

Methodological innovations in the National Survey of American Life

JAMES S. JACKSON,¹ HAROLD W. NEIGHBORS,² RANDOLPH M. NESSE,¹ STEVEN J. TRIERWEILER,¹ MYRIAM TORRES¹

1 Research Center for Group Dynamics, University of Michigan, USA

2 Program for Research on Black Americans, University of Michigan, USA

Abstract

This paper provides an overview of the conceptualization and methods used in the National Survey of American Life (NSAL). The objectives of the NSAL are to investigate the nature, severity, and impairment of mental disorders among national samples of the black and non-Hispanic white ($n = 1,006$) populations in the US, including African American ($N = 3,570$), and Afro-Caribbean ($N = 1,623$) immigrant and second and older generation, populations. National multi-stage probability methods were used in generating the samples and race/ethnic matching of interviewers and respondents were employed in the largely face-to-face interview, lasting on average 2 hours and 20 minutes. Two methodological approaches are described for addressing sampling coverage of individuals attached to, but not residing in, selected households at the time of the study. The paper also describes two approaches used to address concerns about the interpretations of standard symptom probe information in assessing serious mental disorders. This included a clinical reappraisal study designed to ascertain differences in symptom responding and ascertainment of cases ($N = 677$) in a subset of the same NSAL respondents. Finally, an abbreviated, novel method for estimating the prevalence of mental disorders in first-degree family members is described and the preliminary results from this new approach are reported.

Key words: race and ethnicity, DSM-IV disorders, incarceration, clinical validation, family history

Introduction

Several methodological issues must be considered when conducting surveys on black and other ethnic minority populations (Jackson and Williams, 2003). Some of the most important concerns include: 1) ensuring proportional representation among members of the target populations for sampling (for example, gaining adequate population coverage); 2) understanding similarities and differences in the connotative meaning of various constructs across ethnic and racial groups; and 3) studying efficiently the importance of familial contributions to mental disorders within and across racial and ethnic groups. The National Survey of American Life (NSAL) conducted from 2001 to 2003 by the Program for Research on Black Americans

(PRBA), part of the Research Center for Group Dynamics, Institute for Social Research, at the University of Michigan, added special methodological innovations to address these and other concerns. The NSAL is the most comprehensive and detailed study of mental disorders and the mental health of Americans of African descent ever completed (Jackson, Torres, Caldwell, Neighbors, Nesse, Taylor, Trierweiler and Williams, 2004).

This paper describes the overarching conceptualization, methodological approaches and innovations used in the NSAL. In the first section, we present two methodological approaches for addressing sampling coverage of individuals attached to but not residing in selected households at the time of the study. In the

second section, the paper describes two approaches used to address concerns about the interpretations of standard probe information in disorders such as major depression, and in psychotic symptoms that might affect interpretation of responding to items in these domains. This included a clinical reappraisal study designed to ascertain differences in symptom responding and ascertainment of cases in a subset of the same NSAL respondents. In the third section, we describe an abbreviated, novel method for estimating the prevalence of mental disorders in first-degree family members and the preliminary results from this new approach.

Disproportionate institutionalization effects on the NSAL sample and prison pilot study

People who reside outside of a household (in prison/jail, the military, homeless, and so forth) are by definition not part of household probability samples. A disproportionate number of African Americans are at some time in their lives imprisoned or otherwise institutionalized (for example, jails/prisons, military, and so forth). Most notably, there is a disproportionate absence of younger males (18 to 34) in households among these populations affecting not only sample representations but also distributions on major substantive variables of interest, and thus poorly representing the total population. The US Department of Justice estimated that over 28% of black men are institutionalized at any given point in time (Braithewaite, 1996; Bonczar and Beck, 1997). The absence of incarcerated individuals may especially distort estimates of prevalence rates of health and psychiatric morbidity, stressful events, and use of coping resources. Thus, it was vital to consider the potential biases introduced by excluding these individuals. It was beyond the scope and budget of the NSAL to obtain interviews with institutionalized populations.

