Date: March 29, 2012

Subject: Report on 2nd Round Pretests (March 2012) for Child Maltreatment Quality of Life

To: CDC and OMB

From: RTI International (Derek Brown, Charles Strohm, and Sarah Arnold)

INTRODUCTION AND METHODS

In response to the Office of Management and Budget's (OMB) request, RTI conducted a series of in-depth "cognitive interviews" to evaluate and improve the quality of the questionnaire for the child maltreatment quality of life study (hereafter, CMQoL). Cognitive interviews are one-on-one interviews that allow researchers to understand the four main steps involved in answering survey questions. The four steps are (1) comprehension of the response task, (2) retrieval of relevant information, and (3) formulating a judgment based on retrieved information, and (4) providing a response. During a cognitive interview, a respondent completes the questionnaire. The interviewer then probes the respondent about various dimensions of the questionnaire such as the clarity of instructions and questions, cognitive burden, question format, and ease of recall. Respondents also volunteer feedback to the interviewer about the questionnaire. Using this information, researchers can understand the cognitive processes and burdens involved in completing the questionnaire. Understanding how respondents answer survey questions allows researchers to identify and resolve problems with the questionnaire.

RTI conducted nine cognitive interviews in March 2012. Respondents were drawn from the target population for the full CMQoL study: adults age 18 and over. Respondents were recruited by a market research firm in Raleigh, NC (L&E Research) and were paid an $80 cash incentive for their participation in the interviews, which averaged one hour. A total of 11 participants were recruited (R1-R11) because R1 and R7 failed to show and were not reachable by L&E or RTI.

Table 1 details the demographic characteristics of respondents. Respondents represented a range of demographic groups, including young adults, those with low educational attainment, and racial/ethnic minorities. In accordance with RTI's Institutional Review Board (IRB), all respondents provided informed consent before participating in the study. Respondents were also asked for their consent to be observed by other RTI researchers and audio recorded. All but one respondent consented to be observed and recorded.

Table 1.Respondent Characteristics

Respondent	Gender	Education	Race/Ethnicity
R2	Female	College graduate	African American
R3	Male	Post-college	White
R4	Female	High school	White
R5	Male	College graduate	White
R6	Female	Post-college	White
R8	Female	High school	White
R9	Male	Some college	African American
R10	Male	College graduate	African American
R11	Female	High school	Alaskan Native

Dr. Charles Strohm, a survey methodologist from RTI's Survey Research Division (SRD), conducted the interviews. One member from RTI's Public Health Economics Program (PHEP), either Dr. Derek Brown, the principal investigator (PI) of the overall CMQoL study, or Ms. Sarah Arnold, task leader and associate economist on the CMQoL study, observed each interview. Interviews were conducted in RTI's Cognitive Laboratory in Research Triangle Park, North Carolina. The interviews were conducted using paper-and-pencil methods on a subset of the questionnaire identified by OMB as areas of focus for pretesting to assess the clarity and understanding by respondents. (The modules which measure child maltreatment history (the Child Trauma Questionnaire (CTQ)) and other adverse experiences as a child were not administered.)

To guide the interviews, Dr. Strohm used a semi-structured interview protocol that contained questions about various dimensions of the questionnaire such as the frequency labels for the health-related quality of life (HRQoL) questions and valuation tasks (e.g., always, often, sometimes, rarely, never), the clarity of instructions for the valuation task, cognitive burden in completing the valuation task, and questionnaire format. Respondents were also encouraged to volunteer feedback about other aspects of the questionnaire. The semi-structured nature of the interview allowed us to systematically collect data about certain aspects of the questionnaire, but also provided enough flexibility to tailor the interview to specific respondents. In addition to probing respondents about the questions, several respondents (detailed below) were shown two versions of the same question. The interviewing team held debriefings after each interview. After all the interviews were completed, notes from the interviewing team were consolidated and the data were analyzed to identify consistent themes across interviews. Dr. Strohm led the analysis and the reporting because he led the cognitive interviews and has formal training and more experience in survey development and testing.

Note: below we refer to "HRQOL questions," which are the section of questions on the first X pages/screens of the survey; these questions are strictly responses about one's current or past health status and quality of life. The "valuation" questions are the somewhat longer middle section of the survey and are a series of comparisons between profiles A and B, where A is described by one set of characteristics and B a different set. These valuation questions are a form of discrete choice experiment (DCE) or conjoint analysis (CA).

RESULTS

1. HRQoL Questions

Respondents did not report difficulties understanding the HRQoL items/domains. (For example, no respondent said something like "What does 'interfering with daily activities' mean?") All respondents demonstrated knowledge of the health domains when they were probed about their choices in the valuation task.

