International Journal of Methods in Psychiatric Research
Int. J. Methods Psychiatr. Res. 21(4): 311–320 (2012)
DOI: 10.1002/mpr.1371

Diagnostic validity across racial and ethnic groups in the assessment of adolescent DSM-IV disorders

JENNIFER GREIF GREEN,[1] MICHAEL J. GRUBER,[2] RONALD C. KESSLER,[2] JULIA Y. LIN,[3] KATIE A. MCLAUGHLIN,[4] NANCY A. SAMPSON,[2] ALAN M. ZASLAVSKY[2] & MARGARITA ALEGRIA[5]

1 School of Education, Boston University, Boston, MA, USA
2 Department of Health Care Policy, Harvard Medical School, Boston, MA, USA
3 Cooperative Studies Program Coordinating Center, VA Palo Alto Healthcare System, Mountain View, CA, USA
4 Division of General Pediatrics, Children's Hospital Boston, Harvard Medical School, Boston, MA, USA
5 Center for Multicultural Mental Health Research, Cambridge Health Alliance, Somerville, MA, USA

Key words
National Comorbidity Survey Replication Adolescent Supplement, validity, disparities

Correspondence
Ronald C. Kessler, Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA 02115, USA.
Telephone (+1) 617-432-3587
Fax (+1) 617-432-3588
Email: ncs@hcp.med.harvard.edu

Received 31 March 2011; revised 20 December 2011; accepted 25 March 2012

Abstract

We examine differential validity of the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) diagnoses assessed by the fully-structured Composite International Diagnostic Interview Version 3.0 (CIDI) among Latino, non-Latino Black, and non-Latino White adolescents in comparison to gold standard diagnoses derived from the Schedule for Affective Disorders and Schizophrenia for School-age Children (K-SADS). Results are based on the National Comorbidity Survey Replication Adolescent Supplement, a national US survey of adolescent mental health. Clinicians re-interviewed 347 adolescent/parent dyads with the K-SADS. Sensitivity and/or specificity of CIDI diagnoses varied significantly by ethnicity/race for four of ten disorders. Modifications to algorithms sometimes reduced bias in prevalence estimates, but at the cost of reducing individual-level concordance. These findings document the importance of assessing fully-structured diagnostic instruments for differential accuracy in ethnic/racial subgroups. *Copyright © 2012 John Wiley & Sons, Ltd.*

Introduction

Racial/ethnic disparities in treatment of mental disorders are well-documented (Cuffe *et al*., 1995; Kataoka *et al*., 2002) and may be due at least in part to differences in the identification of mental disorders across racial/ethnic groups. Interviews designed to diagnose mental disorders are usually developed in general population samples rather than designed to be as effective in minority samples, with the result that these instruments are often less effective in identifying racial/ethnic minorities with mental health problems as compared to respondents in the majority population (Alegria *et al*., 2011; Alegria *et al*., 2009; Breslau *et al*., 2008; Eisen *et al*., 2006; Johnson *et al*., 2007). These differences may also help explain discrepancies in epidemiological studies of racial/ethnic differences in disorder prevalence. The current report presents results of a psychometric study designed to investigate this issue in a large epidemiological survey of mental disorders among youths in the United States. Previous studies of race/ethnic differences in the mental health of US adolescents have yielded inconsistent results (Anderson and Mayes, 2010; McLaughlin *et al.*, 2007), with some studies finding significant group differences in disorder prevalence (Minsky *et al.*, 2006; Roberts *et al.*, 1997) and others not (Angold *et al.*, 2002). It is important to examine the extent to which lack of equivalence in the sensitivity or specificity of measures of mental health across racial/ethnic groups are involved in explaining these results for both research and service planning purposes.

There are several reasons to expect that the diagnostic validity of fully-structured diagnostic interviews may be lower among racial/ethnic minority adolescents than other adolescents. First, there may be racial/ethnic differences in comfort with the interview process, rapport built with interviewers (potentially related to interviewer race/ethnicity), and general familiarity with the task of responding to questions about mental health symptoms (Kessler and Ustun, 2004; Komiti *et al.*, 2001). A second concern is that minority adolescents may have different interpretations of the structured interview questions, many of which are complex, use vague terminology, and include multiple clauses (Kessler and Ustun, 2004). Third, cultural context may influence adolescents' and parents' subjective interpretation of behaviors, particularly in judging whether symptoms reflect maladjustment or result in impairment (Alegria *et al.*, 2009). Fourth, culturally-grounded conceptions of time (Alegria *et al.*, 2009; Kessler and Ustun, 2004) may influence the accuracy of retrospective reports across racial/ethnic groups, particularly when asked questions about events that happened many years ago or when interviewers request information about the duration and recency of symptoms. Finally, it is possible that the phenomenology of disorders vary across racial/ethnic groups, with some cultures viewing behaviors as pathological that others consider normative (Lewis-Fernández *et al.*, 2010). These variations may influence both endorsement of symptoms and reports of related impairment.

In this study we examine the validity of the Composite International Diagnostic Interview Version 3.0 (CIDI), administered in the US National Comorbidity Survey Replication Adolescent Supplement (NCS-A; Merikangas *et al.*, 2009), for racial/ethnic minority adolescents. Prior research in the NCS-A using the aggregate youth sample showed strong concordance of individual-level CIDI diagnoses with diagnoses determined by blinded interviews carried out by clinicians using the Schedule for Affective Disorders and Schizophrenia for School-age Children (K-SADS; Kessler *et al.*, 2009c). Area under the ROC (receiver operating characteristic) curve (AUC) ranged from 0.84–0.94 for classes of disorders and was 0.87 for diagnosis of any disorder.

Although overall CIDI validity is quite good at identifying mental health needs for the general population of adolescents, it is unclear whether validity differs by race/ethnicity and, if so, whether it is possible to improve measurement precision for racial/ethnic minority youths by changing diagnostic thresholds. Prevalence estimates in the NCS-A indicate some significant racial/ethnic differences, particularly lower prevalence of externalizing disorders among minorities than non-Latino white youths, even when controlling for socio-economic status and other demographic covariates (Kessler *et al.*, 2012). Given inconsistencies in the literature about variation in disorder prevalence by race and ethnicity (Anderson and Mayes, 2010; McLaughlin *et al.*, 2007), we investigate whether group differences in the NCS-A may be influenced by lack of measurement equivalence for certain racial or ethnic groups.

