Received: 12 February 2020 | Accepted: 12 April 2020

DOI: 10.1002/mpr.1828

SPECIAL ISSUE ARTICLE

WILEY

Clinical reappraisal of the Composite International Diagnostic Interview Version 3.0 in the Saudi National Mental Health Survey

Ronald C. Kessler[1] | Majid Al-Desouki[2] | Andrew J. King[1] | Nancy A. Sampson[1] | Abdullah S. Al-Subaie[3,4] | Abdulhameed Al-Habeeb[5] | Lisa Bilal[4,6,7] | Mona K. Shahab[8,9] | Maggie Aradati[4,6,7] | Yasmin A. Altwaijri[4,6,7]

[1]Department of Health Care Policy, Harvard Medical School, Boston, Massachusetts

[2]Psychiatry Unit, King Saud University Medical City, Riyadh, Saudi Arabia

[3]Edrak Medical Center, Riyadh, Saudi Arabia

[4]SABIC Psychological Health Research & Applications Chair (SPHRAC), College of Medicine, King Saud University, Riyadh, Saudi Arabia

[5]National Center for Mental Health Promotion, Ministry of Health, Riyadh, Saudi Arabia

[6]King Salman Center for Disability Research, Riyadh, Saudi Arabia

[7]Biostatistics, Epidemiology and Scientific Computing Department, King Faisal Specialist Hospital and Research Centre, Riyadh, Saudi Arabia

[8]Clinical Epidemiology, Leiden University Medical Center; Faculty of Social and Behavioural Sciences, Clinical Psychology Department, Leiden University, Leiden, The Netherlands

[9]i-psy interculturele psychiatrie, Parnassia Groep, The Hague, The Netherlands

Correspondence
Ronald C. Kessler, Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA 02115.
Email: kessler@hcp.med.harvard.edu

Funding information
Abraaj Capital; King Abdulaziz City for Science and Technology (KACST); King Faisal Specialist Hospital and Research Center; King Saud University; Ministry of Economy and Planning, General Authority for Statistics; Ministry of Health (Saudi Arabia); Saudi Basic Industries Corporation (SABIC)

Abstract

Objectives: The DSM-IV diagnoses generated by the fully structured lay-administered Composite International Diagnostic Interview Version 3.0 (CIDI 3.0) in the Saudi National Mental Health Survey (SNMHS) were compared to diagnoses based on blinded clinical reappraisal interviews.

Methods: Telephone follow-up interviews were administered using the clinician-administered non-patient edition of the Structured Clinical Interview for DSM-IV (SCID) in separate sub-samples of SNMHS respondents who screened positive for four disorders that are of special importance in Arab countries: obsessive–compulsive disorder, separation anxiety disorder, social phobia, and major depressive episode.

Results: Initial diagnoses based on the CIDI were found to have higher prevalence than those based on the SCID for all four disorders. For reasons having to do with respondent denial of symptoms in the SCID reported in the CIDI, we interpreted these differences as due more to under-diagnoses in the SCID than over-diagnoses in the CIDI. Nonetheless, CIDI diagnostic thresholds for three of the four disorders were increased to make sure prevalence estimates based on the CIDI were conservative. The procedures used to implement these recalibrations are described in this paper.

Conclusions: The CIDI interviews used in the SNMHS generated valid but conservative diagnoses of common mental disorders in the Saudi population.

KEYWORDS

clinical reappraisal, Composite International Diagnostic Interview (CIDI), Saudi National Mental Health Survey (SNMHS), Structured Clinical Interview for DSM-IV (SCID), WHO World Mental Health (WMH) Survey Initiative

1 | INTRODUCTION

This paper presents the results of a series of disorder-specific clinical reappraisal studies carried out in conjunction with the Saudi National Mental Health Survey (SNMHS). As detailed elsewhere in this issue (Al-Subaie, Al-Habeeb, & Altwaijri, In press), SNMHS is a community epidemiological survey of the prevalence and correlates of common mental disorders carried out in a nationally representative household sample in the Kingdom of Saudi Arabia (KSA). SNMHS is part of the World Health Organization (WHO) World Mental Health (WMH) Survey Initiative (Alonso, Chatterji, & He, 2013; Kessler & Üstün, 2008; Scott, de Jonge, Stein, & Kessler, 2018). The diagnostic interview used in the SNMHS, the WHO Composite International Diagnostic Interview Version 3.0 (CIDI 3.0; Kessler & Üstün, 2004), has been validated previously in conjunction with earlier WMH surveys (Ghimire, Chardoul, Kessler, Axinn, & Adhikari, 2013; Haro et al., 2006; Kimerling et al., 2014; Lu, Huang, Liu, & Cao, 2015; Montoya Gonzalez et al., 2016). Results of these studies show that diagnoses based on the CIDI have generally good concordance with independent diagnoses based on clinical reappraisal interviews in which the clinicians were blinded to CIDI diagnoses. The clinical reappraisal interview used in these validation studies was the Axis I research version, non-patient edition of the Structured Clinical Interview for DSM-IV (SCID; First, Spitzer, Gibbon, & Williams, 2002). We carried out a similar series of disorder-specific clinical reappraisal studies of particular DSM-IV/CIDI diagnoses in the SNMHS despite the earlier evidence of CIDI validity based on considerations discussed below involving the special relevance of certain diagnoses to Arab countries.

The version of CIDI used in the WMH surveys, CIDI 3.0, was developed to optimize the validity of the CIDI for use in community surveys (Kessler & Üstün, 2004). These developments were based on insights gained from clinical reappraisal studies carried out with earlier versions of the CIDI (Kessler et al., 1998; Wittchen, 1994; Wittchen, Kessler, Zhao, & Abelson, 1995; Wittchen, Zhao, Abelson, Abelson, & Kessler, 1996). All these diagnoses use a stem-branch structure that focuses initially on lifetime prevalence, which requires survey respondents who do not currently meet criteria for the disorder to engage in active memory search in which they recall past emotional experiences. The revisions implemented in CIDI 3.0 were designed to use knowledge from cognitive psychology about the most effective memory search strategies to motivate and guide respondents in engaging in effective active memory search, with the goal of maximizing sensitivity (i.e., the proportion of true cases that endorsed the diagnostic stem questions). Subsequent disorder-specific questions were then used to distinguish true positives from false positives. Recency questions administered to respondents designated to be true positives, finally, were used to estimate current (in the 30 days before interview) and recent (in the 12 months before interview) prevalence and to obtain basic information about course of illness (e.g., age-of-onset, age-of-recency among respondents who did not have an episode in the 12 months before interview, and number of years in episode between age-of-onset and age-of-recency).

Past CIDI clinical reappraisal studies showed that a major reason for the discrepancies found between lifetime diagnoses based on the CIDI and the SCID involved instances in which CIDI diagnoses of lifetime prevalence among remitted cases were classified in the SCID as lifetime non-cases (Kessler & Üstün, 2004). Debriefing suggested two reasons for this discrepancy: that survey respondents become aware of the stem-branch structure of the CIDI in the course of the interview and consequently recognize that they can shorten the follow-up interview by denying the diagnostic stem questions in the clinical reappraisal follow-up interview that they endorsed in the earlier CIDI interview; and that clinical interviewers are much more used to assessing patients who present to them with current problems than probing for past lifetime occurrences of remitted disorders (Edelbrock, Crnic, & Bohnert, 1999; Kessler & Üstün, 2004).

As discussed in more detail elsewhere (Haro et al., 2008; Kessler & Üstün, 2004), we addressed the first of these problems in CIDI 3.0 by developing a lifetime review section early in the interview in which we had respondents engage in active memory search and respond to lifetime diagnostic stem questions for all diagnoses before probing any diagnoses. We addressed the second problem in the WMH clinical reappraisal studies by partially unblinding clinical interviewers to respondent endorsement of diagnostic stem questions. That is, we focused on the subset of WMH survey respondents who reported that they sometime in their life experienced one or more of the core criteria for the disorder under investigation (e.g., in the case of major depressive episode, having a time lasting two or more weeks when the respondent experienced dysphoria or anhedonia most of the day nearly every day) by telling the clinical interviewer that the respondent reported this experience. The clinical interviewer then began the assessment of that disorder by informing the respondent that they now wanted to ask some additional questions about one of the things the respondent reported in their earlier interview, repeating the diagnostic stem question(s) that the respondent endorsed in the earlier CIDI interview and probing that positive response.

A concern might be raised that this partial unblinding would bias results, but we adjusted for this possibility by enriching the clinical reappraisal sample for SNMHS respondents who screened positive on the diagnoses but failed to meet DSM-IV/CIDI criteria. Importantly, as detailed below, substantial proportions of respondents in the SNMHS (as in community epidemiological surveys elsewhere in the world) endorsed diagnostic stem questions but failed to meet full criteria for disorders. The real challenge for clinical interviewers given this fact is to distinguish screened positives who meet full diagnostic criteria from those that do not. That was the focus of our clinical reappraisal interviews. We recognize that this leaves open the possibility of underdiagnosis due to the CIDI screens missing some people who meet diagnostic criteria. As noted above, the CIDI screens were developed to address this possibility by maximizing sensitivity (i.e., maximizing the proportion of true cases that endorse a diagnostic stem question). Nonetheless, the prevalence estimates in the SNMHS should be considered conservative because of the possibility of residual false negatives.