In general, respondents provided retrospective data (since age 18, during ages 12-17, or during ages 5-11) with little difficulty or hesitation. Some respondents took more time to decide on an answer for these time periods compared to the past 30 days. In the "since age 18" questions, R8 said this was "big time span"

and R9 said it seems like "a lifetime." But these respondents were indeed able to provide answers comfortably. Most respondents said they thought of an "average" response—as intended—because their health states varied throughout the time period. No respondent reported having difficulty doing this.

Some slight misunderstanding of the "since age 18" question was observed. Two respondents (R10, R11) reported data for the time period *around age 18* rather than from age 18 to the present. These three respondents appeared not to either read or comprehend the word "since" and simply reported about around age 18 (e.g., R10 said "post-high school years"). Because people's HRQoL generally improved over time, this reporting error biased responses towards worse HRQoL.

Recommendation: In first HRQoL question, revise instruction to "Now please think about your health from age 18 to the present day." In subsequent questions, change "Since age 18" to "From age 18 to the present day."

2. Valuation Task Instructions

Respondents were probed extensively about their comprehension of the valuation task instruction. This section presents the results of this probing. We begin with respondents' overall understanding of the valuation task and then move to more specific areas of the valuation task instructions.

Overall
Out of the first five respondents, three (R2, R3, R4) did not fully understand the valuation task after reading the instructions. Two respondents (R5, R6) understood the task very well without problems. All respondents who did not fully understand the task thought the instructions were telling them to describe their actual health experiences rather than their preferences for health. For example:

- R2 said "it's a little bit confusing," and added that she thought the instructions are asking her to describe herself. R2 said she thought this because she was just asked the HRQoL questions about her actual health. After a short description by the interviewer, however, the respondent understood the task. The respondent suggested using more direct language such as "which would you prefer, not which one would you want."
- R3 indicated "I don't feel like I have a choice I can make here" because he thought he was being asked "which one do I identify with." R3 later noticed the word "prefer" in the instructions and said that he "read that too fast" and did understand that the question was about preferences. The respondent suggested emphasizing the word "prefer" or saying more directly "which would you rather experience?"
- R4 said "it's really confusing." She said that instructions such as "If I had to pick one, which one would it be?" would be clearer. After a short description by the interviewer, however, the respondent understood the task.

In recognition of these problems, we added a separate page between the HRQoL and valuation instructions. This page used clearer language to indicate to respondents that the subsequent survey questions will ask about preferences, not actual experiences. This addition considerably improved

respondents' understanding of the task. With one exception (R11), the other respondents (R8, R9, R10) understood the valuation task well after having read the two instruction pages.

Text and Table
Respondents differed in whether they thought the text or the table describing the profiles was more helpful in explaining the valuation task. Three respondents (R2, R3, R9) thought the text was more helpful, but three respondents (R4, R6, R10) thought the table was more helpful. These results suggest that keeping both the text and table are important in explaining the task to respondents.

R3 suggested adding heavier borders for the columns to show that the table is organized by columns.

Recommendation*:* Use a heavier font for the column borders compared to the row borders.

Response Selection
One respondent (R9) wasn't sure where to select his response. R9 started to circle the individual domains he preferred rather than selecting one profile or the other because the respondent did not see the "which would you prefer" at the bottom.

Recommendation*:* No action. The Knowledge Networks version of the survey looks clearer here. The "Which would you prefer" row and the first row of headings have a slightly darker shading with bold font.

"Profile" Terminology
Some respondents were also probed about their understanding of the term "Profile." All of these respondents (R3, R5, R6, R8, R10) understood this term, and none of these respondents suggested a better word. The term "Profile" did not appear to confuse any of the other respondents.

One respondent had difficulty understanding that the profiles for each question were distinct from the other profiles. R9 thought that Profiles A and B in the first adult comparison were continuations of the Profiles of A and B from the instructions. That is, he thought that new domains were simply being added to the profiles from the instructions.

Recommendation*:* Indicate that these are new profiles (e.g., Compare the following two new health profiles for an adult.")

Repetition of Instructions
Respondents were split about whether to repeat the valuation task instructions before the teenage and child valuation sections, with four respondents (R2, R3, R6, R8) in favor and three (R4, R5, R10) not in favor. We decided to keep the instructions to reinforce the idea of the valuation task.

3. Valuation Task Observations

This section describes results of probing respondents about how they completed the valuation task.