Methods

The NCS-A sample

The NCS-A was carried out between February 2001 and January 2004. Adolescents (ages 13–17) were interviewed face-to-face in dual-frame household and school samples (Kessler *et al.*, 2009a, 2009b). The household sample included 904 adolescents (25 who had dropped out of school) from households that participated in the National Comorbidity Survey Replication (NCS-R), a national survey of adults (Kessler and Merikangas, 2004). The conditional (on NCS-R participation) adolescent response rate was 86.8%. The school sample included 9244 adolescents from a representative sample of 320 schools in the NCS-R counties. The conditional (on school participation) adolescent response rate was 82.6%. One parent or surrogate (henceforth referred to as *parents*) was asked to complete a self-administered questionnaire (SAQ) about the participating adolescent's developmental history and mental health. The parent SAQ conditional (on adolescent participation) response rate was 82.5% in the household sample and 83.7% in the school sample. This report focuses on the 6483 adolescent–parent pairs with data from both adolescent interviews and SAQs.

Although the proportion of initially selected schools that participated was low (28.0%), these were replaced with matched replacement schools. Comparison of household sample respondents who attended non-participating schools with school sample respondents from replacement schools found no evidence of bias in estimates of either prevalence or correlates of disorders (Kessler *et al.*, 2009a).

Once the survey was completed, cases were weighted for variation in within-household probability of selection (household sample) and residual discrepancies between sample and population socio-demographic and geographic distributions. These weighting procedures are detailed elsewhere (Kessler *et al.*, 2009a, 2009b). The weighted composite sample socio-demographic distributions of the NCS-A closely approximate those of the 2000 Census population, with 65.6% of respondents being non-Latino White, 15.1% non-Latino Black, 14.4% Latino, and 5.0% Other. The Others were not included in the analyses reported here because of their small numbers.

Written informed consent was obtained from parents or guardians before approaching adolescents to participate in the survey. Written adolescent assent was then obtained from adolescents before surveying either adolescents or parents. Each respondent was given $50 for participation. These recruitment-consent procedures were approved by the Human Subjects Committees of both Harvard Medical School and the University of Michigan.

The clinical reappraisal sample

A clinical reappraisal study was completed by telephone with a quota sample of 347 adolescent respondents from the school sample and their parents (Kessler *et al.*, 2009c). Adolescents who met Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV)/CIDI criteria for one or more relatively uncommon disorders (e.g. bipolar I or II, agoraphobia) were oversampled relative to respondents who met criteria only for more common disorders and those with no disorders were sampled at an even lower rate to acquire a large enough sample of adolescents with each disorder to conduct disorder-specific analyses of concordance between diagnoses based on the CIDI and the K-SADS. The sample included 241 non-Latino Whites, 56 non-Latino Blacks, 32 Latinos, and 18 Others. Each adolescent and parent respondent was given a $50 incentive for participation in the clinical reappraisal survey (over and above the $50 incentive for participation in the main survey). (For more details on the NCS-A clinical reappraisal study see Kessler *et al.*, 2009c).

Measures

CIDI diagnostic assessment

Adolescents were administered a modified version of the CIDI, a fully-structured interview designed for use by trained lay interviewers (Kessler and Ustun, 2004). The DSM-IV disorders assessed include mood disorders (major depressive disorder [MDD] or dysthymia, bipolar I–II disorder [BPD]), anxiety disorders (panic disorder with or without agoraphobia, agoraphobia without history of panic disorder, social phobia, specific phobia, generalized anxiety disorder [GAD], post-traumatic stress disorder [PTSD], separation anxiety disorder), behavior disorders (attention deficit hyperactivity disorder [ADHD], oppositional-defiant disorder [ODD], conduct disorder [CD], intermittent explosive disorder [IED], and eating disorders [anorexia nervosa, bulimia nervosa, binge-eating disorder]), and substance disorders (alcohol and drug abuse, alcohol and drug dependence with abuse).

In addition, parent questionnaires assessed disorders for which parent reports have previously been shown to play a large part in diagnosis: behavior disorders (Grills and Ollendick, 2002; Johnston and Murray, 2003) and depression/dysthymia (Braaten *et al.*, 2001). Parent and adolescent reports were combined at the symptom level using an “or” rule (i.e. the symptom was considered present if endorsed by either respondent). All diagnoses were made using DSM-IV organic exclusion rules. All but two diagnoses were made using DSM-IV diagnostic hierarchy rules. The exceptions were ODD, which was defined with or without CD, and substance abuse, which was defined with or without dependence. Prevalence was assessed in three time frames: lifetime, past 12 months, and past 30 days. Our analyses of diagnostic concordance are focused on the concordance of lifetime diagnoses, as the K-SADS administration focused on a lifetime time frame. In order to address the problem that lifetime prevalence is under-estimated in retrospective cross-sectional surveys of adults (Kessler *et al.*, 1998; Moffitt *et al.*, 2010) and youth (Cohen *et al.*, 1993; Costello *et al.*, 2003; Kim-Cohen *et al.*, 2003), special probing procedures were used to increase accuracy of memory search (Kessler and Ustun, 2004). These procedures have been shown experimentally to increase recall of lifetime mental disorders among adults (Kessler *et al.*, 1998).

K-SADS diagnostic assessment

The K-SADS is a semi-structured research diagnostic interview that is designed to be administered by trained clinical interviewers. K-SADS questions are designed to elicit rich verbal responses that form the basis of interviewer ratings about the presence or absence of symptoms. The standard K-SADS was modified by deleting disorders not assessed in the NCS-A, focusing only on a lifetime time frame, and streamlining the initial screening section of the interview to include information about respondent endorsement of diagnostic stem questions in the earlier CIDI interview (for description, see Kessler *et al.*, 2009c).

