

Same Questionnaire, Multiple Evaluation Methods—Do the Results Line Up?

Paul Beatty

National Center for Health Statistics, USA

1. Introduction

Numerous methods are available to survey researchers to evaluate questionnaires. Cognitive interviewing, one of the most common, relies on self-reports about thought processes, interpretations, and experiences to evaluate how well the questionnaire is performing. Through vignettes, researchers use respondent reactions to carefully constructed hypothetical scenarios to understand nuances of their thought processes. Response latency measures indicate how long respondents take to answer questions; these measures can be used as an indicator of the complexity of the response process. All of these methods have some theoretical roots in cognitive sciences, yet they vary considerably in terms of assumptions they are based upon, the data they produce, and the manner in which data may be used to reach conclusions. But when applied to the same instrument, would they point to the same overall conclusions about its strengths and weaknesses? This paper will explore the rationales behind the various methods, discuss what was discovered in a recent application using each, and explore what each did and did not accomplish. It will also consider the extent to which findings overlapped—and when they didn't, attempt to explain why.

Recently, all three methods were applied to a questionnaire tested at the National Center for Health Statistics (NCHS). This questionnaire, the Uniform Blood Donor History Questionnaire, is a screening instrument used to determine the eligibility of blood donors in the United States. While it is not a survey, its questions are similar to those asked on many health surveys. The instrument includes 47 questions on topics such as current health and medication usage, recent medical events that could affect blood safety (transfusions, grafts, transplants, etc.), risk behaviors, travel outside the United States, and the presence of diseases such as HIV, malaria, and others that could be transmitted through the blood. The questions are all “yes/no” but vary considerably in terms of memory challenges, with some asking about recent single incidents and others asking about cumulative behavior over many years.

Prior to any empirical evaluation efforts, the questionnaire was reviewed by methodological and subject matter experts. This analysis revealed that some questions were needlessly complicated, that others were included based on obsolete regulations, and that the instrument on the whole was confusingly organized. A number of questions were simplified or dropped as a result of these technical reviews, and the questions were organized based on reference period (with questions with recent reference periods asked before questions with longer ones). The revised instrument (which appears in Appendix 1) was then tested in three rounds of cognitive interviews.

2. Cognitive interviewing

Each round of cognitive interviews was based upon approximately 12 participants (total n=35), each of whom shared a certain key characteristic of interest. Participants in the first round had never donated blood before but were eligible to do so as far as they knew. These were “naïve users” who brought minimal experience to their interpretations of the questions. Participants in

the second round had actually been deferred from an attempted blood donation in the past. The rationale for including this group was to obtain greater response variety. Generally, people answer “no” to most or all of the questions on the instrument, but clearly people who had been deferred would have a greater chance of answering “yes” to one or more items. Participants in the third group were younger and less educated, chosen under the assumption that they would be the most likely to misunderstand the questions.

Since the questionnaire is self-administered in practice, we asked participants in the laboratory to fill it out completely on their own prior to any discussions about the meaning of question or their response processes. When they finished this task, NCHS staff members conducted debriefings with in-depth probes. The probes were designed not only to explore participant interpretations of questions, but also to explore their experiences in detail in an effort to identify any errors (especially false negatives or under-reports of risk factors).

The cognitive interviews uncovered several reporting errors that seemed attributable to characteristics of the questions. For example, consider the question: “In the past 4 weeks, have you had any shots or vaccinations?” Although this seemed straightforward, we identified several situations in which participants initially forgot to report shots (in one instance, a shot that was received only the day before). Based on responses to probes, it became clear that vaccinations dominated participant recall, possibly at the expense of other shots. We therefore recommended revising the question to read: “in the past 4 weeks have you had any vaccination, or any other kind of shot.” The revision puts more emphasis on non-vaccinations and may facilitate recall or a greater variety of shots.

Cognitive interviews also revealed several questionnaire design features that were seen as illogical and confusing to participants. For example, one question asked “in the past 12 months, have you had an accidental needle-stick or come into contact with someone else’s blood?” To at least one participant, the implication that stepping on a used hypodermic needle was as significant as touching a child’s skinned knee was difficult to understand. Similarly, another question asked “in the past 12 months, have you had an ear or skin piercing, including acupuncture.” The potential problem with such juxtapositions is that respondents might second-guess the intention of the questions—e.g., deciding that the question must not “really” be about reputable acupuncture, or that only “dangerous” contact with other blood should warrant a “yes” response. Such assumptions would be incorrect. While clustering these concepts into single questions seemed economical to questionnaire designers, doing so has a potential downside in terms of respondent interpretations. We therefore recommended asking separate questions about piercings and acupuncture, and separate questions about any contact with blood and needle-sticks.

