Cognitive Interviewing Methods for Questionnaire Pre-Testing in Homeless Persons with Mental Disorders

Carol E. Adair, Anna C. Holland, Michelle L. Patterson,
Kate S. Mason, Paula N. Goering, and Stephen W. Hwang,
At Home/Chez Soi Project Team

ABSTRACT *In this study, cognitive interviewing methods were used to test targeted questionnaire items from a battery of quantitative instruments selected for a large multisite trial of supported housing interventions for homeless individuals with mental disorders. Most of the instruments had no published psychometrics in this population. Participants were 30 homeless adults with mental disorders (including substance use disorders) recruited from service agencies in Vancouver, Winnipeg, and Toronto, Canada. Six interviewers, trained in cognitive interviewing methods and using standard interview schedules, conducted the interviews. Questions and, in some cases, instructions, for testing were selected from existing instruments according to a priori criteria. Items on physical and mental health status, housing quality and living situation, substance use, health and justice system service use, and community integration were tested. The focus of testing was on relevance, comprehension, and recall, and on sensitivity/acceptability for this population. Findings were collated across items by site and conclusions validated by interviewers. There was both variation and similarity of responses for identified topics of interest. With respect to relevance, many items on the questionnaires were not applicable to homeless people. Comprehension varied considerably; thus, both checks on understanding and methods to assist*

Adair is with the Department of Community Health Sciences and Psychiatry, University of Calgary, Calgary, AB, Canada; Holland, Mason, and Hwang are with the Centre for Research on Inner City Health, The Keenan Research Centre in the Li Ka Shing Knowledge Institute of St. Michael's Hospital, Toronto, ON, Canada; Holland and Hwang are with the Department of Medicine, Division of General Internal Medicine, University of Toronto, Toronto, ON, Canada; Patterson is with the Faculty of Health Sciences, Simon Fraser University, Vancouver, BC, Canada; Goering is with the Centre for Addiction and Mental Health, Toronto, ON, Canada; and with the Department of Psychiatry, University of Toronto, Toronto, ON, Canada.

Correspondence: Carol E. Adair, Department of Community Health Sciences and Psychiatry, University of Calgary, Calgary, AB, Canada. (E-mail: ceadair@ucalgary.ca)

The At Home/Chez Soi Project Team includes Jayne Barker, Ph.D., VP Research Initiatives, Mental Health Commission of Canada; Cameron Keller, M.C., Director At Home/Chez Soi; and approximately 40 investigators from across Canada and the USA. In addition, there are five site coordinators (one for each city where the study is carried out) and numerous lead service and housing providers. CEA is lead for quantitative measurement and data collection at the national level; MLP and KSM are site research coordinators; SWH is lead investigator for the Toronto site; ACH was an undergraduate medical research student for the summer of 2009; and PNG is the national research lead.

comprehension and recall are recommended, particularly for participants with acute symptoms of mental illness and those with cognitive impairment. The acceptability of items ranged widely across the sample, but findings were consistent with previous literature, which indicates that “how you ask” is as important as “what you ask.” Cognitive interviewing methods worked well and elicited information crucial to effective measurement in this unique population. Pretesting study instruments, including standard instruments, for use in special populations such as homeless individuals with mental disorders is important for training interviewers and improving measurement, as well as interpreting findings.

KEYWORDS *Mental disorders, Homelessness, Survey methods, Cognitive interviewing, Pretesting*

BACKGROUND

Homelessness is a significant social, economic, and health problem in Canada, especially in major cities.¹ Compared with the general population, homeless people experience higher rates of physical and mental illness, substance use, victimization, violence, criminal justice involvement, and mortality.¹⁻⁴ The health, social, and economic consequences of homelessness translate into significant societal costs.^{2,5} There is little information on the most cost-effective approaches to addressing the broad health and social needs of homeless individuals with mental disorders at both service and policy levels.⁶

We use the term “homelessness” to describe those sleeping rough outdoors or residing in temporary accommodations such as emergency shelters. However, we also invoke the word to describe something more abstract: the absence of belonging, both to a place and with the people settled there. Indeed, the term homelessness is used to encapsulate a variety of phenomena including social dislocation, extreme poverty, itinerant work, and unconventional and marginalized ways of living. We presume that people who are homeless share a unique experience and relate to the larger social order in ways that are different from the general population.⁷⁻⁹ However, reliably capturing the effects of various services in a fragmented system in meeting the long-term housing, health, and social needs of homeless individuals has been a challenge.^{10,11}

Many jurisdictions have launched policy and program development initiatives to address the complex issues of homelessness. Research on the effectiveness of interventions for various subpopulations among the homeless is crucial before programs are widely disseminated in order to justify and mobilize public spending. Several interventions have been studied, including the Housing First model,¹²⁻¹⁴ which provides permanent housing (typically through rent subsidies for scattered-site housing units) and support, without requiring sobriety or willingness to engage in treatment. Housing First-type interventions have been shown to be cost-effective, to increase residential stability, and to reduce institutional service needs among homeless individuals with serious mental illness.¹⁵⁻¹⁷ However, questions remain about the effectiveness of Housing First among specific subpopulations, such as those with concurrent substance use disorders,¹⁸ and the model’s applicability to individuals from diverse ethnocultural backgrounds is largely untested. In 2009, a large ($n \sim 2,300$), multisite randomized controlled trial of Housing First for homeless individuals with mental disorders was initiated in five Canadian cities (Vancouver, Winnipeg, Toronto, Montreal, and Moncton). The trial is examining

the effectiveness of a Housing First intervention combined with Assertive Community Treatment (ACT) for individuals with very high needs and Intensive Case Management (ICM) for individuals with moderately high needs, each compared with usual care. The research is also designed to examine effectiveness (including cost-effectiveness) in individuals with serious concurrent substance use issues, to test ethnoculturally relevant service components, and to study new primary care and vocational service components.