In order to address these issues within the budgetary and design limitations of the NSAL, a pilot study was conducted to test: 1) whether adult respondents from average households would divulge information on family members who lived away from home in a variety of institutionalized (military, nursing homes, jail/prisons, college, and so forth) and non-institutionalized settings (for example, the homeless), and; 2) whether close relatives of the most sensitive household non-presence, jail/prison inmates would acknowledge a person in such institutionalized settings. With the cooperation of

authorities from a local urban, county jail, telephone numbers were obtained from inmate community households or households of their closest relatives. A brief telephone interview (about 8 minutes in length) was developed that asked about relatives who may or may not be in a wide variety of institutionalized and non-institutionalized settings. The inmate telephone numbers ($N = 40$) were distributed among a set of telephone numbers generated randomly ($N = 82$). Telephone interviewers, blind to the purposes of the study and the source of the telephone numbers, called each household. When a household was reached by phone, the interviewer asked for a household listing, then asked about close relatives (spouse, child, parents, sibling) who were away from home at school or college, working abroad, in the military, in a long-term care facility or nursing home, homeless, or in jail or prison. If respondents indicated that they had a family member who fit into one of these categories, they were asked how many. First name or initials, gender, age and relationship to respondent were ascertained for each family member indicated. A total of 23 of the 40 inmate households responded and 34 of the 82 randomly selected households agreed to participate. Table 1 displays the results of the pilot test and data from the NSAL.

The results of this pilot study revealed that people are, in general, willing to provide information on relatives away from home. To assess how many people in these types of living arrangements were considered part of the household sampled for NSAL, questions based upon the results of the pilot study were included in the adult instrument ascertaining whether any persons were missing from the household, and if so, how many. As shown in Table 1, the NSAL respondents were also willing to report on individuals attached to the household but who were not eligible to be listed and screened. Thus, it is possible to estimate potential individuals lost to particular household listings.

In addition, respondents in the pilot study were asked about their own history of incarceration. Based upon work on non-response revealing that the most difficult to reach respondents are more like those not obtained (Jackson and Williams, 2003), it was hypothesized that respondents with more recent incarceration experiences would be more similar to individuals currently incarcerated, but not listed or included in the household-based study. Using this logic, we can estimate whether there are significant differences on the main variables of interest among

Table 1. Pilot and NSAL results

Relative is:	Pilot		NSAL*
	General sample (N = 34) 68% African-American	Inmate sample (N = 23) 70% African-American	
in jail/prison	15%	44%	9.3%
at school/college	18%	17%	13.5%
working abroad	3%	9%	–
in military	15%	26%	7.5%
in long care facility	12%	4%	3.2%
homeless	–	–	0.9%

*Unweighted figures; include African-Americans and Afro-Caribbeans

individuals with no incarceration history, a history of incarceration but not currently on parole, on parole, and those recently released from jail/prison. For example, unweighted estimates indicate that a total of 26.4% of all NSAL respondents have been arrested in their lifetime, and 11.2% have had some incarceration history (jail or prison) or current circumstance of incarceration.

Race and ethnic differences in construct meaning and reliability

Random probe approach

When making statistical comparisons across diverse groups (race, gender, age), there is always a risk that individuals will understand and interpret the same survey questions in different ways. Thus, population group differences in results could indicate either true variation or inconsistencies in the measurement of various constructs across groups (Jackson, Tucker, and Bowman, 1982; Caldwell, Jackson, Tucker, and Bowman, 1999). This is particularly likely to occur when respondents differ from researchers in cultural characteristics, level of education, or life chances. The random probe technique, as developed by Schuman (1966) and used in the 1980 National Survey of Black Americans (NSBA) (Jackson, 1991) and in the NSAL, is an efficient and cost-effective method for examining the fit between respondents' survey answers and their personal conceptualization of the construct on which they are reporting. After responding to particular survey items, respondents were given standardized open-ended follow-up probes, such as 'could you tell me what you mean by [RESPONSE]?' This elicits a more detailed description of respondents'

thought processes and definitions of the terms and concepts at hand.

Prior to the start of the NSAL interviewing period, 10 close-ended questionnaire items (a list of these items are available from the authors) were selected for use in a modified version of Schuman's random probe technique. Items chosen covered a wide variety of topics, including religiosity, spirituality, mental and physical health, work, family, ethnicity, discrimination, and views on social and economic issues. These items were chosen for random probing based on the value of their overall contributions to the research and especially because prior studies had demonstrated the potential for cultural misinterpretations among different race and ethnic groups.

These 10 questionnaire items produced a potential pool of approximately 62,000 close-ended responses to probe (6,200 respondents \times 10 items). Each respondent was randomly preassigned to receive one of the 10 random probes during his or her interview. Thus, this random probe procedure provided a 10% sample of the pool, or approximately 6,200 probed elaborations, and each of the 10 probes was administered to approximately 620 respondents. This number was somewhat lower for whites because one of the probes was explicitly for African American and Afro-Caribbean respondents only. Use of the computer assisted survey instrument allowed for complete randomization of the probe administration, insuring the unbiased application of the random probes across all respondents and across all the questionnaire items designated for probing.

