The development and implementation of the National Comorbidity Survey Replication, the National Survey of American Life, and the National Latino and Asian American Survey

BETH-ELLEN PENNELL,[1] ASHLEY BOWERS,[1] DEBORAH CARR,[1] STEPHANIE CHARDOUL,[1] GINA-QIAN CHEUNG,[1] KARL DINKELMANN,[1] NANCY GEBLER,[1] SUE ELLEN HANSEN,[1] STEVE PENNELL,[1] MYRIAM TORRES[1]

1   Institute for Social Research, University of Michigan, Ann Arbor MI, USA

Abstract

This paper provides an overview of the development and implementation of the Collaborative Psychiatric Epidemiology Surveys (CPES): the National Comorbidity Survey Replication (NCS-R), the National Survey of American Life (NSAL), and the National Latino and Asian American Study (NLAAS). It describes the instrument development and testing phases, the development of training and other project materials, interviewer recruitment and training activities, and data collection procedures and outcomes. The last section offers recommendations for other researchers who undertake similar studies and who might benefit from the experiences learned in the development and operation of NCS-R, NSAL and NLAAS.

Key words: mental disorders, psychiatric epidemiology, Composite International Diagnostic Interview, questionnaire development, field procedures

This paper details the design and implementation of data collection procedures and systems for the Collaborative Psychiatric Epidemiology Surveys (CPES). The CPES consist of three national surveys funded primarily by the National Institute of Mental Health. These studies are: the National Comorbidity Survey Replication (NCS-R) (Kessler et al., 2004); the National Survey of American Life (NSAL) (Jackson et al., 2004); and the National Latino and Asian American Study (NLAAS) (Alegria, Takeuchi et al., 2004). This work was conducted by the Survey Research Center (SRC) of the Institute for Social Research at the University of Michigan from early 2000 through the end of 2003. The paper is organized into 10 sections. Section 1 provides a summary of the CPES' designs; Section 2 discusses questionnaire development; and Section 3 presents an overview of the development of study-specific materials. Section 4 describes interviewer recruitment and Section 5 details

interviewer training protocols. Section 6 addresses respondent recruitment and Section 7 highlights the organizational structure of field and central data collection staff, data collection procedures and sample outcomes. Sections 8, 9 and 10 focus on quality control procedures, data processing, and final conclusions and recommendations, respectively.

Overview

The primary objective of the CPES was to collect data about the prevalence of mental disorders, impairments associated with these disorders, and their treatment patterns from a representative sample of the adult population of the US. Secondary goals were to obtain information about language use and ethnic disparities, support systems, discrimination and assimilation in order to determine whether and how closely various mental health disorders are linked to social and

cultural issues. The core CPES questionnaire was based largely on the World Health Organization's (WHO) expanded version of the Composite International Diagnostic Interview (CIDI) developed for the World Mental Health (WMH) Survey Initiative, the WMH-CIDI (Kessler and Üstün, 2004).

The NCS-R was designed to collect data from a representative sample of English-speaking adults aged 18 or older living in the non-institutionalized civilian household population of the coterminous 48 states of the US (excluding Alaska and Hawaii). In 75% of the sampled households, one adult was selected at random to be interviewed. In the remaining 25% of households, two adults were randomly selected to be interviewed. Between February 2001 and April 2003, 9,836 interviews were conducted.

A nationally representative sample of black Americans of African and Caribbean descent was interviewed for the NSAL. In each sample household, one adult (aged 18 and older) was randomly selected to be interviewed. The NSAL sample yielded 6,199 adult interviews, including 1,006 white respondent interviews taken in predominantly black neighbourhoods. Interviewing started in early 2001 and was completed in the spring of 2003.

The National Latino and Asian American Study (NLAAS) began in May 2002 and ended in December 2003 and collected information from nationally representative samples of adults (aged 18 and older) from four Latino groups (Cuban, Mexican, Puerto Rican, and other adults of Latino descent); four Asian groups (Chinese, Filipino, Vietnamese, and other adults of Asian descent); and a control group of non-Hispanic, non-Asian white respondents. Interviews were completed with 4,864 adults. 2,095 Asian respondents, 2,554 Latinos, and 215 non-Hispanic, non-Asian white respondents participated in the study. In order to reduce costs, the interviewing of white controls was discontinued several months into the field period. Also, as part of an effort to address budgetary concerns, a design change was made mid-way through data collection to allow interviews with a second adult in Latino and Asian American households.

The primary mode of data collection for all three studies was in-person interviewing in the respondent's home. Telephone interviews were permitted if requested by the respondent. These requests were generally in response to respondent concerns about privacy; specifically, fear that the interview could be overheard by other members of the household, or concern about the presence of a stranger in the house (the interviewer). Telephone interviews were also used to complete long interviews that took two or more sessions to complete, or where costs to travel to the respondent's home were prohibitive.

The sample for all three studies consisted of primary sampling units selected with probabilities proportional to size. Data collection for the three studies was conducted in a total of 252 geographic areas or primary sampling units across the country. Only 50 of these areas were shared by all three studies, those representing the most densely populated areas of the country. These latter areas were selected into all three studies with certainty. In addition, there were 52 areas unique to NSAL and 18 areas unique to NLAAS. These unique areas were introduced to reflect the particular racial and ethnic focus of those studies. Figure 1 shows the distribution of the primary sampling units for each of the three studies and whether they were unique or common to each study. See Heeringa et al. (2004), for a detailed description of the sample design for each study.

Although each project was managed by a separate staff, the data collection model was the same. Project managers and support staff were located at SRC in Ann Arbor, Michigan, and there were teams of interviewers located throughout the US who were supervised by team leaders and regional field managers. The staff of the three projects worked closely together and whenever possible used similar procedures and materials.

Questionnaire development

All three projects used a modified version of the WMH-CIDI, which had been developed over the course of more than a year by an international group of collaborators, as described by Kessler and Üstün (2004). The design of the WMH-CIDI involved modifications and additions to the existing WHO-CIDI. The CIDI was an expansion of the Diagnostic Interview Schedule (DIS), the first standardized psychiatric diagnostic interview developed for administration by lay interviewers. The CIDI was designed to produce diagnoses based on WHO International Classification of Disease (ICD) criteria, while diagnoses from the DIS could only be made based on American Psychiatric Association (APA) Diagnostic and Statistical Manual (DSM) of Mental Disorders criteria. As summarized by Wittchen (1994), Kessler and Üstün (2004), DIS and CIDI validity studies suggest that there is a significant correlation between
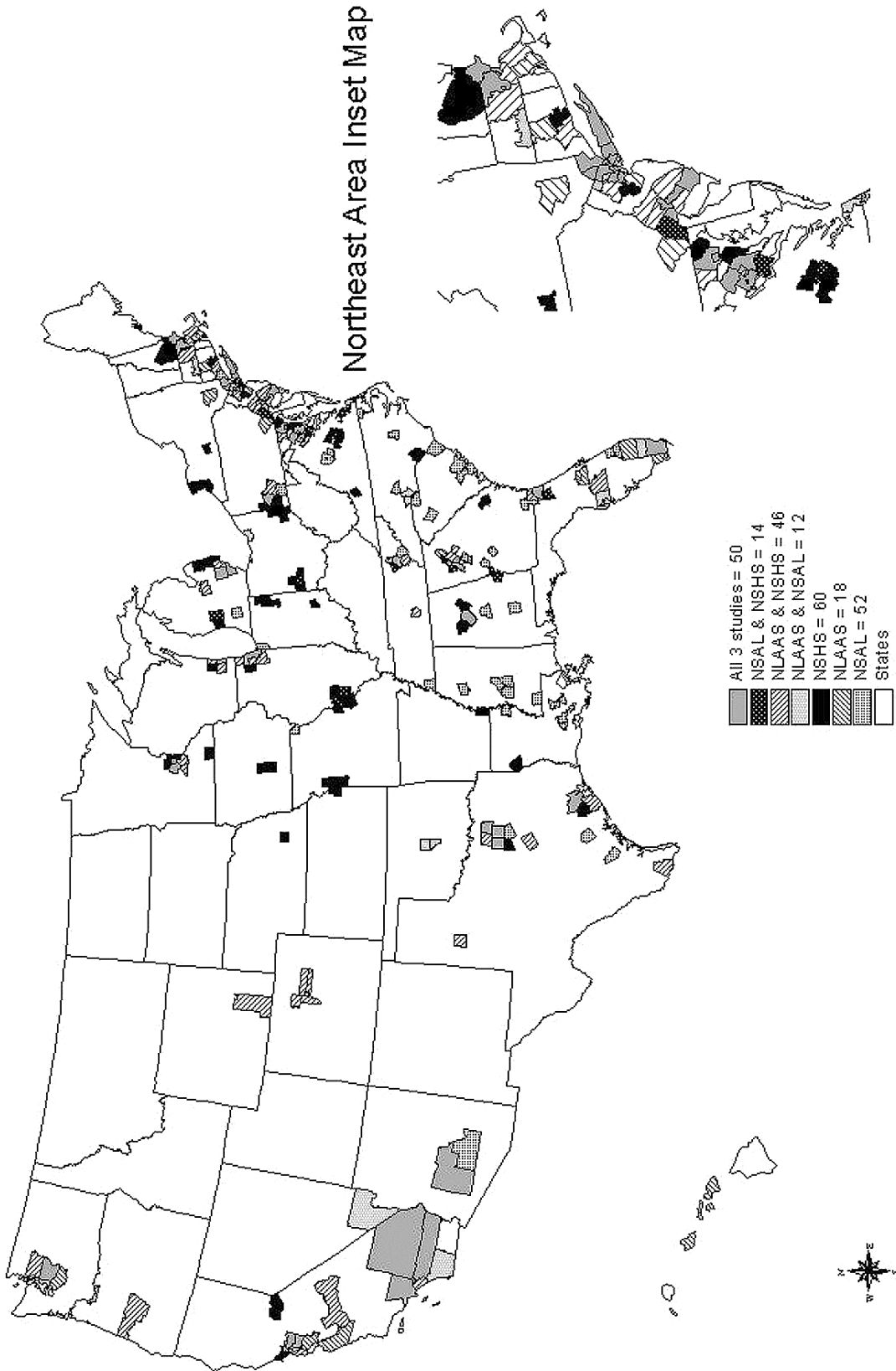
Figure 1. Sample distribution map for NCS-R, NSAL and NLAAS.

diagnoses based on data collected using the DIS and CIDI and diagnoses made by clinicians who reinterview a sample of respondents. Researchers who were involved in the design of the WMH-CIDI were interested in continuing to enhance the quality of survey measurement. As part of the WMH-CIDI development process, debriefing interviews were conducted with CIDI respondents and several methodological improvements were made to the instrument based on findings from the debriefing interviews. See Kessler et al. (2000) Kessler and Üstün (2004) for a detailed discussion of the content additions and methodological enhancements that were incorporated in the WMH-CIDI instrument.

The WMH and CPES questionnaires were administered using computer-assisted interviewing (CAI). The instruments were programmed using Blaise, a CAI software package developed by Statistics Netherlands and used by many government statistical agencies and large survey research organizations worldwide (Blaise Survey Processing System, Version 4.5). Blaise software is specifically designed to accommodate very complicated questionnaire skip patterns and sub-sampling algorithms. Potential drawbacks of using Blaise include its cost and the requirement for highly trained programmers to write the code for complex surveys.

The developmental work on the WMH questionnaire continued with the preparation of the NCS-R and NSAL instruments. The various instruments were created in a modular fashion. Each section progressed through the developmental lifecycle of initial question drafts, expert review, programming, and then testing of the module by both programmers and substantive experts. At each step, revisions were made and tracked. Since multiple programmers worked on the project, developing and adhering to common programming standards and conventions was important to facilitate the eventual linking of sections. Once the modules were concatenated, the instrument was again reviewed and tested by substantive experts before proceeding to traditional pretests involving professional interviewers and recruited respondents.