Researchers examining housing interventions for homeless populations have generally adopted measures and outcome domains from the general mental health services literature (e.g., mental health symptoms, functioning, and service utilization) and then added residential stability.^{16,17,19} In addition to these outcomes, the Canadian multisite trial includes measures of community integration, recovery, and quality of life. Instruments for the larger study were chosen based on a consensus measurement framework and specific literature reviews for validated instruments. However, no instrument for several of the new outcomes had been designed for or used in the study population. In addition, most of the information needed for these concepts required self-report responses, collected in one-on-one interviews. We shared concerns expressed by other authors about the applicability and validity of some instruments, items, and self-report responses given the marginalized life circumstances of homeless individuals.^{20–22} While much of the concern in the survey methods literature surrounds inadequate recall due to the effects of physical and mental illness, and underreporting of sensitive information due to social desirability bias, Rosen et al. found that most of the inconsistencies between administrative records and self-report were attributable to simple confusion of terms.²² In keeping with this finding, we considered pretesting to be essential for better understanding of the meaning of items and responses in context, since meaning is affected by the situation as well as a range of social and cultural factors of the communicants.²³ In order to maximize validity of measurement, questions and response options need to be based on this shared understanding of language and context.²³

Cognitive interviewing is one technique of Cognitive Aspects of Survey Methodology, involving a systematic, in-depth approach to assessing the validity of questionnaire content and instructions.^{24,25} The approach is based on a theory which specifies four stages of cognitive response to questioning: comprehension, retrieval/recall, estimation/judgment, and response.²⁶ Cognitive interviewing uses “think-aloud” and “probing” methods to examine all of these stages of question answering. The results of cognitive testing can inform item selection and adaptation or framing and instructions. Where change to standardized items is not possible, it can also assist with interpretation of results.

To date, only one published study has used cognitive interviewing methods in a homeless sample to examine the appropriateness of questionnaire items. Matter and colleagues tested a bank of proposed items in the development of a new pain questionnaire.²⁷ Most draft items were found to be problematic and considerable modifications were needed to ensure relevance and clarity. No study has been published using cognitive methods to test the utility and relevance of items from existing, standardized instruments in homeless individuals with mental disorders. In the current study, cognitive interviewing methods were used for focused pretesting of questionnaire items for which validity was in question with regard to the population of interest for the large, multisite trial. The focus of our testing was on relevance. We adapted standard cognitive interviewing probes for all of these foci if testing. The cognitive interviewing testing occurred in the context

of a broader pretest process which also included the assessment of overall face validity of instruments, administration time, and flow.

METHODS

Study Sample

Eligibility criteria for the main trial are legal adult status, homelessness, not currently receiving services similar to ACT or ICM, and the presence of one or more serious DSM-IV Axis I mental disorders.* Participants are not excluded based on the presence of substance use disorders as long as a co-occurring mental disorder is present. For the pretesting study, mental health and homelessness status (no fixed accommodation for the previous seven nights) were loosely defined based on report from the referral agency. Individuals were recruited from a variety of locations including shelters, drop-in centers, mental health agencies, and directly from the street in three of the five study sites (Winnipeg, Vancouver, and Toronto) and were given a cash honorarium for participation. Institutional Research Ethics Boards reviewed and approved the study at all three study sites.

Thirty participants were recruited through a convenience sampling process whereby staff at agencies serving homeless adults nominated known individuals or prospective participants were approached directly on the street by interviewers. The goal of this sampling strategy was to recruit a variety of homeless individuals with a high likelihood of having mental disorders who might have some difficulty with the items due to cognitive impairment but that would still be capable of reflecting and commenting on the questionnaire items and instructions. Sampling was purposively diverse in terms of gender and ethnic background. One person who was approached for an interview was not able to continue past the informed consent process due to substantial difficulties understanding the task and communicating with the research assistant.

The first interview schedule was administered to 16 participants (mean administration time=50 minutes [range, 34–63]) and the second to 14 participants (mean administration time=52 minutes [range, 31–76]). The sample consisted of 20 males and 10 females (mean age, 44 years [range, 25–66]); 12 were from the Vancouver site, 11 from the Toronto site, and 7 from the Winnipeg site. Referral sources included homeless drop-in centers ($N=11$), shelters ($N=8$), mental health or ethnocultural resource centers ($N=7$), and direct recruitment from the street ($N=2$; referral information was

*In the main study, homelessness is defined as being “absolute” (having no fixed place to stay for at least the previous seven nights and little likelihood of getting a place in the upcoming month) or “precariously housed.” No fixed place to stay includes living rough in a public or private place not ordinarily used as a regular sleeping accommodation for a human being (e.g., outside on the streets, in parks or on the beach, in doorways, in parked vehicles, squats, or parking garages), as well as those whose primary night-time residence is a supervised public or private emergency accommodation (e.g., a shelter or hostel). Those currently being discharged from an institution, prison, jail, or hospital with no accommodation also qualify as absolutely homeless. Precarious housing is defined as having a room in a single room occupancy facility, a rooming house, or hotel/motel as a primary residence, and two or more episodes of being absolutely homeless in the past year. The criteria for presence of a mental disorder includes two of five observed behaviors, one of five functional impairment items, written documentation of a diagnosed disorder or psychiatric inpatient admission, and/or an indication on the Mini International Neuropsychiatric Interview²⁸ of the presence of current major depression, bipolar disorder, posttraumatic stress disorder, panic disorder, or psychotic disorder (more details available from the authors).

not available for two participants). Current living situations were shelters ($N=10$), the street ($N=6$), single room occupancy hotel or hotel ($N=9$), and a mix of unstable circumstances including hostels, transitional housing, and or couch surfing ($N=4$; living situation was missing for one). Nearly half ($N=13$) reported having gotten “extra help in school” (an indicator of possible cognitive impairment), not including two who reported being unsure whether they had.