2 | METHODS

2.1 | The main sample

As described in more detail elsewhere in this issue (Mneimneh, Heeringa, Lin, Altwaijri, & Nishimura, In press), the SNMHS is a nationally representative household survey of Saudi citizens ages 15–65 excluding the two out of 13 administrative areas in KSA (Jazan & Najran) that were having politcal conflict at the time of the survey. The survey was based on a stratified multistage cluster area probability sample of 4,302 households in the remaining 11 administrative areas. The estimated response rate using the American Association of Public Opinion Research RR2 definition (American Association for Public Opinion Research, 2016) was 61%, based on a household screening rate of 84% and a conditional interview response rate of 73%. A total of 4,004 interviews were completed. We attempted to interview one randomly selected male and one randomly selected female in households that contained both males and females in the age range 15–65. Only one respondent was randomly selected in households in which eligible residents were either all male or all female.

All interviews were carried out face-to-face by trained lay interviewers. The interview schedule and all training materials were translated using a standardized WHO translation and back-translation protocol (Harkness et al., 2008; Shahab et al., 2019). Interviewer training procedures and field quality control procedures were used consistent with those in other WMH surveys (Heeringa et al., 2008; Pennell et al., 2008). Interviewers followed a strict fieldwork protocol to guarantee data quality. Details of these quality assurance and quality control procedures are described elsewhere (Hyder et al., 2017). Study procedures conformed to the international standards set by the Declaration of Helsinki. Written informed consent was obtained from respondents prior to beginning each interview. These consent procedures were approved by the Institutional Review Board at the King Faisal Hospital and Research Center.

2.2 | The clinical reappraisal samples

The clinical reappraisal samples were selected after the SNMHS was completed and the data processed to generate DSM-IV/CIDI diagnoses. We had originally planned not to carry out clinical reappraisal interviews in the SNMHS sample based on prior clinical appraisal studies consistently finding good concordance between diagnoses based on the CIDI and diagnoses based on blinded SCID clinical reappraisal interviews. However, SNMHS prevalence estimates for 3 diagnoses were considerably higher than expected, leading us to have concerns about the validity of the CIDI for assessing them in KSA. The three were obsessive–compulsive disorder, adult separation anxiety disorder, and social phobia. We also carried out a clinical reappraisal of CIDI diagnoses of major depressive episode given the importance of this disorder in recent discussions of the global burden of disease (GBD 2016 Disease and Injury Incidence and Prevalence Collaborators, 2017).

The time between the CIDI interviews and the SCID clinical reappraisal interviews was quite long both because of the protracted nature of the SNMHS (Altwaijri et al., In press; Mneimneh et al., In press) and because the decision to carry out disorder-specific clinical reappraisal studies was not made until after we had completed data cleaning, coding, weighting and had made initial estimates of prevalence. (Table 1) This complication led us to select a separate clinical reappraisal sample for each of 4 diagnoses and to administer a version of the SCID in each of these samples that focused only on the single diagnosis under investigation. This allowed us to keep the length of the SCID interviews relatively short, which was important because the SCID interviews were administered by telephone.

Telephone administration is now widely accepted in clinical reappraisal studies based on evidence of comparable validity to in-person administration (Kendler, Neale, Kessler, Heath, & Eaves, 1992; Rohde, Lewinsohn, & Seeley, 1997; Sobin et al., 1993). However, the quality of telephone interviews decreases as the length of interview goes beyond 20–30 min (Groves & Kahn, 1979). This made it important to administer disorder-specific SCID reappraisal interviews, which could be kept within this time range. A great advantage of telephone administration compared to face-to-face administration, in comparison, is that a centralized and closely supervised clinical interview staff can carry out the interviews throughout the entire sample area without the geographic restriction that is typically required for face-to-face clinical assessment. Although SNMHS respondents were interviewed initially face-to-face, information was obtained about their phone numbers for future recontacts to classify initial survey respondents.

2.3 | The clinical reappraisal study design

The clinical reappraisal studies were designed to determine the extent to which the diagnostic classifications made on the basis of the CIDI would have been different if the surveys had been carried out entirely by carefully trained clinical interviewers using the SCID rather than by trained lay interviewers using the CIDI. As the entry questions (i.e., the diagnostic stem questions) in the CIDI and SCID are very

TABLE 1 Distributions of time (in months) between carrying out the Saudi National Mental Health Survey and conducting SCID clinical reappraisal interviews

	Median	[IQR]	Min	Max
Obsessive–compulsive disorder	47	[38–54]	24	69
Adult separation anxiety disorder	45	[31–50]	25	69
Social phobia	46	[29–57]	25	71
Major depressive episode	52	[39–60]	29	74

Abbreviations: IQR, inter-quartile range; SCID, Structured Clinical Interview for DSM-IV.

similar, the distinction between the two types of interview hinges on two things: first, CIDI-SCID differences: differences in the ability to elicit endorsement of diagnostic stem questions based on CIDI yes-no questions versus the more flexible open-ended probing in the SCID; and, second, differences in symptom assessments among respondents who endorse diagnostic stem questions based on fully-structured CIDI questions versus the more conversational probes in the SCID. We felt that SCID procedures were almost certainly superior to CIDI procedures in eliciting clear information about symptom characteristics. It was less clear, though, whether the SCID was superior to the CIDI in eliciting endorsement of diagnostic stem questions or symptom reports, as our previous work carrying out CIDI-SCID comparisons documented a number of cases in which respondents were more comfortable admitting embarrassing feelings and behaviors to lay interviewers than to clinical interviewers (Kessler et al., 1998).

A major impediment to making accurate CIDI-SCID comparisons of the sort described in the last paragraph is that respondents are inconsistent in their reports over time. Indeed, our own previous experience and that of other researchers shows that respondents in community surveys tend to report less and less as they are interviewed more and more (Bromet, Dunn, Connell, Dew, & Schulberg, 1986; Edelbrock et al., 1999; Kessler & Üstün, 2004). Part of this pattern is a tendency for respondents to endorse a smaller number of diagnostic stem questions in follow-up interviews than in initial interviews (Kessler et al., 1998), leading to the biased perception that initial structured interviews over-estimate prevalence compared to second clinical interviews. Based on this observation, we modified the conventional blinded clinical re-interview design in two important ways in the SNMHS. First, as noted in the introduction, we unblinded the clinical interviewers to whether the respondents endorsed diagnostic stem questions in the CIDI, but not to the final CIDI diagnoses. Second, we encouraged respondents to endorse diagnostic stem questions in the clinical reappraisal interviews by reminding respondents who endorsed the CIDI stem question in their initial interview of this fact. We noted in the introduction that even though this partial unblinding of interviewers might be seen as introducing a bias, that turns out not to be the case because, as shown below, the majority of community survey respondents who endorse CIDI stem questions do not go on to meet full CIDI criteria for the associated disorder.

The stem question reminder process had a substantial effect on the completeness of respondent reports in clinical re-interviews. Respondents were told at the beginning of their clinical re-interview that they will be asked some of the same questions as in their earlier interview. They were also told that this was being done to test the interview and not to test their memory, so they should answer without trying to remember what they said to the earlier interviewer. Respondents were then taken through the clinical interview in the usual fashion, with the exception that the sections of the clinical re-interview in which they endorsed a diagnostic stem question in the CIDI were started with the introduction: *During the first interview, you said (FILL). Has that happened in the past 12 months?* The FILL was a presentation of the lifetime diagnostic stem question or questions endorsed in the initial CIDI interview. Re-interview respondents could still deny that they reported a diagnostic stem question in the initial interview, which, as detailed below, did happen. In cases where the respondent had not endorsed the CIDI stem question in the original interview, the SCID probing for a diagnostic stem endorsement was carried out in the conventional fashion to discover false negative responses in the CIDI, but the latter were extremely rare.

2.4 | Clinical interviewer training and supervision

Clinical interviewers were five clinical psychology interns who were trained and closely supervised throughout the fieldwork period by a senior psychiatrist who was a certified SCID trainer (Dr. Majid Al-Desouki). An expanded version of the training program created by the developers of the SCID (Gibbon, McDonald-Scott, & Endicott, 1981) was used for interviewer training. This program included (a) use of the standard SCID training tapes and manuals, which required 30 hr of self-study, followed by (b) 40 hr of in-person group training. In order to maximize the reliability of clinical ratings, the interviewers were supplied with textual extracts from the DSM-IV manual that described specific study diagnostic criteria. Quality control monitoring included clinical supervisor review of all hard copy completed SCID interviews, re-contact of respondents whenever the clinical supervisor felt that more information was needed to make a rating, and weekly interviewer-supervisor meetings to prevent drift.

2.5 | Analysis methods

As detailed below, the main focus of our recalibration work was on increasing the severity of the CIDI diagnostic thresholds to reduce

over-estimation of cases. This focus was due to our initial finding that prevalence estimates based on the CIDI for the disorders considered in the clinical reappraisal studies were generally higher than prevalence estimates based on the SCID but that the majority of SCID cases were captured by the CIDI. Bias in estimating prevalence was evaluated using McNemar χ^2 tests to evaluate the statistical significance of differences in the proportions of respondents who were false positives versus false negatives.