The disorders assessed in this version of the K-SADS included six DSM-IV anxiety disorders (panic disorder with or without agoraphobia, agoraphobia without a history of panic disorder, GAD, specific phobia, social phobia, PTSD), three mood disorders (bipolar spectrum disorder, MDD, dysthymic disorder), three disruptive behavior disorders (ADHD, CD, ODD), and four substance use disorders (alcohol abuse with or without dependence, illicit drug abuse with or without dependence, alcohol dependence with a history of abuse, illicit drug dependence with a history of abuse).

Socio-demographics

Adolescents were assigned to the following racial/ethnic groups based on self-reported responses to race and ethnicity questions: non-Latino White, non-Latino Black, Latino, and Other. As noted earlier, adolescents classified as "Other" were excluded from these analyses because of their small number and heterogeneity.

Analysis methods

Analyses of concordance between diagnoses based on the CIDI and those based on the K-SADS were carried out after weighting the clinical reappraisal sample data to adjust for the under-sampling of CIDI non-cases and those with only common CIDI disorders. We investigated whether there were racial/ethnic differences in concordance with binary logistic regression equations that predicted K-SADS diagnoses from the main effects and interaction of race/ethnicity with CIDI diagnoses. Significant interactions indicated that concordance between diagnoses based on the CIDI and K-SADS differ significantly by respondent race/ethnicity. In cases of this sort, we investigated the nature of the discordance by testing whether CIDI sensitivity (SN) and specificity (SP) differed significantly by race/ethnicity based on the Rao–Scott χ^2 test.

For each disorder with significant racial/ethnic differences in either SN or SP, we calculated McNemar χ^2 tests to evaluate bias in prevalence estimates and area under the ROC curve (AUC) (Hanley and McNeil, 1982) and Cohen's κ (Cohen, 1960) to evaluate individual-level concordance between diagnoses based on the CIDI and K-SADS. We used the κ because it is the most commonly used measure of diagnostic concordance. But we also used AUC as an alternative measure because κ has the disadvantage of varying across populations that differ in prevalence even when the populations do not differ in SN or SP. The AUC does not have this problem. Although the AUC was developed to study the association between a continuous predictor and a dichotomous outcome, it can be used in the special case where the predictor is a dichotomy, in which case AUC equals (SN + SP)/2.

The final phase of analysis involved modifying the strategies for combining CIDI symptom-level data to generate diagnoses that would improve concordance with K-SADS diagnoses for different racial/ethnic groups. This was done by using regression analysis to predict K-SADS diagnoses from CIDI symptom-level items. Where possible, we used results from previous analyses of K-SADS/CIDI concordance in the full clinical reappraisal sample to guide the selection of criteria for CIDI modification. When the interaction of a CIDI item and the race/ethnicity of the respondents significantly improved prediction of K-SADS diagnosis, we modified the CIDI diagnostic algorithm and then compared modified CIDI diagnoses to the K-SADS diagnoses to see if they had higher concordance than did the original CIDI diagnoses.

Results

Racial/ethnic differences in CIDI/K-SADS diagnostic concordance

With the exceptions of one low AUC value for Latinos (0.66 for ADHD) and three low values for non-Latino Blacks (0.53–0.65 for panic disorder, PTSD, and ODD), AUC values are quite comparable across diagnoses among Latinos (0.75–1.0), non-Latino Blacks (0.73–1.0), and non-Latino Whites (0.76–0.94) (Table 1). Disaggregation of AUC into its two main components, SN and SP, finds significant race/ethnic differences in SN for four disorders (agoraphobia, panic disorder, PTSD, and ADHD) and in SP for one (agoraphobia). These significant differences encompass all cases where AUC is meaningfully lower among minorities with the exception of ODD. In the latter case, the low AUC among non-Latino Blacks (0.53) is not associated with significant race/ethnic differences in either SN or SP due to the small number of respondents with this disorder.

For agoraphobia, SN is significantly lower for non-Latino Whites than others ($\chi^2_1 = 4.0$, $p = 0.045$) and SP is significantly lower for non-Latino Blacks than others ($\chi^2_2 = 8.0$, $p = 0.018$). For panic disorder and PTSD, SN is significantly lower for non-Latino Blacks (panic disorder $\chi^2_1 = 9.1$, $p = 0.003$; PTSD $\chi^2_1 = 5.6$, $p = 0.018$), whereas ADHD SN is significantly lower for Latinos than those in other race/ethnic groups ($\chi^2_2 = 7.8$, $p = 0.020$). These four diagnoses were the focus of subsequent analyses to investigate how to enhance the algorithms for detecting these disorders so as to remove differences in concordance between diagnoses based on the CIDI and K-SADS across racial and ethnic groups.

CIDI/K-SADS diagnostic concordance for agoraphobia

For agoraphobia, SP was strong for all race/ethnic groups; however, the relatively low SP for non-Latino Black youths (95.9% versus 97.3–99.4% for Latinos and non-Latino

Table 1 Race/ethnic differences in the sensitivity, specificity, and area under the ROC curve (AUC) of CIDI/DSM-IV diagnoses compared to the clinical gold standard diagnoses in the NCS-A clinical reappraisal sample ($n = 347$)

	AUC in sub-samples[1]			Significance of race/ethnic differences	
				Sensitivity	Specificity
Disorder	NLW	LA	NLB	χ^2_2	χ^2_2
Agoraphobia	0.85	0.99	0.95	4.0*	8.0*
GAD	0.76	1.00	0.87	0.8	1.0
Social phobia	0.81	0.75	0.82	4.6	2.9
Specific phobia	0.94	0.96	0.94	2.8	3.2
Panic disorder	0.90	1.0	0.65	9.1*	0.5
PTSD	0.78	0.99	0.65	5.6*	0.9
MDD/dysthymia	0.89	0.86	0.78	5.5	4.8
ADHD	0.81	0.66	0.73	7.8*	3.1
ODD	0.90	0.92	0.53	0.9	3.4
Any substance	0.93	1.0	1.0	—	5.0
(*n*)	(241)	(32)	(56)		

[1] NLW, non-Latino White; LA, Latino; NLB, non-Latino Black.