Procedure

Instrument items selected for testing came from six instruments:

- The Colorado Symptom Index (CSI);²⁹
- The GAIN Substance Problems Scale (GAIN SPS)³⁰
- The Vocational Time-Line Follow-Back Questionnaire (VTLFB)³¹
- The Comorbid Conditions (CMC) list^{32–34}
- The Community Integration Scale (CIS)^{35–37}
- The Health, Social, Justice Service Use (HSJSU) inventory.

The CSI is a 14-item scale designed to measure the past month frequency of symptoms of major mental illnesses. Ratings are made on a five-point scale from 0 (not at all) to 4 (at least every day).²⁹ The CSI has established reliability (test–retest above 0.70), internal consistency (alpha above 0.90) and convergent and content validity.^{38,39} The GAIN SPS is a 16-item subscale from a comprehensive assessment instrument that measures problems resulting from alcohol and other drug use (including street drugs and nonmedical use of prescription drugs).³⁰ Response options are: “past month,” “2–12 months ago,” “1 or more years ago,” and “never.” Internal consistency of this scale is reported to be 0.90 and there is substantial documentation of psychometric characteristics for the broader instrument.³⁰ The VTLFB elicits the recent history of employment and work-related information, including income and education, in the past 3 months.³¹ The questionnaire was developed to gather vocational and related information for a prior study. For the CMC, we pretested a list of terms for 30 medical conditions lasting 6 months or longer, such as epilepsy and hepatitis, which we compiled from several sources.^{32–34} On this questionnaire, the respondent is asked about the presence of any of the conditions and responses are simply “yes,” “no,” “don’t know,” or “declined.” The CIS was a set of items from three shorter scales that measured aspects of three domains of community integration (physical, social, and psychological).^{35–37} The HSJSU covers health services (including visits with a health professional, outpatient, and emergency room visits), social services (including such things as visits for income-support services, food banks, and drop-in centers), and justice services (including police contacts, arrests, and court appearances). It was necessary to develop this questionnaire specifically for the study since none of the service use inventories we reviewed was sufficient for our research questions and population, and as such it had not undergone psychometric evaluation prior to pretesting.

We also report herein on two other scales which were not pretested but for which illustrative issues arose in later piloting—the Recovery Assessment Scale (RAS)⁴⁰ and the Quality of Life Interview (QoLI).^{41,42} The RAS is a 22-item questionnaire that taps current perceptions of personal recovery items such as “I can handle what happens in my life.” The five-point response scale anchors range from “strongly disagree” to “strongly agree.” Reliability and validity are reported to be good by Corrigan et al.⁴⁰ The QoLI is a condition-specific instrument for individuals with mental disorders that measures quality of life in domains such as social relationships,

finances, safety, and general life satisfaction.⁴¹ We used the 20-item version of the QoLI, with reliability and validity confirmed using item response theory methods.⁴² We focused on the six instruments listed above for pretesting because they had little prior use in this population, were newly developed, or had content that was considered to be high risk for problems of relevance, comprehension or recall, or sensitivity/acceptability. Two researchers independently selected items from the six instruments that were considered high risk for problems. Differences were reconciled through discussion, which included a third researcher and consensus decisions were made on a final list of items to be tested. Table 1 provides details for each instrument, including prior use in this population, the number of items and probes used, the number of participants tested, and example items and probes for each instrument. Due to the large number of items to be tested, the interview content was divided into two standard interview schedules; each designed to take about 1 hour. Interviews were conducted in spaces that allowed for privacy in local shelters or drop-in facilities or in some cases outside (according to participant preference), and the schedules were administered on an alternating basis. All interviewers had previous experience conducting interviews with homeless populations and were trained in cognitive interviewing methods. Training included specific practice with the interview schedules including role play. Interviewers were also involved in two iterations of refinement of the interview schedules.

The schedule included introductory text on the purpose of testing, instructions on how to respond, and relevant scripted probes for each item.²⁵ Initial interview questions solicited demographic information including age, gender, current living situation, and referral source. An item intended to serve as an indicator of possible cognitive impairment: “Did you ever get extra help with learning in school?” was also included.

Two cognitive interviewing techniques were used.²⁵ First, participants were asked to “think aloud” while responding to the item (i.e., to talk about how they interpreted the question and how they came up with their answer). Second, each test item was read aloud and participant’s direct responses were recorded followed immediately by probing questions for more in-depth exploration. Probes were selected or adapted from those recommended in the cognitive interviewing literature to assess particular issues of concern about item content, construction, and possible participant reactions for each questionnaire.²⁵ We opted for concurrent probing (probe questions being presented immediately after administration of each item) to maximize the respondents’ recollection of their thoughts at the time. Additional optional probes were also provided for further exploration as needed and the interviewers were also trained to use spontaneous probes as needed. Questions were also designed to elicit responses about general item construction and suitability of language and in one case, opinions about the feasibility of data collection processes were solicited.

One-on-one interviews were used to best accommodate lower literacy levels and to reflect the planned administration mode of the main study. Interviewers took extensive notes and recorded most comments made by participants verbatim. After the interviews were over, interviewers recorded additional and general observations.