Individual-level CIDI-SCID diagnostic concordance was next evaluated by calculating area under the receiver operator characteristic curve (AUC; Hanley & McNeil, 1982). Although Cohen's κ(Cohen, 1960) is a much more widely used measure of concordance in validity studies of psychiatric disorders, κ is dependent on prevalence and consequently is often low in situations where there appears to be high agreement between low-prevalence measures (Byrt, Bishop, & Carlin, 1993; Cook, 1998; Kraemer et al., 2003). An important implication is that κ varies across populations that differ in prevalence even when the populations do not differ in sensitivity (SN; the percent of true cases correctly classified by the CIDI) or specificity (SP; the percent of true non-cases correctly classified). As sensitivity and specificity are the fundamental parameters in defining concordance, comparison of κ across populations that differ in prevalence cannot be used to evaluate the cross-population performance of a test.

Critics of κ prefer to assess concordance with measures that are a function of SN and SP. The odds-ratio (OR) meets this requirement, as OR is equal to $[SN \times SP]/[(1-SN) \times (1-SP)]$ (Agresti, 1996). However, the upper end of the OR is unbounded, making it difficult to use the OR to evaluate the extent to which CIDI diagnoses are consistent with clinical diagnoses. Yules Q has been proposed as an alternative measure to resolve this problem (Spitznagel & Helzer, 1985), as Q is a bounded transformation of OR [$Q = (OR - 1)/(OR + 1)$] that ranges between −1 and + 1. Q can be interpreted as the difference in the probabilities of a randomly selected clinical case and a randomly selected clinical non-case that differ in their classification on the CIDI being correctly versus incorrectly classified by the CIDI. However, "tied pairs" (i.e., clinical cases and non-cases that have the same CIDI classification) are excluded, which means that Q does not tell us about actual prediction accuracy.

The AUC is a measure that resolves this problem, as AUC can be interpreted as the probability that a randomly selected clinical case will score higher on the CIDI than a randomly selected non-case. Although developed to study the association between a continuous predictor and a dichotomous outcome, the AUC can be used in the special case where the predictor is a dichotomy, in which case AUC equals (SN + SP)/2. As a result of this useful interpretation, we focus on AUC in our evaluation of CIDI-SCID diagnostic concordance. We also report SN and SP, the key components of AUC in the dichotomous case, as well as positive predictive value (PPV; the proportion of CIDI cases that are confirmed by the SCID), negative predictive value (NPV; the proportion of CIDI non-cases that are confirmed as non-cases by the SCID), and χ^2.

3 | RESULTS

3.1 | Distinguishing DSM-IV/CIDI disorders from other screened positives

We noted in the introduction that the clinical reappraisal studies were carried out exclusively among SNMHS respondents who screened positive for the disorders. We also noted that we purposefully developed broad diagnostic screening questions designed to cast a wide net and maximize the proportion of true cases that responded positively. Evidence that we succeeded in this task can be found in the fact that 27–34% of respondents across disorders endorsed the diagnostic screening questions for the four disorders considered in the clinical reappraisal studies. These are much higher proportions than are suggested by previous epidemiological studies to meet lifetime criteria for these disorders (Kessler et al., 2007). Consistent with these prior epidemiological data, only minorities of the screened positives (32.2–41.3% across diagnoses) went on to meet full diagnostic criteria for the disorders in the CIDI. (Table 2) The real challenge for clinical interviewers given this fact is to distinguish screened positives who meet full diagnostic criteria from those that do not.

3.2 | Obsessive–compulsive disorder (OCD)

3.2.1 | CIDI assessment

The CIDI diagnostic stem question for lifetime OCD began with the following preamble: Some people have repeated unpleasant thoughts,

TABLE 2 Proportion of Saudi National Mental Health Survey respondents who endorsed CIDI diagnostic stem questions[a] for the DSM-IV disorders included in the clinical reappraisal studies

	Total sample	Among screen positives	Total sample
	Screen positive[a]	Original CIDI	Original CIDI
	(%)	(%)	(%)
Obsessive–compulsive disorder	27.1	41.3	11.2
Adult separation anxiety disorder	33.9	32.2	10.9
Social phobia	30.6	37.2	11.4
Major depressive episode	28.2	39.2	11.0

Abbreviations: CIDI, Composite International Diagnostic Interview.
[a]See the text for the definition of screening positive for each diagnosis.

I. Obsessions: Some people have repeated unpleasant thoughts, images, or impulses that they can't get out of their heads. For example, some people have the idea that their hands are dirty no matter how much they wash them. Did you ever have a time in your life when you were bothered by any of the following:

a. A recurrent, persistent concern about dirt, germs, or contamination?

b. A recurrent, persistent concern about harming someone, or being responsible for things going wrong?

c. A recurrent, persistent concern about having things symmetrical, lined up, or ordered in exactly the right way, or a recurrent urge to count or touch things?

d. A recurrent, persistent concern about having to save or keep things, even if they have little monetary or sentimental value?

e. Any another disturbing thought that kept entering your mind, such as concerns about doing something terrible or morally wrong, sexual thoughts that you found disturbing and unpleasant, or some other repeated, upsetting thought, image, or impulse?

II. Compulsions: Some people feel driven to do certain behaviors over and over, either physically or in their mind. For example, some people check the stove in their home again and again, many times a day, no matter how many times they see that the stove is turned off. Did you ever have a time in your life when you repeatedly carried out any of the following behaviors:

a. Repeatedly washing, cleaning, or decontaminating?

b. Repeatedly checking things like locks or stoves, or repeatedly making sure that no harm or injury was done to yourself or someone else?

c. Repeatedly straightening, lining up, arranging, counting, or touching things, or doing things in an exactly defined order?

d. Always having to save things, to the point where you could not throw away things that you no longer needed or cared about?

e. Any other repetitive behaviors that you felt driven to do, such as going over and over a moral argument in your mind, or praying over and over for forgiveness, or some other physical or mental act you felt you had to do repeatedly?

Abbreviations. CIDI, Composite International Diagnostic Interview.

FIGURE 1 The obsessions and compulsions asked about in the CIDI

images, or impulses that they cannot get out of their heads. For example, some people have the idea that their hands are dirty no matter how much they wash them. Did you ever have a time in your life when you were bothered by any of the following? This preamble was followed by five yes-no questions about obsessions (Figure 1), at least one of which had to be endorsed for the respondent to be queried about DSM-IV Criteria A-E OCD symptoms associated with obsessions. All respondents (not limited to those that reported obsessions) were then asked about compulsions. We began with the following preamble: Some people feel driven to do certain behaviors over and over, either physically or in their mind. For example, some people check the stove in their home again and again, many times a day, no matter how many times they see that the stove is turned off. Did you ever have a time in your life when you repeatedly carried out any of the following behaviors? This preamble was followed by five yes-no questions about compulsions (Figure 1), at least one of which had to be endorsed for the respondent to be queried about DSM-IV Criteria A-C of OCD associated with compulsions.

The diagnostic stem question threshold used for purposes of our clinical reappraisal study was the endorsements of at least one of the 10 yes-no questions about obsessions or compulsions included in the CIDI. As shown in Table 2, 27.1% of SNMHS respondents passed this screen. After the additional CIDI questions were administered to establish lifetime prevalence, age-of-onset, and course of illness, respondents who reported that any of their 10 obsessions or compulsions continued to exist at any time within 12 months of the interview were administered a self-report version of the Yale-Brown Obsessive–Compulsive Scale Second Edition (YBOCS-II) focused on the 12 months before the CIDI interview. The YBOCS-II is the gold standard clinician-administered severity scale for OCD symptoms (Storch et al., 2010). But because it is very time-consuming to administer, several self-report versions were developed that have very good concordance with the clinician-administered version (du Mortier et al., 2019; Hauschildt, Dar, Schröder, & Moritz, 2019; Hiranyatheb et al., 2015). We converted one of these to a lay interviewer-administered format for the WMH surveys.

3.2.2 | The clinical reappraisal sample

As noted in Table 2, only a minority (41.3%) of the SNMHS respondents who endorsed at least one OCD diagnostic stem question went on to meet full DSM-IV criteria for lifetime OCD in the CIDI. But that resulted in a lifetime prevalence estimate (11.2%) that was much higher than estimates based on previous community epidemiological surveys (Mathes, Morabito, & Schmidt, 2019; Ruscio, Stein, Chiu, & Kessler, 2010), leading us to launch a clinical reappraisal study to evaluate CIDI diagnoses against diagnoses based on blinded clinical follow-up interviews. We attempted to administer a SCID OCD clinical reappraisal interview by telephone for this purpose to a probability sample of 50 SNMHS respondents who met lifetime DSM-IV/CIDI criteria for OCD and a probability sample of 25 SNMHS respondents who had positive diagnostic stem question responses but were classified by the CIDI as not having lifetime OCD. Subsequent cases (n = 69) and controls (n = 63) were subsequently added to the sample based on the low response rate among originally selected respondents and on additional issues we decided to probe after completing the first iteration of clinical appraisal interviews and evaluations. We ended up completing 103 SCID OCD clinical reappraisal interviews (61 with CIDI cases and 42 with CIDI non-cases). As shown in Table 1, these SCID clinical reappraisal interviews were completed an average of 47 months after the CIDI interviews.

