Cross-cultural comparisons of International Tobacco Control Policy Evaluation (ITC) survey questionnaires: A multinational, multicultural and multilingual cognitive interviewing study

James Thrasher, PhD, MA, MS
Department of Health Promotion, University of South Carolina, USA
Departamento de Investigación sobre Tabaco, Instituto Nacional de Salud Pública, México

Anne Quah, PhD
Department of Psychology, University of Waterloo, Canada

Ashlee Watts, MA, and Gregory Dominick, MA
Department of Health Promotion, University of South Carolina, USA

Ron Borland, PhD
Cancer Council of Victoria, Melbourne, Australia

Ana Dorantes Alonso
Instituto Nacional de Salud Pública, México

Warwick Hosking, PhD
School of Psychology, Victoria University, Melbourne, Australia

Rahmat Awang, PharmD, Maizurah Omar, PhD and Ahmad Shalihin Mohd Samin, MSc.Ed
National Poison Centre, Universiti Sains Malaysia, Malaysia

Buppha Sirirassamee, PhD
Institute for Population and Social Research, Mahidol University, Thailand

Marcelo Boado, PhD
Departamento de Sociología, Universidad de la República, Montevideo, Uruguay

Kristen Miller, PhD
National Center for Health Statistics, USA

Abstract

Research to identify the factors that promote or inhibit tobacco use often involves comparison of survey data from disparate groups. Conclusions regarding which factors explain such differences are strengthened when researchers can rule out the influence of systematic measurement error due to cultural differences across the populations of interest. Cognitive interviewing techniques are increasingly used to assess the type and extent of measurement error for particular questions within populations. However, existing literature is only suggestive on the use of this methodology across cultural and linguistic contexts. This presentation will describe a cognitive interviewing protocol developed based on tobacco survey questions that were taken from the International Tobacco Control Policy Evaluation Project (ITC) surveys in six (i.e., US, Australia, Malaysia, Thailand, Mexico, Uruguay) of the 14 participating countries. The ITC Project aims to evaluate the impact of policies promoted by the World Health Organization's Framework Convention on Tobacco Control (FCTC) as well as psychosocial mechanisms that explain policy impact. The ITC survey is based on questions with well-established validity in some, if not all, Anglo countries that participate in the ITC Project although some English vocabulary had to be adapted to each of the participating English-speaking country (i.e., Australia, Canada, New Zealand, US and UK). However, when the ITC project expanded to countries that do not have English as their language of communication, the surveys were translated into the respective languages (e.g., Malay, Thai, Spanish, Chinese, Korean, Hindi). Because countries have gradually joined this international project, coordination across translation efforts has been difficult, if not impossible, potentially compromising the comparability of question comprehension across countries. The cognitive interviewing protocol we describe focuses on four linguistic groups in six countries, with the aim of determining differential comprehension and meaning of a number of selected tobacco survey questions among adult smokers within and across these countries. In describing the protocol development, we will discuss the importance of: 1) definitions of study constructs that make sense across the cultural settings in which the survey will be administered; 2) multiple bi-lingual partners in each linguistic/cultural group involved; 3) developing structured probes that anticipate concerns about question

comprehension and meaning; 4) audio recording and transcription of interviews to capture open-ended responses; 5) addressing translation issues, both when translating the cognitive interview questionnaire and when translating participant responses into a primary language for cross-country analysis; 6) coordination of analysis. We will also discuss results from our analyses of cognitive interviews with convenience samples of 20 adult smokers in each of the six participating countries. In so doing, we will describe participants' responses to tobacco survey questions that aimed to address key tobacco control concepts, such as addiction and the social acceptability of smoking. The protocol we describe can inform future attempts to ensure valid comparative analyses across cultural and national contexts, a critical step towards identifying the foci for efforts to reduce tobacco use.

Introduction

Cross-national, cross-cultural, and cross-linguistic research is often done under the unexamined assumption that question meaning, comprehension, and measurement properties are equivalent across cultural groups (Bollen, Entwisle, & Alderson, 1993; T. W. Smith, 2004). However, cross-cultural differences in language, social conventions, cognitive abilities, and response styles may cause systematic measurement error that biases results in unpredictable ways (Fiske, Kitayama, Markus, & Nisbett, 1998; Harkness, Van de Vijver, & Mohler, 2003). Apparent differences found across socio-cultural groups may be merely due to measurement artifacts, such as systematic group differences in the meanings ascribed to the same question, whether phrased in the same or different languages. Conversely, true differences may be obscured by such factors as the differential influence of social desirability or the exclusion of items that are important indicators of study constructs in one cultural context but not in another. Whereas the implications of these issues appear most obvious for international and multi-lingual comparative research, they may also impede our understanding of different social groups that share the same language and country. In the end, valid cross-cultural comparison demands that measurement error is minimized across the settings and groups of interest (Bollen et al., 1993; T. W. Smith, 2004).