Analysis

Simple analysis of the content of text-based responses was used to get a small sample sense of the prevalence of particular types of problems. Responses were summarized and compared across respondents and sites for each item, noting similarities,

TABLE 1 Instruments, previous use in the target population, example items, testing goals, and cognitive interviewing probes

Instrument/numbers of items, probes and participants	Prior use in the homeless population?	Example items	Purpose(s) of testing	Example probes
Colorado Symptom Index (CSI)	Developed specifically for homeless populations	In the past month, how often have others told you that you acted “paranoid” or “suspicious”?	Comprehension	Can you repeat the question in your own words? Who do you think counts as “others” in this question?
6 Items		In the past month, how often have you felt like seriously hurting someone else?	Recall	Can you remember what time period the question was asking about?
23 Probes			Sensitivity/Acceptability	How would most people you know respond to this question?
16 Participants				
Global Assessment of Individual Need—Substance Problem Scale (GAIN SPS)	Developed for individuals with substance use and mental health issues more broadly and some use in homeless populations	When was the last time that your alcohol or other drug use caused you to feel depressed, nervous, suspicious, uninterested in things, reduced your sexual desire or caused other psychological problems?	Comprehension of complex item stems	Can you tell me what you think they are trying to get at in this question?
7 Items ^a			Sensitivity/acceptability	Do you think it is OK to talk about in an interview, or is it too uncomfortable?
20 Probes				
14 Participants				
Vocational Time-Line Follow-Back (VTLFB)	Developed for individuals with mental disorders broadly; not used in homeless populations to date	Have you worked at any job for a week or more (including volunteer jobs and paying jobs) during this period? For the month of ____ how much was your total income? I now want to ask about some ways people living on the street have said they get income—I’ll read a list and I’d like to know for each one if you feel it is a way that people get income on the street in this community.	Recall	How well do you remember this?
7 Items			Recall	Are there any other sources of income on the street that I didn’t list?
16 Terms			Relevance	
10 Probes				
16 Participants				

TABLE 1 (Continued)

Comorbid Conditions List (CMC) 2 Items 30 Terms 6 Probes 16 Participants	Source items came from a general population survey and a study of individuals with mental health issues; not used in homeless populations	Do you currently have... [List of conditions, e.g., asthma, TB, migraine headaches, dental problems, high blood pressure, cancer]	Comprehension Relevance	Are there any health problems that you have or people you know have that I didn't ask about? Are you familiar with this condition? Do you have a better word for this condition?
Health, Social, Justice Service Use (HSJSU) 16 Items (5 in depth) 33 Terms 26 Probes 14 Participants	Source items from a range of service use inventories; some specific to individuals with mental health issues but none specific to homeless populations	You said you had some services at a hospital (not including ER visits) but you didn't stay overnight. How many times did this happen in the past 6 months? You mentioned that you have taken prescription medications in the past 6 months. Do you carry any of your medications with you?	Recall Availability of medication packages/prescriptions Acceptability of a potentially sensitive process	How was it to remember for the past 6 months? Would it be easier to remember for 3 months? What about the past month? Do you think other people would be comfortable if we asked them to bring their medication bottles to the interview?
Community Integration Scale (CIS)	Original scales were used with individuals with serious mental illness but not homeless populations	<i>Physical integration:</i> In the past month, have you visited a park or museum? In the past month, have you gone for a walk? <i>Social integration scale:</i> In the past month, have you received a ride from a neighbor? In the past month, have you discussed with a neighbor such things as home repairs, gardening, or other matters related to improving a home? <i>Psychological integration scale:</i> I feel at home on this block. I expect to live on this block for a long time	Relevance Comprehension relevance	I want you to tell me if the item even applies to you or your friends—that is if you and your friends would EVER do that? What does the word “block” mean to you? If you had the choice, would you keep the word “block” or change it to a different word?

^aFour of these items were about substance use but not directly from the GAIN

differences, and frequencies of types of responses. Recommendations were made for item revisions and/or adjustments to the administration process. Interviewers reviewed and commented on the findings and recommendations. In the case of highly standardized instruments, items were not changed but issues were noted to assist with interpretation of subsequent trial results. Where appropriate, item alternatives or revisions were discussed with instrument authors.

RESULTS

Results are presented in terms of the three primary areas of inquiry: (1) relevance, (2) comprehension and recall, and (3) sensitivity/acceptability. Although comprehension and recall are different cognitive stages in theory,²⁶ we have combined them in reporting our results because they were so closely related in our participants' responses.

Relevance

According to our participants, many of the items from the standard scales were not relevant to their circumstances. The community integration scale items were particularly problematic in this regard. For example, the set of items meant to measure physical integration covers activities such as going for walks, seeing movies, visiting parks or museums, etc. Although many participants said that they had visited these locations, it was stated to be out of necessity (i.e., to seek food or shelter) rather than for the implied measurement intention—seeking greater involvement in the community. Two items: “going for a walk” and “going to a store” were so universally endorsed that, in this population, they would be unlikely to provide useful information. Other items were seen by participants as never happening for them or the people they know (e.g., “attending a sporting event”). The wording of one item: “attending a church or place of worship” was not perceived to be sufficiently inclusive of diverse spiritual practices, particularly for Aboriginal participants, resulting in a recommendation for rewording.

Many of the items designed to assess social integration were also reported by our participants as having little relevance to their life circumstances. For example, “borrowing things from a neighbor such as books, magazines, dishes, tools, recipes” and “discussing home repairs” with a neighbor were not common practices for our participants. Many of these items seemed to reflect more middle-class assumptions about interactions with neighbors. Some participants appeared to be annoyed by these items, perhaps because they reflected the extent of their marginalization from mainstream society. The terms “neighbor” and “neighborhood” were also variable in interpretation by our participants. Many reported sleeping in one place and spending their day in a different part of the city, so neighborhood identification was nonspecific. Furthermore, the language in some items was seen to be dated by some (e.g., church bazaar). Due to lack of relevance and potential interpretation problems, these items were not included in the larger study.

The final set of items tested assessed the psychological domain of community integration (i.e., sense of belonging). When asked if they “feel at home on their block,” many participants were unsure what the question meant. Probing revealed that the identification with a specific “block” was difficult for many. For example, one participant stated “I don’t live on the same block—it changes a lot. ‘Area’ might be better.” Some respondents interpreted the word “block” to mean the immediate surroundings or the street while others interpreted it to mean a larger area or

broader community. Most participants reported that the items in this group were vague and did not adequately assess a sense of belonging. Some suggested being more direct: “Ask it straight up. Do you feel you belong here?” The items were modified accordingly.