Data collected in the random probe procedure can be used quantitatively and qualitatively. To analyse the data quantitatively, probed responses can be coded

using Schuman's recommended five-point scale. Coders first read individual responses to the probes, and then use those responses to predict their answers to the original close-ended questions. Points are granted to each respondent based on the predictive ability of his or her response to the probe. These point values can be averaged across all respondents who received the probe for a particular questionnaire item, thus evaluating the overall level of understanding of that particular close-ended item. The same coding system and point scale can be used for all probed items in the same manner.

To evaluate the random probe data qualitatively, codes will be developed based on the substantive content of responses to each probe. The distribution and variation of responses within these codes can then be compared to the researchers' original intents and meanings of the questions. Lack of congruence between researchers' intended meaning of terms or questionnaire items and respondents' interpretation of these items can be uncovered and evaluated. If there are considerable discrepancies in perceived meaning, further analysis of these close-ended variables should proceed with caution.

Both quantitative point values and qualitative codes can also be assessed across various groups of respondents (for example, race, gender, age), and potential differences in question interpretation and meaning can be reviewed. It will be particularly important to assess interpretations of probed questions across racial and ethnic groups: blacks, non-Hispanic whites, and blacks of Caribbean descent may bring diverse cultural perspectives to their interpretations of questionnaire items. If these groups have differential understandings of the constructs used in the probed questions, this should be taken into account when conducting further analyses with these close-ended questionnaire items. Recognizing these potential differences will also be catalytic in the refinement of future question wording and new research areas, and will be instrumental in increasing awareness of cultural sensitivity in survey research.

We are especially interested in those areas of the questionnaire that addressed psychiatric morbidity. The probe flow of the WHO-CIDI requires positive responses to continue (or negative responses to stop); interpretations of items in this flow, either positive or negative, could create incorrect estimates of the extent and nature of psychiatric morbidity. This issue has long

been of concern in ascertaining 'caseness' and impairment in culturally different groups (for example, Neighbors, Jackson, Campbell, and Williams, 1989) and was a major source of concern in the NSAL. For example, if groups differ in their conception of 'sad' or 'depressed' then estimates of major depression could be affected by these interpretations. Although considerable time during pre-testing was devoted to assessing the cultural equivalence of the stimulus materials, the random probe procedure provides a methodology for gathering considerable data on possible cultural differences during production interviewing.

Clinical reappraisal study

An important set of questions concerns definitions of mental disorders, as manifested in various types of instruments. For example, there is some evidence that changes in diagnosis across two measurement occasions is more common for African American psychiatric patients than for whites (Neighbors, Trierweiler, Ford, and Muroff, 2003). Other evidence suggests that clinicians may interpret symptom information differently for patients of different ethnic backgrounds (Trierweiler, Neighbors, Munday, Thompson, Binion, and Gomez, 2000). The clinical reappraisal study offers an opportunity to explore these and other issues in some depth in a community epidemiological sample.

No absolute measurement criterion exists for externally validating a psychiatric diagnosis. Therefore, it has been standard practice in the development of lay-administered diagnostic instruments to validate instrument-based diagnoses against a reasonable compromise: clinician-based diagnoses (for example, Robins, Helzer, Croughan and Ratcliff, 1981). In these clinical reappraisal studies, clinicians administer a series of symptom questions comparable to the structured instrument but, unlike the lay interviewers, they are expected to exercise professional judgement in arriving at a diagnosis (Robins, Helzer, Ratcliff, Seyfried, 1982; Helzer, Robins, McEvoy, Spitznagel, Stoltzman, Farmer, and Brockington, 1985). In general, measures of concordance between these two diagnostic modalities have been small to moderate (kappas averaging in the 0.50s) with more behaviourally described disorders, such as eating disorders or alcohol abuse, faring somewhat better.

Similar results have been found in recent research. During the 1990s, Kessler and colleagues refined a

DSM-III-R diagnostic instrument using data from the National Comorbidity Survey (NCS) (Kessler, McGonagle, Zhao, Nelson, Hughes, Eshleman, Wittchen and Kendler, 1994). That work involved a modification of the WHO-CIDI, which is the immediate precursor of the lay-administered diagnostic survey used in the present study. A more recent article on major depressive disorder from the National Comorbidity Survey Replication (NCS-R) compared CIDI and Structured Clinical Interview for DSM-IV (SCID) diagnoses, finding kappas on the order of 0.40 and 0.59 for lifetime and as low as 0.20 for 12-month prevalence rates (Kessler, Bergland, Demler, Jin, Koretz, Merikangas, Rush, Walters, and Wang, 2003). Thus, available data suggest a low to moderate limit on the extent to which structured, lay-administered instruments can generate diagnoses comparable to clinician-based diagnoses.