The design of the NLAAS instrument benefited from the previous work completed for NCS-R and NSAL, but the translation of the instrument into Spanish, Vietnamese, Tagalog (Filipino native language) and Chinese (Mandarin), and the subsequent integration of those languages into the CAI application, contributed to an equally challenging task. The integration of Vietnamese and Chinese translations presented additional obstacles during the development phase. Although Blaise accommodates the integration of multiple languages very well when using Latin-based alphabets (American National Standards Institute (ANSI) standard using 1-byte, 8-bit representation with 256 unique characters), additional steps, which varied depending on the language, were necessary to display non-Latin alphabets in Blaise (Groeneveld, 2003).

Typically, the process of integrating various languages into the application involves a 'cut-and-paste' procedure. Since the paper version of the NLAAS instrument was more than 800 printed pages, such a labour-intensive procedure would have taken months to complete. To expedite the process, SRC developed a foreign language merge utility to integrate the languages (Cheung and Liu, 2003). This utility also provided additional modifications that were necessary for the Blaise programming syntax as well as Chinese language encoders.

Each step in the development process involved many instrument modifications. To facilitate the questionnaire development and review process, paper versions of the questionnaires were maintained in parallel to the CAI versions.

To facilitate communication and document sharing among the project team members at the four participating institutions (University of Michigan, Harvard University, Cambridge Health Alliance, and University of Washington), a secure Web-based collaborative knowledge management system was used. The system provided distributed team members with direct access to the current and previous versions of all questionnaire materials. Each change made to one of the questionnaires or related materials was tracked by the system. Every night the system would query the Web server for any modifications and automatically generate an email notifying all project staff of the changes.

Once all of the questionnaire modules were concatenated and advanced through the second developmental phase, the sample management system, which had been developed in parallel to the questionnaire, was tested and integrated. The sample management system, SRC's proprietary system SurveyTrak, delivers sample addresses and questionnaire versions to the field staff and allows interviewers to transmit completed questionnaires back to the central office along with information about call attempts and sample outcomes and interviewer reports of time and expenses.

Initial questionnaire design and testing

The questionnaire design and testing phase for each project spanned approximately one year. The initial goal was to finalize the NCS-R questionnaire first, particularly the psychiatric disorder sections, in order to make these sections available to NSAL and NLAAS. However, the time required to finalize the NCS-R questionnaire was lengthened because development of the NCS-R questionnaire and the World Mental Health Survey Initiative questionnaire occurred in tandem, resulting in many revisions to the core questionnaire over this period. Management of this questionnaire development activity was made somewhat more manageable by the SRC's involvement as the Data Coordination Center for the WMH Survey Initiative and by drawing upon the same core technical staff for WMH and the CPES studies.

During the instrument development phase, frequent telephone conference meetings were held between staff from NIMH and each of the participating institutions. One of the primary objectives of this early phase was to decide on common content areas across the studies. Content shared across the studies occurred primarily in the modules measuring the psychiatric disorders. Table 1 displays the sections included in each study with the sections that were common across multiple studies shaded in grey. The average time in minutes to administer each section is also provided in Table 1.

Study-specific modules added

The content of the core sections of the instrument remained essentially the same across the studies with the exception of additional questions about service use that were added to each diagnostic module in NSAL and NLAAS. The demographic sections were also comparable across the studies. Each study decided independently whether to keep, omit, or change non-core sections. Both NSAL and NLAAS added content areas located before and after the disorder sections, which addressed issues unique to the populations of those two studies.

Questionnaire testing and pretesting

The goal of the questionnaire testing phase was to identify wording and comprehension problems, logical inconsistencies, formatting issues, and programming errors and to measure the time required to administer each section. Figure 2 shows the flow of the instrument development and testing process for NCS-R and NSAL.
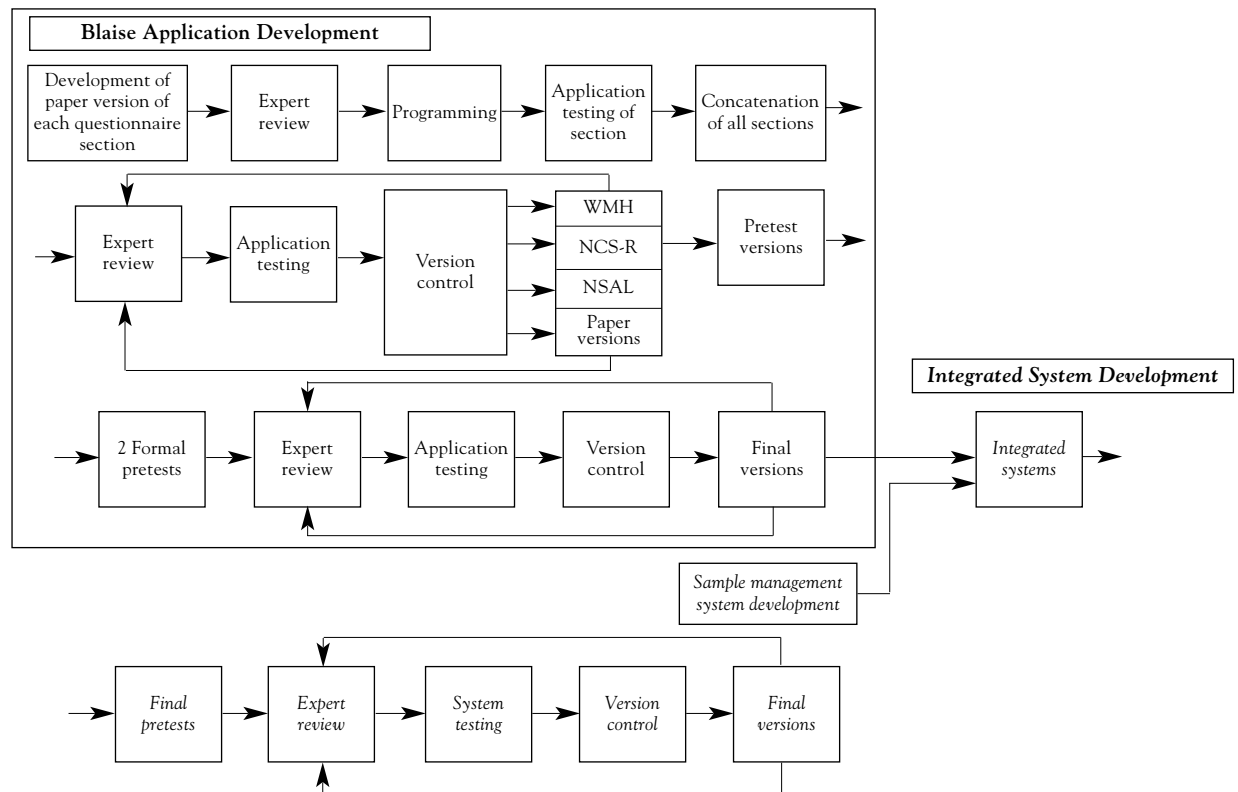

Figure 2. Instrument development and testing process for NCS-R and NSAL.

Table 1. Questionnaire sections and average time to administer section by study

NCS-R		Time	NSAL		Time	NLAAS		Time
1.	Household Listing	5:12	0.	Household Listing	4:41	1.	Household Listing	n/a
2.	Screening	16:55	8.	Screening	9:21	3.	Screening	18:52
3.	Depression	8:04	9.	Depression	6:45	4.	Depression	10:20
4.	Mania	5:52	10.	Mania	4:58			
5.	Irritable Depression	2:43				5.	Irritable Depression	5:41
6.	Panic	4:55	11.	Panic	4:51	6.	Panic	5:08
7.	Specific Phobia	7:13						
8.	Social Phobia	7:45	12.	Social Phobia	8:42	7.	Social Phobia	9:32
9.	Agoraphobia	6:46	13.	Agoraphobia	7:11	8.	Agoraphobia	8:02
10.	Generalized Anxiety Disorder	5:33	14.	Generalized Anxiety Disorder	4:46	9.	Generalized Anxiety Disorder	6:22
11.	Intermittent Explosive Disorder	3:07				10.	Intermittent Explosive Disorder	3:33
12.	Suicidality	0:49	15.	Suicidality	0:39	11.	Suicidality	0:37
17.	Substance Use	6:33	16.	Substance Use	5:23	17.	Substance Use	5:43
13.	Services	4:58	32.	Services	3:16	13.	Services	4:04
14.	Pharmacoepidemiology	3:03	17.	Pharmacoepidemiology	2:31	14.	Pharmacoepidemiology	4:59
15.	Demographics	3:06				15.	Demographics	6:32
16.	Personality	5:15	18.	Personality	5:21	16.	Personality	1:46
18.	Post-Traumatic Stress Disorder	10:05	19.	Post-Traumatic Stress Disorder	9:06	18.	Post-Traumatic Stress Disorder	10:33
19.	Chronic Conditions	12:02				20.	Chronic Conditions	14:27
20.	Neurasthenia	1:02				19.	Neurasthenia	0:49
21.	30-Day Functioning	6:55				21.	30-Day Functioning	8:34
22.	30-Day Symptoms	7:27	20.	30-Day Symptoms	3:14			
23.	Tobacco	3:46	21.	Tobacco	0:36			
24.	Eating Disorders	1:28	22.	Eating Disorders	1:07	22.	Eating Disorders	1:20
25.	Premenstrual Syndrome	2:26	23.	Premenstrual Syndrome	2:04	24.	Premenstrual Syndrome	2:15
26.	Obsessive-Compulsive Disorder	2:53	24.	Obsessive-Compulsive Disorder	1:55			
27.	Psychosis	2:38	25.	Psychosis	2:19	25.	Psychosis	2:48
28.	Gambling	3:15	26.	Gambling	1:14			
29.	Worries and Unhappiness	3:29						
30.	Employment	10:06	6.	Employment	5:34	26.	Employment	11:44
31.	Finances	3:52				27.	Finances	5:53
32.	Marriage	4:44				28.	Marriage	4:21
33.	Children	2:30				29.	Children	3:36
34.	Social Networks	2:47				30.	Social Networks	2:49
35.	Adult Demographics	6:30				31.	Adult Demographics	7:02
36.	Childhood Demographics	3:95				33.	Childhood Demographics	0:34
37.	Childhood	9:08						
38.	Attention Deficit Hyperactivity Disorder	3:09	28.	Attention Deficit Hyperactivity Disorder	3:23			
39.	Oppositional-Defiant Disorder	2:01	29.	Oppositional-Defiant Disorder	2:30			
40.	Conduct Disorder	3:02	30.	Conduct Disorder	2:55	34.	Conduct Disorder	3:08
41.	Separation Anxiety Disorder	4:24	31.	Separation Anxiety Disorder	5:07			
42.	Family Burden	2:31						
43.	Perceptions of the Past	2:57						
44.	Terror	1:06	38.	Terror				
45.	Respondent Contacts	n/a						

Table 1. (contd)

NCS-R	Time	NSAL	Time	NLAAS	Time
46. Interviewer Observations	n/a	43. Interviewer Observations	n/a	41. Interviewer Observations	n/a
47. Dementia – paper only	n/a			42. Dementia – paper only	n/a

<table>
<tr><td colspan="6" align="center">Study-specific questionnaire sections</td></tr>
<tr><td>NCS-R</td><td>Time</td><td>NSAL</td><td>Time</td><td>NLAAS</td><td>Time</td></tr>
<tr><td></td><td></td><td>1. Neighbourhood</td><td>3:36</td><td>2. Language of Interview</td><td>2:14</td></tr>
<tr><td></td><td></td><td>2. Religion</td><td>7:04</td><td>12. Service Use</td><td>1:16</td></tr>
<tr><td></td><td></td><td>3. Digit Symbol and SPMSQ</td><td>4:14</td><td>23. Nervous Attacks</td><td>2:55</td></tr>
<tr><td></td><td></td><td>4. Psychological Resources and Health</td><td>18:46</td><td>32. Social Position</td><td>1:42</td></tr>
<tr><td></td><td></td><td>7. Family and Friends</td><td>6:57</td><td>35. Migration Status and History</td><td>0:31</td></tr>
<tr><td></td><td></td><td>27. Family History</td><td>3:09</td><td>36. Language Proficiency and Preference</td><td>1:17</td></tr>
<tr><td></td><td></td><td>33. Use of Help Resources</td><td>3:18</td><td>37. Family Cohesion</td><td>2:35</td></tr>
<tr><td></td><td></td><td>34–35. Group and Personal Identity, Discrimination</td><td>12:08</td><td>38. Discrimination</td><td>3:05</td></tr>
<tr><td></td><td></td><td>36–41. Personal Data</td><td>15:20</td><td>39. Context of Exit Questions</td><td>2:08</td></tr>
<tr><td></td><td></td><td>42. Technology and Roots</td><td>3:19</td><td>40. Acculturative Stress</td><td>2:31</td></tr>
</table>

Notes: The number associated with each section represents the sequential order of that section within the questionnaire for the study. Sections that are highlighted represent 'core' diagnostic sections that were administered across multiple studies.