While most of the content on the standardized CSI worked well, even this instrument, which was designed specifically for homeless populations, contained terms that did not resonate with many of our participants. For example, participants interpreted the term “psychological and emotional difficulties” (from the CSI instructions) in different ways, including “acting crazy,” psychotic symptoms, low mood, or general problems with functioning. When given a list of alternative terms, most participants preferred the term “mental health,” however, there was no agreement on the use of the words “issue” or “problem.” We concluded that the term “psychological and emotional difficulties” was too ambiguous for homeless individuals with mental disorders. Given that a variety of terms for mental disorders were used across the instruments in our battery, we recommended that the more clearly understood and favored term “mental health problems” be used consistently across the full battery of instruments, although there was not complete consensus on this term among our participants.

The original list of income sources from the VTLFB did not include several sources that our participants reported to be common among homeless people (e.g., bottle collecting and recycling, cleaning car windshields, busking, and panhandling). Based on specific feedback about these other income sources, as well as illegal means of obtaining income, and with the original author’s permission, we added this content to the VTLFB.

On the HSJSU, most services listed were reported to be relevant by participants, however, there was some confusion around terms used for various service providers (e.g., “tenant support worker,” “life skills worker”) as well as certain service locations (e.g., “drop-in centres,” “specialized clinics”). Types of services offered by different professionals (e.g., “counseling,” “case management,” and “help with daily living”) were not well discriminated. With regard to medications, most participants could name their medications but could not specify the dosage. Many indicated that they do not carry their medication containers because of the risk of loss or theft. Given the importance of capturing information related to medication use in the study, we opted to collect this information including requesting medication packaging in the main study for a 2-month pilot period. The pilot confirmed the poor feasibility of collecting medication information via self-report and packaging and alternative sources of data (administrative data) for this information were identified for the trial.

Comprehension/Recall

Some of the items and instructions tested were poorly understood by our participants. Comprehension problems were reported and observed for items with lengthy stems and items with higher level vocabulary as well as items with multiple alternative phrases. For example, in the CSI, one of the questions asks “In the past month, how often have you felt nervous, tense, worried, frustrated or afraid?” Respondents found the list of adjectives to be confusing and one respondent said, “Having all these words is frustrating, it’s overkill.” Another item that participants were observed to struggle with was “In the past month, how often did you have trouble thinking straight, or concentrating on something you needed to do like worrying so much, or thinking about problems so much that you can’t remember or

focus on other things?” Recommended remedies for this problem included training interviewers to slow down when presenting complex stems and partitioning them if necessary.

In addition, there were terms used in the CMC list that were not easily understood by our participants. Many were unfamiliar with the following medical terms: “fibromyalgia,” “urinary incontinence,” “bowel disorder,” and “anemia.” As such, some items representing conditions considered to be low frequency in this population such as “fibromyalgia” were dropped from the list of medical conditions, while alternative terms preferred by respondents were used for “urinary incontinence” (“inability to hold urine”), “bowel disorder” (“bowel problem such as Crohn’s disease or colitis”), and “anemia” (“low iron in the blood”).

In general, recall was variable, depending on the item and its salience in the participants’ lives. For example, nearly all participants reported that they could easily recall the age at which they first got drunk or started using drugs. Recall of details related to work (e.g., hours worked per week and income) was much more variable. Reports of ability to remember various health, social, and justice services over a 6-month period were also wide ranging. For salient, low-frequency events (e.g., an occasional emergency room visit) recall was reported and observed to be quite good; whereas for routine, high-frequency events (e.g., visits from outreach workers, use of community meal programs) it was reported to be quite poor, and recall over long periods of time was often reported to be too difficult. As one participant stated, “Ya, you forget, maybe a month is too long, some of the guys in the shelter can’t remember what they did yesterday, its hard work keeping track of your life, all the bits and pieces, a lot don’t work at it.” Instructions were revised to include precise definitions of all terms and time frames were clearly emphasized. In few instances, shorter recall periods were used to sample the frequency of events rather than attempting to collect total frequencies over long time periods.

In contrast, reporting details related to substance use-related problems in the GAIN SPS and related question was observed and reported to be quite good. Most participants reported that they were familiar with substance-use related terms and various consequences of use (e.g., hepatitis, “the shakes”). Many confirmed that they had little problem formulating responses about frequency of use and amount of money spent per month on substances. Some comments were “you just know” and “I’ll never forget—it’s a hard life.”

Sensitivity/Acceptability

One important goal of pretesting was to ensure effective handling of sensitive content, in keeping with the broader person-centered philosophy of the intervention being trialled, and to prevent attrition that could result from invasive or offensive content. For example, we examined our participant’s responses to items about suicidal and homicidal ideation from the CSI. It was reassuring to find that the item about suicidal ideation was generally acceptable to participants. Moreover, as one participant commented, “It is okay. It is a sensitive thing, but if a person is planning suicide, he would definitely need help.” There were a few stronger reactions to the item about homicidal ideation, but these were still a minority. “They will answer it but will they give you an honest answer? That’s the question.” “It depends on the person, there are secretive people, quiet people, or talkative ones, some people could flip out or take it offensively.” These items were retained in the study, and recommendations focused on revisions to the preamble to these questions including

acknowledging their personal nature, reiterating the option to decline responses and assurances of confidentiality.