3.2.3 | Concordance of diagnoses based on CIDI and SCID interviews

Concordance of lifetime diagnoses based on the CIDI and SCID interviews was poor, with only 10 of the 61 CIDI cases confirmed by the SCID. But was this due to over-diagnosis on the part of the CIDI, under-diagnosis on the part of the SCID, or some combination of the two? We investigated this issue by comparing lifetime SCID diagnoses with 12-month diagnoses based on the YBOCS in the CIDI interviews (either severe or moderate cases as defined by the standard YBOCS scoring system). Only five of the 18 YBOCS cases (15 moderate, 3 severe) were classified by the SCID as meeting full DSM-IV criteria for OCD at any time in their life. (Table 3) But even more striking was the fact that another 11 of the 18 YBOCS cases were classified by the SCID as never having had any lifetime OCD symptoms. Investigation of the SCID transcripts shows that was due to these SCID respondents denying all SCID OCD symptom questions despite reporting significant symptoms in the earlier YBOCS. We concluded from this result that the SCID was under-diagnosing OCD and that this was occurring because of respondent inconsistency rather than because of differences in the way clinical interviewers rated symptom severity compared to the YBOCS ratings.

3.2.4 | Recalibrating the CIDI by focusing on YBOCS rather than SCID ratings

Based on these results, we decided to use the YBOCS reports in the CIDI interviews rather than the subsequent SCID ratings as the basis for recalibrating the CIDI OCD diagnoses. A comparison of 12-month OCD diagnoses based on the CIDI with those based on the YBOCS in the total Part II SNMHS sample found that the CIDI picked up 64% of all YBOCS cases (compared to the SCID lifetime diagnoses picking up only 28% of YBOCS cases). But the CIDI over-diagnosed compared to the YBOCS (PPV = .134). We consequently focused on the associations of CIDI symptom severity measures with 12-month YBOCS diagnoses among the 159 12-month CIDI OCD cases in the total Part II SNMHS sample.

All YBOCS cases but only a minority of 12-month CIDI cases reported both at least one obsession and one compulsion from the lists of obsessions and compulsions presented in the CIDI. (Figure 1) In addition, among respondents who reported both obsessions and compulsions, virtually all YBOCS cases but only a minority of 12-month CIDI cases endorsed six lifetime CIDI symptom questions and four 12-month symptom questions. (Figure 2) Based on these results, we recalibrated CIDI diagnoses to require the original DSM-IV/CIDI criteria in addition to endorsement of these other lifetime and 12-month symptom questions. Concordance of CIDI 12-month diagnoses with YBOCS diagnoses increased markedly when the CIDI diagnostic criteria were tightened in this way. (Table 4) AUC of the recalibrated CIDI increased from .760 to .803 based on a substantial reduction in false positives (i.e., an increase in SP from 89.7 to 96.9%) and no change in true positives (i.e., SN = 63.6% in both coding schemes). In addition, the discrepancy between prevalence estimates based on the recalibrated CIDI and the YBOCS (χ^2_1 = 32.9, p < .001) was substantially less than the discrepancy between prevalence estimates based on the original CIDI and the YBOCS (χ^2_1 = 166.7, p < .001) even though the CIDI prevalence was higher than the YBOCS prevalence in both cases.

The lifetime CIDI diagnostic criteria were tightened in the same way as the 12-month criteria by requiring respondents to endorse the lifetime symptom severity questions associated with 12-month YBOCS diagnoses but not requiring any of the 12-month symptom questions to be endorsed. As YBOCS was assessed only for the 12 months before interview, we had no lifetime YBOCS diagnoses to go on to use a different calibration rule for the lifetime CIDI. This means that our scoring of lifetime prevalence might be conservative, as non-persistent cases could have had fewer lifetime symptoms than 12-month cases. However, this scoring decision might be anti-conservative in estimating the persistence of lifetime OCD (i.e., 12-month prevalence among lifetime cases) for the same reason. It will consequently be important in future analysis of the SNMHS data to recognize that a large proportion of the KSA population has subthreshold symptoms of OCD and to evaluate the clinical significance of these subthreshold cases.

TABLE 3 Concordance between 12-month symptom ratings based on the patient self-report YBOCS-II scale in the CIDI interview and diagnoses of lifetime DSM-IV OCD based on the SCID obtained an average of 4 years later in the clinical reappraisal interview

	12-month YBOCS-II severity							
	Non/subclinical		Mild		Moderate		Severe	
	%	(*n*)	%	(*n*)	%	(*n*)	%	(*n*)
Lifetime DSM-IV/SCID diagnosis								
No symptoms	78.9	(45)	89.3	(25)	60.0	(9)	66.7	(2)
Criteria A only	5.3	(3)	—	(0)	—	(0)	—	(0)
Criteria A and B	3.5	(2)	3.6	(1)	13.3	(2)	—	(0)
DSM-IV OCD	12.3	(7)	7.1	(2)	26.77	(4)	33.3%	(1)
(n)	(57)		(28)		(15)		(3)	

Abbreviations: CIDI, Composite International Diagnostic Interview; OCD, obsessive–compulsive disorder; SCID, Structured Clinical Interview for DSM-IV; YBOCS-II, Yale-Brown Obsessive–Compulsive Scale Second Edition.

I. Lifetime
   a. Experienced both one or more lifetime obsessions and one or more lifetime compulsions
   b. Obsessions caused emotional upset or distress
   c. Tried to resist obsessions or push them out of mind when they occurred
   d. Obsessions were sometimes so strong that they could not be pushed out of mind no matter how hard tried
   e. Compulsive behaviors were recognized as excessive or unreasonable
   f. Tried to resist engaging in compulsive behaviors
   g. Urges to engage in compulsive behaviors were sometimes so strong that they could not be resisted no matter how hard tried
   h. Engaging in compulsive behaviors sometimes reduced emotional distress

II. Twelve-month
   a. On average, at least 1 hour per day occupied by obsessions
   b. Obsessions caused upset or anxiety
   c. Tried to resist obsessions or turn attention away from them as they entered mind
   d. On average, at least 1 hour per day spent on compulsive behaviors

Abbreviations. OCD, obsessive compulsive disorder, YBOCS-II, Yale-Brown Obsessive-Compulsive Scale Second Edition; CIDI, Composite International Diagnostic Interview.

FIGURE 2 The lifetime and 12-month OCD symptom questions found to be more common among respondents in the OCD clinical reappraisal sample who met 12-month diagnostic criteria based on the YBOCS-II than based on the original CIDI coding scheme (n = 103)

TABLE 4 Concordance of original and recalibrated CIDI diagnoses of 12-month DSM-IV OCD compared to diagnoses based on the self-report YBOCS-II in the total Part II SNMHS sample (n = 1,981)[a]

	Original CIDI		Recalibrated CIDI	
	Est	(SE)	Est	(SE)
Prevalence[b]	11.2	(0.7)	4.1	(0.4)
SN	63.6	(8.4)	63.6	(8.4)
SP	89.7	(0.7)	96.9	(0.4)
PPV	9.5	(2.0)	25.6	(4.8)
NPV	99.3	(0.2)	99.4	(0.2)
AUC	76.7	—	80.3	—
χ^2_1	166.7		32.9	

Abbreviations: AUC, area under the receiver operator characteristic curve; CIDI, Composite International Diagnostic Interview; NPV, negative predictive value; OCD, obsessive–compulsive disorder; PPV, positive predictive value; SE, standard error; SN, sensitivity; SNMHS, Saudi National Mental Health Survey; SP, specificity; YBOCS-II, Yale-Brown Obsessive–Compulsive Scale Second Edition.
[a]The calculations are based on the weighted Part II sample.
[b]YBOCS prevalence is 1.7% (0.3).

3.3 | Adult separation anxiety disorder

3.3.1 | CIDI assessment

The CIDI assessment of separation anxiety disorder (SAD) began with a preamble and question about pediatric-onset SAD: Some young kids get very upset when they are separated from their mother or the person they are most attached to emotionally. Examples include getting very upset when they are away from these people, worrying a lot that something bad will happen to separate these people from them, or wanting to stay home from school or not wanting to go other places without them. Did you ever feel this way for a month or longer when you were a child or adolescent? We then asked all respondents a separate diagnostic stem question about adult SAD as follows: Some adults have difficulties with separation from family members, romantic partners, or close friends. Examples include getting very upset when they are away from this person, worrying a lot that this person might leave them, and being too "clingy" or dependent. Did you ever have a time lasting 1 month or longer as an adult when you had problems like that? Respondents who endorsed the first of these questions were probed to assess the eight Criterion A symptoms of pediatric-onset SAD along with questions about age-of-onset (Criterion B), duration (Criterion C, which could have persisted into adulthood) and clinically significant distress or impairment caused by these symptoms (Criterion D). Respondents who endorsed the diagnostic stem question for adult SAD were asked similar questions to assess symptoms in adulthood.

3.3.2 | The clinical reappraisal sample

As noted in Table 2, only a minority (32.2%) of the SNMHS respondents who endorsed at least one adult SAD diagnostic stem question went on to be defined as having lifetime adult-onset SAD using all DSM-IV criteria other than the requirement in Criterion C that onset was before age 18. DSM-IV requires three or more of eight Criterion A symptoms, whereas we required only one to define a respondent as screening positive even though we required three or more for a CIDI diagnosis. But that resulted in a lifetime prevalence estimate (10.9%) much higher than estimates based on previous community epidemiological surveys (Silove et al., 2015), leading us to carry out a clinical reappraisal study to evaluate CIDI diagnoses against diagnoses based on blinded clinical follow-up interviews. We attempted to administer a SCID interview to assess adult-onset SAD by telephone for this

purpose with a probability sample of 50 SNMHS respondents who met lifetime DSM-IV/CIDI criteria and a probability sample of 25 SNMHS respondents who had positive diagnostic stem question responses but were classified by the CIDI as not having lifetime adult-onset SAD. We were able to interview only 14 CIDI cases and eight controls out of this target sample of 75, but the results were sufficiently clear that no attempt was made to expand the sample.