More generally, interviews yielded information about a number of potentially ambiguous concepts. For example, a number of questions asked whether respondents had “had sex” with people who met various criteria. Such questions were intended to capture multiple forms of sexual contact, but were sometimes interpreted as referring only to sexual intercourse. Another example was a question that asked whether an individual had “come into contact” with the saliva of someone with hepatitis; some people were uncertain as to whether this had happened and what sorts of contact were to be included.

We were particularly interested in using in-depth probing to identify whether any initial responses to the questions appeared to have been incorrect. Such probing identified false positive responses much more commonly than false negatives. Some false positives appeared to reflect forward

telescoping (i.e., respondents including events within the reference period of the question, when they actually happened longer ago—see Neter and Waksburg, 1964). Others reflected overly broad interpretations of terms (i.e., including ibuprofen in reports of aspirin usage). Both forms of false positives might reflect participant tendencies to err on the side of caution—in other words, when in doubt as to the most appropriate answer, they might tend to answer yes. False negatives are much more serious from a public health perspective, as they reflect a risk factor not captured, but were also rather rare in this study. Some false negatives reflected either a forgotten incident, such as the shot mentioned earlier. Others reflected conceptual vagueness: when asked whether they had any “problems with your heart and/or lungs,” several participants who had asthma answered no, thinking that this should not apply. Another participant answered a question about travel outside the U.S. negatively, but probing revealed that he had been on a cruise with stops in Mexico and the Caribbean. While false negatives appeared to be isolated incidents in the testing, it was often possible to recommend modifications to questions that would improve their clarity (e.g., modification of the question about shots and vaccinations, as described above). For ambiguities expected to be very infrequent, we sometimes recommended that supplemental material should be available when requested, leaving the actual question wording unchanged.

3. Extensions of the questionnaire evaluation

Cognitive interviews were the dominant method used in the evaluation of this questionnaire. But as this phase of the work came to a close, some limitations of the work became clear. One problem was that sample limitations did not allow us to explore a variety of uncommon circumstances. Ideally, we would have interviewed at least some people who had all of the conditions or experiences mentioned on the questionnaire. While it was easy to find participants who experienced relatively common events (travel outside the U.S., receiving tattoos, having hepatitis), we had fewer participants who had engaged in certain high-risk behaviors, and none who had some of the more exotic experiences (receiving a brain graft, or having had diseases such as babesiosis). In addition to interviewing some participants who had experienced each of these events, we would ideally interview people in ambiguous situations—that is, where it is not immediately apparent whether their situation warrants a “yes” response. These are situations where more complex judgment processes are taking place, and presumably the potential for error is higher.

Another problem was that the cognitive interviewing study allowed for only very crude assessments of the magnitude of problems with a question. We made note of the number of false positive and false negative responses, but these numbers were too small for distinctions across questions to be meaningful—and furthermore, it is not clear whether these counts represent the real extent of underlying problems, or merely varying interviewer ability to identify them. This is not to denigrate the usefulness of qualitative findings from cognitive interviewing. Certainly in this case, they provided valuable insight into the potential problems of various wordings. However, regulatory groups charged with oversight of this questionnaire were interested in any quantitative measures of question performance that could be provided.

4. Vignettes

A brief supplemental study using vignettes attempted to compensate for sample limitations through administration of hypothetical scenarios to research participants. The basic rationale for vignettes is that while hypothetical reactions are imperfect, data about these situations might not be attainable in any other way (see Martin, 2004). Although vignettes do not fully recreate the

thought processes involved in autobiographical recall, they did require participants to apply definitions rather than simply interpreting terms outside of any particular context. The scenarios in this study covered situations that we did not see in the actual interviews, and that were designed to be deliberately ambiguous. For example, one vignette read: “Kim has a boyfriend who has used a needle to inject illegal drugs at least once. They have not had sexual intercourse although they have had oral sex together.” After participants read the vignette, we asked them to evaluate how “Kim” should answer the questionnaire item: “In the past 12 months, have you had sex with anyone who has ever used needles to take drugs or anything else not prescribed by their doctor?” This vignette was designed to explore participant conceptualizations of the meaning of “have sex”—specifically, whether having oral sex should be deemed sufficient to respond positively. In the cognitive interviews, we did not encounter any participants who reported having been in such a situation, but responses to probing suggested that the term “have sex” could be interpreted in different ways.