On the other hand, there were some items that were pretested which may seem benign, yet that elicited some negative reactions. For example, some participants felt that the items related to jobs and being part of a community were very sensitive and a few responded that questions about contact with friends and family raised some issues for them. One participant in particular became very upset when being asked about community activities and required a break before moving on in the interview because the items reminded her of previously happy times, now lost to her. Some participants also reported that terms used in questions about ways of obtaining income were sensitive. For example, many participants did not approve of the term “begging” and one participant said, “Begging sounds cruel—panhandling is a nicer way to put it.”

Other items that were expected to be sensitive and confirmed by some participants to be so were those related to criminal justice activity (i.e., arrests, charges, and incarcerations). “Most people won’t answer it, [they would] want to know what [you were] getting at if [they were] on the street.” Other comments were “some might lie” and “they might think its pushy.” It was interesting, though, that even as several respondents felt that others would *not* report these events, they themselves provided detailed and seemingly honest responses about their own experience. These questions were retained because of their importance to the research questions but strategies including explanations about the purpose of such questions, prior notice of the line of questioning, acknowledgement of sensitivity and reiterated assurances of confidentiality, and the right to decline responses were used in the interview guide for the main study. Agreement between some of these items and administrative data sources will also be examined.

Questions about substance use (most from the GAIN SPS) were also more favorably received than predicted. Participants generally reported that not only were these questions acceptable, but even important to ask in order that the study findings would ultimately help individuals with their substance use issues. As one participant stated: “It’s okay to me. The more research that can be done, the better. It’s not a fun way to live.”

DISCUSSION

We found that pretesting of questionnaire items using cognitive interviewing methods was a feasible and useful way of capturing information to inform instrument framing and instructions, item inclusions and revisions, as well as interpretation of results for items that could not be changed or dropped. Matter and colleagues had a similar experience with these methods in a homeless sample in Seattle.²⁷

It was our impression that many issues that were identified would not have been identified without the explicit solicitation of feedback and specific probes employed. For example, most participants agreed that the adjective “mental health” should be used but suggested a variety of different nouns including: issues, problems, concerns, difficulties, and symptoms. Perhaps this is not surprising, given that the term “mental health issues” is greatly affected by personal experience and broader social attitudes. In fact, homeless individuals may internalize social discourse around mental illness and homelessness such that their preference of terminology may not be the most inclusive and least discriminating.

Our testing process confirmed that it was feasible to administer sensitive and complex questions on the CSI and GAIN SPS, and to ask about chronic health conditions on the CMC in this study population with minor adjustments to questionnaire preambles or terms used and specific interviewer training. The testing was essential for valid data collection on the VTLFB and HSJSU, which were newly developed questionnaires that solicit factual information about life events. Our finding resulted in many revisions to these questionnaires, including to instructions, terms, and recall periods. The community integration scale items required the most extensive revisions for our population given that their content covered activities of everyday living that were developed for more conventional life circumstances. Fortunately, these items came from scales for which revisions were possible.

While we did not set out to compare and contrast various cognitive interviewing approaches, we offer some general comments about these options based on our experience. First, we involved our interviewers in reviewing and interpreting findings based on the belief that direct observation of reactions to particular questions was important in capturing subtleties that go beyond verbal responses. In addition, given that our analytic approach was relatively simple, we sought to validate conclusions via interviewer review and feedback. Some approaches formalize the involvement of interviewers through systematic interviewer debriefing.²⁵ While we did not go that far, we do feel that the interviewers played a valuable role for the intended purposes in the study.

Second, because of the very large number of instruments and items in the full instrument battery for the main study, instead of a completely data-driven approach (i.e., assessing all items in multiple rounds of testing),²⁵ we used expert consensus to select priority items for testing in advance. Our mistaken assumptions about sensitivity affected the items we selected for pretesting. Instruments with content initially considered benign and positive: the RAS (with its recovery-based content) and the QoLI (with its quality of life content) were not slated for pretesting. During the subsequent pilot, we noted that despite the positive wording, the process of providing repeated low ratings was very demoralizing for respondents because they reinforced the daily physical and emotional struggles of life on the street and hardships such as alienation from family. After observing this phenomenon in the pilot, the preambles to these questionnaires were modified accordingly. Kavanaugh and colleagues stress the importance of avoiding assumptions in research with vulnerable participants and “focusing on what a participants’ situation means to him or her, as opposed to what it means to the researcher.”⁴³ One way that this result could have been prevented would to have preceded the process with a formal expert appraisal process.²⁵ A second way would have been to test more instruments and items with fewer participants for each, although that approach may have reduced our ability to observe the full diversity of responses and to generalize recommendations. One of the many strengths of cognitive interviewing is the flexibility to mix methods to achieve the right balance for a given project and context.²⁵

Our third observation relates to the balance between “think aloud” and “probing” techniques. For our participants, direct probing solicited more substantial feedback than the “think aloud” approach. Edwards et al. also found the “think aloud” technique to be challenging in a street-involved population.⁴⁴ For homeless individuals with mental health issues and who also often have cognitive impairment, it is not surprising that the metacognitive skills required for the “think aloud” approach were quite challenging.

Our specific findings on recall periods were very similar to other reports in homeless populations²⁰ and other vulnerable populations.⁴⁴ They were not distinctly different from what is known about recall in other survey populations.⁴⁵ While recommendations for questionnaire construction to enhance recall of past events may be no different in this population, the process provided valuable information about recall for specific information and the maximum measurement periods for which valid responses could be expected.

While cognitive interviewing pretesting did have value in addressing comprehension, recall, and sensitivity of items, the greatest impact was in the realm of relevance. A common observation was that many items lacked relevance because of prior normative assumptions. Frequently, items and response options simply did not apply to the life circumstances of these individuals. Our participants often laughed uncomfortably at these types of questions or expressed exasperation with them. Matter and colleagues had similar findings, noting poor fit between items associated with home ownership and a conventional middle class life and their participants' circumstances.²⁷ Even after omitting such items, we still recommended the addition of "don't know" or "declined" answer options for many questions. Without these response options, participants may find the interview process to be a demoralizing experience because their answer may not be among those offered, which may imply that they are abnormal.

