67.9 %)	49 (67.1 %)	0.09 ns
Externalizing disorders	40 (62.5 %)	50 (62.5 %)	0.00 ns	40 (75.5 %)	49 (67.1 %)	1.03 ns
Other disorders	50 (78.1 %)	53 (66.3 %)	2.46 ns	39 (73.6 %)	51 (69.9 %)	0.21 ns

* Significance (two sided), $p < 0.05$, ** Significance (two sided), $p < 0.01$, *** Significance (two sided), $p < 0.001$

Table 4 Frequencies of total agreements of psychiatrists and psychologists with and without DAWBA information

	Frequency of total diagnostic agreement by clinicians with DAWBA information ($n = 144$)			Frequency of total diagnostic agreement by clinicians without DAWBA information ($n = 126$)		
	Psychiatrists ($n = 49$)	Psychologists ($n = 95$)	χ^2	Psychiatrists ($n = 45$)	Psychologists ($n = 81$)	χ^2
Any disorders	37 (75.5 %)	73 (76.8 %)	0.03 ns	31 (68.9 %)	64 (79.0 %)	1.60 ns
Internalizing disorders	39 (79.6 %)	72 (75.8 %)	0.27 ns	34 (75.6 %)	5 (61.7 %)	2.10 ns
Externalizing disorders	32 (65.3 %)	58 (61.1 %)	0.25 ns	31 (68.9 %)	58 (71.6 %)	0.10 ns
Other disorders	32 (65.3 %)	71 (74.7 %)	1.41 ns	32 (71.1 %)	58 (71.6 %)	0.03 ns

* Significance (two sided), $p < 0.05$, ** Significance (two sided), $p < 0.01$, *** Significance (two sided), $p < 0.001$

In addition, it should be noted that the pattern of diagnoses provided by the participating clinicians in the present study very much reflects the composition of a referred sample rather than a community sample. There was a large proportion of externalizing disorders, with hyperkinetic disorder being substantially commoner than internalizing disorders and developmental disorders. This is strikingly different from the findings from a representative community study that had been performed some fifteen years previously in the same area, showing that anxiety disorders were commonest, followed by ADHD and a rather small proportion of oppositional defiant disorders [28].

In this clinic sample, the inter-rater reliability of the DAWBA expert diagnoses was very satisfactory in predicting the presence of any diagnosis as well as for the presence of an internalizing, externalizing or other diagnosis. This is in line with previous findings on the reliability of the DAWBA in community samples [6, 7, 15]. In the present study DAWBA expert raters and independent clinical raters did not differ significantly in their total frequency of diagnoses, though clinical raters typically made just one diagnosis and DAWBA raters were more likely to make multiple diagnoses. A tendency for clinicians to focus on the presenting problem and miss comorbidity has been noted previously [13]. Diagnostic agreements were fair to moderate for any disorder as well as for internalizing

and externalizing disorders. Furthermore, diagnostic agreements were good for some diagnoses from the category of other disorders such as selective mutism and eating disorders.

The disclosure of DAWBA information before clinical decision-making had an impact only on the total agreement regarding internalizing disorders, which was significantly increased compared to the condition when DAWBA diagnoses were not disclosed. As a consequence, after disclosure Cohen's kappa for internalizing disorders increased to a moderate degree. More specifically, the kappa coefficients of PTSD and depression increased considerably. In contrast to internalizing disorders, we did not detect a significant difference between the disclosure and non-disclosure condition regarding the total diagnostic agreement on externalizing disorders and other disorders. Furthermore, across all diagnostic categories, the rate of positive agreements was equal or higher when DAWBA information was disclosed. This result suggests that access to DAWBA information may be more likely to prompt clinicians to add an extra diagnosis than to prompt them to withdraw a diagnosis that they would otherwise have made.

The moderate agreements for internalizing disorders, externalizing disorders and other disorders are in line with previous studies comparing SDI and clinical judgments, finding similar κ values of 0.29 for externalizing and 0.28

for internalizing disorders [26]. In agreement with these findings, an initial study by Goodman et al. [13] analyzed 39 clinical cases from Manchester and London, and compared DAWBA findings and diagnoses from case notes. This study found similar convergence between the two types of diagnoses (κ values of 0.48 for internalizing disorders, 0.40 for disruptive behavior disorders and 0.64 for hyperkinetic disorders). The agreement between computer diagnosis and clinician diagnosis in the parallel UK study by Ford et al. (submitted) was all in all comparable to the present study under both the disclosure and the non-disclosure condition.

