Response to OMB Questions regarding Kazakhstan ICR

1. Please add a discussion to Section A of the Supporting Statement that characterizes the epidemiological analysis (exposure and thyroid disease) to which this dose reconstruction will ultimately be applied. That discussion should include a description of and justification for the epidemiological design that will be used so that the type of data needed is clear. In some places it seems you're collecting data for an ecological comparison among villages, yet in others it seems the data are being interpreted as providing the basis for a retrospective cohort analysis at a group level. Is this the most powerful approach to this question? The added discussion should address the strengths and weaknesses of the design chosen. (**Response: please see new material in paragraphs 2-4 of A.1.)** In formulating that discussion please address the following concerns.

    - If the dose reconstruction effort is being conducted on an aggregate (rather than an individual) level, it is important that NIH lay out the justifications for the "groups" chosen at the beginning of the application. We assume that the choice of "groups" is based on a-priori assumptions about differences in exposure, however, the reasons for the choice of age at exposure, gender, and ethnicity as stratification variables is not clearly articulated. How are you testing whether sex, age, and ethnic heritage are the best surrogates for indoor/outdoor time use and milk consumption? (i.e., how will you determine whether you are just replacing one set of assumptions with another set vs. making more "accurate assumptions" (as stated in the Justification.)

    - If the dose reconstruction is at the individual level, is the plan to assign ambient radiation exposure levels geographically (e.g., distance from the site, taking account meteorology and geography), and then within each geographic area assign different scaling factors to individuals by age and sex and ethnicity? Have you conducted an analysis to determine that these are the most sensitive parameters in the exposure model? See prior question about whether these are the best surrogates for time use and diet.

    Response to above 2 bullets: We were remiss in not emphasizing that the most important part of dose reconstruction, estimation of fallout deposition levels, and therefore of ambient levels of radioactivity by geographic location and over time, was estimated by a binational group of experts from the US and Russia with extensive experience with nuclear bomb testing. These estimates are state-of-the-art and most likely the best that can be done. As now discussed in A.1, paragraphs 3-4, the proposed investigation is concerned with factors that influence the extent to which deposited radioactivity was transferred to individual thyroid glands.

    Regarding surrogates, except for individual responses to the questionnaire interviews administered in1998, which will figure in the analysis to some extent, essentially all we know about individual subjects, other than their screening results, is where they were screened, their residential histories, date of birth, gender, and ethnic group. Fortunately, these factors are related to dwelling type, time spent outdoors, and diet, and it is feasible to estimate the relationships based on focus group data. Unfortunately, there isn't anything else we could use in this way.

    - Is there any way to get a handle on between group vs. within group variability in dairy consumption and time use? This would seem to be particularly important.

      Response: The focus group approach is specifically designed to characterize central values and distributions of factors contributing to radiation doses associated with consumption of milk and milk products, as well as time spent outdoors and shielding provided by dwellings and other buildings.

- Do you assume all individuals of a given sex are the same with respect to ingestion and time indoors vs. outdoors? Are you assuming away within group variability? How will this affect your estimates?

  Response: We do not assume that all individuals in any group are the same with respect to most of the attributes of concern in the proposed study. Rather, we assume that a particular individual attribute value is uncertain but can be characterized by a particular uncertainty distribution (e.g., normal with mean 6, and standard deviation 3) which we deduce from the focus group data and other sources. In the analysis, we sample randomly from that distribution for each individual in the group, do the data analysis, and store the numerical result (e.g., an estimate of excess relative prevalence as a function of radiation dose, sex, age at exposure, etc.) We repeat the process multiple times, so that the combined numerical results define an empirical distribution which constitutes our final estimate with uncertainty bounds. This description is a sketchy and over-simplified description of a complex and computationally intensive process that would not have been feasible a few years ago but is becoming the standard for radiation dose-response analyses based on uncertain dose estimates.