Measurement equivalence in cross-cultural survey research is likely to be enhanced through the following steps: 1) Assess whether the conceptual definitions and theoretical frameworks that orient the study reasonably apply across the survey contexts, as these definitions should inform subsequent stages of question selection, development, adaptation, and assessment; 2) Select and/or develop questionnaire items that reflect conceptual definitions and have good

measurement properties, whether in a single source language or in multiple languages; 3) Adequate translation of items to other languages, with the "committee approach" increasingly seen as superior to back-translation techniques (Brislin, 1970; Forsyth, Kudela, Levin, & Lawrence, 2007; Harkness, Pennell, & Schoua-Glusberg, 2004; Harkness & Schoua-Glusberg, 1998)}; 4) consider and attempt to address systematic differences in "response styles," such as social desirability (Johnson & Van de Vijver, 2004; Marlow & Crowne, 1960), extreme responding (P. Smith, 2004), and acquiescence (Knowles & Condon, 1999); 5) pre-test the resulting questionnaire amongst the populations to be surveyed, in order to determine question comprehension issues, using any of a variety of methods including focus groups (Kreuger & Casey, 2000), cognitive interviewing (Willis, 2005), or behavioral coding (Fowler, 1995; Johnson et al., 2006; Van der Zouwen & Smit, 2004); 6) assess measurement properties post-hoc using statistical techniques, such as structural equation modeling (Bollen, 1989), multi-trait multi-method (Saris, 2003a), multidimensional scaling (Fonatine, 2003), and item response theory approaches (Saris, 2003b). For the purpose of this paper, we focus on cognitive interviewing methods for pre-testing questions, providing an example of how they can be extended to assess measurement error across linguistic and cultural settings.

What is cognitive interviewing?

Cognitive interviewing aims *"to study the manner in which targeted audiences understand, mentally process, and respond to materials presented, with an emphasis on breakdowns in that process"* (Willis, 2005) pp 3. The technique is

recommended before a survey is fielded, so that surveys can be changed to address comprehension issues and thereby improve measurement properties of questions. However, the method can also be used in post-hoc fashion to uncover systematic differences in the meanings attributed to the "same" questionnaire item, whether examining either different linguistic variants of the item or the same linguistic variant across different social contexts {Miller, in press #1974}. Cognitive interviewing follows from research on the cognitive processes involved in responding to survey questions (Willis, 2005). The question response process generally involves comprehension (i.e., meaning of terms and perceived intent of question), retrieval from memory (i.e., availability of and strategies to access relevant information), judgment processes (i.e., motivation to respond and to respond truthfully), and mapping the internally generated response to the question onto the response categories provided. Each step along this pathway may involve the introduction of measurement error. Cognitive interview techniques attempt to reveal such error by prompting study participants to provide information on break downs across the response process. Although this methodology has been suggested for the assessment of translated questionnaires (Harkness, Van de Vijver, & Johnson, 2003), existing literature is only suggestive on how to apply this methodology across cultural and linguistic contexts. Efforts besides the present one are currently under way to develop such a protocol {Miller, 2009 #1975}.

What is behavioral coding?

Another promising tool for assessing measurement equivalence is behavior coding, a methodology originally developed to evaluate interviewer performance (Cannell CF, 1975) but which is now also employed to evaluate questions and respondent cognitions (Fowler, 1995; Johnson et al., 2006). This approach usually entails the audio, and more recently the video, recording of respondent-interviewer interactions during field interviews. Recordings are systematically reviewed, with standardized codes assigned to both respondent and interviewer behaviors associated with each question (Van der Zouwen & Smit, 2004). Respondent behaviors can also be recorded in the course of administering the interview {Esposito, 1992 #1976; Hughes, 2004 #1977}. In addition to providing indications of interviewer performance (e.g., reading questions exactly as worded, using non-directive probing), behavior codes can also be employed to identify respondent comprehension difficulties with specific questions (e.g., asking for questions to be repeated or clarified, expressions of uncertainty about question meaning in general or within a particular context, rephrasing of a question prior to answering). A key advantage of behavior coding is its ability to provide systematic and objective information that can be compared quantitatively across cultural groups or questionnaire languages. A potential drawback of its application for cross-cultural research is the likelihood that respondents with varying cultural backgrounds may also vary in the likelihood that they will express the types of overt behaviors that can be captured by this technique.

What is the ITC project?