3.3.3 | Concordance between initial CIDI and SCID diagnoses

The SCID interviews initially classified only one respondent in the clinical reappraisal sample as having a lifetime history of adult-onset SAD. This respondent was also classified as a CIDI case. Eleven of these other 13 were classified by the SCID as never having had any SAD symptoms. Investigation of the SCID transcripts showed that these respondents did, in fact, respond in the negative to all SCID symptom questions. Yet it is clear from their CIDI responses (which were made an average of 45 months before the SCID reappraisal interviews) that these respondents met criteria for adult-onset SAD, as they all reported a minimum of three Criterion A symptoms that lasted for a minimum of 3 years and caused both clinically significant distress and clinically significant impairment.

It seemed clear from this inconsistency that the 11 SCID respondents who denied all adult SAD symptoms were false negatives, but to investigate this issue more directly we re-contacted four of these 11 with a different SCID interviewer and described to these four respondents several of the adult SAD symptoms they reported in their CIDI interview. Two of the four respondents acknowledged these symptoms and went on to complete full SCID assessments that resulted in the SCID interviewers rating them as meeting full criteria for adult SAD. The other two respondents continued to deny any adult SAD symptoms in their second SCID interview.

3.3.4 | Recalibrating the CIDI to increase concordance with SCID ratings

We took a conservative approach in setting a CIDI diagnostic threshold in light of these clinical reappraisal results by comparing CIDI SAD severity measures between the three clinical reappraisal respondents who were classified as adult SAD cases by the SCID and the 11 other CIDI adult SAD cases who were classified by the SCID as not meeting lifetime criteria. A series of nine CIDI SAD symptom questions were endorsed more often by the SCID cases than non-cases. (Figure 2) All the SCID cases and about half the CIDI cases endorsed at least four of these eight CIDI symptom questions and also reported having times when their emotional distress about separation was "so severe that nothing could cheer you up or calm you down." We recalibrated the CIDI diagnoses to require these responses (i.e., at least four of the eight symptoms in Figure 3 and a report of at least sometimes having inconsolable distress) to generate a conservative estimate of adult-onset SAD.

Concordance of CIDI diagnoses with SCID diagnoses among clinical calibration study respondents increased substantially when CIDI diagnostic criteria were tightened in this way, from an AUC of .711 to one of .894. (Table 5) 100% of SCID cases were confirmed in this sample both in the original CIDI and recalibrated CIDI (i.e., SN = 1.0) and the number of false positives (i.e., the inverse of SP) was cut by nearly two-thirds (i.e., from 1–.421 = .579 to 1–.789 = .211). The number of respondents classified as cases by the CIDI remained significantly higher than the number classified as cases by the SCID even after recalibration (χ^2_1 = 4.0, p = .046). However, we judge this to reflect under-estimation in the SCID due to respondents who reported clearly significant symptoms in the CIDI interviews denying ever having such symptoms in the SCID. It is noteworthy in this regard that all the respondents classified in the SCID as not meeting criteria despite admitting at least some symptoms to SCID interviewers were classified as not meeting criteria in the recalibrated CIDI.

The next questions are about difficulties adults face with separation from a family member, romantic partner, or close friend. As an adult how often did you have any of the following difficulties lasting one month or longer – often, sometimes, rarely or never?

1. How often did you get very sad, worried, or upset whenever you had to be apart from [person]?
2. How often did you fear that [person] might be seriously injured in an accident or die or that some other terrible thing might happen to them?
3. How often did you worry that something bad was going to happen to you that would separate you from [person]?
4. How often did you want to stay home or not go places so that you could stay near (HIM/HER/this person)?
5. How often did you plead with [person] to stay with you or to take you along with them when they needed to leave you for even a short period of time?
6. How often did you get sick to your stomach, have headaches, or have other physical symptoms when you had to be apart from [person]?
7. How often did you feel like you could not go to sleep at night unless [person] was near you?
8. How often did you have repeated nightmares about [person] being harmed or about something happening that would separate you from one another?

Abbreviations. CIDI, Composite International Diagnostic Interview; SCID, Structured Clinical Interview for DSM-IV.

FIGURE 3 The CIDI symptom questions that distinguished respondents in the clinical reappraisal who met the original CIDI criteria for adult-onset separation anxiety disorder (SAD) depending on whether they were also classified as cases in the subsequent independent SCID clinical reappraisal interviews

TABLE 5 Concordance of original and recalibrated CIDI diagnoses of lifetime adult-onset separation anxiety disorder (SAD) with diagnoses based on the SCID in the clinical reappraisal sample (n = 22)

	Original CIDI		Recalibrated CIDI	
	Est	(SE)	Est	(SE)
Prevalence	63.6	10.5	31.8	(10.2)
SN	100.0	(0.0)	100.0	(0.0)
SP	42.1	(11.6)	78.9	(9.6)
PPV	21.4	(11.2)	42.9	(19.1)
NPV	100.0	(0.0)	100.0	(0.0)
AUC	71.1	—	89.5	—
χ^2_1	11.0		4.0	

Abbreviations: AUC, area under the receiver operator characteristic curve; CIDI, Composite International Diagnostic Interview; NPV, negative predictive value; PPV, positive predictive value; SCID, Structured Clinical Interview for DSM-IV; SE, standard error; SN, sensitivity; SP, specificity.

3.4 | Social phobia

3.4.1 | CIDI assessment

The CIDI assessment of social phobia began by presenting respondents with a visual list of common social or performance situations (e.g., meeting new people, going to parties or other social gatherings, speaking up in a meeting or class, talking to people in authority) and instructing them to look at the list as we asked how they felt in these situations. We then asked two questions. The first was: Was there ever a time in our life when you felt very afraid or really shy with people, like meeting new people, going to parties, going on a date, or using a public bathroom? The second was: Was there ever a time in your life when you felt very afraid or uncomfortable when you had to do something in front of a group of people like giving a speech or speaking in class? Respondents who endorsed either of these diagnostic stem questions were then asked specifically about fear of 14 specific social or performance situations and then probed to assess additional symptoms of social phobia, including Criteria A (marked-persistent fear of acting in a way that will be humiliating or embarrassing), B (exposure to the situation invariably provoking anxiety), C (recognition that the fear is excessive or unreasonable), D (avoidance or endurance with intense anxiety or distress), E (the avoidance, anxious anticipation, or distress causes clinically significant impairment or marked distress about having the phobia), and F (duration of at least 6 months).

3.4.2 | The clinical reappraisal sample

As noted in Table 2, 37.2% of the SNMHS respondents who endorsed at least one of the two social phobia diagnostic stem question went on to be defined as having lifetime DSM-IV/CIDI social phobia. Based on the fact that 30.6% of SNMHS respondents endorsed one or both of these stem questions, that conditional probability resulted in a lifetime social phobia prevalence estimate of 11.4%, which is somewhat higher than the estimates obtained in previous WMH surveys in other countries (Stein et al., 2017), leading us to carry out a clinical reappraisal study to evaluate CIDI diagnoses against diagnoses based on blinded SCID clinical follow-up interviews. We attempted to administer a SCID interview by telephone to assess lifetime social phobia with a probability sample of 50 SNMHS respondents who met lifetime DSM-IV/CIDI criteria and a probability sample of 25 SNMHS respondents who endorsed at least one diagnostic stem question but were classified by the CIDI as not having lifetime social phobia. We were able to interview 24 CIDI cases and 9 controls out of this target sample of 75.

3.4.3 | Concordance between initial CIDI and SCID diagnoses

The SCID classified nine respondents as having a lifetime history of social phobia. Six of these nine were also classified as cases by the CIDI (SN = 0.667). However, the CIDI also classified 18 other respondents as being cases. The SCID classified 11 of these 18 as never having had any social phobia symptoms, two others as meeting only Criterion A, an additional three as also meeting Criteria B-C, and the other two as also meeting either Criterion D or E but not both. Investigation of the SCID transcripts showed that the 11 respondents who were classified by the SCID as having no symptoms did, in fact, deny all SCID symptom questions even though they reported these symptoms in their earlier CIDI interview (which, as noted above in the section on the clinical reappraisal samples, took place an average of 46 months before the SCID reappraisal interview). It is clear from the CIDI responses that these respondents met all lifetime criteria for social phobia, as they reported situational fears that invariably provoked distress, were recognized as excessive or unreasonable, led to avoidance or endurance with intense anxiety or distress, resulted in clinically significant impairment or marked distress (about having the phobia), and persisted beyond 6 months.