Other vignettes addressed situations related to feeling well, having shots, coming into contact with someone’s blood, and sexual contact (see Appendix 2). We believed that a reasonable case could be made for answering yes or no to the questions related to each of them. For example, the vignette on “feeling well today” was based on very slight cold symptoms and balanced out with high energy levels and good spirits.

The vignettes were administered to the 11 participants from our third round of cognitive interviewing. Participants responded to the vignettes after completing the self-administered questionnaire, but prior to the debriefing or any probing.

In spite of our efforts to construct ambiguous vignettes, most participants thought that the hypothetical respondent should answer yes to each of the associated questions. The most ambiguous scenario involved Vignette 3, about coming into contact with another person’s blood. In that vignette, the subject clearly did touch another person’s blood, but the exposure was minimal and the subject took immediate safety precautions. Eight participants thought that such exposure warranted a yes response to the question, but three did not. For most other vignettes, nine of eleven participants thought the subject should answer yes. The vignette addressing oral sex described at the beginning of this section (Vignette 4), turned out to be the least ambiguous of all, with all eleven participants expressing the belief that the yes response was appropriate.

Justifications for yes responses tended to focus on pragmatic assessment of risks involved. For example, while “Kim’s” behavior in Vignette 4 might or might not qualify as sex in other contexts, her behavior did entail elevated risk of exposure; on that basis participants thought she should answer yes in the context of the screening questionnaire.

It is difficult to say for certain how well these results line up with cognitive interview findings, as the vignettes addressed situations we did not observe in cognitive interviews. Taking the vignette results at face value would suggest that respondents will be likely to err on the side of caution in reporting about ambiguous situations (something we observed from time to time with self-report responses as well). However, there are several reasons to be cautious with such interpretations. One obvious reason is that answering based on vignettes does not draw upon autobiographical memory in the same way that self-reporting does. The decision processes involved in responding could be quite different. Another concern is that even if respondents do include situations beyond “literal truth” in their answers, in order to include activities that they perceive as risky, respondents might make incorrect judgments regarding which behaviors are sufficiently risky to

count. In general, however, the findings suggest that respondents take into account the perceived purpose of the question in determining what to report. In addition, the use of vignettes to fill in gaps of observation has advantages over other approaches, such as asking participants their opinion about the effectiveness of particular questions. Participants are not questionnaire design experts. It seems more logical to spent research effort focusing on judgments about scenarios that participants could reasonably envision rather than to asking their opinions about measurement quality.

5. Time spent responding to questions as an evaluative measure

The availability of video recordings made it possible for us to perform a new analysis based on time spent responding to each of the questions. Several studies have used a related measure of response latency (the time between administration of the question and response) as a survey diagnostic tool. For example, a study by Bassili (1996) suggested that longer response latencies are associated with both structurally complex questions and unstable attitudes. Also, Draisma and Dijkstra (2004) found that respondents answered factual questions more quickly when their responses were correct (as determined by externally verifiable information); as response latency increased, so did the probability of giving an incorrect answer. The measures in this study are slightly different, as they include time spent reading and responding to a question, but the data nevertheless provide the opportunity to evaluate whether response times correspond with the qualitative results already considered. That is, are the questions with the longest response times the same ones we flagged before as having notable problems?

Response time measures for this study were recorded manually using a stopwatch from video recordings of participants completing the questionnaire. A stopwatch measure was taken each time a participant marked a response. For example, the time spent responding to Q2 was computed as the time between marking a response to Q1 and marking a response to Q2. This measure included the time spent reading, thinking about, and responding to the question. Unfortunately, we could not reliably gauge when participants began answering the first question on the instrument, nor the first question on subsequent pages of the three-page questionnaire. Consequently, response times for some questions were not recorded. (However, re-ordering of the questionnaire at several points in the study meant that different questions appeared at the top of pages 2 and 3 on different versions of the instrument, resulting in at least some usable data for all questions other than Q1). We also excluded response times that included interruptions, such as participant queries about the meaning of questions. Fortunately, such interruptions were very infrequent. Only one question was interrupted by three participants making queries about question meaning or intent, and three more were interrupted twice. Most were never interrupted.