- Regardless of the design and the level of aggregation, the key issue is whether there will be sufficient variability in exposure (at a group or individual level) to allow differences to be document. Please provide power calculations that illustrate the extent to which you hypothesize that you will end up with sufficient power to identify meaningful differences. Within this context, we are particularly concerned about managing expectations. What types of results is the community (local and international level) expecting?

  Response: We are rather sure that we will end up with statistically significant dose responses, because we did in our preliminary dose-response analysis which did not take dosimetric uncertainty into account. The values of the estimated dose-response coefficients may change, but their statistical significance should not be affected greatly. Regarding expectations, because of sample size limitations, our conclusions pertain to prevalence of thyroid nodules, and only indirectly to thyroid cancer. Studies based on exposure to radioactive iodine from the Chernobyl accident, indicate that such radiation does increase risk of thyroid cancer. Thus, we now know that we have to worry about both types of radiation. Our study reinforces that conclusion and provides additional information which we would like to refine.

  It would be helpful to provide a description of the exposure model itself and the sensitivity analysis that you conducted to determine which parameters are driving the results, and the degree to which [the focus group provided] assumptions about exposure would need to differ (from the default) to provide conceptually significant changes to the overall results.

  Response: We believe that the exposure model should now be clearer given that its most important component, deposition of radioactive materials, is not the subject of the present proposal. In our current dose reconstruction model, we have assumed that, during the period of testing, dwellings in the predominantly Kazakh villages to the south and southeast of the test site were of adobe construction, with a shielding factor of 1/13, due to the absence of a convenient source of timber, and that the predominantly Russian villages to the northeast of the test site were predominantly of timber construction, with a shielding factor of 1/3. We have also assumed that milk from horses and sheep, which have a higher concentration of dietary iodine compared to cows' milk, made up a higher proportion of the diet in the Kazakh compared to the Russian villages, based on individual questionnaire responses and a summary report from the IRME. These assumptions result in a decrease in external dose and an increase in internal dose in the Kazakh villages compared to the Russian villages. Our sensitivity analyses, based on systematic reductions and increases in the two types of dose for individual villages,

suggested that our estimates are sensitive to the assumptions just described and that the assumptions need additional validation.

2. It seems unlikely that NIH will be able to get the level of precision required for an epidemiological analysis from focus groups, let alone those that are being asked to recall their milk consumption patterns and play activity patterns some 50 years back. Please add a discussion to Section A of the supporting statement that discusses the limitations of the methodology in general and for meeting the specific objectives set out. Such a discussion should address:

   Response: Discussion on limitations was added. See response to specific questions below.

   - The effect of recall bias – 50 years is a long time. Is there any published literature on the effectiveness of focus groups on reducing recall bias? How will you gauge recall bias?

     Response: The focus group methodology was selected because of the difficulty of the recall task for this study. Previous research indicates that focus groups may stimulate in-depth individual participant responses more than individual interviews (Edmonds, 2005). Participant interaction is a unique and compelling feature of focus groups (Edmonds, 2005; Kitzinger, 1995; Lakshman et al., 2000; McLafferty, 2004; Rabiee, 2004; Twinn, 1998) where participants share their experiences to describe the range of experiences in a group as well as the reasons for differences among participants. The individuals in these villages know each other very well and often cared for each other's children. Therefore, they may question each other if a given response does not match their own recall. In this way, the participants provide a sort of check against one another.

   - We understand that NIH has chosen the focus group approach as an aid in recall conditions 50 years ago (i.e, in a group format, participant discussion can provide memory triggers). However, it seems unlikely that NIH will be able to get the level of precision required to differentiate among demographic groups through this method.

     Response: We use several additional methods to prompt participant long-term recall in the focus group sessions. Facilitators used drinking bowls typically used at the time to give participants choices for the quantity and volume of milk consumed. Observation of focus group interviews and debriefing sessions with facilitators supports the utility of the discussion prompts in aiding recall. We also use wall charts designed to lead participants through a typical day both in terms of time of day and typical daily activities (i.e., meal times). This multi-method approach to prompting recall was tested in pilot studies and further refined during subsequent facilitator training sessions. The combination of group discussion, the use of drinking bowls, multi-method wall charts, and facilitator training to probe recall seemed effective in reducing recall bias as much as possible.