Response Strategies
Respondents used a variety of strategies to complete the valuation task. The most common strategy was to decide whether physical or mental health issues were more harmful, and then chose the Profile with the fewer number of physical or mental health issues. For example, R4 said she can take a lot more physical pain than other people she knows, so she was less concerned about feeling limited, but more concerned about mental health problems. R11 avoided the physical issues: she said it's possible to work on emotional problems, but it's much more difficult to address physical issues. R6 reported having religious beliefs that make mental health problems much more serious than physical health problems.

While some respondents considered multiple domains simultaneously, other respondents searched for a single domain that made one profile less desirable than another. R9 characterized this strategy as searching for the "deal breaker." R10 said "everyone's going to have a bullet that they can or cannot live with." For R6, when there was a difference between the profiles in the risky choices domain, she always chose the domain with fewer risky choices.

While completing the valuation task, respondents noted that some of the domains were related to each other, or were correlated. R6 noticed that illness and feeling limited "go together." R9 observed that physical and mental health both affect each other. But he thought that physical health affects mental health more than vice versa. Therefore, he avoided physical health problems. R2 noted a discrepancy within one profile that had *always angry* and *rarely emotions out of control*. She said that anger is an emotion. And because emotions are rarely out of control, then she was less worried about being always angry. Therefore, she did not consider the *always angry* characteristic to be negative.

Finally, some respondents paid more attention to certain frequency labels rather than others. In particular, respondents generally appeared to weigh *always* and *never* frequencies more heavily than *often*, *sometimes,* or *rarely*. One respondent (R5) said "I don't like the always," and chose the profile with the fewest number of *always*. In doing so, he paid less attention to the specific domains and more attention to identifying the profile with the "lowest average score," which he determined as the one with the fewest number of *always*.

All of these observations are commonly observed when conducting pretests for DCE or conjoint analysis studies. The fact that different individuals evaluate the questions using different styles is to be expected and the observations here do not represent problems. Rather, these are notes about the cognitive processes that individuals use when completing a DCE style question.

Teenager and Child Valuation: Overall Preferences
In general, respondents felt more comfortable about accepting worse health for an adult than for a teenager or child. For example, R3 said he had a high pain tolerance, so he could deal with some pain if necessary, but said that he wouldn't want to make a teenager feel pain. Most respondents took slightly longer to answer the teenage and child valuation tasks (versus the adult task) and felt less comfortable assigning poor health to these age groups—which is manifested as willing to trade off more time (below)

to avoid such states. Although based on a small sample, this observation is consistent with other published findings in health economics research and also supports the survey design in accounting for different preferences for different ages.

All respondents reported having different HRQoL preferences for an adult, teenager, and child. For example, R10 said "some things that affect teenagers … don't affect kids." In particular, respondents emphasized certain HRQoL domains that were particularly serious for teenagers and children relative to adults. Three respondents (R3, R5, R9) said that adults can accept more pain than teenagers can. One respondent (R8) said that some mental health problems are normal for teenagers and therefore was not concerned about teenagers having out of control emotions. But another respondent (R2) thought that mental health problems were more serious for teenagers and children compared to adults. In short, respondents had strong preferences about what HRQoL domains were worse for adults, teenagers, and children. But a clear pattern of which domains they preferred for the different age groups did not emerge from the small sample in these pretests. This is a common finding in DCE research, and the mixed logit (random parameters) econometric model that will be used to estimate the valuation data estimate the degree of "preference heterogeneity" within the sample.

Teenager and Child Valuation: Time Reference
When completing the teenage and child valuation tasks, respondents generally made their choices using their knowledge and experience of an adult (i.e., as an adult, I think teenagers should not be depressed). They were not adopting their teenage mindset when making the choices (i.e., they were not saying, as a teenager, I would have preferred Profile A). This is also as intended in the instrument design, since the goal is to measure preferences of the sampled adults (18+) for health states experienced over the three time periods (child, teen, and adult).

Description of Domains
For R2 – R6, we showed respondents two versions of the valuation task table. The first version included abbreviated descriptions (e.g., *anger*), as were used in the first round of pretests in May 2011. The second version included the full descriptions from the HRQoL questions. Three respondents (R2, R5, R6) said the full descriptions were better than the abbreviated descriptions. R2 said the full description "helps you think more of the answers you gonna [sic] give." No respondents said the abbreviated descriptions were better than the full descriptions.