Despite some general consistency in responses, there was heterogeneity among our participants in opinions about questions and wording. This variation in response was particularly true for sensitivity of items. Because of individual histories and life circumstances, some reactions were counter-intuitive. We not only learned that our assumptions about some items being benign and others being invasive were often wrong, but further that acceptability could neither be completely predicted nor completely guaranteed. The training of our interviewers included enhancing awareness of this phenomenon and preparedness for the unexpected. Overall, findings were consistent with previous literature, which indicates that "how you ask" is as important as "what you ask."^{20,21,45}

The limitations of our study include the use of a relatively small, convenience sample which may not reflect the full range of subpopulations of homeless individuals with mental disorders. Although the sample size is in keeping with what is recommended for cognitive interviewing methods²⁵ and was similar to Matter and colleagues²⁷ given the heterogeneity nature of the homeless population, a larger sample size would have perhaps yielded more diverse results. That being said, we made an effort to draw the sample from a variety of locations (Toronto, Vancouver, and Winnipeg) and from a variety of places (drop-ins, shelters, and the street). And, even with the small sample, responses were reasonably congruent across sites and participants for many of the items tested. Another important limitation is the use of English language interviews and items only; our methods did not permit examination of linguistic differences. Finally, because of timelines for the larger study, we used only one round of testing and revision; a second or even third round would have allowed us to validate the instruction and item changes made.

CONCLUSIONS

In this study, cognitive interviewing methods were used to systematically test targeted questionnaire items from a battery of quantitative instruments selected for a large multisite trial of supported housing interventions for homeless individuals with

mental disorders. Most of the instruments had no published psychometrics in this population. Much was learned about the suitability and acceptability of items and instruments for the larger multisite trial. Cognitive interviewing methods worked well and elicited information crucial to effective and respectful data collection in this unique population. Pretesting study instruments, including standard instruments, for use in special populations, such as homeless individuals with mental disorders, is important for optimizing measurement as well as interpreting findings. It is also critically important that research instruments be designed, in the first place, to be appropriate to vulnerable and underserved populations.

ACKNOWLEDGMENTS

Appreciation is extended to Kimberley Lewis-Ng and Rebecca Godderis who assisted with the planning stages of this study. Susan Mulligan, Verena Strehlau, and Melinda Markey are also thanked for conducting some of the interviews. ACH was supported for this work through the University of Toronto "Comprehensive Research Experience for Medical Students." The results of the study were presented at the University of Toronto Medical Student Research Day in February 2010. This, and the main study, was made possible through a financial contribution from Health Canada. The Mental Health Commission of Canada is the oversight organization for the study. The views expressed herein solely represent those of the authors and not of the afore-named organizations.

REFERENCES

1. Hwang SW. Homelessness and health. *Can Med Assoc J*. 2001; 164: 229–233.
2. Canadian Institute for Health Information. *Improving the Health of Canadians: Mental Health and Homelessness*. Ottawa, Canada: CIHI; 2007.
3. Goering P, Tolomiczenko G, Sheldon T, Boydell K, Wasylenki D. Characteristics of persons who are homeless for the first time. *Psychiatr Serv*. 2002; 53(11): 1472–1474.
4. Aubry T, Klodowsky F, Hay E. *Panel Study on Persons Who Are Homeless in Ottawa: Phase 1 Results*. Centre for Research on Community Services. Ottawa, Canada: University of Ottawa; 2003.
5. Patterson M, Somers JM, McIntosh K, Shiell A, Frankish CJ. *Housing and Support for Adults with Severe and/or Mental Illness in British Columbia*. British Columbia, Canada: Centre for Applied Research in Mental Health and Addiction, Simon Fraser University; 2008.
6. Pinkney S, Ewing S. *The Costs and Pathways of Homelessness: Developing Policy-Relevant Economic Analyses for the Australian Homelessness Service System*. Melbourne, Australia: Institute for Social Research, Swinburne University of Technology; 2006.
7. Hopper K, Baumohl J. *Redefining the Cursed Word: a Historical Interpretation of American Homelessness*. Westport, CT: Oryx; 1996.
8. Boydell K, Goering P, Morrell-Bellai T. Narratives of identity: re-presentation of self in people who are homeless. *Qual Health Res*. 2000; 10(1): 26–38.
9. Kessell ER, Bhatia R, Bamberger JD, Kushel MB. Public health care utilization in a cohort of homeless adult applicants to a supportive housing program. *J Urban Health*. 2006; 83 (5): 860–873.
10. Hwang SW, Ueng JJM, Chui S, et al. Universal health insurance and health care access for homeless persons. *Am J Public Health*. 2010; 100: 1454–1461.
11. Wright JD. *Address Unknown: the Homeless in America*. New York, NY: Transaction Publishers; 2009.