Questionnaire testing

Individuals familiar with the content of the questionnaire from SRC and the other participating institutions conducted much of the preliminary testing. The testing process was structured to ensure that each section of the questionnaire and each path within each section were thoroughly reviewed. At least two staff members working independently tested each module. Any problems in the flow and skip logic were recorded by the testers and subsequently reviewed by the testing coordinator in order to diagnose the source of the problem and to formulate solutions. When necessary, modifications were made. Each module was tested as many times as necessary using the same procedures to identify and resolve all problems.

Several features of the questionnaires' architecture made the testing procedures particularly challenging. First, the core questionnaire included many intricate skip patterns, which controlled the flow of questions.

Second, some sections of the core questionnaire were randomly administered to only a subgroup of respondents in order to reduce the overall length of the interview and respondent burden (NCS-R) (Kessler et al., 2004; Kessler and Üstün, 2004). To verify that these skip sequences worked correctly, each possible path had to be tested multiple times.

Questionnaire pretesting

Three formal pretests of the questionnaire were conducted for NCS-R and NSAL from September to November 2000. The first two pretests were performed using the computerized questionnaires as stand-alone applications; the third pretest tested the integrated sample management application. For convenience, all NCS-R and NSAL pretests were conducted near SRC's offices in Michigan. Pretest respondents were identified through random household selection in specific neighbourhoods in the greater Detroit, Michigan area where demographic characteristics of the area

accommodated the requirements of each study. Patient populations with particular mental health disorders were also included in the pretest to ensure that all sections of the questionnaire were rigorously tested.

Two formal pretests were also completed for NLAAS. The first pretest was conducted in Fall 2001 in the Detroit area and in several locations in the southwest US where a higher proportion of Spanish-speaking Latino respondents reside. The second pretest was implemented in Spring 2002 in selected areas in California among Asian respondents who spoke Chinese, Vietnamese, or Tagalog. Probability methods were not used to sample households for the pretest, although respondents within households were selected randomly. Pretest interviewers were instructed not to conduct interviewing in any of the neighbourhoods in which the sample for the main study was located.

A total of 30 to 50 pretest interviews were completed during each round of pretesting for each of the three surveys. Pretest interviewers, programmers, researchers, and project management staff participated in debriefing meetings after each pretest. These discussions proved to be valuable in identifying problems with wording, interviewer instructions, and programming errors.

Beyond testing for wording problems, logical inconsistencies, and programming errors, the pretests were also used to obtain section timings so that an overall average administration time could be estimated. Since the time required to complete the questionnaire was a function of the number of items endorsed in the screening section, it was important to recruit respondents who would be routed into each section in order to obtain stable section-specific time estimates. The administration time for each section was then weighted by the anticipated percentage of general population respondents who would be routed into each section in order to obtain an overall estimate of the administration time for each study.

The initial length of time required to complete the interview was much longer than expected. In order to bring administration times more in line with budget assumptions, reductions in questionnaire content were made after each pretest. Deciding what content to cut was a lengthy process because of the breadth and scope of the research goals. Programming these extensive revisions after each pretest was also a time-consuming task.

For NLAAS, the testing and revision process was complicated by the need to make post-translation changes in five different language versions of the instrument. Figure 3 displays the instrument development process for NLAAS.

Revisions to questionnaires after pretesting
After formal pretesting ended, revisions to questionnaire content continued in all three studies. Some of these late revisions were based on the experiences of other countries that were part of the World Mental Health Survey Initiative. Interviewer training dates were fixed so training materials had to be developed and finalized while questionnaire revisions and additional testing continued.

In the end, it took on average approximately 2 hours (median = 2.0) to administer the NCS-R instrument and 2.4 hours (median = 2.2) and 2.7 hours (median = 2.4) for NSAL and NLAAS, respectively. The source code for the NCS-R Blaise instrument has in excess of 72,000 lines of programming code. The complexity of the program is also reflected in the size of the data files, which contain 5,774 output variables for NCS-R, 3,893 for NSAL and 4,317 for NLAAS.

Translation procedures and issues
The principal investigators for NLAAS took primary responsibility for translating the English version of the questionnaire into each of the non-English target languages used in this study. Once the translated versions had been drafted and provided to SRC, native speakers as well as faculty at the University of Michigan who teach these languages were hired to review the questionnaire and conduct additional testing and evaluation. The results of each review were prepared in written form and forwarded to the NLAAS principal investigators, who made appropriate modifications. Two papers discuss the development of the NLAAS instrument in more detail (see Alegria, Takeuchi et al., 2004; Alegria, Vila et al., 2004).

Interviewer materials
Table 2 provides an overview of the data collection materials that were developed for the CPES studies. Materials were standardized across the studies whenever possible. All NLAAS study materials were translated into Spanish, Vietnamese, Tagalog, and Chinese.

Interviewers
Each project required a national field staff of several hundred interviewers. The following section describes
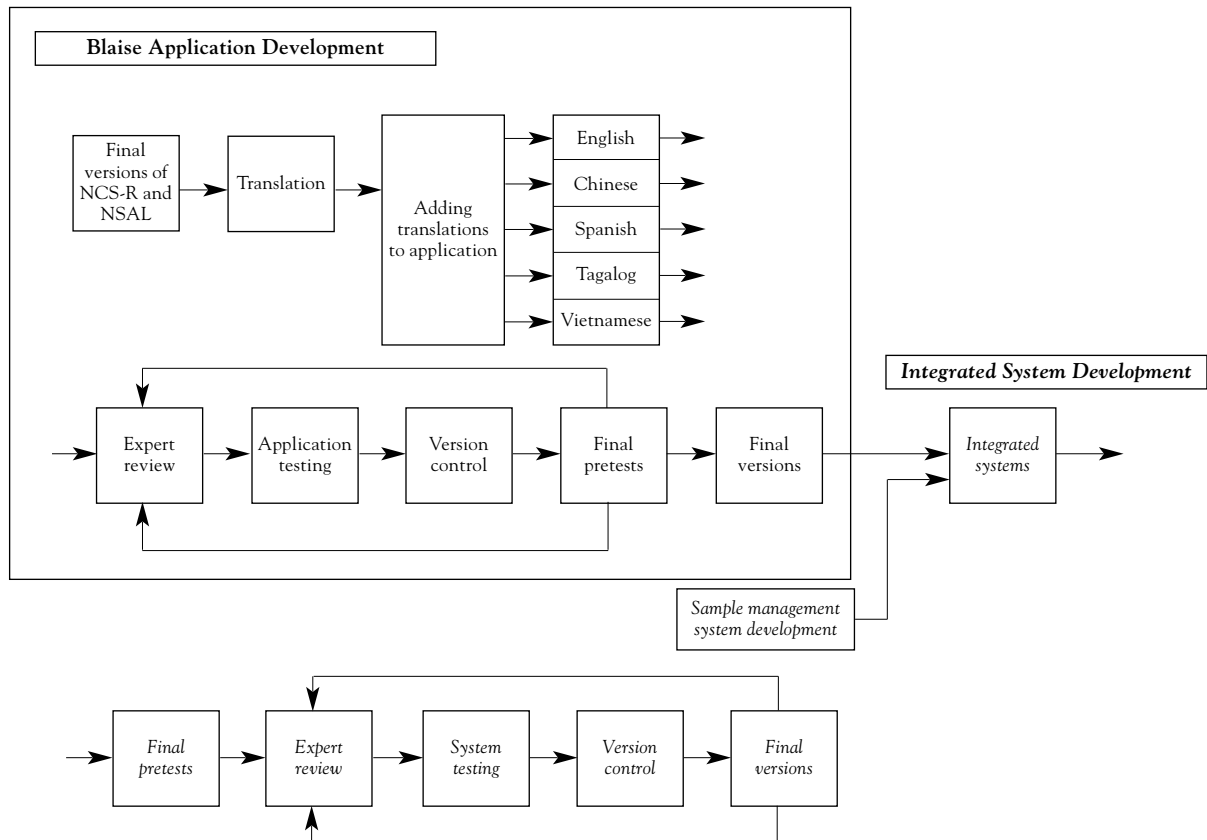
Figure 3. Instrument development and testing process for NLAAS.

interviewer recruitment and briefly summarizes the characteristics of interviewers working on the projects. Interviewer remuneration is also discussed.

Recruitment
More than 900 interviewers were recruited and trained to work on the three studies. The geographic locations of the NSAL and NLAAS samples, coupled with the specific race and language requirements, created the need for an unusually high number of newly hired interviewers.

Hiring needs
Staffing requirements were determined by examining the sample locations of each project and assessing the availability of the SRC's on-staff interviewers. For NSAL, an attempt was also made to match interviewer and respondent race, whereas NLAAS had the goal of accommodating the language preferences of sample households (which was frequently tantamount to

matching ethnicity). In addition, both NSAL and NLAAS included oversamples of target populations in areas of the country outside the SRC's national sample frame, thus adding to the need to recruit new interviewers to work in these areas.

Recruitment procedures
Supervisory staff living in or near sample areas were responsible for meeting recruitment goals in those areas. Standard SRC recruitment procedures included placing newspaper ads and posting flyers in strategic locations, such as community centres, libraries, college campuses and employment offices. Given the timing of the study – just after the completion of the US decennial census – it was often possible to recruit interviewers who worked for the US Census Bureau.

For NSAL and NLAAS, standard recruitment methods were supplemented with more targeted procedures. For NSAL, this included circulating flyers at churches in predominantly African American and Caribbean

Table 2. Interviewer materials for CPES studies

Study material	Description

Field interviewer manual	Contents:
	• Study overview (e.g., research goals, SRC organizational structure, funding sources for study)
	• Study-specific interviewing procedures
	  ○ Guidelines on contacting households and encouraging cooperation
	  ○ Eligibility screening procedures
	  ○ Importance and necessity of informed consent and consent procedures
	• Interviewer instructions for administration of the questionnaire
	• Other miscellaneous topics (e.g., working with computer hardware and software, quality control measures, procedures for reporting payroll, expenses and productivity)

Interviewer training workbook	Contents: • Training slides and exercises • Summary sheets or checklists for complex procedures • Copies of respondent letters, consent documents, and incentive payment receipt

Sample information	• Detailed maps and listing of addresses within each area to assist in locating households
	• A coversheet for each address contained the following information:
	  ○ Sample household address, unique identification number, interviewer to whom case was assigned
	  ○ Study introduction
	  ○ Screening and respondent selection instructions
	  ○ Informed consent statement (for NLAAS)
	  ○ Space for date, time and outcome of every call attempt
	  ○ Notes to assist in future calls to the household

Identification badge	Picture identification badge for interviewers that confirmed their employment with SRC
Various letters	• Introductory letter and study brochure sent prior to initial contact attempt • 'Not at home' letter for households that could not be reached after multiple visits • Letter for local law enforcement agencies that notified the agency that interviewing staff would be working in the area • Press release for the local news media • Letters of support from the director of NIMH • Project updates highlighting preliminary findings
Question by question instructions	• Instructions or definitions for interviewers that were provided on-screen or accessible by pressing a function key and accessing online help file • Instructions were also included in the field interviewer manual

Respondent booklet	• Reference for respondents to use during the interview. Booklet contained the following information:
	  ○ Response categories for questions with complex response options
	  ○ Response categories for long lists of options
	  ○ Visual aids to assist respondent's understanding of response options

neighbourhoods. For NLAAS, recruitment ads and flyers were translated into each study language and advertisements were placed in newspapers targeting the Latino, Vietnamese, Chinese and Filipino populations.