*AUC varies significantly across the different race/ethnic sub-samples at the 0.05 level, two-sided test.

Whites) led to over-identification of the disorder (Table 2). When based on the K-SADS, the estimated prevalence of agoraphobia among non-Latino Blacks was 2.6%; the estimate more than doubled when it was based on the CIDI (6.5%; $\chi^2_2 = 19.5$, $p < 0.001$). The CIDI also significantly over-estimated agoraphobia among Latino youths (CIDI 4.4% versus K-SADS 1.8%, $\chi^2_2 = 7.2$, $p = 0.007$).

To improve specificity, we tested strategies to tighten diagnostic criteria and lower CIDI prevalence of agoraphobia for non-Latino Blacks and Latinos. We selected questions measuring agoraphobia that were not included in our initial diagnostic algorithms (e.g. inability to leave home without a family member or friend) and entered each of these symptoms into a regression model with CIDI diagnosis and non-Latino Black race (versus other race/ethnicity) to predict the K-SADS agoraphobia diagnosis. Of the symptoms tested, only one significantly interacted with non-Latino Black race to improve K-SADS prediction: “Was there ever a time in your life when you felt badly about yourself or were upset, worried, or disappointed with yourself because of your fear (or avoidance)?” This item was added as a requirement to the original CIDI diagnosis for non-Latino Black and Latino youths.

With the addition of this symptom to the diagnostic algorithm for agoraphobia, CIDI prevalence estimates improved for non-Latino Black (CIDI prevalence became 4.0% versus K-SADS 2.6%; $\chi^2_2 = 3.6$, $p = 0.06$) and Latino adolescents (CIDI 2.1% versus K-SADS 1.8%; $\chi^2_2 = 0.1$, $p = 0.71$). SP also improved. However, SN decreased from 100% to 47.5% for Latinos and from 93.8% to 69.8% for non-Latino Blacks, leading to a decline in κ for Latinos (0.56 to 0.42) and in AUC (from 0.99 to 0.73 for Latinos and from 0.95 to 0.84 for non-Latino Blacks).

CIDI/K-SADS diagnostic concordance for panic disorder

For diagnoses of panic disorder, SN was significantly lower for non-Latino Black (SN = 31.5%) than non-Latino White and Latino youths (SN = 81.4%, 100.0%, respectively), indicating that the CIDI failed to identify a large proportion of non-Latino Black adolescents with the disorder ($\kappa = 0.34$, AUC = 0.65). Despite differences in SN, CIDI and K-SADS prevalence estimates were similar (CIDI 1.7%, K-SADS 2.2%; $\chi^2_2 = 0.5$, $p = 0.49$). The small number of non-Latino Black adolescents in the clinical reappraisal sample with a K-SADS diagnosis of panic disorder prevented us from investigating CIDI modifications that could improve SN.

CIDI/K-SADS diagnostic concordance for PTSD

PTSD SN was significantly lower for non-Latino Black youths (SN = 32.9%) than others (SN = 56.9% for non-Latino White, 100.0% for Latino youths) resulting in low κ (0.36) and AUC (0.65). The CIDI PTSD prevalence estimate was slightly lower than the K-SADS estimate (3.4% versus 4.6%), but this difference was not statistically significant ($\chi^2_2 = 1.7$, $p = 0.19$). As with panic disorder, the small number of non-Latino Black adolescents

Table 2 Race/ethnic differences in the diagnostic concordance of original and modified CIDI/DSM-IV diagnoses compared to the clinical gold standard diagnoses in the NCS-A clinical reappraisal sample ($n = 347$)

	Prevalence					Concordance[1]					
	K-SADS		CIDI			SN		SP			
	Percentage	SE	Percentage	SE	χ^2	Estimated	SE	Estimated	SE	κ	AUC
Original agoraphobia											
Non-Latino White	1.2	0.6	1.4	0.7	1.2	71.4	15.6	99.4	0.4	0.65	0.85
Latino	1.8	1.1	4.4	2.8	7.2*	100	0.0	97.3	1.8	0.56	0.99
Non-Latino Black	2.6	1.4	6.5	2.9	19.5*	93.8	6.8	95.9	1.9	0.52	0.95
Modified agoraphobia[2]											
Non-Latino White	1.2	0.6	1.4	0.7	1.2	71.4	15.6	99.4	0.4	0.65	0.85
Latino	1.8	1.1	2.1	2.0	0.1	47.5	30.8	98.7	1.2	0.42	0.73
Non-Latino Black	2.6	1.4	4.0	2.4	3.6	69.8	22.8	97.7	1.5	0.53	0.84
Original panic disorder											
Non-Latino White	2.3	0.7	2.7	0.7	3.3	81.4	11.5	99.1	0.3	0.74	0.9
Latino	0.7	0.7	1.2	1.2	1.3	100	0.0	99.5	0.5	0.75	1.0
Non-Latino Black	2.2	1.3	1.7	1.1	0.5	31.5	21.3	98.9	0.7	0.34	0.65
Original PTSD											
Non-Latino White	3.9	1.2	4.0	1.1	0.1	56.9	16.1	98.1	0.8	0.54	0.78
Latino	5.4	2.9	7.4	3.8	5.2*	100	0.0	98.0	1.2	0.84	0.99
Non-Latino Black	4.6	2.1	3.4	1.7	1.7	32.9	19.3	98.1	1.0	0.36	0.65
Original ADHD[3]											
Non-Latino White	6.0	1.4	7.2	1.4	5.8*	64.7	12.2	96.5	1.2	0.56	0.81
Latino	7.4	5.3	2.4	2.3	11.5*	32.4	31.0	100	0.0	0.47	0.66
Non-Latino Black	20.4	7	16.9	7.3	3.5	53.8	20.1	92.6	5.4	0.50	0.73
Modified ADHD[4]											
Non-Latino White	6.0	1.4	7.2	1.4	5.8*	64.7	12.2	96.5	1.2	0.56	0.81
Latino	7.4	5.3	7.4	5.3	.	100	0.0	100	0.0	1.0	1.0
Non-Latino Black	20.4	7.0	21.5	8.5	0.3	53.8	20.1	86.8	8.0	0.40	0.7

[1] SN, sensitivity; SP, specificity; κ, Cohen's kappa; AUC, area under the receiver operating characteristic curve; SE, standard error.