On the one hand, when considered solely in terms of instrument validation, low kappas might indicate poor validity for a diagnostic measure (Helzer et al., 1985) and raise questions about the extent to which emphasis should be placed on lay-administered surveys in mental health epidemiology and in the development of mental health policy (Brugha, Bebbington, Jenkins, Meltzer, Taub, Janas, and Vernon, 1999). On the other hand, for community studies, structured epidemiological instruments offer considerable advantages in terms of cost efficiency and usability with large national samples. Moreover, efforts to improve the structured instruments by enhancing the objectivity and reliability of self-reports have had the salutary effect of clarifying diagnostic process and self-report issues that were left unclear when diagnosis depended primarily on vaguely specified clinical judgements (Wittchen, Üstün, and Kessler, 1999). Thus, the concordance rates obtained in clinical reappraisal studies like Kessler et al. (2003) and the one conducted here are not merely indicators of the match between interview results and an established standard, but rather reflect a complex and theoretically interesting relationship between diagnostic self-reports and clinical judgements. From a broader psychometric perspective, structured survey diagnoses and clinician-based survey diagnoses are related but not identical indicators of mental disorder; that is, they exhibit convergent and discriminant validity in mapping the diagnostic system onto population self-reports (Messick and Mackie, 1989) within a larger nomological network of diagnostic issues and relationships (Cronbach

and Meehl, 1965). Some of these issues involve the overall scientific accuracy of the diagnostic system itself (for example, Regier et al., 1998). The clinical reappraisal portion of our study is of particular interest because the data address diagnosis in racially and ethnically distinct subpopulations.

The goal for NSAL has been to examine mental disorders both in terms of the impact of social and economic contextual stressors on prevalence rates and in terms of the cultural and social issues that may affect self-reports. In addition to careful sampling procedures designed to approximate true national random probability samples of the African American and Afro-Caribbean groups, our clinical reappraisal data were collected with carefully trained clinicians completely blind to the results of the previously collected structured instrument self-reports and diagnoses. Unlike any previous study, our concordance data will estimate true population values for the instruments used within our study groups. Comparisons of our concordance results among study samples, which includes a white sample carefully calibrated to the demographic characteristics of the African American population, and those of the NCS study, will suggest how well the structured instruments work with the two ethnic minority populations and allow us to examine self-report issues that may affect prevalence estimates for study populations in some detail. In the end, we will have the first large sample estimates of the influences of ethnicity, race and culture on the representation of mental disorders in community samples using state-of-the-art self-report measures with African American, white and Afro-Caribbeans.

Work on the clinical reappraisal study spanned all three years of the NSAL. The NSAL clinical reappraisal study was designed to assess the degree to which the estimates of mental disorder derived from the NSAL WHO-CIDI agreed with disorder estimates using a methodology that relies on the judgement of experienced clinicians trained in the use of the Structured Clinical Interview for DSM-IV (SCID; Spitzer et al, 1992). Data collection for this part of the study began about four months after the interviewing started on the main study. The initial plan for the NSAL clinical reappraisal study was for NCS-R's reappraisal clinicians to carry out NSAL's reappraisal study. In January 2002, NCS-R's study staff decided they could not take on the burden of interviewing the 700 respondents for the NSAL study; thus, an independent

Michigan team was established for this purpose. The reappraisal questionnaire designed by the NCS-R study staff was modified for use with our populations, selection criteria were developed, and six professional clinical interviewers were hired and trained, as well as a phone scheduler to set up the interviews. Clinical interviewers, all MSW social workers with years of experience performing diagnostic interviews, received 40 or more hours of SCID training. Telephone interviewing, consistent with the NCS-R procedures, began in June 2001. Interviewing for the reappraisal study continued throughout the NSAL field period. The plan was to interview 10% of the NSAL sample (final sample size is 677). A data entry procedure was designed and data entry has been completed.