When NCS-R and NSAL were first fielded, unemployment rates in the US were at historically low levels, which significantly affected recruitment efforts. Where appropriate, labour market analyses were

conducted and interviewer starting pay rates in major labour markets were reviewed and increased to attract a larger applicant pool and also to recognize the complex nature of these studies.

Interviewer selection procedures

Interviewer candidates were instructed to call a toll-free telephone number or to access a Web page to request an employment application. Based on review of the candidates' qualifications and performance in a telephone screening interview, the best candidates were invited to a face-to-face interview. During the recruitment process, the rigours of the job were emphasized to help applicants evaluate whether they would be able to meet the responsibilities of the position. Before an applicant was offered a position, a criminal background check was conducted.

Language certification procedures

Language fluency was evaluated for all NLAAS interviewer candidates before an offer of employment was made. A candidate's English speaking ability was evaluated during the recruitment process by SRC staff. Non-English fluency among bilingual candidates was initially evaluated by a language testing service and later by on-staff personnel previously certified by the language service. Spoken and written language fluency was evaluated for accuracy, complexity and ease of use. Only candidates who scored in the passing range on the language assessment were hired.

Interviewer remuneration

As is the norm with all studies undertaken at SRC, interviewers were paid by the hour, rather than by the number of interviews completed. This payment approach recognizes the variability in each interviewer's work assignment (travel distances, varying levels of respondent cooperation, varying levels of access, and so forth). This method of payment also helps to motivate interviewers to work the most challenging cases and to take the time to use appropriate techniques (probing, seeking clarification) during the interview. In these studies, compensation based on hours worked also took into consideration the need to screen for eligible respondents as well as variations in the length of interview (from less than 1 hour to more than 10 hours). For these reasons, the standard payment approach was also followed in the WMH Survey Initiative.

Hourly pay rates for interviewers varied across the country due to local labour market conditions. Within a labour market, pay rates could differ depending on the experience of the interviewer, and for NLAAS, whether the interviewer was bilingual.

Training reimbursement and pay

All travel expenses incurred as part of an interviewer's training were covered, including transportation, hotel room, and meals. Due to expected high attrition rates just before training ended but before any interviews were conducted, trainees were paid approximately half of their regular hourly rate after training ended and received the balance as a 'bonus fee' once they were certified (see below) and completed five production interviews.

Production reimbursement and pay

As shown at the bottom of Table 3, interviewer hourly pay rates ranged from $8.50 USD to $19.90 USD for NCS-R and NSAL, and $9.00 USD to $25.00 USD for NLAAS. Interviewers who were asked to travel into areas outside of their assigned region received a 'travel differential' of $2.00 USD per hour for the time worked while on overnight travel status. In an attempt to encourage productive interviewers on NSAL to travel more extensively to other areas (generally to urban areas where resistance was highest), the differential was increased to $5.00 USD per hour. Interviewers were also reimbursed for expenses associated with their work including mileage, supplies, postage, and telephone charges.

Interviewer characteristics

As shown in Table 3, the interviewer workforce was primarily female. Nine out of 10 interviewers had completed some college coursework; and more than a third were college graduates. The diversity of the study populations for NSAL and NLAAS is also reflected in the interviewer workforce.

Attrition

Interviewer turnover rates for NCS-R, NSAL, and NLAAS were high. For NCS-R, the attrition rate was approximately 55%, for NSAL approximately 75%, and for NLAAS 61%. During this same time period, attrition rates on other SRC studies ranged from 25% to 40%.

There were many factors that contributed to the attrition rates experienced, including historically low unemployment during the initial hiring phase, the need to hire a large number of new interviewers, a

Table 3. Interviewer characteristics by study

	NSAL(%)	NCS-R(%)	NLAAS(%)
Gender			
Male	23.8	17.3	28.6
Female	76.2	82.7	71.4
Education			
Some high school	0.8	0.3	0.0
High school graduate	8.6	10.0	9.4
Some college	38.9	40.2	40.0
College graduate	30.3	31.5	35.7
Master's degree	13.9	14.0	12.7
PhD	1.6	1.9	1.3
Missing data	5.7	2.4	0.7
Ethnicity			
White	24.1	75.3	43.5
Black	61.9	16.1	14.9
American Indian	0.4	0.6	1.0
Asian	0.4	1.2	14.3
Hawaiian	0.0	0.0	8.4
Other	6.1	3.9	23.4
None given	7.0	3.8	1.6
Latino	6.1	6.0	36.7
Pay range			
Low	$8.50 USD	$8.50 USD	$9.00 USD
High	$19.90 USD	$19.70 USD	$25.00 USD
Total interviewers	(329)	(342)	(275)

lengthy field period, and the demanding requirements of the position. The concentration of sample in major metropolitan areas, where recruiting and retaining staff is always more difficult, and where the environments are generally more challenging for interviewing, also probably played a role in the high attrition rates.

Training programme

Table 4 summarizes the training sessions each study held for field staff. A total of 342 interviewers were trained in one of six training sessions for NCS-R. There were four training sessions held for NSAL field staff, with a total of 329 interviewers trained. For NLAAS, a total of 275 interviewers completed training in one of four training sessions. Study-specific training lasted 5 to 7 days, depending on the study and which components were being covered. In addition, new interviewers received two additional days of training in general interviewing techniques (GIT) (Guenzel et al., 1983). These techniques included basic instruction on standardized interviewing

procedures and general administrative tasks. All training sessions were held at conference hotels in Ann Arbor, Michigan.

Training session overview

Study-specific training sessions consisted of five main components: (1) instruction on household eligibility and respondent selection procedures; (2) questionnaire training, which included a section-by-section review of each module of the questionnaire, followed by question and answer sessions and two hour practice sessions; (3) computer training and practice sessions; (4) review of interview procedures and study materials; and (5) mock interviews in which interviewing and administrative tasks were integrated to model realistic interviewing experiences. To better convey the content and to engage the training participants, trainers used a variety of formats, including large and small group lectures, round-robin practice sessions, mock interviews and one-on-one help sessions. Participants were given homework

Table 4. Training schedule and number of trainees by study

Study	Date	Number of trainees
NCS-R Training #1	January 2001	134
NCS-R Training #2	April 2001	57
NCS-R Training #3	October 2001	33
NCS-R Training #4	March 2002	70
NCS-R Training #5	July 2002	27
NCS-R Training #6	January 2003	21
	NCS-R total interviewers	342
NSAL Training #1	January 2001	92
NSAL Training #2	March 2001	102
NSAL Training #3	October 2001	49
NSAL Training #4	March 2002	86
	NSAL total interviewers	329
NLAAS (English and Spanish)	April 2002	94
NLAAS (English, Mandarin, Cantonese, Vietnamese, and Tagalog)	June 2002	83
NLAAS (English and Spanish)	October 2002	64
NLAAS (English second adult phone interviews)	February 2003	34
	NLAAS total interviewers	275

assignments, which the trainers reviewed to identify interviewers who were having problems with the computer hardware or software.

For later trainings, experienced interviewers served as trainers for the two days of general interviewer training. The new interviewers benefited from the descriptions of the experiences of these interviewers, who were able to provide tested and concrete suggestions on how best to handle all aspects of the job.

Specialized training considerations
All three studies provided training in sensitivity to cultural, racial and socioeconomic diversity that would be encountered while conducting face-to-face interviews. Additional training was also provided on how to interview on sensitive or potentially embarrassing topics. Finally, because some of the questionnaire topics covered subjects that could reveal information about pending harm to the respondent or others, interviewers were trained on their legal obligations and on how to handle these rare but critical situations.

Additional instruction was given for completing the complex screening procedures for NSAL and NLAAS. Finally, all NLAAS interviewers were instructed on how to use the translated materials.

Training certification procedures
Interviewers from all three studies were certified before they were permitted to contact sample households. The certification process assessed whether an interviewer could successfully apply interviewing principles and practices and administrative procedures in a mock interview with a supervisor (team leader) acting in the role of respondent. Depending on their performance during this mock interview, interviewers were certified to work immediately, given the opportunity to retake the certification test after self-study, or removed from the project. The majority of interviewers were certified in person at the conclusion of training.

The number of interviewers who were unable to meet the certification requirements was minimal. No more than 12 trainees failed certification in any of the three studies.

Field management staff training

Supervisors are usually recruited from SRC interviewing staff who perform at a high level of effectiveness and efficiency and display leadership qualities and training abilities. Supervisors also participate in a training programme that covers their responsibilities and tasks. For NCS-R and NSAL trainings, interviewer supervisors (team leaders) received training at the same time as the interviewers. NLAAS conducted a separate team leader training before interviewer training and provided a second, more in-depth training session for new team leaders. NSAL also provided a three-day training session for team leaders part way through the field period to assist them in helping interviewers as they encountered the more challenging resistant and difficult to reach cases.

Team-leader training was two to three days in length and included in-depth discussions of study goals and procedures; effective methods for managing a team and coaching interviewers; review of production and administrative reports; quality control procedures; personnel management; time management skills; and training in the use of spreadsheet, word processing, and email programs.

Respondent recruitment

As discussed in detail in this section, determining eligibility, which involved a complex procedure of sorting household members into various categories defined by age, gender, race and ethnicity, was often a challenge for interviewers. A number of response enhancement strategies used in the three studies are also described in this section.

Screening

After introducing the study, the interviewer requested information about each adult living at the address on that day in order to determine his or her eligibility. Following a set of procedures that reflected each study's eligibility criteria (see Table 5), the interviewer would identify which, if any, adults in the household were eligible for the study.

One respondent (or additional respondents depending on the study design) was randomly selected from a sequential listing of the eligible household members using the Kish selection procedure (Kish, 1965). It should be noted that the Kish procedure produces a small bias for households with more than six adults. The procedure underrepresents young females in these large households. However, it only affects a very small number of households (approximately one household in 1,000) (Kish, 1965). Household listing and respondent selection information was subsequently transcribed from paper coversheets into the Blaise interviewing application.

Once a respondent was selected, no substitutions were allowed. Every effort was made to interview the selected respondents. Appointments were scheduled at times convenient to the respondent, the respondent's language preference was accommodated, and interviewers returned to the household as many times as needed to complete an interview.