The International Tobacco Control Policy Evaluation Project, or the ITC Project, aims to evaluate the impact of policies promoted by the World Heath Organization's Framework Convention on Tobacco Control (FCTC), as well as the

psychosocial mechanisms that explain policy impact (Fong et al., 2006; Thrasher et al., 2006). Using a quasi-experimental research design, data from cohorts of adult smokers are compared in countries with and without particular policies of interest (Thompson et al., 2006). The ITC project began in 2002 with four Anglo countries, but since then has expanded to include 10 additional non-Anglo countries in Europe, Asia and Latin America as well as two ITC pilot studies in India and Sudan.

ITC survey questions have relatively well-established validity in some, if not all, Anglo countries that participate in the ITC Project. Nevertheless, some questions were modified to suit particular purposes, while others were altered to reflect common phrasing in each Anglo country (e.g., "factory made" vs. "package" cigarettes). With the expansion of the ITC Project to other countries, the original survey has so far been translated and administered in 11 additional languages (e.g., French, Dutch, German, Malay, Thai, Chinese, Korean, Spanish, Arabic, Hindi and Marathi). Because countries have gradually joined this international project, coordination across translation efforts has been difficult, if not impossible, potentially compromising the comparability of question comprehension across countries.

Research aims:
The research that we will now describe involved the development of a pre-testing protocol with existing ITC questions, drawing insights form cognitive interviewing and behavioral coding techniques, in order to: assess equivalence of comprehension and response errors across countries; Identify probable sources of response error; and Improve question comprehension and equivalence in later waves of data collection. Be increasing the validity of measures and reducing

systematic response bias, this research ultimately aims to increase as confidence in conclusions based upon comparative data analyses; however, we believe that it is important to describe the protocol we followed so that we might engage with other researchers who have similarly aimed to minimize systematic measurement error across settings.

Methods:

Protocol development:

In developing the protocol for this study, we followed the steps outlined in Figure 1. We started by identifying candidate questions from the survey for inclusion in the protocol, because likely response burden would not allow the inclusion of all ITC survey questions. Investigators in each country were queried to identify problematic questions, as well as their rationale for picking these questions. Results from cognitive interview pre-testing of the ITC survey in one country (Mexico) had turned up some issues whose resolution appeared to demand restructuring the source language question. Finally, distributions of question responses across countries were examined to determine if any questions were more likely to provoke non-response in some countries than in others.

Once a suitable list of questions was established, we developed follow-up questions or "probes" to target the specific assessment issues raised. Some of these probes allowed for open responses (e.g., *What does it mean to say that something is "addictive"?*), whereas others were closed (e.g., *Can you think of anything else besides tobacco that is addictive?* Yes; no; don't know). Structured probes were chosen over more flexible, intuitive probing techniques because

of concerns about standardization of the protocol across countries. Structured probes would help ensure that interviewers were following the same protocol, independent of their varied experience with the methodology. Each original survey question was immediately followed by two or three behavioral codes that interviewers would assess based on respondent behavior.

Once the draft protocol was ready, we developed a "translator's guide" to provide a definition of the concept that original survey questions presumably measured, the rationale for each probe that was used to elucidate potential problems with the question, and, if applicable, an indication of the country-specific nature of the problem (see Table 1). Bilingual partners from the ITC project in each country were provided the interview protocol and this guide, and they provided comments on its further elaboration and refinement. The original survey questions were included in the translated protocol exactly as phrased in the corresponding survey instrument for each country, with the guide to assist with the translation of the structured probes and behavioral codes that followed each question. The protocols across countries were adjusted across countries as translation issues were raised and addressed.

Training with country project coordinators and cognitive interviewers took place through a telephone call and accompanying PowerPoint presentation and video spots of interview administration. Cognitive interviewers in each country then piloted the protocol with two different participants, tape recording the interview and entering responses into an Excel spreadsheet. The project coordinator reviewed the audio and resulting data, communicating concerns and needs to adjust to appropriate coordinators and interviewers. Once interviewers could administer the protocol in the

desired manner, participant recruitment began. All interviews were audio recorded to capture open-ended responses and to verify data entry.

Sample:

Convenience samples of participants were recruited from 20 adult smokers in six countries (US, Australia, Uruguay, Mexico, Thailand, Malaysia) in four languages (English, Spanish, Thai and Malay). Recruitment was limited to people 18 years or older, who had smoked 100 or more cigarettes in their lives and who had smoked at least one cigarette in the previous week, which are the same inclusion criteria used for the ITC Project. In all countries but Thailand, participants were recruited from a single metropolitan area (Columbia, SC, USA; Melbourne, Australia; Montevideo, Uruguay; Cuernavaca, Mexico; Minden, Penang, Malaysia). Of these 5 sites, four used intercept survey techniques in public areas, while the other (Montevideo, Uruguay) randomly selected participants from among the adult smokers who participated in the ITC survey there. In Thailand, 5 participants were recruited in similar fashion in Bangkok; however, recruitment was also extended to the other four ITC Project regions, i.e. north-eastern (Nong Khai and Nakhon Ratchasima), north (Chiangmai and Phrae), southern (Nakhon Si Thammarat and Songkhla) and central (Samut Sakhon and Nakhon Pathom), because ITC survey data indicated that smokers in these regions responded quite differently from those in Bangkok.