However, two subsequent studies found even higher rates of agreement between DAWBA expert diagnoses and clinician diagnoses [1, 24]. Using the Arabic version of the DAWBA, Alyahri and Goodman [1] found κ values of 0.63 for internalizing disorders and of 0.69 for externalising disorders. Finally, Mullick and Goodman [24] found higher convergence of DABWA and clinical diagnoses in a clinical sample of Bangladesh (any disorder $\kappa = 0.81$, internalizing disorder $\kappa = 0.94$, hyperkinetic disorder $\kappa = 0.63$ and disruptive behaviour disorder $\kappa = 0.64$). Perhaps the lower agreement in the present study is attributable to larger teams and more varied trainings than in the two low-income countries that were previously studied.

In a previous study that was performed in the same clinic as the present analyses, the inter-rater reliability of ICD-10 diagnoses was clearly superior in senior child and adolescent psychiatrists as compared to junior child and adolescent psychiatrists and clinical psychologists [27]. However, in the present study no significant differences were found in comparing diagnoses made by experienced versus non-experienced clinicians or psychiatrists versus psychologists in their agreement with DAWBA diagnoses. The difference between the present and the previous study was unexpected and would warrant further exploration.

The present study showed that the total agreement regarding internalizing disorders increased significantly after the DAWBA diagnosis had been disclosed. This was not true for externalizing disorders. The diagnoses of affective and anxiety disorders are strongly influenced by clinicians' assumptions and beliefs [25], potentially leading to missed or inappropriate treatment. Knowing that access to DAWBA information increases the diagnosis of emotional disorders does not in itself prove that the resultant diagnosis is more accurate or leads to more effective treatment—though previous studies of standardized diagnostic interviews suggests that this is a realistic hope [17].

Finally, without detailed knowledge of the exact processes it is difficult to explain why the impact of DAWBA information on clinical decision-making was so limited.

Firstly, practitioners may have been right in overruling some DAWBA diagnoses by applying the ICD-10 criteria that symptoms may have been better explained by another diagnosis. Secondly, some diagnoses by the clinicians may have been based on information that was not available in the DAWBA, e.g., direct observations and psychological testing. Thirdly, some of the previously mentioned barriers in clinicians regarding the use of SDI may have also affected the present findings [16, 21]. Thus, future research should be aiming for a more detailed analysis of clinical decision-making under ordinary mental health service conditions.

Limitations

There was no strict control of the procedure of diagnostic decision-making and it was not practicable to test the reliability of diagnoses within clinical teams. It was also not realistic to try to decide which source was right when the DAWBA and clinical diagnoses differed. While comparable information was collected by all DAWBA assessments, the length and focus of the clinical assessments was too varied to permit valid consensus diagnoses. Furthermore, the sample sizes for some of the more specific disorders were relatively small.

Conclusions

The present study used the DAWBA under normal clinic conditions in a large group of patients assessed by a heterogeneous group of clinicians. Like other SDI, the DAWBA diagnoses show only fair to moderate agreement with the diagnoses made by clinicians. This may partly reflect the well-recognized low reliability of clinical diagnoses based on unstructured interviews. The use of the DAWBA resulted in increased diagnosis of emotional disorders, and it could potentially also allow clinicians to avoid the trap of focusing excessively on the presenting problem, thereby missing significant comorbidity. The primary purpose of including the DAWBA or any other SDI in a clinical assessment is to make psychiatric diagnoses more accurate. Future studies should explore the causes of discrepancies between SDI and clinical diagnoses and investigate as to why clinicians' diagnoses are influenced only to such a small extent by the disclosure of SDI data.

Conflict of interesrt None.