An assumption of using the abbreviated version is that respondents will remember the full description of the domain from the HRQoL task. That is, they will remember that *anger* means *so angry that you feel like throwing things...* To evaluate this assumption, we asked respondents with the abbreviated version to describe each domain. Although the respondents described the major components of the domain (e.g., R3 said risky behaviors included unsafe sex, drugs, and alcohol; R4 said anger included getting mad, hitting, and punching people), respondents also included other behaviors that weren't part of the domain. For example, R3 said anger was a "quick temper or a quick fuse" (a milder version of the anger domain), and R4 said risky behaviors included not wearing a seat belt and driving 100 mph.

This evidence led us to include the full descriptions of the domains after R6. Although the interactive version to be fielded by Knowledge Networks was to include a "mouse-over" in which full descriptions

were shown if a respondent moved their pointer over these labels, this requires an extra step for respondents. Based on Dr. Strohm's experience, the mouse-over would be less reliable and is not recommended over the full descriptions. Although the short labels were consistent with some aspects of good practices for DCE design, it should be noted that other HRQOL applications of DCEs have most often used fuller descriptions.

Labeling of Frequencies
Also for R2 – R6, we showed respondents two versions of the frequency labels in the valuation task table. One version simply listed the frequencies (e.g., *never, rarely, sometimes, often*). The other version included a short description of the domain after the frequency (e.g., *never angry, rarely angry, sometimes angry, often angry*). One respondent (R4) preferred the frequency by itself (without the domain label). R2 said that it's easier to compare the profile rows without the extra domain label. R4 said the extra labels make the table harder to read. Although R6 didn't voice a preference, she did mention that the frequencies alone made the table look more "clean" and helps to compare across rows. However, R5 preferred the domain label (e.g., never angry) because it forced him to think more carefully about the domain itself: "you have to think about each individual domain."

Due to the results, as well as our desire to minimize cognitive burden on respondents, we removed the labels after R6 and simply presented frequencies (e.g., *never, rarely, sometimes, often*). This is also consistent with best practices for DCE design.

"Average Month" Terminology
Several respondents commented about the "average month" terminology. R5 said he paid attention to the "average month" language. He wasn't sure whether this meant that he would have the profile continuously for the rest of his life, or just for one month out of the rest of his life. R6 didn't pay any attention to the "average month" terminology. R8 was also confused by the "average month" language: She said "… the question is asking for one month out of your adult life, which one would you like to have." After reflecting on this feedback, we determined that the "average month" language was not necessary and potentially complicated the task. The language was removed after R8.

4. Number of Response Options

A guiding principle of the CMQoL questionnaire is that the same response options must be used in the HRQoL assessment and in the valuation tasks. Using the same response options simplifies the cognitive process for respondents, has advantages for the statistical analysis, and ensures that the valuation data appropriately map back to the HRQOL. Before the start of these pretests, Dr. Strohm discussed the five category responses with Dr. Brown. We realized that these had not been specifically probed or tested during the first round of pretests in May 2011. Therefore, while not specifically requested by OMB, RTI and CDC felt that it was consistent with an overall goal of maximizing the instrument quality to further explore the responses and potential alternatives, if necessary. We tested four different sets of response options. The results for the HRQoL and valuation task are described below.

- *Five categories (Always, Often, Sometimes, Rarely, Never)*. These categories were tested for R2 – R6. Respondents did not generally have problems with the five categories for the HRQoL questions: all but one made clear distinctions between the five categories. For example, R6 said "Sometimes indicates to me that it happens on some regular basis, even if it's a limited basis … Rarely is it happens but there's no rhyme or reason.".

  However, respondents had difficulty with five categories in the valuation task. Some respondents could not remember the order of the frequency labels. For example, R4 said it was hard to remember "which one is which," adding that *sometimes* and *often* were "pretty much the same thing." R2 was not sure whether *sometimes* was more or less frequent than *often* and *rarely*. She suggested adding a scale at the bottom of the page reminding respondents about the order of the frequency labels. More generally, respondents found the valuation task somewhat difficult to complete, in part because of the cognitive effort in weighing five levels of frequency for the seven domains.

- *Three categories (Always, Sometimes, Never).* These categories were tested for R2 – R6. Two respondents (R5, R6) mentioned that they would prefer to have more categories. R5 mentioned that his valuation task choices might be different if *rarely* was used instead of *never*. R6 said that "sometimes, always, and never isn't that much [sic] option," adding that more response options "makes it more realistic." The respondent elaborated that profiles with five categories seemed more like real people, while the profiles with three categories seemed to be more like caricatures.