Analysis

All interviews were audio-taped, transcribed, and entered into an Excel spread sheet. Responses from non-English language countries were translated to English for analysis. Analysis then proceeded through five general stages (Miles & Huberman, 1994). Data associated with the 120 responses to each original ITC question, corresponding behavioral codes, and follow-up probes were entered into a single table in Excel. First, responses to the original survey question, behavioral codes, and structured probes with closed responses were examined, testing for country-level differences using t-tests or Fisher's exact tests. Second, graduate research assistants read open responses to the probes, familiarizing themselves with their content and developing a coding system to characterize this content, including the development of a definition for each code. Third, she re-read each of these narrative segments, applying these codes to them, allowing for multiple codes to be applied to a single segment. The fourth stage involved data display, in which she developed matrices for each code, with matrix rows containing relevant narrative segments from each interview. In the fifth stage, she examined data within each matrix to determine whether the primary concepts could be further subdivided or, perhaps, combined with data given a different code. This data reduction process involved grouping and condensing similar narrative material in order to identify, describe, and contextualize these concepts, including their relationships with other attributes associated with the respondent, such as sociodemographics, responses to the original ITC survey, behavioral codes, or responses to other probes associated with that question. These decisions and the composition of the final tables and associated codes were discussed amongst the research group. Finally, research assistants and bi-lingual primary authors interpreted the data in light of the research aims and the relevant literature.

Results:

Sample population:

Twenty adult smokers participated from each of the six countries, yielding a total of 120 participants. The percentage of

male participants generally reflected the gender distribution of smoking in each country, with more male participants in

Malaysia and Thailand. The age of participants ranged from 18 to 75, with a mean age of 31 in Malaysia and 40 in

Uruguay.   Levels of education were generally comparable across countries, except for the case of Thailand, which was

the only country where rural populations were recruited. The percentage of daily smokers in each country was generally

equivalent across countries (i.e., 80% to 90%), as was the mean number of cigarettes smoked each day (13 to 15 / day),

except for higher frequency of smoking in the US sample (23 / day).

Example results:

One of the questions that our project assessed was the phrase "Tobacco is addictive", which had a 5-point Likert scale

response format indicating extent of agreement. Participant responses were generally comparable across countries, with

a low of 80% of US participants and a high of 100% of Australian participants agreeing or strongly agreeing with the

statement. Table 3 shows the results from the behavioral codes, with no more than one person in any country

experiencing any issues with this question, and no statistically significant differences in the three behavioral codes across

countries. For some of the other questions examined, the behavioral codes registered more issues, which in some cases

were similarly frequent across countries, whereas others differed significantly across countries. Issues registered by

behavioral coding was not limited to Western countries, as issues signaled by behavioral coding were more frequent for some questions in the Asian countries than in the other countries.

The first probe aimed to assess the semantic universe of addictive substances and behaviors, asking participants: *Can you think of anything else **besides tobacco** that might be addictive?* There were more people who could not think of anything else in Mexico and Malaysia than in the other countries (p=0.027). Those who could think of something addictive were asked to name what those addictive things are. Illegal drugs, such as cocaine and heroine, stood out as the most prevalent response across countries, and there was no significant difference the frequency of mentioning this response. Alcohol use was similarly prevalent in the US (n=14), Australia (n=14), Uruguay (n=14) and Mexico (n=10), but mentioned less often in Malaysia (n=3) and Thailand (n=3), perhaps reflecting religious prohibitions against alcohol use in those countries. Non-alcoholic drinks, such as tea, coffee, and mate were mentioned more frequently in Mexico and Uruguay than in the other countries (p=0.046). Legal drugs (such as prescription drugs), food, and sex were mentioned more often in the US and Australia than in Malaysia, Thailand, Mexico and Uruguay, except in the case of comparably frequent mentioning of food in Uruguay, sex in Mexico and legal drugs in Malaysia.