[2] The modified agoraphobia diagnosis required Latino and non-Latino Black youth to meet the original diagnostic criteria and also endorse an item that indicated that the respondent felt upset with him/herself because of the fear/avoidance.

[3] The original ADHD diagnosis was based on parent report only and required parents to report 10+ symptoms of ADHD and indicate that symptoms interfered "a lot" or "extremely" in at least one area of functioning.

[4] The modified ADHD diagnosis was based on parent report only and either required parents to report 10+ symptoms of impairment and indicate symptoms that interfered "a lot" or "extremely" in at least one area of functioning (as in the original diagnosis) or to endorse 6+ symptoms of AD or HD and indicate that these symptoms interfered "a lot" or "extremely" in at least one area of functioning.

*The prevalence estimate based on the CIDI differs significantly from the estimate based on the K-SADS at the 0.05 level, two-sided test.

with PTSD in the clinical reappraisal sample prevented investigating CIDI modifications that could improve CIDI diagnostic algorithms.

CIDI/K-SADS diagnostic concordance for ADHD

The CIDI significantly over-estimated ADHD prevalence among non-Latino Whites (CIDI 7.2% versus K-SADS 6.0%; $\chi^2_2 = 5.8$, $p = 0.016$) and under-estimated prevalence among Latinos (CIDI 2.4% versus K-SADS 7.4%; $\chi^2_2 = 11.5$, $p = 0.001$). For Latinos, SN was consequently quite low (SN = 32.4%) and, although somewhat higher, SN was only moderate for non-Latino Blacks (SN = 53.8%) and non-Latino Whites (SN = 64.7%). Because of the low SN, we attempted to loosen ADHD diagnostic criteria to improve CIDI identification of the disorder among minority youths.

In a previous study (Green *et al.*, 2010), we determined that the best CIDI estimate of K-SADS ADHD in the total NCS-A sample was based on parent report only. Further, parents over-estimated K-SADS ADHD, and we therefore tightened diagnostic criteria, requiring 10 or more ADHD symptoms (in any combination) and indication that symptoms interfered "a lot" or "extremely" in at least one area of functioning. To improve disorder identification for racial/ethnic minorities, we revisited this diagnostic algorithm and examined several modifications to the criteria. We found that the best modification required either meeting the criteria for ADHD described earlier, or for parents to endorse (a) six or more attention deficit (AD) or hyperactivity disorder (HD) symptoms, and (b) indicate that symptoms, at their worst, interfered with home life, friendships, or school/work "a lot" or "extremely." With these modifications, the CIDI was able to identify Latino adolescents with K-SADS ADHD diagnoses perfectly (κ increased from 0.47 to 1.0; AUC increased from 0.66 to 1.0). Estimated prevalence for non-Latino Black adolescents also became closer to the K-SADS value (CIDI 21.5% versus K-SADS 20.4%; $\chi^2_2 = 0.3$, $p = 0.59$); however, SN remained the same and SP dropped (from SP = 92.6% to 86.8%), decreasing the overall κ (from 0.50 to 0.40) and AUC (from 0.73 to 0.70).

Discussion

We found that the sensitivity of the CIDI varies by race/ethnicity for four diagnoses out of the 10 considered here (ADHD, agoraphobia, panic disorder, and PTSD) and that specificity of the CIDI varies by race/ethnicity for one diagnosis (agoraphobia). Although it is not clear why these four rather than the other six disorders were particularly discrepant in their concordance with clinical diagnoses, it is noteworthy that all four have a component of physiological hyperarousal and reactivity, which previous research suggests may be particularly sensitive to the cultural context of racial/ethnic minority youth (Pina and Silverman, 2004; Varela *et al.*, 2007).

In the case of PTSD, which has previously been identified as problematic for the CIDI to diagnose accurately in racial/ethnic minority groups (Alegria *et al.*, 2009), it has been suggested that fully-structured instruments like the CIDI are differentially biased for minorities because they are less able than semi-structured clinical interviews to interpret the cultural context of trauma and trauma-related symptoms (Alarcon, 1995). However, we were unable to investigate this possibility in the NCS-A clinical reappraisal study because the number of minority youths with PTSD was too small to allow modifications of diagnostic criteria to be evaluated with adequate precision. The same was true for panic disorder.

The situation was different for agoraphobia and ADHD, where we were able to make modifications to improve diagnostic criteria. In the case of agoraphobia, we found that tightening the diagnostic algorithm for Latino and non-Latino Black adolescents reduced the inflated CIDI prevalence estimates in these sub-samples. In particular, we added the requirement that racial/ethnic minority youth indicate feeling badly about or disappointed in themselves because of their fear or avoidance. We speculate that this item may tap into perceptions of the cultural acceptability of symptoms. However, this change in diagnostic criteria also substantially decreased CIDI sensitivity to detect agoraphobia among racial/ethnic minority youths, leading to a decrease in individual-level concordance between diagnoses based on the CIDI and clinical diagnoses. The end result was that we did not implement any changes in the CIDI diagnosis of agoraphobia. A general absence of data on the accuracy of assessments of agoraphobia for racial/ethnic minority youths (Lewis-Fernandez *et al.*, 2010) suggests that this is an important area for future research.

Modifications to the diagnostic algorithm for ADHD were more successful. In an earlier paper, we modified CIDI diagnostic criteria in the total sample to account for parent over-reporting of ADHD symptoms (Green *et al.*, 2010). Here, the finding that diagnostic criteria needed to be loosened to improve concordance for Latino and non-Latino Black parents suggests that parents of racial/ethnic minority youth are *less* likely to over-endorse ADHD symptoms. This finding is consistent with prior research indicating that, given comparable levels of hyperactivity, parents of racial/ethnic minority youth less often endorse symptoms than parents of non-Latino White children (Hillemeier *et al.*, 2007). By modifying CIDI criteria to allow more flexibility in parent symptom report, we were able to perfectly identify Latino youths with ADHD and improve the accuracy of prevalence estimates for non-Latino Black adolescents, although these benefits occurred at the expense of a slight decrease in the specificity of diagnoses for non-Latino Black adolescents.