References

1. Alyahri A, Goodman R (2006) The validation of the Arabic SDQ and DAWBA. East Mediterr Health J 12:S138–S146

2. Angold A, Erkanli A, Copeland W, Goodman R, Fisher PW, Costello EJ (2012) Psychiatric diagnostic interviews for children and adolescents: a comparative study. J Am Acad Child Adolesc Psychiatry 51:506–517

3. Basco RM, Bostic JQ, Davies D, Rush AJ, Witte B, Hendrickse W, Barnett V (2000) Methods to improve diagnostic accuracy in a community mental health setting. Am J Psychiatry 157:1599–1605

4. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc Series B 57:289–300

5. Duffy A, Doucette S, Lewitzka U, Alda M, Hajek T, Grof P (2011) Findings from bipolar offspring studies: methodology matters. Early Interv Psychiatry 5:181–191

6. Fleitlich-Bilyk B, Goodman R (2004) Prevalence of child and adolescent psychiatric disorders in southeast Brazil. J Am Acad Child Adolesc Psychiatry 43:727–734

7. 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 42:1203–1211

8. Garb HN (2005) Clinical judgment and decision making. Annu Rev Clin Psychol 1:67–89

9. Garb HN (1998) Studying the clinician: judgment research and psychological assessment. American Psychological Association, Washington

10. 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

11. Goodman R (2001) Psychometric properties of the strengths and difficulties questionnaire. J Am Acad Child Adolesc Psychiatry 40:1337–1345

12. Goodman R (1997) The Strengths and Difficulties Questionnaire: a research note. J Child Psychol Psychiatry 38:581–586

13. Goodman R, Ford T, Richards H, Gatward R, Meltzer H (2000) The development and well-being assessment: description and initial validation of an integrated assessment of child and adolescent psychopathology. J Child Psychol Psychiatry 41:645–655

14. Goodman R, Slobodskaya H, Knyazev G (2005) Russian child mental health—a cross-sectional study of prevalence and risk factors. Eur Child Adolesc Psychiatry 14:28–33

15. Heiervang E, Stormark KM, Lundervold AJ, Heimann M, Goodman R, Posserud MB, Ullebo AK, Plessen KJ, Bjelland I, Lie SA, Gillberg C (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 46:438–447

16. Jensen-Doss A, Hawley KM (2010) Understanding barriers to evidence-based assessment: clinician attitudes toward standardized assessment tools. J Clin Child Adolesc Psychol 39:885–896

17. 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

18. 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

19. Jewell J, Handwerk M, Almquist J, Lucas C (2004) Comparing the validity of clinician-generated diagnosis of conduct disorder to the diagnostic interview schedule for children. J Clin Child Adolesc Psychol 33:536–546

20. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33:159–174

21. Martin AM, Fishman R, Baxter L, Ford T (2010) Practitioners' attitudes towards the use of standardized diagnostic assessment in routine practice: a qualitative study in two child and adolescent mental health services. Clin Child Psychol Psychiatry 16:407–420

22. Meltzer H, Gatward R, Goodman R, Ford T (2000) The mental health of children and adolescents in Great Britain. The Stationery Office, London

23. Meltzer H, Lader D, Corbin T, Goodman R, Ford T (2004) The mental health of young people looked after by authorities in Scotland. The Stationery Office, Edinburgh

24. Mullick MS, Goodman R (2005) The prevalence of psychiatric disorders among 5–10 year olds in rural, urban and slum areas in Bangladesh: an exploratory study. Soc Psychiatry Psychiatr Epidemiol 40:663–671

25. Potts MK, Burnam MA, Wells KB (1991) Gender differences in depression detection: a comparison of clinician diagnosis and standardized assessment. Psychol Assess J Consult Clin Psychol 3:609–615

26. Rettew D, Doyle Lynch A, Achenbach TM, Dumenci L, Ivanova MY (2009) Meta-analyses agreement between diagnoses made from clinical evaluations and standardized diagnostic interviews. Int J Methods Psychiatr Res 18:169–184

27. Steinhausen HC, Erdin A (1991) The inter-rater reliability of child and adolescent psychiatric disorders in the ICD-10. J Child Psychol Psychiatry 32:921–928

28. Steinhausen HC, Metzke CW, Meier M, Kannenberg R (1998) Prevalence of child and adolescent psychiatric disorders: the Zurich Epidemiological Study. Acta Psychiatr Scand 98:262–271