The first step in our analysis was to examine which questions had the shortest and longest response median times (medians being used because extreme outliers sometimes had a dramatic effect upon the means). This information appears in Table 1.

Table 1: Questions with shortest and longest median response times

		Median (Low High)		
<u>Longest</u>				
28.	Since 1980 spent time that adds to 5 years in Europe	5.87	(2.34	31.96)
18.	In past 12 mo had sex with anyone used needle for drugs	5.48	(0.51	29.60)
6.	In past 36 hrs taken aspirin or anything w/aspirin in it	4.95	(2.31	15.75)
22.	In past 12 mo been given hepatitis immune globulin	4.89	(1.29	9.75)
34.	Ever had a positive HIV test	4.50	(2.22	10.21)
<u>Shortest</u>				
39.	Ever had babesiosis	1.89	(0.29	6.92)
38.	Ever had Chagas' disease	1.66	(0.63	3.70)
36.	Ever had hepatitis	1.49	(0.74	8.54)
44.	Ever been to Africa	1.33	(0.43	3.82)
37.	Ever had malaria	1.13	(0.63	9.17)

Note that the shortest median response times are associated with questions that require only basic judgments—specifically, whether certain unusual events have ever happened. The questions can be comprehended almost instantly, and most people can quickly judge that the events in question are not applicable to them. In contrast, the longest median response times are associated with questions that have more complex reference periods, concepts, or judgments. Q28, with the longest median response time, requires that respondents understand a specific reference period (from 1980 to the present), grasp that the response task involves cumulative travel time, and recognize the focus on European travel in particular. Even if the question is not applicable, respondents must invest a certain amount of cognitive effort to understand the premise of the question. If the question has any applicability (i.e., respondents judge that it is possible that the answer is yes), then additional effort will be required for memory and judgments related to responding—remembering times spent abroad, consideration of whether sites were in Europe or not, and so on.

The other questions with long median response times also present some combinations of structural, definitional, or temporal complexities. Only Q34 (regarding HIV tests), a seemingly straightforward question, seems out of place on this list. This might be because the meaning of positive test is not completely intuitive—it could take a few seconds to grasp that a “positive test” indicates an undesirable result. Also, some cognitive interview participants had expressed uncertainty as to whether they had received an HIV test as part of a battery of other blood tests. This uncertainty, along with the possibly counter-intuitive meaning of “positive,” might have made the question more difficult than it initially appears.

While median response time might serve as a reasonable measure of how much cognitive effort is required for a typical respondent to answer, it is not necessarily a useful indicator of which questions had the most serious cognitive or conceptual problems. As an illustration, consider the question about “contact with the blood of another person.” For the majority of participants, the nuances of the question’s meaning are irrelevant—they can respond quickly because they can easily recognize that the question does not apply. However, the relatively short median response time might provide no clue that a subset of participants spend considerable time struggling with meaning. And of course, response times provide no evidence at all about whether participants understood the meaning of the question correctly.

Thus, questions flagged as having noteworthy problems based on cognitive interviewing might not have particularly long response times in general. Unfortunately, exploring the relationship between cognitive interview problems and response times is difficult, because there is no clear measure of the severity of problems found in cognitive interviews. One potential criterion for “serious problem” is that a particular question failed to accurately capture a response that should have been “yes” for one or more participants. For example, as noted earlier, one participant initially failed to report a shot received the day before the interview; this error was only discovered through subsequent probing. Another more subjective criterion of a question having a problem could be that the cognitive interviewing team judged problems to be serious enough that a wording change was recommended. A related criterion is that the cognitive interviewing team recommended supplemental definitions or explanations (available to screening personnel, but not uniformly read to donors—the actual wording of the questions would not change) to address some ambiguity or other problem.

To explore whether each of these criteria were reflected in response times, we assigned a binary variable to each question regarding whether or not these criteria applied. We then compared the means of the response times for questions with and without each criterion. Results appear in Table 2, below. In each case, response times were longer when there was some indication of a problem, but in most cases the differences were slight. However, response times were significantly longer for questions when there was a recommendation for a wording change. This makes sense if recommended wording changes are generally driven by problems with comprehension and clarity. The same sorts of problems that cause participants to take longer to answer are those that we attempt to address through wording changes.