On average, it took 7.1 contact attempts (median = 5) to complete the primary respondent interview for NCS-R, and another 4.7 (median = 3) to complete the second respondent interview. A contact attempt included in-person visits as well as telephone calls. For NSAL, the average number of contacts

Table 5. Study eligibility criteria for NCS-R, NSAL, NLAAS

| Study	Age	Language	Race and ethnicity
NCS-R	18 years and older	English only	No race or ethnicity criteria
NSAL	18 years and older	English only	• Blacks • Non-Hispanic whites (Subsample) NOTE: Only Caribbean black respondents were eligible in specified Caribbean sample areas.
NLAAS	18 years and older	English, Spanish, Chinese, Tagalog, Vietnamese	• Latinos • Asians • Non-Hispanic, non-Asian whites (subsampled early in field period)
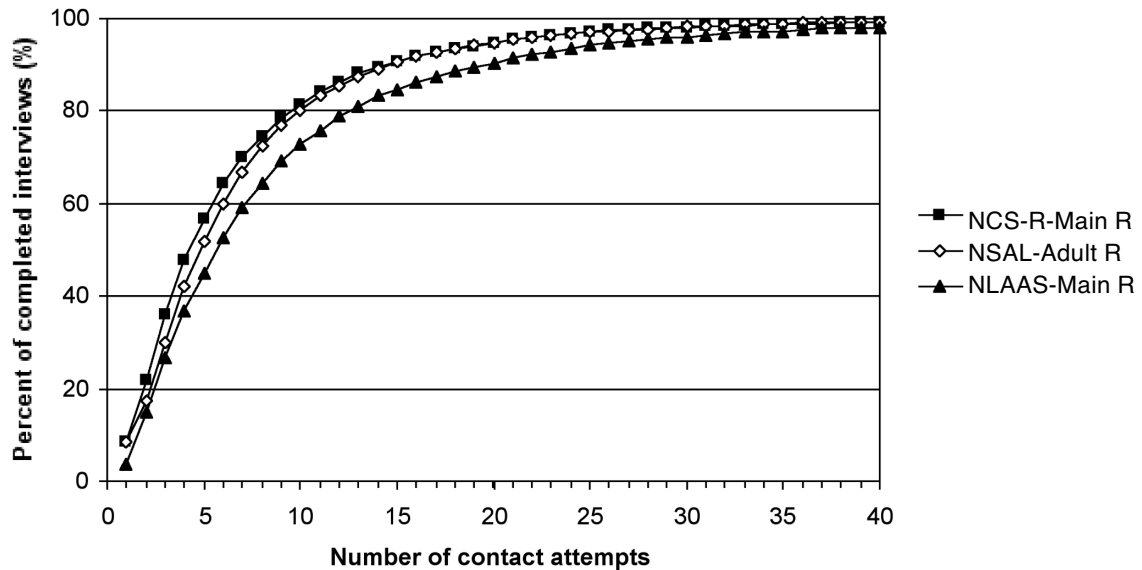
Figure 4. Percentage of completed interviews by contact attempt for NCS-R, NSAL and NLAAS.

needed to complete an interview was 7.4 (median = 5). For NLAAS, an average of 9.2 contacts (median = 6) were necessary to complete the first respondent interview and another 11.6 (median = 8) to complete the second respondent interview. To reduce costs, a higher portion of the second adult interviews for NLAAS were conducted by phone, if convenient for the respondent.

As Figure 4 illustrates, 90% of the interviews for NCS-R and NSAL were completed in 15 or fewer attempts. The figure was similar though slightly lower for NLAAS (85%). By the 40th contact attempt, more than 98% of the interviews for each study had been conducted. It should be noted that in rare cases in each of the three studies, 70 or more attempts were made to complete an interview.

Obtaining informed consent
Written informed consent was required for NLAAS, while NCS-R and NSAL only required oral consent before an interview. The consent form for NLAAS was translated into all of the study languages. As adult respondents were being interviewed, the Institutional Review Board (IRB) allowed either the respondent or the interviewer with respondent permission to sign the consent form.

Written informed consent for taping an interview was required for all three studies. A sample of each interviewer's completed interviews were recorded to evaluate interviewer performance.

When approaching a household for the initial household screening, interviewers read the following script before requesting household information:

> Hello, I'm (interviewer's full name) calling from the University of Michigan's Survey Research Center [SHOW ID BADGE IF IN PERSON]. We're conducting a study about people's health and health issues. This address was selected as part of the study's national sample, and I may need to interview someone here.

Once screening was completed and a respondent selected, the interviewer read SRC's standard informed consent statement before beginning the interview as follows:

> Before we start, I would like to assure you that this interview is confidential and completely voluntary. If we should come to any question which you do not want to answer, just let me know and we will go on to the next question.

Procedures for reducing non-response
A wide variety of tools and procedures were developed at the beginning of each project to maximize respondent participation. Additional refusal conversion

strategies and protocols were devised and implemented as the study progressed.

Interviewer training
Interviewers were provided with initial and ongoing training on the importance of and techniques for reducing non-response. During interviewer training, in-depth explanations of why non-response bias was an important issue to address were presented. This information helped interviewers develop a deeper understanding of non-response and how it could adversely affect survey estimates.

As part of initial study-specific training, techniques for gaining cooperation were discussed. These included maintaining professionalism, establishing a dialogue, building knowledge of the project and the respondent, anticipating common questions, using active listening techniques, treating each contact as a unique interaction, and ending interactions on a positive note.

Interviewers were given additional materials including 'frequently asked questions' that they were likely to encounter from respondents (for example, 'what is the purpose of the study?') and common respondent objections (for example, 'I'm not interested', 'I'm too busy') along with possible responses to address these questions or concerns. Interviewers were encouraged to tailor these responses for specific situations.

After training, interviewers were given opportunities to practice their skills in obtaining respondent cooperation during team conference calls and in individual coaching sessions with supervisors. Weekly team calls often involved discussion of challenging situations and effective solutions to address them. Refusal conversion techniques that were working well for interviewers were shared in the weekly interviewer newsletter.

Tailoring
Groves and Couper (1998) discuss the importance of 'tailoring' in interviewer-respondent interactions. In this context, interviewers use different approaches in introducing the survey request based on what they observe about the sample household or the neighbourhood more generally. For example, interviewers may use different language to present the survey request based on their perception of the level of education or socioeconomic status of the potential respondent. Interviewers who excel in gaining cooperation have a wide array of these 'scripts' that they use to tailor their interactions with respondents. Interviewers can tailor these scripts from the cues they receive over successive interactions with respondents. Experiences over a variety of interactions provide interviewers with additional successful scripts, thereby increasing the likelihood of securing cooperation.

Respondent incentives
In all three studies, respondents were offered financial incentives for survey participation. Respondents were initially paid $50 USD upon completion of the interview (most respondents received payment by cheque, however interviewers had the option to pay in cash if they thought it would help to secure the interview). Incentives were later increased near the end of the project in an effort to boost response rates (see discussion of responsive design below).

Persuasion letters
Persuasion letters or refusal conversion letters were another strategy used extensively to encourage participation. Standardized letters were designed to address common concerns (for example, 'I'm too busy') and individually written letters were prepared for situations that warranted a tailored response. These letters were sent to households by team leaders after consultation with the interviewer.

Special respondent recruitment efforts
During the course of data collection, special letters were sent to either the entire remaining sample in a final attempt to maximize response or to targeted groups (for example, managers of locked buildings) based on specific issues that were affecting data collection progress (for example, difficulty gaining access to a locked building or gated community). In addition to these letters, project staff developed a team of 'locked building specialists' who would meet with building managers or residential boards to explain the study, garner their support for the project, and gain access to the building or gated community.

Distinctive mailings
Many different study-related letters might be sent to a selected household, so it was important to make mailings distinctive and noticeable. Some approaches that were used in these studies included handwriting addresses and applying stamps, using different sized envelopes and small token gifts (for example, magnets, pencils), and sending letters by priority mail or courier service.

Toll-free respondent telephone line

Toll-free telephone numbers were available to respondents who wanted to verify the legitimacy of the project or to contact the interviewer about an appointment. For NLAAS, telephone numbers were maintained for each study language. The staff members answering the toll-free lines were trained in interviewing and refusal conversion, and were highly effective in clarifying basic facts about the study and convincing reluctant respondents to schedule an interview. The ability of staff manning the toll-free line to transfer calls from respondents directly to interviewers in the field helped to avoid missed opportunities for interviews. When data collection was at its peak, the toll-free line staff was handling approximately 30 calls a day, and over the course of the three studies more than 6,000 calls were answered.

Use of travelling interviewers

Travelling interviewers were used in several capacities in these studies. Due to the long field periods and interviewer attrition, some areas were understaffed at various points. At times it was more cost effective to send travelling interviewers to work in these areas than it was to hire and train new staff. Interviewers travelling to these locations were quite effective in convincing resistant respondents to participate by emphasizing how far they had come to obtain an interview. Experienced travellers were also used to team up with local interviewers who were inexperienced or meeting with unusual resistance in a particular area.

Responsive design approaches

Responsive design refers to the application or incorporation of survey design features and decision rules that optimize the trade-off between survey cost and the various sources of survey error, such as that potentially produced by non-response (Heeringa and Groves, 2004; Heeringa et al., 2004). The methods use empirical data that accumulate during the course of data collection to determine the best strategies for reducing error. SRC's daily receipt of process and production data, such as detailed information on calls per case, the status of all outstanding sample, and interviewer performance, allowed staff to make timely decisions about interventions and approaches to optimize production.

Near the end of each project, several responsive design strategies were implemented. These included increased respondent incentives, interviewer incentives, offering an abbreviated interview schedule, and subsampling active cases.

Increased respondent incentives

An effective strategy for reducing non-response is to offer an increased monetary incentive (Singer, Groves, and Corning, 1999; Singer, Van Hoewyk, Gebler, Raghunathan and McGonagle, 1999). An increased incentive also has a positive effect on interviewer morale, as it gives the interviewer something new (and better) to offer sample members who up to this point were reluctant to participate.

Both NLAAS and NCS-R employed respondent incentive increases, although the exact timing, size of the increases, and target sample varied across the studies. Incentives of up to $150 USD were offered on these two studies.

Abbreviated interview schedule

Another option that can be effective in maximizing response rates is to offer a shortened interview. As with increasing monetary incentives, this also supplies interviewers with a new approach to attempt to convince potential respondents to participate. The downside, of course, is that decisions must be made as to which items are to remain in the questionnaire.

Only NCS-R elected to introduce a shortened interview. Near the end of the data collection period, an abbreviated interview schedule was presented to reluctant respondents. The short questionnaire took on average less than 60 minutes to administer and included the screening section, premenstrual dysphoric disorder diagnostic section, and sections that collected sociodemographic information (for example, employment, finances, marriage). The shortened interview allowed more interviews to be completed by telephone, which lessened the burden on respondents and reduced interviewer travel costs at the end of the project when the remaining sample was widely dispersed.

Interviewer incentives

A third approach to increase response rates is to provide monetary incentives to interviewers. This is often implemented at the end of the project when most of the remaining sample is resistant or difficult to reach. Various interviewer bonuses were put into effect late in the data collection period for all three studies.

Subsampling

Each study implemented subsampling plans in an attempt to maximize response rates given resource constraints. These are discussed fully in Heeringa et al. (2004).

Institutional review board considerations

These studies presented a number of unique challenges related to human subjects' protection and institutional review boards (IRBs). For NCS-R and NLAAS, the principal investigators were at different universities and thus the initial design and any changes in study protocol had to be approved by the IRB committee at each institution. At times, the IRB committees had varying requirements, making it necessary to negotiate changes among the different organizations until materials and protocols met with approval from all committees. This led to some delays in implementing changes, and difficult negotiations when, for example, one committee required a certain wording in a letter that the other committee did not want to approve. The level of respondent incentives proved another point of contention as the institutions sometimes disagreed on the amount of money that was appropriate to pay respondents for participating in these projects. As an additional complicating factor, the IRB regulations at the different institutions changed over the course of the data collection period.

Effectively addressing these complexities meant that the management staff at the participating universities and primary funding agency had to work closely to coordinate IRB submissions and had to plan major design changes far in advance of their implementation. This often hampered efforts to react quickly to changing needs and requests during data collection.

Considerable resources were spent on tracking protocol changes and obtaining IRB approvals from all participating institutions. These changes also necessitated the adoption of extensive version control procedures to ensure that all interviewers were using the current version of each letter, form, booklet, questionnaire and revised procedures.

Data collection

This section describes the field structure and organization, data collection processes, and production results for NCS-R, NSAL and NLAAS.

Field organization

Figure 5 shows the organizational structure of the field and central data collection staff for NCS-R, NSAL, and NLAAS. The field staff for each project was divided into teams of 6 to 12 interviewers. Each team was supervised by a team leader. Approximately 3 to 4 teams formed a workgroup, which was supported by a team leader coordinator. Each workgroup was assigned to a regional field manager, who was responsible for the workgroup's interview production efforts, quality control and personnel management. Whenever possible, teams were comprised of groups of interviewers from the same region. Every effort was made to assign interviewers to teams prior to training so that interviewers on the same team would be able to work together during training.