   - Given that the most important goal seems like it would be documenting the extent of variability in exposure among groups (assuming that we know the key explanatory factors to identify those groups), is there any concern that the focus group approach will dampen the characterization of variability by encouraging consensus?

     Response: As discussed below, the focus group interviews will elicit both individual- and group level data to capture the extent of variability among individuals and groups.

   - This general concern regarding 'dampening' of the variability' by using focus groups is exacerbated by the format of the questions in the Focus Group Guide. The questions seem designed to elicit concrete response at a group level rather than a

community-wide characterization of variability in milk consumption patterns and play activity patterns. Are those who remember most concretely most likely to be most correct in their memory?

Response: The majority of the questions in the women's guide are designed to elicit responses from each individual participant. Each individual participant is asked to report on her children's consumption of milk and milk products. The group level questions pertain to village-level practices such as whether boys and girls or Kazakhs and Russians differed in their consumption. The question of time spent indoors is also asked at the group level so as to reduce the time burden on the individual participants. For those questions which seek group response, we will try to obtain responses from multiple participants and not merely defer to the few who answer most definitively.

- We don't usually think of focus groups as providing representative results. Rather, focus groups are usually convenience samples. However, in this case you seem to be implying that the focus group participants will provide a picture of exposures for the entire community. If this is a case, do you think that a more rigorous selection process is needed? How will you address the likelihood that you will end up with survivor bias among the focus group participants or lack of representation among less advantaged members of the population?

  Response: Although qualitative data may not be generalized to larger populations (Edmunds, 1999; Merton, Fiske, & Kendall, 1990; Morgan & Krueger, 1998), there may be occasions when it is the only reasonable alternative to conducting a large-scale field study. In this instance, there are very few people who have personal knowledge of daily life practices in rural Kazakhstan villages in the 1950s. It may be that participants available for the current study had different daily life practices in the 1950s than parents of 1998 cohort members. There is no way to completely discount this potential bias but questions were designed to ask about their own life practices and follow-up probes asked if their experience was different for others in the village they knew. There is also the potential that village residents with more resources have moved out of the village either to another village or to Semipalatinsk (the closest city). However, it is our understanding that there was generally minimal variability of resources among village residents in the 1950s and that most village residents lived in poor conditions.

  The screening process proposed for this study restricts participation to men and women who lived in the village at the time and who were of age to have personal experience with dairy consumption and daily life practices at the time of the tests. Because there are so few of these individuals remaining in the villages, we anticipate that eligible focus group participants may in fact represent the entire population of surviving parents of children during the time of the tests. There is no way of discounting survivor bias that results from our recruiting process. Many of the questions ask about a participants own experience, typical experiences during the 1950s, and probes to determine if their experiences were different from typical experiences. When possible, multiple groups will be conducted among selected participant groups to serve as a validation from one group to the other and identify any biases that may exist.

- For all of these reasons, we do not see how the focus group approach will improve reliability and validity of responses. We are uncomfortable with the references on page four to "more objective model" and on page five to "accurate dose."

  The focus group results will in fact result in a "more objective model" if data merely confirm assumptions previously made in the model. In fact, many of the group level

estimates will provide valuable information that will lead to a more objective model or accurate does estimates. For example, use of information learned about time spent outdoors by village and for time periods when school was in or out of session, animal grazing practices, and supplemental feed information will remove uncertainty from the model.

3. Whereas we think this study has the potential to yield interesting qualitative information for looking at a specific set of past exposures, we do not see how scientists will be able to apply this knowledge to current threats. This study seems very limited in application due to the lack of both accuracy and precision in the data to be collected, the differences in building structure and dietary practices between Kazakhstan in the 1950s and the US today, and the uncertainty in exposure estimates from this study. As such, we do not see how this study would lead to 'altering current scientific understanding of biological effectiveness of internal and external radiation exposures' as a justification for this study.