Several limitations in the design of the NCS-A and of the clinical reappraisal study may have influenced our results. First, the NCS-A sample excluded school dropouts, the homeless, and non-English speakers; all of which are groups where racial/ethnic minority youth are disproportionately represented. Second, there were high rates of individual non-response and school non-response, although analysis of effects of non-response in the NCS-A found little evidence of bias (Kessler *et al.*, 2009a). No data were collected on the race/ethnicity of non-respondents, so we do not know whether non-response rates differ across race/ethnic groups. Third, the K-SADS was administered by telephone (in contrast to the face-to-face CIDI administration). There is strong evidence that telephone interviews are a valid method for clinical assessment (Aneshensel *et al.*, 1982; Rohde *et al.*, 1997; Sobin *et al.*, 1993) and, in the case of the NCS-A, it provided the only feasible method for this type of large-scale data collection. However, the comparison of in-person CIDI with telephone K-SADS interviews likely make concordance estimates more conservative. Fourth, the design of the clinical reappraisal study, which provided K-SADS clinical interviewers with information about responses to diagnostic stem questions in the CIDI interview, may have influenced racial/ethnic differences. Fifth, all the analyses reported here were based on the untested assumption that diagnoses based on the K-SADS are equally valid for minority and non-minority youth. Sixth, the clinical reappraisal study was not specifically designed to study CIDI validity by race/ethnicity. As a result, the number of racial/ethnic minority youths in the reappraisal study sample was smaller than we would have desired, limiting statistical power to study modifications to diagnostic criteria for the least common disorders.

These findings underscore the importance of testing measurement validity by race and ethnicity. They suggest that, although CIDI diagnostic algorithms appear to function similarly across racial/ethnic groups for some disorders, there are four for which CIDI classifications have lower validity for racial/ethnic minority youth. In these cases, applying a universal framework to assessment may mask racial/ethnic differences, resulting in misleading prevalence estimates and disorder misclassification (Alegria and McGuire, 2003). In some cases we were able to adjust disorder-specific CIDI diagnostic algorithms across racial/ethnic groups to improve estimated prevalence for racial/ethnic minorities, but these benefits were offset by a diminished ability to classify individuals with disorders. In deciding whether to use these modified diagnostic algorithms, researchers should be guided by the specific purposes of diagnostic assessment and, in particular, whether they are emphasizing prevalence estimation, or individual classification.

We recommend that future studies of diagnostic validity similarly attend to the potential for differential validity across racial/ethnic groups. Results have implications for interpreting subgroup comparisons and, further, may suggest qualitative distinctions between groups in the phenomenology of disorders (Alegria *et al.*, 2009).

Acknowledgements

This study was supported by grants from the National Institute of Health National Center on Minority Health and Health Disparities (NIH; RC1MD004588, Alegria) and the NIH National Institute of Mental Health (K01-MH085710, Green). The National Comorbidity Survey Replication Adolescent Supplement (NCS-A) is supported by the National Institute of Mental Health (NIMH; U01-MH60220 and R01-MH66627) with supplemental support from the National Institute on Drug Abuse (NIDA), the Substance Abuse and Mental Health Services Administration (SAMHSA), the Robert Wood Johnson Foundation (RWJF; Grant 044780), and the John W. Alden Trust. The views and opinions expressed in this report are those of the authors and should not be construed to represent the views of any of the sponsoring organizations, agencies, or US Government. A complete list of NCS-A publications can be found at http://www.hcp.med.harvard.edu/ncs. Send correspondence to ncs@hcp.med.harvard.edu. The NCS-A is carried out in conjunction with the World Health Organization World Mental Health (WMH) Survey Initiative. We thank the staff of the WMH Data Collection and Data Analysis Coordination Centers for assistance with instrumentation, fieldwork, and consultation on data analysis. The WMH Data Coordination Centers have received support from NIMH (R01-MH070884, R13-MH066849, R01-MH069864, R01-MH077883), NIDA (R01-DA016558), the Fogarty International Center of the National Institutes of Health (FIRCA R03-TW006481), the John D. and Catherine T. MacArthur Foundation, the Pfizer Foundation, and the Pan American Health Organization. The WMH Data Coordination Centers have also received unrestricted educational grants from Astra Zeneca, BristolMyersSquibb, Eli Lilly and Company, GlaxoSmithKline, Ortho-McNeil, Pfizer, Sanofi-Aventis, and Wyeth. A complete list of WMH publications can be found at http://www.hcp.med.harvard.edu/wmh/.

Declaration of interest statement

Dr Kessler has been a consultant for AstraZeneca, Analysis Group, Bristol-Myers Squibb, Cerner-Galt Associates, Eli Lilly & Company, GlaxoSmithKline Inc., HealthCore Inc., Health Dialog, Integrated Benefits Institute, John Snow Inc., Kaiser Permanente, Matria Inc., Mensante, Merck & Co, Inc., Ortho-McNeil Janssen Scientific Affairs, Pfizer Inc., Primary Care Network, Research Triangle Institute, Sanofi-Aventis Groupe, Shire US zInc., SRA International, Inc., Takeda Global Research & Development, Transcept Pharmaceuticals Inc., and Wyeth-Ayerst; has served on advisory boards for Appliance Computing II, Eli Lilly & Company, Mindsite, Ortho-McNeil Janssen Scientific Affairs, Plus One Health Management and Wyeth-Ayerst; and has had research support for his epidemiological studies from Analysis Group Inc., Bristol-Myers Squibb, Eli Lilly & Company, EPI-Q, GlaxoSmithKline, Johnson & Johnson Pharmaceuticals, Ortho-McNeil Janssen Scientific Affairs., Pfizer Inc., Sanofi-Aventis Groupe, and Shire US Inc.

References

Alarcon R.D. (1995) Culture and psychiatric diagnosis. Impact on DSM-IV and ICD-10. *Psychiatric Clinics of North America*, **18**(3), 449–465.