The second probe aimed to asses the primary connotations of the term addiction for participants: *What does it mean to say that something is addictive?* Each narrative response could be assigned up to three codes, depending on content. The most frequently occurring theme across countries concerned general control over smoking. For example, one 39-year-old Australian man said: *It's got a grip of you and you can't let go. It's a must have, you've gotta have it.*

There was no significant difference across countries in the frequency of providing this kind of response. The next most frequent response included a focus on the higher frequency or quantity of consumption, which was mentioned more frequently in Australia (n=10) and Uruguay (n=8) than in the other countries (p=0.015). A focus on the physiology of addiction was typified by a 27-year old US female: T*he body gets tolerance for it and the body won't function normally with out it.* Indeed, this theme appeared more frequently in the US (n=8) and Australia (n=5) than in the other countries (p=0.001). Other noteworthy themes that were different across countries involved psychological control (*If he doesn't get that thing, he cannot do other things…His mind is always imagining that thing…if he doesn't get it, he cannot sit still.* 26 year old Malaysian male), which was most frequently mentioned in Malaysia (p=0.005) and dangerousness, which was only mentioned in Thailand (n=7) and by one US smoker (p=<0.001).

Conclusions from example:

Behavioral coding indicated that there were no systematic biases in overt, observable problems with this question. Moreover, the most prevalent associations with addiction (i.e., illegal drugs and general control) were generally comparable across countries. However, most of the other addictive substances, behaviors, and connotations of addiction differed across countries. Cross-national comparative analyses of responses to this question may yield biased results, since they could be systematically biased across countries. Nevertheless, differences appeared less pronounced when comparing countries with more similar cultural backgrounds (i.e., US and Australia; Uruguay and Mexico; and Malaysia

and Thailand); hence, analyses of these subgroups may not suffer from this bias. This was not always the case, as Uruguay was in some cases more similar to the US and Australia, perhaps because of its strong European heritage.

It appears that the term "addiction" is too abstract to be used in cross-cultural surveys that span divergent cultural groups. Less systematically biased results may result from using questions that focus on concrete behaviors that indicate addiction, such as control over smoking (e.g., Smokers cannot go long without smoking, even when they don't want to smoke) or control over quitting (e.g., Completely stopping smoking is extremely difficult). Indeed, similar questions are often used in clinical settings to assess tobacco dependence (West, 2004). Nevertheless, any attempts to create new questions with more equivalent measurement properties should be verified through additional pretesting, including another round of cognitive interviews.

Discussion:

Conclusions regarding which factors explain such differences are strengthened when researchers can rule out the influence of systematic measurement error due to cultural differences across the populations of interest. The results from this study suggest that coordinated pretesting across cultural, national and linguistic groups using cognitive interviewing and behavioral coding techniques can provide insights that may be useful in ensuring comparison of comparable stimuli. Moreover, protocols that provide structured probes can be implemented by different interviewers across linguistic and cultural settings. In the future, however, data quality could be increased by better training of interviewers to clarify

ambiguous or repetitive responses. The results nevertheless provided evidence that comprehension of some questions appeared generally comparable while other questions generated systematic response error. Particular adjustments were recommended based on pretesting data, much as is done for cognitive interviewing in monolingual settings.

More research is necessary to ensure that these methods yield comparable information across countries. Furthermore, these pretesting techniques may be more suitable for some cultural settings than others. For example, we were initially concerned that behavioral coding techniques could be less useful in Asian countries, where acquiescence appears to be a larger problem than in other countries than in Western countries (Knowles & Condon, 1999). However, for some questions, the behavioral codes registered some problems in Malaysia and Thailand, but not in the other countries. Although these results indicate that the behavioral codes may pick up issues with comprehension in Asian countries, they do not tell us whether the meaning and gravity of behavioral codes is comparable across countries. Cognitive interviews may also produce differential reactions across cultural settings, especially as participants commonly view this type of information gathering as questioning the veracity of their original responses (Willis, 2005). As has been recommended for cognitive interviewing more generally, we introduced our study by telling participants that we were investigating the quality of the questions, and we were not interested in whether they provided us with the "right" answer or not. Our field experience across countries suggests that this proviso worked. Nevertheless, further investigation of the cross-cultural validity of behavioral coding and cognitive interviewing is warranted.

Our project experience underscored the importance of collaboration from multi-lingual partners in all countries of interest. Across all phases of the project, input from these partners was critical to adequate protocol development, application, and interpretation. Furthermore, we emphasize the importance of the translator's guide, for translation of both the survey and the pretesting protocol. Ideally, this work will be done by translators who are truly "bilingual", recognizing that the term "bilingual" is very much abused (Samuelsson-Brown, 1998) and does not necessarily include the ability to translate. This requires additional skills in order to transfer concepts between languages, having an equal or complete functional competence in both of them, as well as having an equal understanding of both cultures. There are many challenges when trying to standardize the translation process across countries and cultures (Harkness et al., 2004; Hines, 1993; Newmark, 1988). No matter how rigorous the translation methods used, the final result should be pretested to ensure that the populations in which the survey will be administered similarly understand the resulting questions.