12. Tsemberis S. From streets to homes: an innovative approach to supported housing for homeless adults with psychiatric disabilities. *J Community Psychol*. 1999; 27: 225–241.
13. Tsemberis S, Eisenberg RF. Pathways to housing: supported housing for street-dwelling homeless individuals. *Psychiatr Serv*. 2000; 51: 487–493.
14. Pearson CL, Locke G, Montgomery AE, Buron L. *The Applicability of Housing First Models to Homeless Persons with Serious Mental Illness*. Washington, DC: US Department of Housing and Urban Development Office of Policy Development and Research; 2007.
15. Tsemberis S, Gulcur L, Nakae M. Housing first, consumer choice, and harm reduction for homeless individuals with a dual diagnosis. *A J Public Health*. 2004; 94(4): 651–656.
16. Sadowski LS, Kee RA, VanderWeele TJ. Effect of a housing and case management program on emergency department visits and hospitalizations among critically ill homeless adults: a randomized trial. *JAMA*. 2009; 301(17): 1771–1778.
17. Larimer ME, Malone DK, Gardner MD. Health care and public service use and costs before and after provision of housing for chronically homeless persons with severe alcohol problems. *JAMA*. 2009; 301(13): 1349–1357.
18. Kertesz SG, Crouch K, Milby JB, Cusimano RE, Schumacher JE. Housing first for homeless persons with active addiction: are we overreaching? *Milbank Q*. 2009; 87(2): 495–534.
19. Yanos PT, Barrow SM, Tsemberis S. Community integration in the early phase of housing among homeless persons diagnosed with severe mental illness: successes and challenges. *Community Ment Health J*. 2004; 40(2): 133–150.
20. Gelberg L, Sieckel N. Accuracy of homeless adults' self-reports. *Med Care*. 1997; 35(3): 287–290.
21. Goldfinger SM, Schutt RK, Seidman LJ, et al. Self-report and observer measures of substance abuse among homeless mentally ill persons in the cross-section and over time. *J Nerv Ment Dis*. 1996; 184(11): 667–672.
22. Rosen MI, McMahon TJ, Rosenheck RA. Homeless people whose self-reported SSI/DI status is inconsistent with social security administration records. *Soc Secur Bull*. 2007; 67(1): 53–62.
23. Mallinson S. Listening to respondents: a qualitative assessment of the Short-Form 36 Health Status Questionnaire. *Soc Sci Med*. 2002; 54: 11–21.
24. Collins D. Pre-testing survey instruments: an overview of cognitive methods. *Qual Life Res*. 2003; 12: 229–238.
25. Willis GB. *Cognitive Interviewing: a Tool for Improving Questionnaire Design*. New York, NY: SAGE; 2005.
26. Tourangeau R, Rasinski K. *Psychol Bull*. 1988; 103: 299–314.
27. Matter R, Kline S, Cook KF, Amtmann D. Measuring pain in the context of homelessness. *Qual Life Res*. 2009; 18: 863–872.
28. Sheehan DV, Lecrubier Y, Harnett-Sheehan K, et al. The Mini International Neuropsychiatric Interview (M.I.N.I.): the development and validation of a structured diagnostic psychiatric interview. *J Clin Psychiatry*. 1998; 59(suppl 20): S22–S33.
29. Shern DL, Wilson NZ, Saranga Coen A, et al. Client outcomes II: longitudinal client data from the Colorado treatment outcome study. *Milbank Q*. 1994; 72(1): 123–148.
30. Dennis ML, Chan Y, Funk RR. Development and validation of the GAIN short screener for internalizing, externalizing and substance use disorders and crime/violence problems among adolescents and adults. *Am J Addict*. 2006; 15(suppl 1): S80–S91.
31. Latimer EA, Lecomte T, Becker DR, et al. Generalisability of the individual placement and support model of supported employment: results of a Canadian randomised controlled trial. *Br J Psychiatry*. 2006; 189: 65–73.
32. Canadian Community Health Survey 4.1. <http://www.statcan.gc.ca/cgi-bin/imdb/p2SV.pl?Function=getSurvey&SurvId=3226&SurvVer=1&InstaId=15282&InstaVer=4&SDDS=3226&lang=en&db=imdb&adm=8&dis=2>. Accessed August 20, 2010.
33. National Population Health Survey. <http://www.statcan.gc.ca/concepts/nphs-ensp/index-eng.htm>. Accessed August 20, 2010.

34. Mares AS, Rosenheck RA. *HUD/HHS/VA Collaborative Initiative to Help end Chronic Homelessness National Performance Outcomes Assessment: Preliminary Client Outcomes Report*. West Haven, CT: Northeast Program Evaluation Center; 2007.
35. Segal SP, Aviram U. *The Mentally Ill in Community Based Sheltered Care: a Study of Community Care and Social Integration*. New York, NY: Wiley; 1978.
36. Aubry T, Myner J. Community integration and quality of life: a comparison of persons with psychiatric disabilities in housing programs and community residents who are neighbors. *Can J Commun Ment Health*. 1996; 15: 5–20.
37. Chavis DM, Hogge JH, McMillan DW, Wandersman A. Sense of community through Brunswick's lens: a first look. *J Community Psychol*. 1986; 14: 24–40.
38. Boothroyd RA, Chen HJ. The psychometric properties of the Colorado Symptom Index. *Adm Policy Ment Health*. 2008; 35(5): 370–378.
39. Conrad KJ, Yagelka JR, Matters MD, et al. Reliability and validity of a modified Colorado Symptom Index in a national homeless sample. *Ment Health Serv Res*. 2001; 3 (3): 141–153.
40. Corrigan PW, Salzer M, Ralph RO, Sangster Y, Keck L. Examining the factor structure of the recovery assessment scale. *Schizophr Bull*. 2004; 30(4): 1035–1041.
41. Lehman AF. Measures of quality of life among persons with severe and persistent mental disorders. *Soc Psychiatry Psychiatr Epidemiol*. 1996; 31: 78–88.
42. Uttaro T, Lehman A. Graded response modeling of the quality of life interview. *Eval Program Plann Spring*. 1999; 22: 41–52.
43. Kavanaugh K, Moro TT, Savage T, Mehendale R. Enacting a theory of caring to recruit and retain vulnerable participants for sensitive research. *Res Nurs Health*. 2006; 29: 244–252.
44. Edwards M, Thomsen SC, Toroitich-Ruto C. Thinking aloud to create better condom use questions. *Field Methods*. 2005; 17: 183–199.
45. Bradburn NM, Sudman S, Wansink B. *Asking Questions: the Definitive Guide to Questionnaire Design for Market Research, Political Polls, and Social and Health Questionnaires*. San Francisco, CA: Jossey-Bass; 2004.