The clinical reappraisal study was labour intensive. Interviewers/clinicians were closely supervised to ensure reliable, high quality data. At the start of the interviewing period, the first 50 tapes were reviewed to check interviewer coding of the SCID and other scales to ensure inter-rater reliability. Feedback was provided to interviewers when necessary. Once reviewers were confident that criteria were being rated in a consistent manner, the percentage of tapes reviewed was reduced but every interview received a visual review. Approximately 20% of the interviews were given an audio review. In addition to carefully checking completed interviews, reliability and productivity were maintained by holding regular meetings with interviewers to discuss complicated diagnostic issues and review procedures.

In addition to the SCID, the clinical reappraisal questionnaire includes several severity scales that were added to enhance understanding of the respondents' impairment associated with each disorder. These instruments include the Panic Disorder Severity Scale, the Marks Fear Questionnaire for Agoraphobia, the Liebowitz Social Anxiety Scale, the Structured Interview for the Hamilton Anxiety Rating Scale for GAD, the MADRAS Depression Scale for Major Depressive Episode, and the Clinical Global Impression Scale. The final section of the reappraisal questionnaire includes open-ended questions about the observations and inferences clinicians draw upon when making diagnostic decisions. Results from this study will allow us to address questions raised by clinicians and others that population based epidemiological studies tend to overestimate prevalence rates and focus on disorders that are not clinically significant.

Brief family disorder prevalence estimates

Family history measures have been included in large national studies only rarely due to cost constraints and respondent burdens imposed by the lengthening of average interview times. It has been estimated that about half of the between-subject variance in personality characteristics and vulnerability to mental disorders can be attributed to genetic differences (Bouchard and Loehlin, 2001); ranges for specific conditions vary from 20% for mild depression to over 80% for bipolar disorder (Kendler, 1997). In most large surveys of mental disorders, this unmeasured variance substantially inflates error terms and thus severely weakens the possible influence of social and behavioural factors. The definitive family study method of individually interviewing each family member provides accurate information (Andreasen, Endicott, Spitzer, and Winokur, 1977), but is impossible for all except very specialized and expensive genetic studies. Instead, family history information is more often gathered from one or more informants who are asked about possible disorders in other family members one at a time. The sensitivity of these family history methods is lower than that from individual interviews, but the specificity remains high (Andreasen, 1994). Even asking a single informant detailed questions about each of a dozen or more disorders for each identified relative, however, is very time consuming, often taking well over an hour; as has been found in the Family Informant Schedule and Criteria (FISC) (Mannuzza, Fyer, Endicott, and Klein, 1985).

Various methods have been used in an attempt to more efficiently gather family history data. One of these, the Brief Family History Survey, has reduced the time to 10 to 30 minutes by relying on screening questions (Weissman, Wickramaratne, Adams, Wolk, Verdelli and Olfson, 2000). Even this relatively shorter length of time, however, exceeds what is available in many survey studies. Furthermore, many studies do not require specific data on specific relatives, but only an estimate of the prevalence of a mental disorder in other family members. We therefore set the goal of creating an instrument that would provide the best possible estimate of the lifetime prevalence and severity of common mental disorders that could be obtained in an interview averaging five minutes or less.

The time constraint eliminated the possibility of asking individually about the mental health of each

relative. Instead, questions were developed that determined the number of first-degree biological relatives (parents, siblings and children) over a certain age (in this case, age 13) that the subject knew something about. Once this number was ascertained, screening questions for 12 conditions were asked in the form, 'How many of these [N] relatives have had problems with . . .?' The 12 conditions were: depression, mania, panic, agoraphobia, generalized anxiety, social phobia, tobacco use, specific phobia, alcohol use, drug use, psychosis, and suicidality. Most of the screening phrases included only a brief layperson's description rather than a diagnostic term but some included both, for instance, 'How many of these [N] relatives have had problems with depression, that is, periods lasting two weeks or longer when they felt sad, blue, or depressed?' Because prior research has shown that family history data systematically underestimates prevalence (Weissman et al., 2000), questions were intentionally worded broadly. If the answer was one or more family members, three follow up questions were asked to estimate the number who had that condition at various levels of severity: 1) 'How many had [this problem] bad enough to disturb and interfere with their lives at times?' 2) 'How many of them received professional treatment for this?' 3) 'How many were hospitalized for this?' The last condition, number of relatives who tried to commit suicide, had only one follow-up question, 'How many of them actually committed suicide?' Thus, a subject who had no relatives with any mental disorder would answer only 12 questions; those with family members who experienced disorders were asked three additional questions for each condition experienced by one or more family members. Using unweighted data, 42% reported no relatives with any disorder, and 18% reported having relatives with only one or two of the twelve disorders. Thus, the estimates for most respondents (about 60%) required fewer than twenty questions. The mean completion time for this section of the questionnaire was 3 minutes, 9 seconds.