The CPES projects had varying levels of interviewer supervision requirements. These levels were largely determined by the number of interviewers working at any given time and the ratio of newly hired interviewers to experienced interviewers. Over the course of the project, NCS-R had 21 supervisors (team leaders), NSAL had 17, and NLAAS had 25. The ratio of supervisor to interviewer ranged from approximately 1 to 6 immediately after trainings when interviewers needed the highest level of support to approximately 1 to 12 when new interviewers were starting to gain proficiency in their tasks (generally about three months after training). Since supervisory staff also conducted interviews, the supervisory ratios could easily be adapted to these changing needs.

The responsibilities of a team leader involved verifying and reviewing interviews for completeness and accuracy, and coaching interviewers to help them meet production goals and maintain quality standards, while also minimizing costs. Team leader coordinators were responsible for writing weekly production reports summarizing progress toward goals and technical and other issues that may have impacted production. The team leader coordinator also oversaw activities such as supervising a tracking team or coordinating and resolving technical problems for the region.

Meetings were held regularly and frequently during the field period. The regional field managers met with the central office management staff to assess progress, identify and solve problems in the field, and work out details of new procedures. In weekly meetings with their interviewing teams, team leaders reviewed the previous week's progress and distributed information
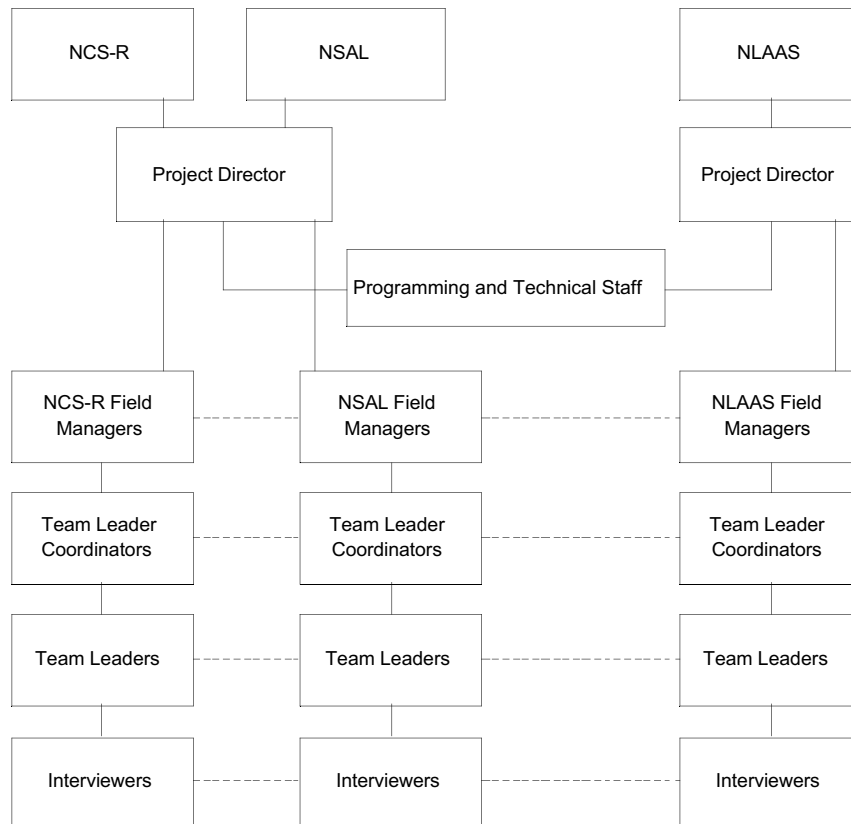
Figure 5. Organizational structure for CPES data collection.

about procedures or upcoming changes to the project. Team leaders also had a weekly one-on-one call with each interviewer to review the sample and talk about issues or problems specific to the interviewer.

Email was used extensively as a communication tool for these projects. Weekly memos and notes from meetings were distributed to keep everyone informed of study progress and changes in procedures.

Interviewer assignments
Team leaders were responsible for assigning sample to interviewers. Initially, the sample was assigned with the goal of minimizing travel time and thus costs. As the available sample diminished and special needs became evident (for example, resistant cases, matching the language of the respondent and interviewer, locked buildings, and so forth), sample addresses were assigned to the interviewers most capable of handling the special situations.

During the NLAAS field period, an additional sample was released to adjust for lower than expected eligibility rates, and second adults who were identified in sample households were automatically added to the available sample. Because of the ever-changing nature of the sample, and due to attrition and hiring and training of new staff at various times, reassignment of work was often necessary to provide the interviewers with work assignments that were appropriate to their availability and skills.

Cost and quality were also considerations in interviewer assignments. As the study progressed, high-quality and low-cost interviewers were assigned additional sample, and interviewers with high cost and/or lower quality were released from the project.

Both high rates of interviewer attrition and implementation of the level of quality control necessary for the significant number of new interviewers hired for these projects were challenges for the field supervisory

staff. These challenges were met by providing strong centralized office support for the supervisory staff.

Data collection results

For NCS-R, a sample of 13,054 addresses was fielded. In 98.1% (11,222) of occupied housing units (11,443), interviewers were able to determine whether or not the household was eligible for inclusion in the study. A total of 10,622 addresses yielded an eligible household. As summarized in Table 6, 9,282 adult interviews were completed: 7,693 interviews with the main respondent and 1,589 interviews with a second adult in the household. As discussed earlier, 554 additional interviews (not included in Table 6) were collected using a shortened form of the instrument with a subsample of non-respondents in an effort to assess non-response bias. The final weighted response rate for NCS-R, excluding the short form interviews, was 70.9% for primary respondents (American Association for Public Opinion Research (AAPOR) Response Rate 3; AAPOR, 2004). The response rate for the second respondent was 80.4%. See Kessler et al. (2004) and Heeringa et al. (2004) for a complete discussion of final sample outcomes and response rates.

In contrast to the relatively high sample yield for NCS-R, in NSAL, 11,634 eligible households were identified from 26,495 randomly sampled addresses. A total of 6,199 adult respondents were interviewed as part of NSAL (1,006 white respondents, 1,623 respondents of Caribbean descent, and 3,570 African American respondents). The overall response rate for the core NSAL national sample was 71.5% (AAPOR

Response Rate 3; AAPOR, 2004). The Caribbean Supplement sample, which was designed to target areas with high concentrations of persons of Caribbean origin, yielded a weighted response rate of 76.4%. See Jackson et al. (2004) and Heeringa et al. (2004) for additional description of the final distribution of sample dispositions and response rates.

The NLAAS project also involved a significant amount of screening. From a total sample of 27,026 addresses, 4,345 eligible main respondents and 1,234 eligible second adult respondents were identified. For NLAAS, 3,620 main respondent and 1,029 second adult interviews were completed. The weighted response rate for NLAAS was 75.7% among main respondents (77.6% for Latinos, 69.3% for Asians) (AAPOR Response Rate 3; AAPOR, 2004). For second respondents, the final response rate was 80.3% (82.4% for Latinos, 73.7% for Asians). As noted earlier, 215 non-Asian, non-Hispanic white controls were interviewed prior to the elimination of the interviewing of white controls due to budget constraints. More detailed information regarding sample outcomes and calculation of response rates can be found in Heeringa et al. (2004). See Table 6 for a general summary of data collection results across the three studies.

Although response rates for these three studies were more than acceptable given the respondent burden and oversampling of minority populations who are often less likely to participate in surveys, it should be noted that it is becoming increasingly difficult to achieve high response rates in household surveys. Refusal rates for US face-to-face government-sponsored surveys,

Table 6. Interviews, response rate, interview length and number of contacts for NCS-R, NSAL, NLAAS

Component	Interviews	Response rate (%)	Average interview length (mins)	Average contacts per interview
NCS-R				
Main respondent	7,693	70.9	126	7.1
Second respondent	1,589	80.4	124	4.7
NSAL				
Adult respondent	4,842 (core)	71.5 (core)	145	7.4
	1,357 (supplement)	76.4 (supplement)		
NLAAS				
Main respondent	3,620	75.7	161	9.2
Second respondent	1,029	80.3	152	11.6

including the National Crime Victimization Survey (NCVS), Current Population Survey (CPS), and the National Health Interview Survey (NHIS), are increasing over time (Groves and Couper, 1998; Groves et al., 2004). Atrostic et al. (2001) highlight rising initial non-response rates for the NCVS, CPS, NHIS, Consumer Expenditure Diary Surveys (CED), Consumer Expenditure Quarterly Surveys (CEQ), and Survey of Income and Program Participation (SIPP) during the 1990s. The initial non-response rate for the CED in 1990 was 16.3%. In 1999, the CED initial non-response rate was 28.0%, a striking 11.7% increase. The median Council of American Survey Research Organizations (CASRO) response rate among states conducting the Behavioral Risk Factor Surveillance System (BRFSS), an ongoing telephone survey designed to produce US state level estimates of health risk behaviours and preventive practices, declined from 68.4% in 1995 to 51.1% in 2001. In assessing response rate experiences in countries outside the US, de Leeuw and de Heer (2002) analysed time series data for government-sponsored surveys from 16 primarily European and North American countries and reported similar declines in response rates over time.

Groves and Couper (1998) identify several possible explanations for the downward trend in response rates. With an increase in the number of single-person households and higher levels of female labour-force participation, it is less likely that an interviewer will find a household member at home. As noted earlier, the growing number of gated communities and locked apartment buildings also has affected contactability. Groves and Couper cite increasing distrust of strangers and reduced levels of civic engagement as potential contributors to lower survey cooperation rates. In recent years, telephone survey participation seems to have been negatively affected by the rise in telemarketing calls.[1] With both adults employed and longer work hours for the average worker, many households also express concerns about being too busy to complete an interview. A number of these factors are beyond the survey organization's control but others can be addressed in

designing and implementing surveys like the CPES. See the section on respondent recruitment, above, for additional discussion of response-enhancement strategies.

Missing data rates

Despite the sensitive nature of many of the survey questions, the proportion of missing responses for most survey items was quite low (<1%). A missing response was defined as a 'don't know' or 'refused' response. The missing data rates for some of the more sensitive NCS-R interview items are presented in Table 7. As is often found in surveys that pose less sensitive questions, income items had the highest rates (6.3% to 10.6%) of item-missing data.

These generally low rates of missing data were probably due, at least in part, to the intensive training that interviewers completed combined with the effective coaching and quality control procedures used by the field supervisory staff. There were also features of the questionnaire, such as explicit follow-up probes for the interviewer to use if the respondent provided an initial 'don't know' response (for instance, 'About how old were you the first time? IF DK, PROBE: Was it before your twenties?') and protocols for recording a 'don't know' response as a 'no' when asking respondents if they had experienced a particular symptom, which probably contributed to low rates of missing data. In addition, as briefly mentioned earlier, all three surveys used an introductory commitment statement that has been shown in methodological research to increase the accuracy and completeness of reporting (Cannell, Marquis, and Laurent, 1977). See also Kessler and Üstün (2004). This question was placed in the first section of the interview after some initial warm-up questions and directly before the diagnostic stem questions. The precise wording for this item (*SC19) was as follows:

(READ SLOWLY) The rest of this interview asks about your physical and emotional well-being and about areas of your life that could affect your physical and emotional well-being. It is important for us to get accurate information. In order to do this, you will need to think carefully before answering the following questions.

Are you willing to do this?

INTERVIEWER: PROBE NEGATIVE RESPONSES BY ASKING IF THERE IS A BETTER TIME TO COME BACK FOR THE INTERVIEW. REPEAT

[1] The US Federal Trade Commission's establishment of the National Do Not Call Registry, which was designed to reduce the number of unwanted telemarketing calls that households receive, may help reverse this trend. However, at this time, the impact on telephone survey response rates is undetermined.