3.4.4 | Recalibrating the CIDI to increase concordance with SCID ratings

We compared the six CIDI cases confirmed by the SCID with the 18 not confirmed by the SCID on a range of symptom severity measures. The SCID cases were more likely than the CIDI-only cases to report four or more of the 14 situational fears and frequent avoidance. When we recalibrated the CIDI to require these as additional criteria, concordance between SCID and CIDI diagnoses (AUC) increased in the clinical reappraisal sample from .458 to .632 (Table 6). This increase occurred despite a small decrease in the number of SCID

TABLE 6 Concordance of original and recalibrated CIDI diagnoses of lifetime social phobia with diagnoses based on the SCID in the clinical reappraisal sample (n = 33)

	Original CIDI		Recalibrated CIDI	
	Est	(SE)	Est	(SE)
Prevalence	72.7	(7.9)	36.4	(8.5)
SN	66.7	(16.0)	55.6	(16.8)
SP	25.0	(9.0)	70.8	(9.4)
PPV	25.0	(9.0)	41.7	(14.4)
NPV	66.7	(16.0)	81.0	(8.7)
AUC	45.8	—	63.2	—
χ^2_1	10.7		0.8	

Abbreviations: AUC, area under the receiver operator characteristic curve; CIDI, Composite International Diagnostic Interview; NPV, negative predictive value; PPV, positive predictive value; SCID, Structured Clinical Interview for DSM-IV; SE, standard error; SN, sensitivity; SP, specificity.

cases detected by the CIDI (i.e., SN, which decreased from 66.7 to 55.6%) because of a substantial decrease in the number of false positives (i.e., the inverse of SP) from 75.0 to 29.2%. The number of respondents classified as CIDI cases was no longer significantly different from the number classified as cases by the SCID after recalibration (χ^2_1 = 0.8, p = .37).

3.5 | Major depressive episode

3.5.1 | CIDI assessment

The CIDI assessment of major depressive episode (MDE) began with a series of three diagnostic stem questions asking respondents if they ever had times lasting several days or longer when "most of the day you felt sad, empty, or depressed," "most of the day you were very discouraged about how things were going in your life," or "you lost interest in most things you usually enjoy like work, hobbies, and personal relationships." Respondents who endorsed one or more of these three questions were then asked if they ever had any of these experiences "most of the day, nearly every day, for two weeks or longer." Only respondents who answered yes to this follow-up question were classified as endorsing the diagnostic stem question for purposes of the clinical reappraisal study. The CIDI interviews went on to ask questions about each of the nine DSM-IV Criterion A symptoms of MDE. Respondents who reported having at least five Criterion A symptoms including either A1 (depressed mood most of the day, nearly every day) or A2 (markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day) were then assessed for Criteria C (clinically significant distress or impairment) and E (the symptoms were not better accounted for by normal bereavement). We did not assess DSM-IV Criteria B (the symptoms were not part of a mixed episode) or D (the symptoms were not due to the direct physiological effects of a substance, medication, or medical condition).

3.5.2 | The clinical reappraisal sample

As noted in Table 2, 28.2% of SNMHS respondents endorsed the MDE diagnostic stem question and 39.2% of these screened positives went on to meet DSM-IV/CIDI criteria for lifetime MDE. We attempted to validate these diagnoses by administering a SCID MDE interview to a probability sample of 50 SNMHS respondents who met lifetime DSM-IV/CIDI criteria and a probability sample of 25 SNMHS respondents who had positive diagnostic stem question responses but were classified by the CIDI as not having lifetime MDE. We were able to interview 18 CIDI cases and nine controls out of this target sample of 75.

3.5.3 | Concordance between initial CIDI and SCID diagnoses

The SCID classified eight respondents in the clinical reappraisal sample as having a lifetime history of DSM-IV MDE and four others as meeting all criteria other than the exclusion for normal bereavement. Five of the eight SCID cases and three of the four classified by the SCID as being bereaved were also classified as cases by the CIDI, with the CIDI confirming that not all of the episodes of the latter three respondents occurred after the death of a loved one. However, the CIDI also classified eight other respondents as being cases. The SCID classified these eight either as having no lifetime Criterion A MDE symptoms (n = 1), having fewer than five such symptoms (n = 4), or not having clinically significant distress or impairment (n = 3).

Investigation of the SCID transcripts and disaggregated CIDI interviews led us to retain all 11 of the CIDI cases classified as SCID non-cases in the final definition of DSM-IV MDE. This decision was based on two observations related to findings obtained when we distinguished between the six cases in the clinical reappraisal sample classified as "severe" by the CIDI (7–9 of the 9 Criterion A symptoms including both A1 and A2 in addition to depressed mood in the worst lifetime episode classified by respondents as either "severe" or "very severe") and the other nine classified as "not severe" (i.e., all other CIDI cases).

The first observation is that the SCID interviews classified six of these seven severe CIDI cases as meeting DSM-IV criteria (including one classified as being excluded only because of normal bereavement, but confirmed in the CIDI as having a number of episodes that occurred at times other than after the death of a loved one). The single exception was a severe CIDI case that was classified by the SCID as meeting Criteria A but not having clinically significant distress or impairment. But this respondent reported in the CIDI having both dysphoria and anhedonia (A1-A2), feelings of worthlessness or excessive or inappropriate guilt (A7), suicidality (A9), "very severe" depressed mood during the worst 2 weeks of the episode, and "almost always" feeling so distressed during those 2 weeks that nothing could cheer him up. It is clear from these reports that the respondent did in fact, experience clinically significant distress in this episode even though

this was not reported in the SCID interview, which occurred more than 4 years after the CIDI interview.

The second observation is that discrepancies between MDE diagnoses based on the CIDI and SCID interviews although much more pronounced for non-severe CIDI cases, appear to be due to inaccuracies on the part of the SCID rather than the CIDI. We drew this conclusion by separately considering each of four different subgroups:

1. Two of the nine non-severe CIDI cases were classified by the SCID as meeting Criteria A and C but were excluded from DSM-IV diagnoses of MDE because they were classified as normal bereavement. However, as with the severe CIDI case noted above, the CIDI interview confirmed that both these two non-severe CIDI cases had MDE episodes at times other than after the death of a loved one.
2. Two other non-severe CIDI cases were classified by the SCID as meeting Criterion A but not having clinically significant distress or impairment. Yet both these respondents reported in the CIDI having had both dysphoria and anhedonia (A1-A2) along with suicidality (A9) in their worst lifetime episode as well as *"sometimes"* during that episode feeling so distressed that nothing could cheer them up. These reports clearly qualify as evidence of clinically significant distress even though this distress was not reported 4 years later in the SCID interviews.
3. Four additional non-severe CIDI cases were classified by the SCID as not meeting Criterion A because of having only 3–4 Criterion A symptoms. But these respondents reported 5–7 Criterion A symptoms in their earlier CIDI interviews, including all of them reporting both dysphoria and anhedonia (A1-A2) in their worst lifetime episode, 2 reporting suicidality (A9) in that episode, and the other 2 reporting "severe" depressed mood during that episode that either "often" or "almost always" led them to be so distressed that nothing could cheer them up. These reports clearly qualify as evidence of meeting full diagnostic criteria for DSM-IV MDE even though these symptoms were not reported 4 years later in the SCID interviews.
4. The final one non-severe CIDI case was classified by the SCID as not having any evidence of lifetime MDE. The SCID transcript for this respondent confirms that he denied ever having symptoms of MDE. Yet this same respondent reported in his CIDI interview over 4 years earlier that he had a history of several different MDE episodes, experienced six Criterion A symptoms, including both dysphoria and anhedonia (A1-A2), during his worst lifetime episode, had "moderate" depressed mood during that episode, and was at least occasionally so depressed during that episode that nothing could cheer him up. This profile clearly qualifies as a case of MDE and certainly indicates that the respondent's failure to report a history of depression in the subsequent SCID interview was a false negative report.

4 | DISCUSSION

The CIDI diagnostic thresholds based on the calibration studies reported here were selected to be conservative. This means that the prevalence estimates reported in the SNMHS should be interpreted as lower-bound estimates of true disorder prevalence. This is true not only because the calibration thresholds in the CIDI were set to be conservative relative to diagnoses based on the SCID but also due to the fact that respondents in community surveys tend to under-report embarrassing characteristics whether they are being interviewed by lay interviewers or clinicians (Gnambs & Kaspar, 2015).

Two features of the SNMHS clinical reappraisal studies are also noteworthy with regard to CIDI prevalence estimates. First, an average of 4 years separated the time of the initial CIDI interviews from the time of the SCID clinical reappraisal interviews. This resulted in a much higher proportion of SCID respondents denying diagnostic stem questions that they endorsed in the CIDI than in CIDI clinical reappraisal samples carried out in conjunction with other WMH surveys. It also created the possibility that new first onsets of the disorders occurred that were detected in the SCID but not the CIDI and were coded as CIDI false negatives even though they were true negatives at the time of the CIDI interviews. We largely ignored the problem of CIDI false negatives because of this possibility. Second, the SCID clinical reappraisal interviews were carried out over the telephone, whereas the CIDI interviews were carried out face-to-face in the homes of respondents. While, as noted in the introduction, it has been shown that telephone interviews constitute a valid mode of clinical assessment (Kendler et al., 1992; Rohde et al., 1997; Sobin et al., 1993), we do not know what would have happened if the same mode of administration had been employed consistently in both interviews.