We also observed that the amount of text devoted to each survey question in our report varied considerably—while a great deal was written about some questions, there was relatively little report material for some others. To explore whether the amount of material written could serve as a crude measure of “cognitive problems” with questions, we explored whether the lines of text written about a question correlated with its medial response time. The Spearman correlation between the two variables was 0.44 ($p < .01$).

Table 2: Comparisons of response times of questions with and without problem criteria

<i>Criterion of interest</i>	<i>Number of questions</i>	<i>Mean response time</i>
errors identified	(n=8)	3.36
no errors identified	(n=37)	3.05
		t= -0.74, not significant
recommended wording changes	(n=20)	3.32
no recommended wording changes	(n=25)	2.55
		t= -4.67, $p < .01$
recommended supplemental info	(n=25)	3.13
no recommended supplemental info	(n=20)	3.07
		t= -0.20, not significant

This suggests that the issues raised in the cognitive interviewing report might have been key factors in participants’ effort level in working through their responses.

In general, median response times for questions may be useful for assessing comprehension and judgment burdens that apply to most participants, but some serious problems are only apparent among a few participants. Most of the events covered in this questionnaire were relatively rare—even the questions about relatively common events (e.g., about receiving recent shots) only applied to a handful of participants. However, the circumstances were sometimes complicated for a subset of respondents. For example, most participants had never had anything that could be characterized as a “problem with your heart and/or lungs” (Q42), but a few wondered if conditions such as asthma or bronchitis qualified. Their struggle with the intent of the question did not affect median response times—however, their individual response times were much higher than the median for that question. Response time outliers are potentially useful if they point to situations worthy of additional attention during the analysis phase. Also, if computer-assisted technology were used to administer questions, and researchers had baseline data about typical response times, the information could be used during live interviews. Atypically long response times could serve as signal for interviewers to follow up with additional probing.

6. Conclusion

As is typical with cognitive interviewing, this project involved a very small sample. It also involved a rather unusual questionnaire that included only yes/no questions about topics that generated very few “yes” responses. Yet without extending the analysis too far, the study did suggest a few things about the complementary value of various questionnaire evaluation methods.

Cognitive interviewing generated an incredible amount of qualitative data about how participants made sense of questions and the various issues they sorted through in trying to answer. Extensive follow-up probing identified several instances in which questions failed to capture important information, and also helped to develop reasonable explanations for these failures. The interviews illuminated a number of question characteristics that seemed unnecessarily complicated or confusing. Arguably, the revision of the questionnaire based on cognitive interviewing eliminated some weaknesses of the original.

At the same time, we have no way of knowing how comprehensive or complete the cognitive interview findings were. Participants had not experienced many characteristics of interest in the questions, some of which were rare phenomena. Furthermore, cognitive interviewing provides no means for objectively quantifying the extent or severity of particular questionnaire problems. Vignettes may help to provide some data on unusual circumstances not likely to be found among a small group of cognitive interviewing participants. Due to the hypothetical nature of the response task, results need to be interpreted with caution, but the data can provide some insights into how participants think about a wider variety of circumstances than those that are likely to be captured in cognitive interviews. Debriefings following the administration of vignettes can help to illuminate how participants made the decisions that they did.

One appeal of recording response times is that the data are quantitative. The data also vary considerably across questions, potentially allowing us to make significant distinctions regarding the level of effort required to answer. However, it is not completely clear what drives the differences in response times, and particularly whether these times reflect question quality in a useful way. It does appear likely that response time is a reasonable measure of cognitive effort spent answering. Questions with longest response times are not necessarily the most problematic items, but might be worthy of additional scrutiny to determine whether simplifications are possible. Individual outliers of response times may also be informative.

Questionnaire testers and evaluators have often described alternative testing methods as complementary, and that seems to be the case with the methods studied here. Rather than contradicting each other, the results combined to provide a multi-faceted picture regarding which issues and attributes of the questionnaire warrant closest consideration for additional developmental work. Cognitive interviewing remains an excellent choice for identifying conceptual flaws with questions. Vignettes might shed additional light on judgments in ambiguous circumstances, especially those that are not likely to emerge in a small sample of cognitive interview participants. Response latency can provide a measure of the relative effort spent reading and responding to questions, which can flesh out the big picture even further.

References

Bassili, J. (1996). "The How and Why of Response Latency Measurement in Telephone Surveys." In N. Schwarz and S. Sudman, eds., *Answering Questions*. San Francisco: Jossey-Bass.