Table 7. NCS-R item missing data rates

Item	Main adult respondent			Second adult respondent		
	n	Don't know (%)	Refused (%)	n	Don't know (%)	Refused (%)
SC21: Depression screener	8082	0.0	0.0	1788	0.0	0.0
SC20: Panic screener	8082	0.0	0.0	1788	0.1	0.0
SC29: Social phobia screener	8082	0.0	0.0	1788	0.1	0.0
SC33.1: Set fires as child	8082	0.1	0.0	1788	0.1	0.1
SU1: Age first drank alcohol	4692	1.0	0.0	1007	0.9	0.1
SU41: Ever used marijuana	4692	0.0	0.3	1007	0.0	0.3
SU43: Improper use of medications	4692	0.0	0.1	1007	0.0	0.1
FN2: Respondent income	5597	1.9	6.2	1612	2.8	7.8
FN10: Family income	5597	2.0	4.3	1612	3.8	4.6

*SC19 AS NECESSARY. R MUST ANSWER AFFIRMATIVELY TO CONTINUE WITH THE INTERVIEW. TERMINATE IF R DOES NOT ANSWER AFFIRMATIVELY.

Almost all respondents (99%) agreed to provide accurate information at this point in the interview. Respondents who answered 'no', 'don't know' or 'refused' to the commitment question were terminated from the interview at that point.

Respondent burden

Clearly, participation in these studies involved a significant time investment for many respondents. The effect of burden on response rate and data quality is important to consider particularly in studies like the CPES that place considerable burden on respondents. The scarce literature on the relationship between burden and survey participation is mixed (see Bradburn, 1978; Sharp and Frankel, 1983; Bogen, 1996), although in a review of the literature Bogen (1996) offers some evidence that lengthier interviews are associated with lower levels of response. Bradburn (1978) suggests that respondents may be more willing to consider a longer interview if they believe that their participation will contribute to an important research purpose. Conveying the value of the research effort and the respondent's participation in it was carefully considered in designing these studies. As described earlier in this section, a commitment statement that emphasized the importance of quality data and then asked respondents to agree to provide accurate information was used to attempt to maximize data quality in these studies.

Another aspect of respondent burden is the emotional distress that may be experienced during and/or after the interview as a result of answering sensitive survey questions. In three mental health surveys that assessed respondent reactions to the interview, the percentage of respondents who indicated distress or who appeared distressed as observed by the interviewer ranged from 3% to 9% (Turnbull et al., 1988; Henderson and Jorm, 1990; Jorm et al., 1994). To attempt to address the potential for emotional distress, the researchers in one of these cited studies mentioned the use of strategies that were employed in the CPES, such as the availability of a clinical support person and providing a listing of community mental health resources to respondents (Turnbull et al., 1988). All CPES respondents were given a 'resource card' that contained a list of various organizations that would be able to offer support services and their contact information, so that respondents could call directly and receive assistance on issues including substance abuse, domestic violence, and mental health. The card was given at the beginning of the interview to avoid any perception that the respondent was receiving it in response to information revealed during the interview. Respondents should also be informed about the potentially sensitive nature of the questions and that they can refuse to answer any question (Turnbull et al., 1988; Henderson and Jorm, 1990). Effective techniques for handling emotional interview situations was a component of interviewer training for the CPES and has been highlighted in the literature as an important consideration in research involving sensitive topics (Turnbull et al., 1988; Jorm et al., 1994; Corbin and

Morse, 2003). In sum, it is likely that some respondents may experience anxiety or distress in answering mental health survey items but the techniques used in the CPES and further supported in the literature can be helpful in minimizing any negative impact on respondents.

A number of survey participants indicated that the interview positively affected them (Turnbull et al., 1988; Henderson and Jorm, 1990; Jorm et al., 1994; Corbin and Morse, 2003). Corbin and Morse (2003) note that interviewees may leave the interview with a sense of relief. Positive interview experiences shared by CPES interviewers appear to support this assertion.

Motivational aspects
To meet the challenges of data collection successfully for these studies, study staffs and regional field managers paid particular attention to the morale of field staff. Fostering a team spirit and effective communication between field and central staff, providing technical support, and offering professional mental health support services, were all crucial in supporting and motivating field staff.

Teamwork and communication
Each project placed an emphasis on teamwork at all levels of the organization. The sense of working toward a common goal was vital in meeting the demands of a prolonged data collection effort.

Another motivational factor related to teamwork was the importance placed on open, honest, and frequent two-way communication between the field and central office staff. This was critical for catching errors and streamlining processes.

A weekly interviewer newsletter was an effective formal communication tool used in these studies. The newsletter congratulated staff on their accomplishments, highlighted innovations that could be used by other interviewers, and kept employees focused on what was most important at different phases of production.

From the beginning of field production it was clear that communication from the principal investigators to the field staff was highly motivating. Communication from the investigators via newsletter and email kept field staff apprised of how data collection efforts were progressing from the researchers' perspective, motivated them at certain key times in production (for example, why refusal conversion was important), and helped them to understand the rationale behind changes made to the study after production had begun.

Field help desk
With a large field staff working remotely with laptops operating complex technical systems, it was critical to provide on-call technical support. This vital role was filled by SRC's field help desk, which comprised six individuals who served as the technical support liaison between field staff and the central office.

In the event that an interviewer experienced technical problems, he or she was instructed to call the toll-free help desk phone number. All requests for assistance and follow-up correspondence were logged in a case-management system and tracked weekly to identify systematic problems. A total of 10,709 help desk calls were received between February 2001 and November 2003 for the CPES studies.

Clinical support
Because of the sensitive nature of the questionnaire subject matter, providing clinical support for interviewers and respondents was an essential part of field operations. This support was provided by a clinical social worker who had prior experience working with interviewers. As part of study-specific training, the social worker led a one-hour session in which she described her role as a clinical support person. She also discussed what to expect during the course of a typical interview and how to handle emotional reactions in the interview setting. Interviewers were coached on handling situations where respondents appeared to be a threat to themselves or others, and what to do if there were signs that abuse was taking place in the home.

During the course of data collection, between five to 10 calls or emails per week were placed or sent to the clinician. The vast majority of calls were cases in which interviewers were concerned about the mental health status of respondents. In these cases, the clinician would determine whether follow-up with the respondent seemed appropriate based on the information provided by the interviewer. A second type of call involved interviewers who had administered part of an interview and wanted to talk with the clinician about the respondent before going back to complete the interview. A third type of call was from interviewers who requested help with a personal problem or work-related situation (for example, how to effectively deal with a particularly distressing interview) or supervisors

who had concerns about an interviewer. Access to professional support services was clearly valued by respondents, interviewers and supervisors.

Quality control procedures

Quality control procedures were an integral part of each project. This section details the procedures used to monitor interviewers during production and the protocols that were implemented for receipt of materials from the field. Instrument development, hiring, and training quality control procedures are covered in those corresponding sections.

Quality control procedures during production interviewing

The quality of an interviewer's performance was assessed by determining compliance with household enumeration and respondent selection procedures; evaluating taped interviews; and verifying interviews.

Review of coversheets

Once a sample case was finalized, the coversheet was mailed back to the central office where the household screening and respondent selection procedures were reviewed to ensure that the eligibility status of the case had been correctly assigned and that the appropriate respondent had been selected. Coversheets were examined for systematic errors and interviewers were given feedback on their completion of coversheet information. Where errors resulted in the selection of an incorrect respondent, attempts were made to obtain an interview from the correct household member. In addition, for NLAAS, all coversheets were checked to verify that the informed consent statement had been signed.

Review of taped interviews

All interviewers were required to tape their interviews for evaluation purposes. An interviewer's first five interviews were recorded, as were regularly scheduled interviews throughout the data collection period (approximately 5% of each interviewer's interviews were recorded and evaluated), to assure that interview quality was maintained. A team leader or a member of the centralized quality control staff critiqued and scored the interviewing performance using a standardized online evaluation form that addressed each of the following points:

- knowledge of study objectives;

- proper use of household enumeration and respondent selection procedures;
- appropriate use of refusal aversion and conversion techniques, when necessary;
- use of general interviewing techniques (for example, reading questions exactly as worded, neutral probing); and
- rapport with the respondent (for example, professional, confident manner, use of feedback).

The data supplied in the online form were saved to a master database that contained information on all interviewer evaluations for the project. The evaluator would prepare a summary and email a copy of the report to the interviewer and the team leader. For a subset of interviews, question responses from tape recorded interviews were also compared to responses keyed into the computer to assess interviewer understanding and facility with the data entry process.

Verifications

The term 'verification' refers to the process of confirming that a sample case was either ineligible to participate in the study or eligible and actually participated in an interview. Both types of verifications involve a team leader or a quality control staff member recontacting sample households: any adult household member in the first case, and the respondent in the second case. Five per cent of each interviewer's ineligible households were recontacted to ascertain whether the interviewer's determination was accurate. If the determination was incorrect, the case would be reopened in order to attempt to conduct an interview with the appropriate respondent. A random 10% of an interviewer's completed interviews were verified. To verify that an interview had been completed, a series of factual questions from the beginning, middle, and end of the interview were readministered. The original interview and reinterview responses were then compared. If anomalies were found, all of an interviewer's cases were then verified. Eligibility and interview verification reports were sent to each interviewer and team leader to keep them apprised of any quality control issues.

Data monitoring

Administrative and respondent questionnaire data were monitored throughout the project to assure high quality data collection.

Field production reports

A variety of reports were developed to assist managers in tracking the performance and progress of each study. These reports provided national, regional, team and individual interviewer level statistics (for example, number of completed interviews, interviewer production hours per interview, distribution of finalized cases by status) that were carefully reviewed by project staff to evaluate data collection progress at these various levels. All reports were updated daily.

Section timings and screening question reports

Periodically, reports were produced that documented information on individual interviewer's data entry patterns and time spent in each section of the instrument. Timing information helped to identify potentially fraudulent interviews (timings that were significantly shorter than average) and to verify timesheet claims of extremely long interviews. Staff reviewed the screening question reports for unusual response patterns that might suggest falsification.

Materials receipt

Each of the three studies required the completion of several paper forms for each selected household. These materials were a combination of items that had further use in the project (for example, consent forms that had to be recorded and stored) and items that needed secure short-term storage and disposal because they contained identifying information about the respondent (for example, coversheets and contact information sheets). Monitoring the receipt of these materials and putting systems in place to retrieve missing materials was a major task given the number of paper forms and the large sample size for each study.

All materials mailed by interviewers were sent to Ann Arbor using a trackable mailing or courier service. Receipt control logging programs were set up to record receipt of packages (tracking numbers) and the contents of those packages. A log of materials that had been sent was compared against a log of materials that had been received each week to identify packages that had not arrived in Ann Arbor.

Data processing

Data transmission

Each day interviewers connected their laptop to a telephone line and dialled up an ISP to submit and receive information from the central office. The purpose of the communication was twofold. First, all work performed since the previous communication was transmitted to the central office and the information updated in the master files. As part of the daily communication, interviewers reported the hours that they worked and their expenses. This information transfer allowed for real-time monitoring of interview production, sample disposition and costs. Second, information from the central office, including programme updates, transfers of sample between interviewers, and newly released sample, was transmitted electronically to interviewers.

Data editing and coding

Many data processing activities that are typically completed when collecting data via paper questionnaires were unnecessary in these studies because the questionnaires were computer administered. For example, making sure that questions were asked in the correct sequence, checking for out-of-range or inconsistent responses, and filling in the appropriate question text based on a respondent's previous answers were all controlled by the interviewing application software. Inconsistent responses that failed the programme's edit checks were brought to the attention of the interviewer who could resolve the inconsistency with the respondent during the interview, improving the quality of data and minimizing the need for back-end editing.

Although these software programmes automatically performed many of the decisions formerly made by interviewers using paper questionnaires, the data for each study did require some additional editing and coding. Editing operations included processing each interview through a series of programming routines that evaluated question responses and assigned codes to indicate the presence or absence of each mental health disorder assessed by the study. A number of other summary variables based on individual question items were also created in preparation for the project's analysis phase. In addition, each study included several open-ended questions, which were coded.