- *Four categories (Always, Often, Rarely, Never).* These categories were tested for R8 only. To balance the cognitive demands of the valuation task with the need for more variation, we then tested four categories. R8 did not react favorably to these options: "This doesn't give you much [sic] choices … Seems like there should be something between rarely and often … Rarely to me means almost the same thing as never." R8 said that out of 30 days, *often* is 15-20 times and *rarely* is 1-2 times, so R8 wanted another option between *rarely* and *often*. Discussion with additional members of RTI's Survey Research Division suggested that these labels would be problematic for many respondents, so additional revisions were made.

- *Four categories (Often, Sometimes, Rarely, Never).* These categories were tested for R9 – R11. We then used a different set of four response options. In this new scale, we removed "always" from the scale for several reasons. First, out of 9 respondents, zero reported "always" for any of the HRQoL items for the adulthood, teenage, and childhood time periods. Second, we thought that *always* was an extreme frequency for some of the domains (i.e., even the most risky people are unlikely to *always* make risky choices). Third, several respondents devoted disproportionate attention to the *always* category: R5, for example, simply added up the number of *always* responses in the valuation task and chose the profile with the fewest number of *always*. The new response options were often, sometimes, rarely, and never. Compared to the previous four categories, these categories provide more evenly-spaced options, particularly in the middle of the distribution (e.g., sometimes, rarely).

Respondents reacted favorably to these options. R9, R10, and R11 were probed extensively about the difference among the four categories, but all demonstrated an awareness of the differences among categories. For example, R10 said that the difference between *rarely* and *sometimes* was central in the decision in one of the TTO tasks. R9 said that *never* is "it don't [sic] happen," *rarely* is "limited … almost not happening," *sometimes* is "a nice proportion," and *often* is "a lot." R11 said that out of the past 30 days, *often* is 25 days, *sometimes* is 10 days, *rarely* is 3-5 days, and *never* is 0 days.

Recommendation: Continue to use the last option (Often, Sometimes, Rarely, Never). Using four categories provides more variation and more realistic profiles than three categories, but also reduces the cognitive demands relative to five categories. Further, these four categories are more appropriate than other options we explored (E.g., Always, Often, Rarely, Never.)

5. Time Trade-Off (TTO) Task Instructions

Out of the 9 respondents, 7 had no difficulty understanding the TTO instructions (R3, R4, R5, R6, R8, R9, R10). These respondents were asked to explain the task to the interviewer in their own words. They were able to do so without error.

Two respondents (R2, R11) did not understand that the Profiles on the TTO instructions referred back to the two Profiles from the previous page. They didn't understand the language "In the profile you chose." After the interviewer explained that these profiles were related to the previous page, the respondents understood the task. However, R11 asked about this issue at every subsequent TTO question.

Recommendation*:* No action. The Knowledge Networks version uses a different design that clearly assigns the number of years (e.g, 4 or 8 years from today) to the appropriate profile.

6. Time Trade-Off (TTO) Task Observations

Response Strategies
Respondents used several approaches to answer the TTO task. Some respondents, such as R4, always chose a longer life. These respondents sometimes mentioned that they would accept worse health for a longer life because they could use depression medications or other methods to reduce the harmful effects of the worse HRQoL. But other respondents said they would prefer to "live a short happy life than a long unhappy life" (R5). R10 said two extra years of life would not be worth being limited physically. As R8 put it: "It wouldn't be worth the 2 years because if you have no friends, if you have just yourself in your depressed, dark little world, to me it wouldn't be very exciting."

R6 used a different approach. R6 said she has an "eternal view" due to her strong religious views. Because the respondent considers her life after she dies, a difference of 1 or 2 years is small. She said that the difference would have to be something like 25 years for the extra years alive to make a difference.

(It should be noted that the 10 year period in the questionnaire is designed with the estimation of health-state utilities. Specifically, the difference between the two periods and profiles corresponds to the health-state utilities on the 0-1 scale. For example, a respondent who is willing to accept a 3-year difference to avoid a worth state answers that the other state is worth at least 0.3 more on the utility scale. The 10-year reference period is the most widely TTO for health-state valuation.)

Time Reference Period
Respondents gave slightly mixed answers about when in the life course they would "gain years" by changing profiles. For example, R5 initially thought that the "extra years" would be added onto his life as an elderly person, not as someone in his late 20s or early 30s. For example, if he switched profiles, he thought he would live until 90 rather than until 86. R5 initially missed the word "today" in the instructions. After being told that the "extra years" would be added from today (i.e., he would live until 36 rather than 32), he said would be more likely to switch profiles. This was because 4 extra years is a greater proportion of life for young person compared to an older person.

Recommendation*:* Underline the word "today" in all TTO instructions.