Draisma, S. and Dijkstra, W. (2004). "Response Latency and (Para)Linguistic Expressions as Indicators of Response Error." In S. Presser, J.M. Rothgeb, M.P. Couper, J.T. Lessler, E. Martin, J. Martin, and E. Singer (eds.), *Methods for Testing and Evaluating Survey Questionnaires*. Hoboken, NJ: John Wiley and Sons.

Martin, E. (2004). "Vignettes and Respondent Debriefing for Questionnaire Design and Evaluation." In S. Presser, J.M. Rothgeb, M.P. Couper, J.T. Lessler, E. Martin, J. Martin, and E. Singer (eds.), *Methods for Testing and Evaluating Survey Questionnaires*. Hoboken, NJ: John Wiley and Sons.

Neter, J. and Waksberg, J. (1964). "A Study of Response Errors in Expenditures Data From Household Interviews." *Journal of the American Statistical Association*, 59, 17-55.

Appendix 1. Donor History Questionnaire (initial version)

	Yes	No
Are you		
1. Feeling well today?	<input type="checkbox"/>	<input type="checkbox"/>
2. Taking any medication on the medication list?	<input type="checkbox"/>	<input type="checkbox"/>
3. Currently taking an antibiotic?	<input type="checkbox"/>	<input type="checkbox"/>
4. Currently taking any other medication for an infection?	<input type="checkbox"/>	<input type="checkbox"/>
5. Female donors: Are you pregnant? Not female <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
In the past 36 hours have you		
6. Taken aspirin or anything that has aspirin in it?	<input type="checkbox"/>	<input type="checkbox"/>
In the past 4 weeks have you		
7. Had any shots or vaccinations?	<input type="checkbox"/>	<input type="checkbox"/>
In the past 8 weeks have you		
8. Given blood, platelets or plasma?	<input type="checkbox"/>	<input type="checkbox"/>
In the past 16 weeks have you		
9. Given double red cells?	<input type="checkbox"/>	<input type="checkbox"/>
In the past 12 months have you		
10. Had a blood transfusion?	<input type="checkbox"/>	<input type="checkbox"/>
11. Received clotting factor concentrates?	<input type="checkbox"/>	<input type="checkbox"/>
12. Had a transplant such as organ, tissue or bone marrow?	<input type="checkbox"/>	<input type="checkbox"/>
13. Had a graft such as bone or skin?	<input type="checkbox"/>	<input type="checkbox"/>
14. Had an accidental needle-stick or come into contact with someone else's blood?	<input type="checkbox"/>	<input type="checkbox"/>
15. Had sex with anyone who has HIV/AIDS or has had a positive test for the HIV/AIDS virus?	<input type="checkbox"/>	<input type="checkbox"/>
16. Had sex with a prostitute or anyone else who takes money or drugs or other payment for sex?	<input type="checkbox"/>	<input type="checkbox"/>

In the past 12 months have you		
17. Had sex with anyone who has ever used needles to take drugs or steroids, or anything else not prescribed by their doctor?	<input type="checkbox"/>	<input type="checkbox"/>
18. Had sex with anyone who has hemophilia or has used clotting factor concentrates?	<input type="checkbox"/>	<input type="checkbox"/>
19. Female donors: had sex with a male who has ever had sex with another male? Not female <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
20. Come into contact with blood or saliva from a person who has hepatitis?	<input type="checkbox"/>	<input type="checkbox"/>
21. Had sexual contact with a person who has hepatitis?	<input type="checkbox"/>	<input type="checkbox"/>
22. Been given hepatitis immune globulin (HBIG)? (<i>Note: this is not the same as hepatitis B vaccine.</i>)	<input type="checkbox"/>	<input type="checkbox"/>
23. Had a tattoo applied?	<input type="checkbox"/>	<input type="checkbox"/>
24. Had an ear or skin piercing (including acupuncture)?	<input type="checkbox"/>	<input type="checkbox"/>
25. Had or been treated for syphilis or gonorrhea?	<input type="checkbox"/>	<input type="checkbox"/>
26. Been in juvenile hall, lockup, jail, or prison?	<input type="checkbox"/>	<input type="checkbox"/>
In the past 3 years have you		
27. Been outside the United States or Canada?	<input type="checkbox"/>	<input type="checkbox"/>
Between 1980 and the present did you		
28. Spend time that adds to five (5) years in Europe?	<input type="checkbox"/>	<input type="checkbox"/>
29. Receive a blood transfusion in the United Kingdom? (England, Northern Ireland, Scotland, Wales, the Isle of Man, or the Channel Islands.)	<input type="checkbox"/>	<input type="checkbox"/>
Between 1980 through 1996 did you		
30. Spend time that adds up to three (3) months or more in the United Kingdom?	<input type="checkbox"/>	<input type="checkbox"/>
31. Spend time that adds up to six (6) months or more on a military base in Europe?	<input type="checkbox"/>	<input type="checkbox"/>
Since 1977 have you		
32. Received money, drugs, or other payment for sex?	<input type="checkbox"/>	<input type="checkbox"/>
33. Male donors: had sex with another male? Not male <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Have you EVER		
34. Had a positive test for the HIV/AIDS virus?	<input type="checkbox"/>	<input type="checkbox"/>
35. Used needles to take drugs, steroids, or anything else not prescribed by your doctor?	<input type="checkbox"/>	<input type="checkbox"/>