It is also noteworthy that even though the word *validation* is often used to characterize the kind of investigation carried out in the SNMHS clinical reappraisal studies, this is not an entirely accurate term because the SCID diagnoses cannot be taken as perfect representations of DSM-IV diagnoses. This is true both because the test–retest reliability of the SCID is far from perfect (Segal, Hersen, & Van Hasselt, 1994), especially in community samples (Williams et al., 1992), and because, as noted above, some respondents in community surveys consciously hide information about their mental or substance problems from clinical interviewers (Kranzler, Tennen, Babor, Kadden, & Rounsaville, 1997). That is why we referred to our work as involving *clinical recalibration* rather than *validation*. As a result, the AUCs for CIDI-SCID concordance should be considered lower bound estimates of CIDI validity. A good illustration of the implications of this issue can be found in the work of Booth, Kirchner, Hamilton, Harrell, and Smith (1998), who compared lifetime diagnoses of major depression based on an earlier version of CIDI with diagnoses based on SCID clinical reappraisal interviews, where κ was .53. However, when the CIDI was compared with more accurate LEAD standard diagnoses (Spitzer, 1983) that used not only the SCID, but also all the clinical information available, to arrive at an improved estimate of clinical diagnoses, κ increased to .67. An additional consideration is that we focused on concordance among CIDI respondents who screened positive for the diagnoses under study. This means that AUC would have increased, in some cases dramatically so, if we had also administered SCID clinical reappraisal interviews to CIDI

respondents who did not screen positive for the disorders (the vast majority of whom would be classified by the SCID as not meeting diagnostic criteria) and calculated AUC in the total sample. Given these considerations, the concordance data reported here are broadly consistent with the results of previous CIDI clinical reappraisal studies in finding generally good agreement between diagnoses based on the CIDI and diagnoses based on clinical assessments.

ACKNOWLEDGEMENTS

The Saudi National Mental Health Survey (SNMHS) is conducted by the King Salman Center for Disability Research. It is funded by Saudi Basic Industries Corporation (SABIC), King Abdulaziz City for Science and Technology (KACST), Abraaj Capital, Ministry of Health (Saudi Arabia), and King Saud University. Funding in-kind was provided by King Faisal Specialist Hospital and Research Centre, and the Ministry of Economy and Planning, General Authority for Statistics. None of the funders had any role in the design of the study, data analysis, interpretation of results, or preparation of this paper. The SNMHS 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 Coordination Centre in the Survey Research Center at University of Michigan and the WMH Data Analysis Coordination Centre in the Department of Health Care Policy at Harvard Medical School for assistance with design, instrumentation, fieldwork, and consultation on data analysis. A complete list of all WMH publications can be found at http://www.hcp.med.harvard.edu/wmh. We also acknowledge with gratitude the work and dedication of the SNMHS staff both current and past for their contributions to the study.

CONFLICT OF INTEREST

In the past 3 years, Dr Kessler received support for his epidemiological studies from Sanofi Aventis; was a consultant for Datastat, Inc, Sage Pharmaceuticals, and Takeda. The other authors declare no conflicts of interest.

ORCID

Ronald C. Kessler https://orcid.org/0000-0003-4831-2305

REFERENCES

Agresti, A. (1996). *An introduction to categorical data analysis*. New York, NY: John Wiley & Sons.

Alonso, J., Chatterji, S., & He, Y. (Eds.). (2013). *The burdens of mental disorders: Global perspectives from the WHO world mental health surveys*. Cambridge, UK: Cambridge University Press.

Al-Subaie, A. S., Al-Habeeb, A., & Altwaijri, Y. (In press). Overview of the Saudi National Mental Health Survey. *International Journal of Methods in Psychiatric Research*.

Altwaijri, Y., Al-Habeeb, A., Bilal, L., Shahab, M., Pennell, B.-E., Mneimneh, Z., … Al-Subaie, A. (In press). The Saudi National Mental Health Survey (SNMHS): Survey instrument and field procedures. *International Journal of Methods in Psychiatric Research*.

American Association for Public Opinion Research. (2016). Standard definitions. Retrieved from: https://www.aapor.org/AAPOR_Main/media/publications/Standard-Definitions20169theditionfinal.pdf

Booth, B. M., Kirchner, J. E., Hamilton, G., Harrell, R., & Smith, G. R. (1998). Diagnosing depression in the medically ill: Validity of a lay-administered structured diagnostic interview. *Journal of Psychiatric Research*, *32*, 353–360. https://doi.org/10.1016/S0022-3956(98)00031-4

Bromet, E. J., Dunn, L. O., Connell, M. M., Dew, M. A., & Schulberg, H. C. (1986). Long-term reliability of diagnosing lifetime major depression in a community sample. *Archives of General Psychiatry*, *43*, 435–440. https://doi.org/10.1001/archpsyc.1986.01800050033004

Byrt, T., Bishop, J., & Carlin, J. B. (1993). Bias, prevalence and kappa. *Journal of Clinical Epidemiology*, *46*, 423–429. https://doi.org/10.1016/0895-4356(93)90018-v

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

Cook, R. J. (1998). Kappa and its dependence on marginal rates. In P. Armitage & T. Colton (Eds.), *The encyclopedia of biostatistics* (pp. 2166–2168). New York, NY: John Wiley & Sons.

du Mortier, J. A. M., Visser, H. A. D., van Balkom, A. J. L. M., van Megen, H. J. G. M., Hoogendoorn, A. W., Glas, G., & van Oppen, P. (2019). Examining the factor structure of the self-report Yale-Brown obsessive compulsive scale symptom checklist. *Psychiatry Research*, *271*, 299–305. https://doi.org/10.1016/j.psychres.2018.11.042

Edelbrock, C., Crnic, K., & Bohnert, A. (1999). Interviewing as communication: An alternative way of administering the diagnostic interview schedule for children. *Journal of Abnormal Child Psychology*, *27*, 447–453. https://doi.org/10.1023/a:1021979925865

First, M. B., Spitzer, R. L., Gibbon, M., & Williams, J. B. W. (2002). *Structured clinical interview for DSM-IV Axis I disorders, research version, non-patient edition (SCID-I/NP)*. New York, NY: New York State Psychiatric Institute.

GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. (2017). Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990-2016: A systematic analysis for the global burden of disease study 2016. *Lancet*, *390*, 1211–1259. https://doi.org/10.1016/S0140-6736(17)32154-2

Ghimire, D. J., Chardoul, S., Kessler, R. C., Axinn, W. G., & Adhikari, B. P. (2013). Modifying and validating the composite international diagnostic interview (CIDI) for use in Nepal. *International Journal of Methods in Psychiatric Research*, *22*, 71–81. https://doi.org/10.1002/mpr.1375

Gibbon, M., McDonald-Scott, P., & Endicott, J. (1981). Mastering the art of research interviewing. A model training procedure for diagnostic evaluation. *Archives of General Psychiatry*, *38*, 1259–1262. https://doi.org/10.1001/archpsyc.1981.01780360075007

Gnambs, T., & Kaspar, K. (2015). Disclosure of sensitive behaviors across self-administered survey modes: A meta-analysis. *Behavior Research Methods*, *47*, 1237–1259. https://doi.org/10.3758/s13428-014-0533-4

Groves, R., & Kahn, R. (1979). *Surveys by telephone: A national comparison with personal interviews*. New York, NY: Academic Press.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*, 29–36. https://doi.org/10.1148/radiology.143.1.7063747

Harkness, J., Pennell, B. E., Villar, A., Gebler, N., Aguilar-Gaxiola, S., & Bilgen, I. (2008). Translation procedures and translation assessment in the world mental health survey initiative. In R. C. Kessler & T. B. Üstün (Eds.), *The WHO world mental health surveys: Global perspectives on the epidemiology of mental disorders*. Cambridge, UK: Cambridge University Press.

Haro, J. M., Arbabzadeh-Bouchez, S., Brugha, T. S., de Girolamo, G., Guyer, M. E., Jin, R., … Kessler, R. C. (2008). Concordance of the composite international diagnostic interview version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO world mental health surveys. In R. C. Kessler & T. B. Üstün (Eds.), *The WHO world mental*

health surveys: Global perspectives on the epidemiology of mental disorders. Cambridge, UK: Cambridge University Press.

Haro, J. M., Arbabzadeh-Bouchez, S., Brugha, T. S., de Girolamo, G., Guyer, M. E., Jin, R., … Kessler, R. C. (2006). Concordance of the composite international diagnostic interview version 3.0 (CIDI 3.0) with standardized clinical assessments in the WHO world mental health surveys. *International Journal of Methods in Psychiatric Research*, *15*, 167–180. https://doi.org/10.1002/mpr.196

Hauschildt, M., Dar, R., Schröder, J., & Moritz, S. (2019). Congruence and discrepancy between self-rated and clinician-rated symptom severity on the Yale-Brown obsessive-compulsive scale (Y-BOCS) before and after a low-intensity intervention. *Psychiatry Research*, *273*, 595–602. https://doi.org/10.1016/j.psychres.2019.01.092

Heeringa, S. G., Wells, J. E., Hubbard, F., Mneimneh, Z. N., Chiu, W. T., Sampson, N. A., & Berglund, P. A. (2008). Sample designs and sampling procedures. In R. C. Kessler & T. B. Üstün (Eds.), *The WHO world mental health surveys: Global perspectives on the epidemiology of mental disorders*. Cambridge, UK: Cambridge University Press.