Table 1. Example translator's note

Survey question to be probed	Comments

PS1a. Tobacco is addictive
 	
   1 Strongly Disagree$_1$
   2 Disagree$_2$
   3 Neither disagree nor agree$_3$
   4 Agree$_4$
   5 Strongly Agree$_5$
   9 Can't Say (Don't read out)$_9$
	In this question and the question that follows (PS2a), we are concerned with the meanings that people associate with the term "addiction" (or the term used to translate this word). "Addictive" substances are generally those that, once consumed, compel an individual to regularly consume it, and, thereby, cause the individual to lose voluntary control over the consumption behavior. The loss of voluntary control appears most salient when consumption continues despite causing problems or being against the consumer's best interests. As such, the technical use of the term is often limited to consumption of a drug that causes psychological and/or physiological dependence. In many places, the term is associated with behaviors that do not involve drugs but that are habit forming, potentially destructive, or against social norms—such as gambling or sexual pursuits. Furthermore, the term often connotes a behavior that provides pleasure. Hence, the probes for this question attempt to address these connotations:
	
	➤ PS1e. *Can you think of anything else* **<u>besides tobacco</u>** *that might be addictive?*
	
	   • to determine if people can think of any other substances that are addictive.
	• PS1f. *What are some other things that are addictive?* ***[OPEN RESPONSE]***
	
	   • to determine the range of substances/behaviors that people associate with addictive.
	
	• PS1h. *What does it mean to say that something is "addictive"?* ***[OPEN RESPONSE]***
	
	   • to determine primary associations with the term "addictive".

Figure 1. Protocol development steps.

| Identify candidate questions | → | Determine likely biases | → | Determine probes to assess biases | → | Develop translator guides |

| Translate protocol | → | Train interviewers | → | Pilot protocol | → | Review pilot data |

| Adjust protocol | → | Implement protocol | → | Enter data | → | Translate open-ended responses into English |

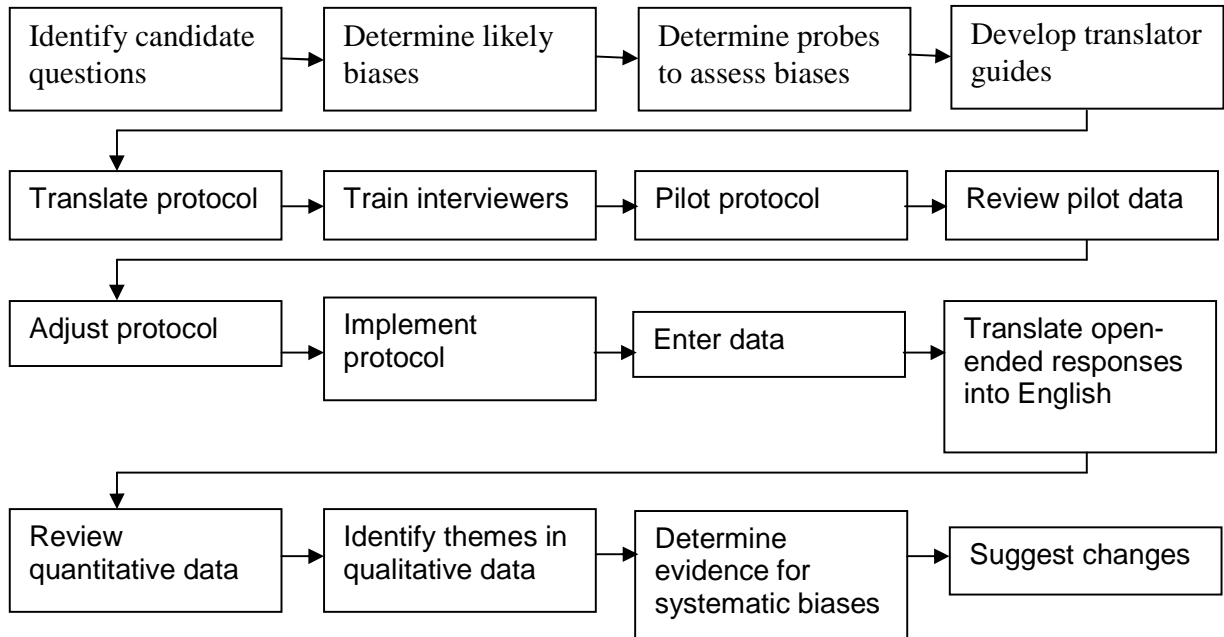| Review quantitative data | → | Identify themes in qualitative data | → | Determine evidence for systematic biases | → | Suggest changes |

Table 2. Sample characteristics (need to integrate Fisher's exact tests)

		USA (n=20)	Australia (n=20)	Uruguay (n=20)	Mexico (n=20)	Malaysia (n=20)	Thailand (n=20)	p-value*
Male		65%	50%	50%	65%	100%	80%	0.005
Age		36	36	40	38	31	39	ttest
Education	**<HS**	5%	30%	30%	20%	10%	85%	
	HS	65%	30%	40%	55%	45%	10%	0.000
	Uni	30%	40%	30%	25%	30%	5%	
Daily smoker		85%	90%	80%	85%	90%	85%	0.948
Average cigs/day		23	15	14	15	13	13	ttest

*categorical variables are assessed with a Fisher's exact test. Continuous variables are assessed with an independent sample t-test.