Data file creation

Data files were extracted to ASCII and converted to SAS format once data collection began and were updated throughout the data collection period. Early in the data collection phase, these files were used by project managers to identify any problems with administration

of the questionnaire and to monitor response trends and patterns. Data sets were produced for the studies' principal investigators on a weekly basis to allow the investigator's staff to perform preliminary analyses.

Data file documentation

A codebook and set of companion instructions and study materials were prepared for each study. The codebook provided the information that users need to associate a variable in the data file with the corresponding question on the questionnaire and documented the characteristics of each variable in the data set, such as its format and response codes. The codebook also contained frequencies for nominal and ordinal variables and a set of basic descriptive statistics for continuous variables. For NLAAS, the HTML compatible codebook included a facility to view each question in any of the five languages.

Conclusions and recommendations

The previous nine sections offered a glimpse into the activities involved in the design and implementation of data collection procedures and systems for the CPES studies. This section reviews some of the challenges and lessons learned at each of the major development and implementation phases.

Questionnaire and materials development

The complexity of the survey instrument and pace of development created several challenges. Systems that facilitated the control of versions were essential given the numerous questionnaire iterations across the WMH and CPES studies and languages. It was also important to allocate sufficient time for senior staff to review and test the instrument. Overlap in key staff across the projects, both at the project management and technical support levels, yielded benefits by making experience gained on one project available to other projects.

After fielding the CPES studies, the SRC developed protocols for instrument specifications to facilitate communication between the instrument designers and programmers to increase programming and testing efficiency. The SRC has also developed tools that automatically produce hard copy of the instrument and allow for online testing of all branching and skips, also increasing testing efficiency.

These studies could not have been fielded without a system for automatic delivery of updated instruments to interviewers. Early review of the production interview data was critical to ensure that the revised version of the programme was performing to specifications.

The various study materials worked well for gathering the required information, although reliance on numerous paper forms created opportunities to lose information. The best solution to this problem is to minimize reliance on paper-based forms. When that is not possible, forms can be combined to reduce the number of separate pieces of paper that need to be completed and tracked. Experienced interviewers can offer guidance on assembling sets of materials and organizing a 'car-based' office for the large quantity of materials. Finally, providing interviewers with small job aids such as checklists for each completed interview can assist in minimizing missing data.

The decision to use a paper-based procedure for screening and respondent selection rather than the electronic alternative was made to address logistical limitations that existed at the time of these studies. The laptops that were used were too heavy (8.2 lbs, 3.7 kg) and lacked the screen clarity necessary for doorstep interviewing. After the completion of these studies, the SRC fielded a number of studies that demonstrate that doorstep screening is possible with appropriately-designed lightweight laptops (3.5 lbs, 1.6 kg) thereby eliminating the need for paper-based respondent selection protocols that are prone to error and take more time to complete.

Interviewer recruitment and training

In recruiting and training interviewers for these studies, a number of enhancements and recommendations were developed that can be applied to many complex surveys. In planning recruitment and training for studies that involve a significant screening effort, it can be more efficient to recruit and train a separate staff to screen households for eligible members rather than have highly trained interviewers spend a significant part of their time visiting ineligible households.

In developing a training approach, sending interviewers materials to read and exercises to complete prior to training can reduce the amount of new content delivered during training. The SRC has subsequently developed an extensive at-home interactive e-learning training package that reduces on-site training time by four days for the CIDI instrument.

For the on-site training component, using a 'community-based' training model with 20 to 30

trainees has the cost advantage associated with large-scale trainings as well as the benefits of small-group settings. In this model, teams of interviewers who will work in the same geographic area can be trained together, thereby establishing an early connection with their teams. The learning burden placed on interviewers during training can be reduced by dividing general and study-specific training into two sessions and holding back training topics that interviewers will not need (for example, advanced training on refusal conversion techniques) until a training session later in the production process.

Data collection
There were a number of important lessons learned as a result of challenges encountered during data collection. Given the wide range in the amount of time required to complete an interview and the complexity of the interviewing task, it was important to set realistic production goals and to revise them to reflect the number and mix of cases in the field.

To maximize response, many strategies and tools that were described earlier are vital. Two additional points merit mention here. First, refusal rates should be monitored early in the field period in order to identify interviewers who are best suited for refusal conversion as well as those who are having problems and need remedial training. Second, a wide variety of 'tools' should be provided to interviewers to allow them to choose the best one for a given situation; for example, offering a different version of a letter or providing a cash or cheque incentive depending on the respondent's preference.

Interviewer safety and administrative issues associated with respondent payment are two areas that may be overlooked in planning for data collection but deserve careful consideration. The production plan should include strategies that minimize risks to interviewer safety. These may include hiring staff to accompany the interviewer or simply providing the interviewer with a cellular phone. For studies that utilize respondent payments, adequate staffing should be devoted to the processing of respondent incentive payments and procedures set up to handle problems with payments.

Data processing
In designing and implementing studies where data will ultimately be shared and analysed across projects, it is important to commit resources to ensure compatibility in data structure across the instruments. These studies were designed to make maximum use of similar data structures and naming conventions to facilitate creation of an integrated codebook combining questions from each study.

The production of codebooks and hard-copy instruments can be time-consuming post survey processing activities. To reduce the time and resources devoted to this activity, SRC has recently developed a data documentation system that produces eXtended Markup Language (XML) documents that represent the aspects of the Blaise CAI instrument, which can then be used to generate codebooks, paper versions of the questionnaire, and other data documentation automatically.

In sum, despite the challenges of these studies, their scope and complexity compelled central office and field staffs to develop new approaches that enhanced the quality of the data and clearly contributed to the success of these projects. It also is important to emphasize the vital role that dedicated interviewers played in each of these studies. Many of the interviewers report that this study was one of their most rewarding interviewing projects. The experience of most of the staff and interviewers can be summed up with a quotation from one of the interviewers:

> It was a great project, frightening at first, but then a pleasure. Can you imagine how it feels to have a long, trying interview, and then when you start to leave, [the respondent] hugs you and tells you how much they enjoyed the study, and they feel so much better? Now that makes you feel like you really have a special job. It doesn't always happen, but enough to make you feel you are doing something worthwhile.

References

Alegria M, Takeuchi D, Canino G, Duan N, Shrout P, Meng X, Vega W, Zane N, Vila D, Woo M, Vera M, Guarnaccia P, Aguilar-Gaxiola S, Sue S, Escobar J, Lin K, Gong F. Considering context, space and culture: The National Latino and Asian American Study. IJMPR 2004; 13(4): 208–20.

Alegria M, Vila D, Woo M, Canino G, Takeuchi D, Vera M, Febo V, Guarnaccia P, Aguilar-Gaxiola S, Shrout P. Cultural relevance and equivalence in the NLAAS instrument: integrating etic and emic in the development of cross-cultural measures for a psychiatric epidemiology and services study of Latinos. IJMPR 2004; 13(4): 270–88.

American Association for Public Opinion Research. Standard Definitions: Final Dispositions of Case Codes and Outcome Rates for Surveys. 3 edn. Lenexa KA: AAPOR, 2004.

Atrostic BK, Bates N, Burt G, Silberstein A. Non-response in US government household surveys: consistent measures, recent trends, and new insights. Journal of Official Statistics 2001; 17: 209–26.

Bogen K. The effect of questionnaire length on response rates – a review of the literature. Proceedings of the Section on Survey Research Methods, American Statistical Association 1996: 1020–5.

Blaise(r) Survey Processing System: Version 4.5. Statistics Netherlands, 2000-2004.

Bradburn NM. Respondent burden. Proceedings of the Section on Survey Research Methods. American Statistical Association 1978: 35–40.

Cannell C, Marquis K, Laurent A. A summary of studies. Vital Health Stat 1977; 2: 69.

Cheung GQ, Liu Y. Displaying Chinese characters in Blaise. Proceedings of the Eighth International Blaise Users Conference 2003, Copenhagen, Denmark.

Corbin J, Morse JM. The unstructured interactive interview: issues of reciprocity and risks when dealing with sensitive topics. Qualitative Inquiry 2003; 9(3): 335–54.

de Leeuw E, de Heer W. Trends in household survey non-response: a longitudinal and international comparison. In R Groves, DA Dillman, JL Eltinge and RJA Little (eds) Survey Non-response. New York: Wiley, 2002, pp. 41–54.

Groeneveld R. Using non-Latin alphabets in Blaise. Proceedings of the Eighth International Blaise Users Conference 2003. Copenhagen, Denmark.

Groves RM, Couper MP. Non-response in Household Interview Surveys. New York: Wiley, 1998.

Groves RM, Fowler FJ, Couper MP, Lepkowski J, Singer E, Tourangeau R. Survey Methodology. New York: Wiley, 2004.

Guenzel PJ, Berckmans TR, Cannell CF. General Interviewing Techniques: A Self-Instructional Workbook for Telephone and Personal Interviewer Training. Ann Arbor, Michigan: Institute for Social Research, Survey Research Center, 1983.

Heeringa SG, Groves RM. Responsive Design for Household Surveys. Ann Arbor: Survey Methodology Program, Institute for Social Research, University of Michigan, 2004.

Heeringa SG, Wagner J, Torres M, Duan N, Adams T, Berglund P. Sample designs and sampling methods for the Collaborative Psychiatric Epidemiology Studies (CPES) IJMPR 2004; 13(4): 221–40.

Henderson AS, Jorm AF. Do mental health surveys disturb? Psychol Med 1990; 20: 721–4.

Jackson JS, Torres M, Caldwell CH, Neighbors HW, Nesse R, 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.

Jorm AF, Henderson AS, Scott R, MacKinnon AJ, Korten AE, Christensen H. Do mental health surveys disturb? Further evidence. Psychol Med 1994; 24: 233–7.

Kessler R, Berglund P, Chiu WT, Demler O, Heeringa S, Hiripi E, Jin R, Pennell B, Walters E, Zaslavsky A, Zheng H. The US National Comorbidity Survey Replication (NCS-R): an overview of design and field procedures. IJMPR 2004; 13(2): 69–92.

Kessler RC, Üstün TB. The World Mental Health (WMH) survey initiative version of the World Health

Organization Composite International Diagnostic Interview (CIDI). IJMPR 2004; 13(2): 93–121.

Kessler RC, Wittchen H-U, Abelson J, Zhao S. Methodological issues in assessing psychiatric disorders with self-reports. In AA Stone, JS Turkkan, CA Barchrach, JB Jobe, HS Kurtzman and VS Cain (eds) The Science of Self-Report: Implications for Research and Practice. Mawah NJ: Lawrence Erlbaum Associates, 200; 229–55.

Kish L. Survey Sampling. New York: John Wiley & Sons, 1965.

Sharp LM, Frankel J. Respondent burden: a test of some common assumptions. Public Opinion Quarterly 1983; 47: 36-53.

Singer E, Groves RM, Corning AD. Differential incentives: beliefs about practices, perceptions of equity, and effects on survey participation. Public Opinion Quarterly 1999; 63: 251–60.

Singer E, Van Hoewyk J, Gebler N, Raghunathan T, McGonagle K. The effect of incentives in interviewer-mediated surveys. Journal of Official Statistics 1999; 15: 217–30.

Turnbull JE, McLeod JD, Callahan JM, Kessler RC. Who should ask? Ethical interviewing in psychiatric epidemiology studies. American Journal of Orthopsychiatry 1988; 58(2): 228–39.

Wittchen H-U. Reliability and validity studies of the WHD Composite International Diagnostic Interview (CICDI): a critical review. J Psychiatr Res 1994; 28(1): 57–84.

World Health Organization. Composite International Diagnostic Interview, Version 1.0. Geneva, Switzerland: World Health Organization, 1990.

World Health Organization. Composite International Diagnostic Interview, Version 2.1. Geneva, Switzerland: World Health Organization, 1997.

Correspondence: Beth-Ellen Pennell, Survey Research Center, Institute for Social Research, University of Michigan, PO Box 1248, Ann Arbor MI, USA 48106-1248.

Telephone (+1) 734-222-8905.

Fax (+1) 734-222-1521.

Email bpennell@isr.umich.edu.