36.	Had hepatitis?	<input type="checkbox"/>	<input type="checkbox"/>
37.	Had malaria?	<input type="checkbox"/>	<input type="checkbox"/>
38.	Had Chagas' disease?	<input type="checkbox"/>	<input type="checkbox"/>
39.	Had babesiosis?	<input type="checkbox"/>	<input type="checkbox"/>
40.	Received a dura mater (or brain covering) graft?	<input type="checkbox"/>	<input type="checkbox"/>
41.	Had any type of cancer, including leukemia?	<input type="checkbox"/>	<input type="checkbox"/>
42.	Had any problem with your heart and/or lungs?	<input type="checkbox"/>	<input type="checkbox"/>
43.	Had a bleeding condition or a blood disease?	<input type="checkbox"/>	<input type="checkbox"/>
44.	Been in Africa?	<input type="checkbox"/>	<input type="checkbox"/>
45.	If yes - While you were there, did you receive a blood transfusion or any other exposure to blood? Not been in Africa <input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
46.	Had sex with anyone who was born in or lived in Africa?	<input type="checkbox"/>	<input type="checkbox"/>
47.	Have any of your relatives ever had Creutzfeldt-Jacob disease?	<input type="checkbox"/>	<input type="checkbox"/>

Appendix 2: Vignettes

Vignette 1

Question: Are you feeling well today?

Jeff has developed a sore throat and is slightly congested– he thinks a cold may be coming on. But he does not have a fever, feels pretty energetic, and is in good spirits.

Based on your understanding of this question and the type of information it is trying to get at, what do you think Jeff's answer should be?

Vignette 2

Question: In the past 4 weeks, have you had any shots or vaccinations?

Maria has not had any vaccinations for several years. She did receive an injection of vitamin B12 about 2 weeks ago. She receives these every month at the recommendation of her doctor.

Based on your understanding of this question and the type of information it is trying to get at, what do you think Maria's answer should be?

Vignette 3

Question: In the past 12 months, have you had an accidental needle-stick or come into contact with someone else's blood?

George has never had an accidental needle stick– but six months ago, his next door neighbor accidentally broke a window and cut his arm. George helped his neighbor bandage the arm, and remembered that a small amount of his neighbor's blood touched his skin at the time. George is sure that the neighbor's blood did not touch any area of open skin, or come into contact with his own blood, and he washed the blood off immediately.

Based on your understanding of this question and the type of information it is trying to get at, what do you think George's answer should be?

Vignette 4

Question: In the past 12 months, have you had sex with anyone who has ever used needles to take drugs or steroids, or anything else not prescribed by their doctor?

Kim has a boyfriend who has used a needle to inject illegal drugs at least once. They have not had sexual intercourse, although they have had oral sex together.

Based on your understanding of this question and the type of information it is trying to get at, what do you think Kim's answer should be?

Vignette 5

Question: In the past 12 months, have you come into contact with blood or saliva from a person who has hepatitis?

Sue knows that one of her good friends has hepatitis. This friend has been over to her house for dinner many times in the past 12 months. At least once, Sue remembers that they drank wine out of the same glass.

Based on your understanding of this question and the type of information it is trying to get at, what do you think Sue's answer should be?

Vignette 6

Question: In the past 12 months, have you had sexual contact with a person who has hepatitis?

Felix has recently been dating a woman who has hepatitis. In the last week they have started deep-kissing (or French kissing), but nothing more than that has happened.

Based on your understanding of this question and the type of information it is trying to get at, what do you think Felix's answer should be?