Table 3. Example of behavioral codes and responses to codes associated with the question "Tobacco is addictive", with Likert response format

| Question type | Code | Country | | | | | | Fisher's exact p-value |
		USA	AUST	URG	MEX	MAL	THAI	
Behavioral codes: *Did the participant…*	*…need you to repeat the question?*	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	5%(1)	>0.1
	…have difficulty with response options?	5%(1)	5%(1)	0% (0)	0% (0)	5%(1)	5%(1)	>0.1
	…ask for clarification or qualify answer?	5%(1)	0% (0)	0% (0)	0% (0)	0% (0)	0% (0)	>0.1
Probe 1: *Can you name some other things that are addictive?*	Alcohol	70%(14)	70%(14)	70%(14)	50%(10)	15%(3)	15%(3)	<.001
	Other drinks	5%(1)	5%(1)	25%(5)	25%(5)	10%(2)	0%(0)	0.046
	Illegal drugs	80%(16)	70%(14)	70%(14)	55%(11)	60%(12)	80%(16)	>0.1
	Legal drugs	35%(7)	35%(7)	0% (0)	10%(2)	25%(5)	0% (0)	<.001
	Inhalants	0% (0)	0% (0)	0% (0)	5%(1)	20%(4)	15%(3)	0.027
	Food	40%(8)	20%(4)	25%(5)	5%(1)	0% (0)	0% (0)	<.001
	Sex	15%(3)	10%(2)	0% (0)	20%(4)	5%(1)	0% (0)	0.085
	Gambling	5%(1)	15%(3)	10%(2)	5%(1)	0% (0)	0% (0)	>0.1
	Other	35%(7)	15%(3)	15%(3)	0% (0)	20%(4)	5%(1)	0.007
	Don't know	5%(1)	0% (0)	5%(1)	25%(5)	15%(3)	0% (0)	0.027
Probe 2: *What does it mean to say that something is addictive*	Control-General	50%(10)	50%(10)	55%(11)	55%(11)	45%(9)	30%(6)	>0.1
	Physiological	40%(8)	25%(5)	5%(1)	15%(3)	0% (0)	0% (0)	0.001
	Control-Psych	10%(2)	10%(2)	0% (0)	0% (0)	30%(6)	0% (0)	0.005
	Control-Quit	0% (0)	10%(2)	5%(1)	10%(2)	10%(2)	15%(3)	>0.1
	Freq-Quantity	5%(1)	50%(10)	40%(8)	15%(3)	20%(4)	30%(6)	0.015
	Danger	5%(1)	0% (0)	0% (0)	0% (0)	0% (0)	35%(7)	<0.001
	Pleasure	5%(1)	0% (0)	5%(1)	0% (0)	5%(1)	5%(1)	>0.1

References:

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons.

Bollen, K. A., Entwisle, B., & Alderson, A. S. (1993). Macrocomparative research methods. *Annual Review of Sociology, 19*, 321-351.

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1*, 186-216.

Cannell CF, L. S., Hausser, DL. (1975). *A Technique for Evaluating Interviewer Performance*. Ann Arbor, MI: University of Michigan.

Fiske, A. P., Kitayama, S., Markus, H. R., & Nisbett, R. E. (1998). The cultural matrix of social psychology. In D. Gilbert, S. Fiske & G. Lindzey (Eds.), *Handbook of Social Psychology* (Vol. 2, pp. 915-981). Boston: McGraw-Hill.

Fonatine, J. (2003). Multidimensional scaling. In J. A. Harkness, F. J. R. Van de Vijver & P. P. Mohler (Eds.), *Cross-cultural Survey Methods*. Hoboken, NJ: Wiley & Sons, Inc.

Fong, G. T., Cummings, K. M., Borland, R., Hastings, G., Hyland, A., Giovino, G. A., et al. (2006). The conceptual framework of the International Tobacco Control Policy Evaluation Project. *Tobacco Control, 15*(Supp 3), iii3-iii11.

Forsyth, B., Kudela, M. S., Levin, K., & Lawrence, D. (2007). Methods for translating an English-language survey questionnaire on tobacco use into Mandarin, Cantonese, Korean, and Vietnamese. *Field Methods, 19*(3), 264-283.

Fowler, F. (1995). *Improving Survey Questions: Design and Evaluation*. Thousand Oaks, CA: Sage.