   Response: The results of this study are not expected to "alter current scientific understanding of biological effectiveness of internal and external radiation exposures". In fact, few if any, epidemiologic studies conducted to-date have had sufficient precision and statistical power to accomplish that on their own. The Kazakhstan epidemiologic study, however, has the opportunity to contribute to what is presently only a very limited understanding on the relative biological effectiveness of internal and external radiation exposures from nuclear weapons fallout radiation an more generally to broadening our experience and understanding in estimating doses from environmental exposure to weapons fallout.

   One only has to review the very limited data on which our understanding of risks from external compared to internal dose is based (see Gilbert et al. 2002) to be convinced that the available data are few and insufficient: 1) studies in Utah relative to fallout from the Nevada test site which were, in general, of insufficient power to make definitive conclusions, and 2) studies of Marshall Islanders which are highly uncertain to due lack of good data on contamination levels immediately following the tests, a poor understanding of intake pathways for Marshallese, and a complex mixture of short- and moderately-long lived isotopes. Because the dose and related health risk data are so sparse and uncertain, our present limited understanding on the relative significance of external exposure versus internal dose due to radioiodine isotopes is primarily based on an extrapolation from other less relevant studies, e.g., (i) releases from the Hanford nuclear site (where doses were also very uncertain, the study only had moderate statistical power, and there many dissimilarities to nuclear weapons fallout since the Hanford emissions were all gaseous), and (ii) medical administrations of ^{131}I (primarily to adults).

   The Kazakhstan study is based on the unique situation where a population was actually exposed to nuclear weapons fallout (rather than any surrogate radiation) - a situation of great security interest today. The Kazakhstan population's exposures included external dose and internal dose through contaminated dairy products. Since the same pathways of exposure exist in the U.S. today (and elsewhere), better understanding of those pathways is useful, particularly as national authorities develop strategies to avert dose and mitigate food contamination from radiologic devices.

   The lack of precision on individual doses in the Kazakhstan study is an acknowledged problem, which is shared in virtually <u>every</u> environmental dose reconstruction. This problem, however, will not prohibit improving our understanding on a variety of specific issues including the effects of dairy management practices, and the effects of varying lifestyles. It should also be understood that the lifestyle in Kazakhstan villages in the 1950s was based on very limited economic opportunities, an extremely limited variety of foods, and limited variations in farming practices. Hence, individual intakes in Kazakhstan may have been, in fact, no more variable than in U.S. situations were lifestyle can extremely variable.

In short, the Kazakhstan study has the opportunity to contribute to, though not necessarily to solve, our limited understanding on the likely doses possible from exposure to radioactive fallout, the related health risks per unit dose (with particular reference to children) and may, in fact, contribute to better understanding the relative risk from external and internal exposure.

4. NIH has not provided any statutory citations for providing assurances of confidentiality this seems particularly worrisome since NIH will not actually ever be in possession of the data. How can NIH assure that an international contractor or collaborator will not release the identifiable data?

   Response: IRB approval has been received from SSMA and our international collaborator will not retain participant data. All participant data, both individual and aggregate, will be maintained by NCI. Additionally, no data will be attached to any one individual. Responses will be captured by participant number to track responses through a focus group interview session as a check for consistency. A participant's name will not be attached to their responses.

5. Part B of the Supporting Statement refers to the primarily Russian and Kazakh villages (2 of each) as being representing a range of moderate to high fallout exposure levels. In other parts of the supporting statement it seemed that you were using ethnicity as a surrogates for exposure based on differences in activity patterns. But here you refer villages themselves as representing a range of exposures. In what way are the villages representative of the range of exposures? Is it geographically representative – i.e, within the fallout plume? If so, what is the broader population across which the four villages are representative sample? Are you extrapolating to other villages or just applying the estimates to the individuals in the village associated with the focus groups? Do you assume all Kazakh's are the same across all villages?