Alegria M., McGuire T. (2003) Rethinking a universal framework in the psychiatric symptom–disorder relationship. *Journal of Health and Social Behavior*, **44**(3), 257–274.

Alegria M., Shrout P.E., Torres M., Lewis-Fernandez R., Abelson J.M., Powell M., Interian A., Lin J., Laderman M., Canino G. (2009) Lessons learned from the clinical reappraisal study of the Composite International Diagnostic Interview with Latinos. *International Journal of Methods in Psychiatric Research*, **18**(2), 84–95, DOI: 10.1002/mpr.280

Alegria M., Carson N.J., Goncalves M., Keefe K. (2011) Disparities in treatment for substance use disorders and co-occurring disorders for ethnic/racial minority youth. *Journal of the American Academy of Child and Adolescent Psychiatry*, **50**(1), 22–31, DOI: 10.1016/j.jaac.2010.10.005

Anderson E.R., Mayes L.C. (2010) Race/ethnicity and internalizing disorders in youth: a review. *Clinical Psychology Review*, **30**(3), 338–348, DOI: 10.1016/j.cpr.2009.12.008

Aneshensel C.S., Frerichs R.R., Clark V.A., Yokopenic P.A. (1982) Measuring depression in the community: a comparison of telephone and personal interviews. *Public Opinion Quarterly*, **46**(1): 110–121, DOI: 10.1086/268703

Angold A., Erkanli A., Farmer E.M., Fairbank J.A., Burns B.J., Keeler G., Costello E.J. (2002) Psychiatric disorder, impairment, and service use in rural African American and white youth. *Archives of General Psychiatry*, **59**(10), 893–901, DOI: 10.1001/archpsyc.59.10.893

Braaten E.B., Biederman J., DiMauro A., Mick E., Monuteaux M.C., Muehl K., Faraone S.V. (2001) Methodological complexities in the diagnosis of major depression in youth: an analysis of mother and youth self-reports. *Journal of Child and Adolescent Psychopharmacology*, **11**(4): 395–407, DOI: 10.1089/104454601317261573

Breslau J., Javaras K.N., Blacker D., Murphy J.M., Normand S.L. (2008) Differential item functioning between ethnic groups in the epidemiological assessment of depression. *The Journal of Nervous and Mental Disease*, **196**(4), 297–306, DOI: 10.1097/NMD.0b013e31816a490e

Cohen P., Cohen J., Kasen S., Velez C.N., Hartmark C., Johnson J., Rojas M., Brook J., Streuning E.L. (1993) An epidemiological study of disorders in late childhood and adolescence — I. Age- and gender-specific prevalence. *Journal of Child Psychology and Psychiatry*, **34**(6), 851–867, DOI: 10.1111/j.1469-7610.1993.tb01094.x

Cohen J. (1960) A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, **20**(1), 37–46, DOI: 10.1177/001316446002000104

Costello E.J., Mustillo S., Erkanli A., Keeler G., Angold A. (2003) Prevalence and development of psychiatric disorders in childhood and adolescence. *Archives of General Psychiatry*, **60**(8), 837–844, DOI: 10.1001/archpsyc.60.8.837

Cuffe S.P., Waller J.L., Cuccaro M.L., Pumariega A. J., Garrison C.Z. (1995) Race and gender differences in the treatment of psychiatric disorders in young adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, **34**(11), 1536–1543, DOI: 10.1097/00004583-199511000-00021

Eisen S.V., Gerena M., Ranganathan G., Esch D., Idiculla T. (2006) Reliability and validity of the BASIS-24 Mental Health Survey for Whites, African-Americans, and Latinos. *Journal of Behavioral Health Services and Research*, **33**(3), 304–323, DOI: 10.1007/s11414-006-9025-3

Green J.G., Avenevoli S., Finkelman M., Gruber M.J., Kessler R.C., Merikangas K.R., Sampson N.A., Zaslavsky A.M. (2010) Attention deficit hyperactivity disorder: concordance of the adolescent version of the Composite International Diagnostic Interview Version 3.0 (CIDI) with the K-SADS in the US National Comorbidity Survey Replication Adolescent (NCS-A) supplement. *International Journal of Methods in Psychiatric Research*, **19**(1), 34–49, DOI: 10.1002/mpr.303

Grills A.E., Ollendick T.H. (2002) Issues in parent–child agreement: the case of structured diagnostic interviews. *Clinical Child and Family Psychology Review*, **5**(1), 57–83, DOI: 10.1023/A:1014573708569

Hanley J.A., McNeil B.J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143**(1), 29–36.

Hillemeier M.M., Foster E.M., Heinrichs B., Heier B., on behalf of the Conduct Problems Prevention Research Group. (2007) Racial differences in parental reports of attention-deficit/hyperactivity disorder behaviors. *Journal of Developmental and Behavioral Pediatrics*, **28**(5), 353–361, DOI: 10.1097/DBP.0b013e31811ff8b8

Johnson M.R., Hartzema A.G., Mills T.L., De Leon J. M., Yang M., Frueh C., Santos A. (2007) Ethnic differences in the reliability and validity of a Panic Disorder Screen. *Ethnicity and Health*, **12**(3), 283–296, DOI: 10.1080/13557850701235069

Johnston C., Murray C. (2003) Incremental validity in the psychological assessment of children and adolescents. *Psychological Assessment*, **15**(4), 496–507, DOI: 10.1037/1040-3590.15.4.496

Kataoka S.H., Zhang L., Wells K.B. (2002) Unmet need for mental health care among U.S. children: variation by ethnicity and insurance status. *The American Journal of Psychiatry*, **159**(9), 1548–1555.