Harkness, J. A., Pennell, B.-A., & Schoua-Glusberg, A. (2004). Survey questionnaire translation and assessment. In S. Pressler, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin & E. Singer (Eds.), *Methods for testing and evaluating survey questionnaires* (pp. 453-473). Hoboken, NJ: Wiley & Sons, Inc.

Harkness, J. A., & Schoua-Glusberg, A. (1998). Questionnaires in translation. In J. A. Harkness (Ed.), *Cross-cultural Survey Equivalence* (pp. 87-126). Mannheim, Germany: ZUMA.

Harkness, J. A., Van de Vijver, F. J. R., & Johnson, T. P. (2003). Questionnaire design in comparative research. In J. A. Harkness, F. J. R. Van de Vijver & P. P. Mohler (Eds.), *Cross-cultural survey methods* (pp. 19-34). Hoboken, NH: Wiley & Sons, Inc.

Harkness, J. A., Van de Vijver, F. J. R., & Mohler, P. P. (Eds.). (2003). *Cross-cultural survey methods*. Hoboken, NJ: Wiley.

Hines, A. M. (1993). Linking qualitative and quantitative methods in cross-cultural survey research: Teachniques from cognitive science. *American Journal of Community Psychology, 21*(6), 729-746.

Johnson, T. P., Cho, Y., Holbrook, A., O' Rourke, D., Warnecke, R., & Chávez, N. (2006). Cultural variability in the comprehension of health survey questions. *Annals of Epidemiology, 16*(9), 661-668.

Johnson, T. P., & Van de Vijver, F. J. R. (2004). Social desireability in cross-cultural research. In J. A. Harkness, F. J. R. Van de Vijver & P. P. Mohler (Eds.), *Cross-cultural Survey Methods* (pp. 195-204). Hoboken, NJ: Wiley & Sons, Inc.

Knowles, E., & Condon, C. (1999). Why people say "yes": A dual-process theory of acquiescence. *Journal of Personality and Social Psychology, 77*, 379-386.

Kreuger, R. A., & Casey, M. A. (2000). *Focus groups: A practical guide for applied research* (3rd ed.). Thousand Oaks, CA: Sage.

Marlow, D., & Crowne, D. P. (1960). *The approval motive*. New York: Wiley.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis* (Vol. 2). London: Sage.

Newmark, P. (1988). *A textbook of translation*. London: Prentice-Hall.

Samuelsson-Brown, J. (1998). *A practical guide for translators.* (3rd ed.). Clevedon, England: Multilingual Matters Ltd.

Saris, W. (2003a). Multitrait - multimethod studies. In J. A. Harkness, F. J. R. Van de Vijver & P. P. Mohler (Eds.), *Cross-Cultural Survey Methods* (pp. 265-274). Hoboken, NJ: Wiley & Sons, Inc.

Saris, W. (2003b). Response function equality. In J. A. Harkness, F. J. R. Van de Vijver & P. P. Mohler (Eds.), *Cross-Cultural Survey Research* (pp. 275-289). Hoboken, NJ: Wiley & Sons, Inc.

Smith, P. (2004). Acquiescent response bias as an aspect of cultural communication style. *Journal of Cross-Cultural Psychology, 35*, 50-61.

Smith, T. W. (2004). Developing and evaluating cross-national survey instruments. In S. Pressler, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin & E. Singer (Eds.), *Methods for Testing and Evaluating Survey Questionnaires* (pp. 431-452). Hoboken, NJ: Wiley & Sons, Inc.

Thompson, M. E., Fong, G. T., Hammond, D., Boudreau, C., Dreizen, P. R., Hyland, A., et al. (2006). The methodology of the International Tobacco Control Policy Evaluation Project. *Tobacco Control, 15*(Supp 3), iii12-iii18.

Thrasher, J. F., Chaloupka, F., Hammond, D., Fong, G. T., Borland, R., Hastings, G., et al. (2006). Evaluación de las políticas contra el tabaquismo en países latinoamericanos en la era del Convenio Marco para el Control del Tabaco. *Salud Pública de México, 48*(Supp 1), S155-S166.

Van der Zouwen, J., & Smit, J. H. (2004). Evaluating survey questions by analyzing patterns of behavior codes and question-answer sequences: A diagnostic approach. In S. Pressler, J. M. Rothgeb, M. P. Couper, J. T. Lessler, E. Martin, J. Martin & E. Singer (Eds.), *Methods for Testing and Evaluating Survey Questionnaires* (pp. 109-130). Hoboken, NJ: Wiley & Sons, Inc.

West, R. (2004). Assessment of dependence and motivation to stop smoking. *British Medical Journal, 328*, 338-339.

Willis, G. B. (2005). *Cognitive interviewing*. London: Sage.