The prevalence estimate for each disorder was calculated by dividing the number of relatives with the disorder by the number of known first-degree relatives. Results were obtained only for individuals with three or more known relatives, since denominators less than three tend to provide unstable estimates. The mean number of relatives the respondent knew something about was 6.4, with 86% of respondents reporting on three or more relatives over the age of 13.

This very brief family prevalence method does not provide results comparable to direct interviews of family members; whether its validity is different from prevalence estimates provided by the brief family history needs to be determined. It does seem clear, however, that the brief approach described here may provide genetic loading estimates that have information helpful in increasing analytic precision. In addition, these loadings should allow respondents with high- versus low-loading scores to be compared and contrasted on symptom profiles and the presence of various psychosocial factors.

Several reporting biases are likely to influence the results. Individuals who have a disorder tend to know or report a larger proportion of relatives with the disorder (Kendler et al., 1991). This can be examined in future studies that compare family prevalence data from siblings with and without disorders. Similarly, the age of a participant will influence the proportion of relatives who have passed through the ages of highest onset, and cohort effects may influence knowledge about relatives or willingness to report about their conditions. In our data, the mean percent of relatives reported to have depression was 9%, but respondents aged under 30 report depression in 20% of their relatives, while participants over age 50 report depression in only 3% of relatives. In addition, having a young age for inclusion of relatives may necessitate adjusting the denominator, for instance, by excluding all children for subjects under age 40 on the presumption that most of their children have yet to enter the significant years of risk for disorder onset.

Because familial variance has been consistently shown to arise from genetic factors, not common environment, the estimate of family prevalence for a disorder can be used as a rough estimate for an individual's genetic loading for vulnerability to a disorder, although it is important to recognize that use of a single rater underestimates heritability (Kendler, Prescott, Jacobson, Myers, and Neale, 2002). Using data from the NSAL, Figure 1 shows the strength of association between familial prevalence and the percent of respondents with depression diagnoses by the WHO-CIDI.

The utility of the Brief Family Prevalence Estimate for Mental Disorders (BFPE) will be clearer after further analyses of the data, indicating whether the expected reduction in variance will reveal the effects of social and behavioural factors that otherwise might