Kessler R.C., Merikangas K.R. (2004) The National Comorbidity Survey Replication (NCS-R): background and aims. *International Journal of Methods in Psychiatric Research*, **13**(2), 60–68, DOI: 10.1002/mpr.166

Kessler R.C., Ustun T.B. (2004) The World Mental Health (WMH) Survey Initiative Version of the World Health Organization (WHO) Composite International Diagnostic Interview (CIDI). *International Journal of Methods in Psychiatric Research*, **13**(2), 93–121, DOI: 10.1002/mpr.168

Kessler R.C., Wittchen H.-U., Abelson J.M., McGonagle K.A., Schwarz N., Kendler K.S., Knäuper B., Zhao S. (1998) Methodological studies of the Composite International Diagnostic Interview (CIDI) in the US National Comorbidity Survey. *International Journal of Methods in Psychiatric Research*, **7**(1), 33–55, DOI: 10.1002/mpr.33

Kessler R.C., Avenevoli S., Costello E.J., Green J.G., Gruber M.J., Heeringa S., Merikangas K.R., Pennell B.E., Sampson N.A., Zaslavsky A.M. (2009a) Design and field procedures in the US National Comorbidity Survey Replication Adolescent Supplement (NCS-A). *International Journal of Methods in Psychiatric Research*, **18**(2), 69–83, DOI: 10.1002/mpr.279

Kessler R.C., Avenevoli S., Costello E.J., Green J.G., Gruber M.J., Heeringa S., Merikangas K.R., Pennell B.E., Sampson N.A., Zaslavsky A.M. (2009b) National comorbidity survey replication adolescent supplement (NCS-A): II. Overview and design. *Journal of the American Academy of Child and Adolescent Psychiatry*, **48**(4), 380–385, DOI: 10.1097/CHI.0b013e3181999705

Kessler R.C., Avenevoli S., Green J., Gruber M.J., Guyer M., He Y., Jin R., Kaufman J., Sampson N.A., Zaslavsky A.M. (2009c) National comorbidity survey replication adolescent supplement (NCS-A): III. Concordance of DSM-IV/CIDI diagnoses with clinical reassessments. *Journal of the American Academy of Child and*

Adolescent Psychiatry, **48**(4), 386–399, DOI: 10.1097/CHI.0b013e31819a1cbc

Kessler R.C., Avenevoli S., Costello E.J., Georgiades K., Green J.G., Gruber M.J., He J.P., Koretz D., McLaughlin K.A., Petukhova M., Sampson N.A., Zaslavsky A.M., Merikangas K.R. (2012) Prevalence, persistence, and sociodemographic correlates of DSM-IV disorders in the National Comorbidity Survey Replication Adolescent Supplement. *Archives of General Psychiatry*, **69**(4), 372–380, DOI: 10.1001/archgenpsychiatry.2011.160

Kim-Cohen J., Caspi A., Moffitt T.E., Harrington H., Milne B.J., Poulton R. (2003) Prior juvenile diagnoses in adults with mental disorder: developmental follow-back of a prospective-longitudinal cohort. *Archives of General Psychiatry*, **60**(7), 709–717, DOI: 10.1001/archpsyc.60.7.709

Komiti A.A., Jackson H.J., Judd F.K., Cockram A.M., Kyrios M., Yeatman R., Murray G., Hordern C., Wainwright K., Allen N., Singh B. (2001) A comparison of the Composite International Diagnostic Interview (CIDI-Auto) with clinical assessment in diagnosing mood and anxiety disorders. *The Australian and New Zealand Journal of Psychiatry*, **35**(2), 224–230, DOI: 10.1046/j.1440-1614.2001.00868.x

Lewis-Fernández R., Hinton D.E., Laria A.J., Patterson E.H., Hofmann S.G., Craske M.G., Stein D.J., Asnaani A., Liao B. (2010) Culture and the anxiety disorders: recommendations for DSM-V. *Depression and Anxiety*, **27**(2), 212–229, DOI: 10.1002/da.20647

McLaughlin K.A., Hilt L.M., Nolen-Hoeksema S. (2007) Racial/ethnic differences in internalizing and externalizing symptoms in adolescents. *Journal of Abnormal Child Psychology*, **35**(5), 801–816, DOI: 10.1007/s10802-007-9128-1

Merikangas K., Avenevoli S., Costello J., Koretz D., Kessler R.C. (2009) National comorbidity survey replication adolescent supplement (NCS-A): I. Background and measures. *Journal of the American Academy of Child and Adolescent Psychiatry*, **48**(4), 367–369, DOI: 10.1097/CHI.0b013e31819996f1

Minsky S., Petti T., Gara M., Vega W., Lu W., Kiely G. (2006) Ethnicity and clinical psychiatric diagnosis in childhood. *Administration and Policy in Mental Health*, **33**(5), 558–567, DOI: 10.1007/s10488-006-0069-8

Moffitt T.E., Caspi A., Taylor A., Kokaua J., Milne B.J., Polanczyk G., Poulton R. (2010) How common are common mental disorders? Evidence that lifetime prevalence rates are doubled by prospective versus retrospective ascertainment. *Psychological Medicine*, **40**(6), 899–909, DOI: 10.1017/S0033291709991036

Pina A.A., Silverman W.K. (2004) Clinical phenomenology, somatic symptoms, and distress in Hispanic/Latino and European American youths with anxiety disorders. *Journal of Clinical Child and Adolescent Psychology*, **33**(2), 227–236, DOI: 10.1207/s15374424jccp3302_3

Roberts R.E., Roberts C.R., Chen Y.R. (1997) Ethnocultural differences in prevalence of adolescent depression. *American Journal of Community Psychology*, **25**(1), 95–110, DOI: 10.1023/A:1024649925737

Rohde P., Lewinsohn P.M., Seeley J.R. (1997) Comparability of telephone and face-to-face interviews in assessing axis I and II disorders. *The American Journal of Psychiatry*, **154**(11), 1593–1598.

Sobin C., Weissman M.M., Goldstein R.B. (1993) Diagnostic interviewing for family studies: comparing telephone and face-to-face methods for the diagnosis of lifetime psychiatric disorders. *Psychiatric Genetics*, **3**(4), 227–233.

Varela R.E., Weems C.F., Berman S.L., Hensley L., de Bernal M.C.R. (2007) Internalizing symptoms in Latinos: the role of anxiety sensitivity. *Journal of Youth and Adolescence*, **36**(4), 429–440, DOI: 10.1007/s10964-007-9168-4