Hiranyatheb, T., Saipanish, R., Lotrakul, M., Prasertchai, R., Ketkaew, W., Jullagate, S., … Kusalaruk, P. (2015). Reliability and validity of the Thai self-report version of the Yale-Brown obsessive-compulsive scale-second edition. *Neuropsychiatric Disease and Treatment*, *11*, 2817–2824. https://doi.org/10.2147/NDT.S90276

Hyder, S., Bilal, L., Akkad, L., Lin, Y. C., Al-Habeeb, A., Al-Subaie, A. S., … Altwaijri, Y. (2017). Evidence-based guideline implementation of quality assurance and quality control procedures in the Saudi National Mental Health Survey. *International Journal of Mental Health Systems*, *11*, 60. https://doi.org/10.1186/s13033-017-0164-0

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., & Eaves, L. J. (1992). A population-based twin study of major depression in women. The impact of varying definitions of illness. *Archives of General Psychiatry*, *49*, 257–266. https://doi.org/10.1001/archpsyc.1992.01820040009001

Kessler, R. C., Angermeyer, M., Anthony, J. C., De Graaf, R., Demyttenaere, K., Gasquet, I., … Üstün, T. B. (2007). Lifetime prevalence and age-of-onset distributions of mental disorders in the World Health Organization's world mental health survey initiative. *World Psychiatry*, *6*, 168–176.

Kessler, R. C., & Üstün, 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*, 93–121. https://doi.org/10.1002/mpr.168

Kessler, R. C., & Üstün, T. B. (Eds.). (2008). *The WHO world mental health surveys*. Cambridge, UK: Cambridge University Press.

Kessler, R. C., Wittchen, H.-U., Abelson, J. M., McGonagle, K. A., Schwarz, N., Kendler, K. S., … 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*, 33–55. https://doi.org/10.1016/j.jad.2006.08.018

Kimerling, R., Serpi, T., Weathers, F., Kilbourne, A. M., Kang, H., Collins, J. F., … Magruder, K. (2014). Diagnostic accuracy of the composite international diagnostic interview (CIDI 3.0) PTSD module among female Vietnam-era veterans. *Journal of Traumatic Stress*, *27*, 160–167. https://doi.org/10.1002/jts.21905

Kraemer, H. C., Morgan, G. A., Leech, N. L., Gliner, J. A., Vaske, J. J., & Harmon, R. J. (2003). Measures of clinical significance. *Journal of the American Academy of Child and Adolescent Psychiatry*, *42*, 1524–1529. https://doi.org/10.1097/00004583-200312000-00022

Kranzler, H. R., Tennen, H., Babor, T. F., Kadden, R. M., & Rounsaville, B. J. (1997). Validity of the longitudinal, expert, all data procedure for psychiatric diagnosis in patients with psychoactive substance use disorders. *Drug and Alcohol Dependence*, *45*, 93–104. https://doi.org/10.1016/S0376-8716(97)01349-5

Lu, J., Huang, Y.-Q., Liu, Z.-R., & Cao, X.-L. (2015). Validity of Chinese version of the composite international diagnostic Interview-3.0 in psychiatric settings. *Chinese Medical Journal*, *128*, 2462–2466. https://doi.org/10.4103/0366-6999.164930

Mathes, B. M., Morabito, D. M., & Schmidt, N. B. (2019). Epidemiological and clinical gender differences in OCD. *Current Psychiatry Reports*, *21*, 36–36. https://doi.org/10.1007/s11920-019-1015-2

Mneimneh, Z., Heeringa, S. G., Lin, Y.-c., Altwaijri, Y., & Nishimura, R. (In press). The Saudi National Mental Health Survey: Sample design and weight development. *International Journal of Methods in Psychiatric Research*.

Montoya Gonzalez, L. E., Restrepo Bernal, D. P., Mejía-Montoya, R., Bareño-Silva, J., Sierra-Hincapié, G., Torres de Galvis, Y., … Gaviria-Arbeláez, S. (2016). Sensitivity and specificity between the composite international diagnostic interview version 3.0 (world mental health, CIDI) and the standardised clinical evaluation version I (SCID-I) in a mental health survey of the city of Medellin, 2012. [Sensibilidad y especificidad entre la Entrevista Diagnóstica Internacional Compuesta versión 3.0 (world mental health, CIDI) con la Evaluación Clínica Estandarizada versión I (SCID-I) en la Encuesta de Salud mental de la ciudad de Medellín, 2012]. *Revista Colombiana de Psiquiatria*, *45*, 22–27. https://doi.org/10.1016/j.rcp.2015.07.001

Pennell, B. E., Mneimneh, Z. N., Bowers, A., Chardoul, S., Wells, J. E., Viana, M. C., … Saiz, G. V. (2008). Implementation of the world mental health surveys. In R. C. Kessler & T. B. Üstün (Eds.), *The WHO world mental health surveys: Global perspectives on the epidemiology of mental disorders*. Cambridge, UK: Cambridge University Press.

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*, 1593–1598. https://doi.org/10.1176/ajp.154.11.1593

Ruscio, A. M., Stein, D. J., Chiu, W. T., & Kessler, R. C. (2010). The epidemiology of obsessive-compulsive disorder in the National Comorbidity Survey Replication. *Molecular Psychiatry*, *15*, 53–63. https://doi.org/10.1038/mp.2008.94

Scott, K. M., de Jonge, P., Stein, D. J., & Kessler, R. C. (Eds.). (2018). *Mental disorders around the world: Facts and figures from the WHO world mental health surveys*. Cambridge, UK: Cambridge University Press.

Segal, D. L., Hersen, M., & Van Hasselt, V. B. (1994). Reliability of the structured clinical interview for DSM-III-R: An evaluative review. *Comprehensive Psychiatry*, *35*, 316–327. https://doi.org/10.1016/0010-440x(94)90025-6

Shahab, M., Al-Tuwaijri, F., Kattan, N., Bilal, L., Hyder, S., Mneimneh, Z., … Altwaijri, Y. (2019). Implementing the TRAPD model for the Saudi adaptation of the World Mental Health Composite International Diagnostic Interview 3.0. *International Journal of Mental Health Systems*, *13*, (1). http://dx.doi.org/10.1186/s13033-019-0267-x.

Silove, D., Alonso, J., Bromet, E., Gruber, M., Sampson, N. A., Scott, K. M., … Kessler, R. C. (2015). Pediatric-onset and adult-onset separation anxiety disorder across countries in the world mental health survey. *The American Journal of Psychiatry*, *172*(7), 647–656. https://doi.org/10.1176/appi.ajp.2015.14091185

Sobin, C., Weissman, M. M., Goldstein, R. B., Adams, P., Wickramaratne, P. J., Warner, V., & Lisch, J. D. (1993). Diagnostic interviewing for family studies: Comparing telephone and face-to-face methods for the diagnosis of lifetime psychiatric disorders. *Psychiatric Genetics*, *3*, 227–234. https://doi.org/10.1097/00041444-199324000-00005

Spitzer, R. L. (1983). Psychiatric diagnosis: Are clinicians still necessary? *Comprehensive Psychiatry*, *24*, 399–411. https://doi.org/10.1016/0010-440x(83)90032-9

Spitznagel, E. L., & Helzer, J. E. (1985). A proposed solution to the base rate problem in the kappa statistic. *Archives of General Psychiatry*, *42*, 725–728. https://doi.org/10.1001/archpsyc.1985.01790300093012

Stein, D. J., Lim, C. C. W., Roest, A. M., de Jonge, P., Aguilar-Gaxiola, S., Al-Hamzawi, A., … WHO World Mental Health Survey Collaborators. (2017). The cross-national epidemiology of social anxiety disorder: Data from the world mental health survey initiative. *BMC Medicine*, *15*, 143–143. https://doi.org/10.1186/s12916-017-0889-2

Storch, E. A., Rasmussen, S. A., Price, L. H., Larson, M. J., Murphy, T. K., & Goodman, W. K. (2010). Development and psychometric evaluation of the Yale-Brown obsessive-compulsive scale—second edition. *Psychological Assessment*, *22*, 223–232. https://doi.org/10.1037/a0018492

Williams, J. B., Gibbon, M., First, M. B., Spitzer, R. L., Davies, M., Borus, J., … Rounsaville, B. (1992). The structured clinical interview for DSM-III-R (SCID). II. Multisite test-retest reliability. *Archives of General Psychiatry*, *49*, 630–636. https://doi.org/10.1001/archpsyc.1992.01820080038006

Wittchen, H. U. (1994). Reliability and validity studies of the WHO—composite international diagnostic interview (CIDI): A critical review. *Journal of Psychiatric Research*, *28*, 57–84. https://doi.org/10.1016/0022-3956(94)90036-1

Wittchen, H. U., Kessler, R. C., Zhao, S., & Abelson, J. (1995). Reliability and clinical validity of UM-CIDI DSM-III-R generalized anxiety disorder. *Journal of Psychiatric Research*, *29*, 95–110. https://doi.org/10.1016/0022-3956(94)00044-r

Wittchen, H. U., Zhao, S., Abelson, J. M., Abelson, J. L., & Kessler, R. C. (1996). Reliability and procedural validity of UM-CIDI DSM-III-R phobic disorders. *Psychological Medicine*, *26*, 1169–1177. https://doi.org/10.1017/s0033291700035893

How to cite this article: Kessler RC, Al-Desouki M, King AJ, et al. Clinical reappraisal of the Composite International Diagnostic Interview Version 3.0 in the Saudi National Mental Health Survey. *Int J Methods Psychiatr Res*. 2020;29:e1828. https://doi.org/10.1002/mpr.1828