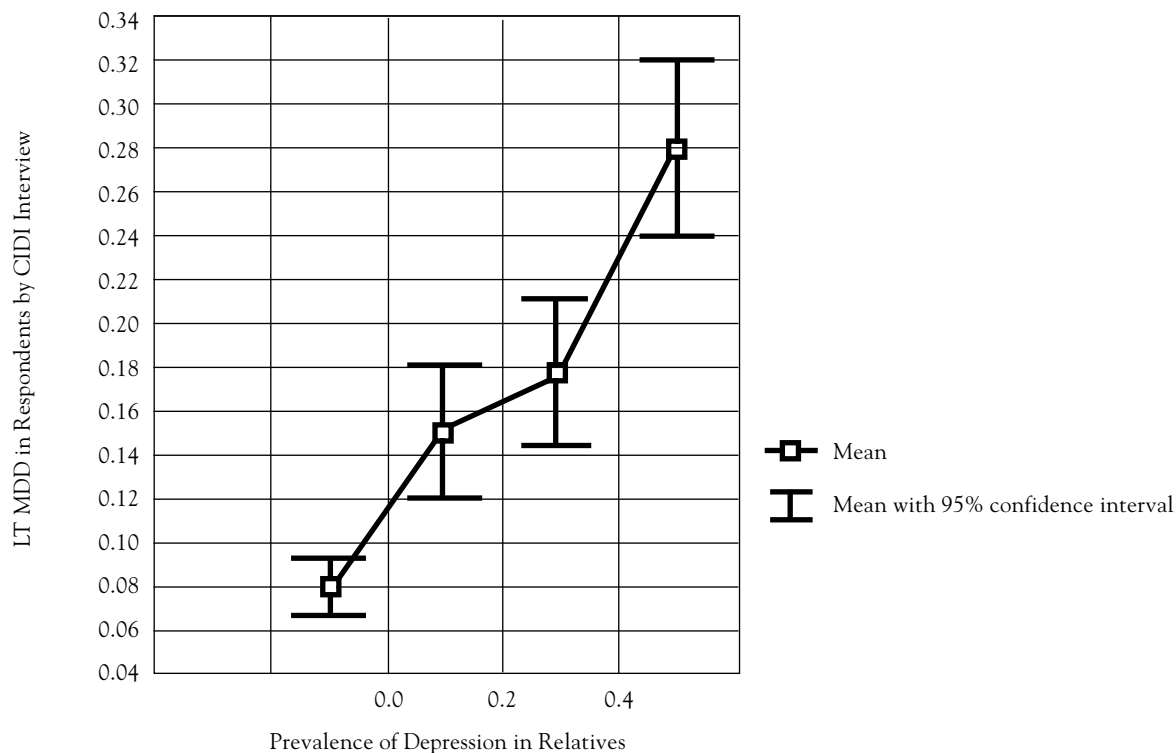


Figure 1. Percentage of NSAL respondents with lifetime MDD diagnosis as a function of depression prevalence in relatives (N = 5,184).

be obscured. Analyses are also needed to determine whether individuals with each disorder have different symptom presentations, depending upon the prevalence of the disorder among family members. The validity of the instrument also remains to be demonstrated in studies that simultaneously administer it with longer family history methods, preferably in settings where data are available from direct interviews with other family members. The BFPE does not aim to provide the same kind of information, however, but only a rough estimate of family prevalence that may reflect genetic vulnerability factors. Even an approximate measure will prove extremely valuable if it permits gathering information on familial risk factors that have so far been excluded of large survey studies.

Conclusions

The purpose of this paper was to describe a few of the methodological innovations introduced to address the challenges in surveying adequately physical health and psychiatric morbidity among large, nationally representative, samples of race and ethnic population

groups. The new NSAL adult cross-section and adolescent data will permit better national estimates of the status and life situation of black Americans across the entire range of socioeconomic and other demographic groupings in the population.

Acknowledgements

The NSAL is supported by the National Institute of Mental Health (NIMH; U01-MH57716) with supplemental support from the Office of Behavioral and Social Sciences Research (OBSSR) at the National Institute of Health (NIH) and the University of Michigan. The authors appreciate the assistance provided in all aspects of the study by the Program for Research on Black Americans faculty, and research assistants and associates (Deborah Coral, Jamie Abelson, Phyllis Stillman, Julie Sweetman, Nakesha Faison and Jennifer Lowe). Special thanks to Dr Tony N. Brown and the postdoctoral fellows of the Program for Research of Black Americans for their invaluable assistance with interviewer training and questionnaire development.

References

- Andreasen NC. The family history approach to diagnosis. In JE Mezzich, MR Jorge, IM Salloum (eds) *Psychiatric*