   Response: These four villages are the most important in terms of estimated deposit of radioactive fallout and size of the exposed population, and represent two distinct lifestyles in terms of ethnicity, housing construction, and agricultural practices. Among the questions of interest is the extent to which ethnicity per se and geographic location (suitability for intensive agriculture compared to herding) can be separated as modifiers of radiation dose.

6. questions about the interview guidelines:

   - Developing Context section: since you are interested in a multi year time frame (1949-1962), "what do you remember feeling at this time" doesn't seem to be a particularly useful probe. Also, to what extent did everyday villagers really know what was going on? How is what they knew relevant to there responses? How is whether the military contacted them relevant? If you are getting at an exposure issue (leaving the home), perhaps that isn't really a context setting question, but one that you want to explore in more detail re: whether they knew of folks who were asked to leave their homes and whether those folks returned.

     Response: In focus group interview methodology, one often uses questions early in the protocol to prepare participants in responding and provide context to aid in the recall of subsequent questions. This is done to help reduce the effect of recall bias. The use of the probe, "What do you remember feeling at this time" appears in the section titled "Providing Context and Stimulating Recall" and is merely intended to help frame the context of how to respond in a focus group interview and to help participants think back to the time of the tests. The purpose of the questions in this section is not to get at external exposure days following the tests for adults living in the village.

     Villagers had an idea of what was going on – some more than others. Results of the pilot test indicate that some villages were evacuated and some village residents witnessed the nuclear detonation. They certainly have an idea of what daily life practices were in the villages during the 1950s since participant screening methods will restrict participation to only those men and women that lived in the village at the time of the testing.

Again, the question, “At the time of the tests, did the military contact you?” appears in the section titled, “Providing Context and Stimulating Recall”, and is merely intended to help frame the context of how to respond in a focus group interview and to help participants think back to the time of the tests.

- We understand that there are many integrated families. In the housing and schooling questions, how are you handling mixed ethnicity families?

  Response: In the pilot study, questions about mixed ethnicity villages and differences between Russian and Kazakh families were asked. Results suggest that there were few or no differences between Russian and Kazakh families living in the same village in dairy consumptions or daily life practices at the time of the tests. Probes in the focus group interview protocol will again ask if there are differences between families based on ethnicity. This is an empirical question that will be examined upon completion of the field study. Pilot study results and information from SSMA collaborators did not lead us to believe that there were very many mixed ethnicity families.

- Why do the questions focus on August through December? Are you referring to this time frame across all of the years of interest? Were tests always conducted at the same time of year EACH year? If ingestion is a key source of exposure, aren’t you interested in ingestion year round – wouldn’t grazing year round affect the rad load in milk? Then also from a participant standpoint, since this time frame overlaps seasons, is it likely that answers might be different for different seasons?

  Response: The most important tests in terms of deposited fallout occurred in August and September. The radioactive decay rate for 131-I, the major contributor to internal dose, is such that exposure would be negligible after 2 or 3 months, whereas the major source of exposure to external radiation sources, 137-Cs, which decays at a slower rate, is affected mainly by weathering and the shielding provided by buildings.

- In the women’s Guide, the reference to “you’ and “your’ children seems interspersed with questions about children in general. See for instance questions 4 and 5, items f-i seem about the more general population, not ‘your’ own children. For the ‘you’ and ‘your’ questions, how do we know whether this particular woman is typical? Perhaps after each “you or your’ question should be followed with a query about how typical this mom’s activities were to others in the village.

  Response: When we ask participants specifically about “your children” we do not know whether they are typical of mothers in the village. On several occasions during pilot testing, we asked participants if they knew specifically about typical practices. Often times, responses were that they knew what their children consumed and possibly a close relative but were less certain what was typically done. Also, data from these estimates will be used to develop individual dose response estimates so variability among participant is necessary in finding a range of responses for dose response calculations.