- Epidemiology: Assessment Concepts and Methods. Baltimore: The Johns Hopkins University Press, 1994, pp. xvii and 612.
- Andreasen NC, Endicott J, Spitzer RL, Winokur G. The family history method using diagnostic criteria. Reliability and validity. *Archives of General Psychiatry* 1977; 34(10): 1229–35.
- Bonczar TP, Beck AJ. Lifetime Likelihood of Going to State or Federal Prison. In Bureau of Justice Statistics, Special Report, 1997. Washington DC: Bureau of Justice Statistics, US Department of Justice.
- Bouchard JTJ, Loehlin JC. Genes, evolution, and personality. *Behavior Genetics* 2001; 31(3): 243–73.
- Braithwaite RL. Prisons and AIDS: a public health challenge. San Francisco: Jossey-Bass Publishers, 1996.
- Brugha TS, Bebbington PE, Jenkins R, Meltzer H, Taub NA, Janas M, Vernon J. Recall bias and major depression lifetime prevalence. *Social Psychiatry and Psychiatric Epidemiology* 1999; 38(6): 290–6.
- Caldwell CH, Jackson JS, Tucker MB, Bowman PJ. Culturally-competent research methods in African American communities: An update. In RL Jones (ed.) *Advances in African American Psychology: Theory, Paradigms, Methodology, and Reviews*. Hampton VA: Cobb & Henry Publishers, 1999, pp. 101–27.
- Helzer J, Robins L, McEvoy L, Spitznagel E, Stoltzman R, Farmer A, Brockington I. A comparison of clinical and diagnostic interview schedule: physician reexamination of lay-interviewed cases in the general population. *Archives of General Psychiatry* 1985; 42: 657–65.
- Jackson JS. *Life in Black America*. Newbury Park: Sage Publications, 1991.
- Jackson JS, Torres M, Caldwell CH, Neighbors HW, Nesse RN, Taylor RJ, Trierweiler SJ, Williams DR. The National Survey of American Life: a study of racial, ethnic and cultural influences on mental disorders and mental health. *IJMPR* 2004, 13(4): 196–207.
- Jackson JS, Tucker MB, Bowman PJ. Conceptual and methodological problems in survey research on black Americans. In WT Liu (ed.) *Methodological Problems in Minority Research*. Chicago IL: Pacific/Asian American Mental Health Research Center, 1982, pp. 11–39.
- Jackson JS, Williams DR. Surveying the Black American Population. In JS House, T Juster, RL Kahn, H Schuman, E Singer (eds) *A Telescope on Society: Survey Research and Social Science at the University of Michigan and Beyond*. Ann Arbor MI: The University of Michigan Press, 2003, pp. 393–438.
- Kendler KS. The genetic epidemiology of psychiatric disorders: a current perspective. *Social Psychiatry Psychiatric Epidemiology* 1997; 32(1): 5–11.
- Kendler KS, Prescott CA, Jacobson K, Myers J, Neale MC. The joint analysis of personal interview and family history diagnoses: evidence for validity of diagnosis and increased heritability estimates. *Psychological Medicine* 2002; 32(5): 829–42.
- Kendler KS, Thornton LM, Gardner CO. Stressful life events and previous episodes in the etiology of major depression in women: an evaluation of the 'kindling' hypothesis. *American Journal of Psychiatry* 2000; 157(8): 1243–51.
- Kessler RC, Berglund P, Demler O, Jin R, Koretz D., Merikangas KR, Rush AJ, Walters EE, Wang PS. The epidemiology of major depressive disorder: Results from the National Comorbidity Survey Replication (NCS-R). *JAMA* 2003; 289(23): 3095–105.
- Kessler RC, McGonagle K, Zhao S, Nelson C, Hughes M, Eshleman S, Wittchen H, Kendler K. Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States results from the National Comorbidity Survey. *Archives of General Psychiatry* 1994; 51: 8–19.
- Mannuzza S, Fyer AJ, Endicott J, Klein DF. Family Informant Schedule and Criteria (FISC). New York: Anxiety Disorders Clinic, New York State Psychiatric Institute, 1985.
- Messick DM, Mackie DM. Intergroup relations. *Annual Review of Psychology*, 1989, pp. 4045–81.
- Neighbors HW, Jackson JS, Campbell L, Williams D. The influence of racial factors on psychiatric diagnosis: a review and suggestions for research. *Community Mental Health Journal* 1989; 25(4): 301–11.
- Neighbors HW, Trierweiler SJ, Ford BC, Muroff JR. Racial differences in DSM diagnosis using a semi-structured instrument: the importance of clinical judgement in the diagnosis of African Americans. *Journal of Health and Social Behavior* 2003; 44: 237–56.
- Robins LN, Helzer J, Croughan J, Ratcliff K. National Institute of Mental Health Diagnostic Interview Schedule: its history, characteristics, and validity. *Archives of General Psychiatry* 1981; 38: 381–9.
- Robins L, Helzer J, Ratcliff K, Seyfield W. Validity of the diagnostic interview schedule, version II: DSM-III diagnosis. *Psychological Medicine* 1982; 12: 855–70.
- Schuman H. The random probe: a technique for evaluating the validity of closed questions. *American Sociological Review* 1966; 31(2): 218–22.
- Spitzer RL, Williams J, Gibbon M, First MB. The structured clinical interview for DSM-III-R (SCID): I. History, rationale, and description. *Archives of General Psychiatry* 1992; 49(8): 624–9.
- Trierweiler SJ, Neighbors HW, Munday C, Thompson EE, Binion VJ, Gomez JP. Clinician attributions associated with the diagnosis of schizophrenia in African American and non-African American patients. *Journal of Consulting and Clinical Psychology* 2000; 68(1): 171–5.
- Weissman MM, Wickramaratne P, Adams P, Wolk S, Verdelli H, Olfson M. Brief screening for family psychiatric history: the family history screen. *Archives of General Psychiatry* 200; 57(7): 675–82.

Wittchen HU, Üstün TB, Kessler RC. Diagnosing mental disorders in the community. A difference that matters? *Psychological Medicine* 1999; 29(5): 1021–7.

*Correspondence: James S. Jackson, Room 5010 Institute for Social Research, 426 Thompson Street, PO Box 1248, Ann Arbor, MI 48106-1248, USA.
Telephone (+1) 734 7632491.
Fax (+1) 734 6153212.
Email jamessj@umich.edu.*