

Supporting Statement B
National Toxicology Program Level of Concern Categories Study
(NIH/NIEHS)

NEW
(OMB NO. 0925-XXXX)

Submitted: Day Month Year

Principal Investigator:

Mary S Wolfe, PhD
Deputy Division Director, Division of NTP
National Institute of Environmental Health Sciences
PO Box 12233
Research Triangle Park NC 27709
Phone: 919-541-7539
Fax: 301-451-5759 and wolfe%3014515759@fax.nih.gov
Email: wolfe@niehs.nih.gov

Co-Principal Investigator:

Kristina Thayer, PhD
Director, Office of Health Assessment and Translation
Division of NTP
National Institute of Environmental Health Sciences
PO Box 12233
Research Triangle Park NC 27709
Phone: 919-541-5021
Fax: 919-541-1994
Email: thayer@niehs.nih.gov

Associate Investigator:

Kyla Taylor, MS
Office of Health Assessment and Translation
Division of NTP
National Institute of Environmental Health Sciences
PO Box 12233
Research Triangle Park NC 27709
Phone: 919-316-4707
Fax: 919-541-1994
Email: kyla.taylor@niehs.nih.gov

TABLE OF CONTENTS

B. COLLECTIONS OF INFORMATION EMPLOYING STATISTICAL METHODS 4

B.1. Respondent Universe and Sampling Methods.....4

B.2. Procedures for the Collection of Information.....5

B.3. Methods to Maximize Response Rates and Deal with Nonresponse.....9

B.4. Test of Procedures or Methods to be Undertaken.....10

B.5. Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing
Data.....11

LIST OF ATTACHMENTS

ATTACHMENT 1: LEVEL OF CONCERN “CARDS” SCENARIOS-

B. Collections of Information Employing Statistical Methods

B.1. Respondent Universe and Sampling Methods

The National Toxicology Program (NTP) will use web-based expert solicitation from five NTP stakeholder sectors (academia, industry, non-government organizations, and federal and state agencies) to collect information. The participants will include experts in toxicology, epidemiology, and risk assessment from each of the five stakeholder sectors. These sectors were chosen because they use Level of Concern (LoC) conclusions from NTP literature-based evaluations, are familiar with NTP and its research and analysis activities, and/or are familiar with the conduct, analysis, interpretation, and use of data from environmental health hazard assessments. Forthcoming studies will evaluate consumer/public perception and interpretation via cognitive testing or focus group sessions with members of the general public and experts from the stakeholder sectors. NTP will use targeted advertisement to recruit participants through communications with federal and state agencies; announcements through professional societies (e.g., Society of Toxicology, American Public Health Association, International Society of Environmental Epidemiology, Association of Occupational and Environmental Medicine), NTP listserv, and outreach to individuals who have served on NTP advisory groups in the past five years or provided comments on NTP's research.

Eligible participants will be men or women who have a Masters of Science, PhD, MD, or equivalent degree and at least 3 years' self-reported employment working in public health or a risk assessment-related activity, such as developing hazard identification conclusions based on human and/or animal data or conducting qualitative or quantitative characterizations of risk. IRB exemption status was granted for this study on January 13, 2015, and all participants will give informed consent.

For Phase 1, NTP will recruit 100 participants (~20 per stakeholder sector). For Phase 2, NTP will ask a second group of 100 participants to sort LoC cards. That group will include 60 new participants and /40 repeat participants from Phase 1. The 40 repeat participants for will be selected from among the Phase 1 experts, based upon their responses in Phase 1 (see below). Because this is a preliminary,

exploratory study that does not aim to directly identify updates to the LoC framework, power calculations have not been conducted. Instead, sizes of 95% confidence intervals were used to determine sample sizes; 20 participants from each of five stakeholder sectors will provide 95% confidence intervals for percentages that are, at most, $\pm 20\%$ for individual sectors and $\pm 5\%$ for the entire sample.

B.2. Procedures for the Collection of Information

NTP will develop 40 hypothetical LoC scenarios (**ATTACHMENT 1 – LEVEL OF CONCERN SCENARIOS**) with each presented in a standard format on an electronic “card”. Each LoC card contains a hypothetical scenario that presents information about toxicity and level of human exposure (without identification of a specific chemical) and asks a question regarding the perceived LoC for a specific population group (e.g., NICU infants, women of childbearing age, men, exposed workers, general population). The cards do not specify a particular chemical. The cards are, by necessity, brief and may not span the entire range of possible scenarios; however, seven scientists with expertise in toxicology, epidemiology, and risk assessment have reviewed the cards to ensure that they adequately cover the scenarios typically encountered in NTP hazard assessment activities.

As shown on the sample card (**ATTACHMENT 1 – LEVEL OF CONCERN SCENARIOS**), the LoC scenario has separate headers for hazard and exposure. It uses the hazard classification scheme (“known,” “presumed,” “suspected,” and “not classifiable”) presented in Rooney et al. (2014). Additional information in support of the hazard category includes human evidence, animal evidence, and mechanistic/other evidence and is presented to aid participants in evaluating the information about toxicity. The exposure section provides information about level of human exposure for participants to integrate with the hazard category to answer the LoC question for a specific population group. Information about exposure in a LoC scenario is not as rigorous and robust as the toxicity information. This is because it is meant to reflect real world situations where information about human exposure is often limited or nonexistent. However, as a point of clarification, relevant exposure studies are identified using systematic review process and the potential bias of those studies can be assessed using relevant

items from the risk of bias tool used to evaluate health outcome studies. The risk of bias tool used by NTP includes questions to evaluate the quality of the exposure measurement. Thus, relevant exposure studies are identified using objective and transparent methodology. Assessing the potential bias in these studies is of course more subjective but the NTP systematic review process has processes in place to address this, i.e., independent assessment by 2-3 individuals. **ATTACHMENT 1** presents the hypothetical scenarios for all 40 LoC cards employed in this study.

NTP will use a web-based application with a user interface (“web-based instrument”) for the LoC card sorting exercises in Phases 1 and 2. NTP will host the web-based instrument on a secure, password-protected website. Each participant will use a unique Username and PIN-code for accessing the website. Both phases of the study will use the same sorting process. Participants will be trained to use the tool and work independently. NTP will monitor the amount of time required for participants to complete each phase, although we estimate it will take approximately 6 hours total per participant for the Phase 1 and 2 sessions (total of 4 sessions). Participants will be able to access the website remotely at any time during the experimental LoC card sorting phases using a variety of Internet browsers. Information collection of participants’ responses will be automated and secure, and stored on NIEHS servers. Responses will not be shared among participants.

Phase 1 will have two trials, A and B, for the purpose of establishing consensus on the number of LoC categories to be used and evaluating the reliability of this number. For Phase 1 Trial A, 100 participants will conduct open sorts into participant-defined categories. This will involve sorting 30 cards defined by hazard identification and level of exposure into undefined LoC categories (where category 1= lowest level of concern; category 10 = highest level of concern) using the web-based instrument. In card sorting tasks, it is best practice to use no more than 30 to 40 cards for an open sort (<https://www.usability.gov/how-to-and-tools/methods/card-sorting.html>). The 30 cards selected from the set of 40 are chosen with the intent that they span the range of variability in potential LoC scenarios, though it is acknowledged in this study that the LoC cards utilized in this project may not represent the universe of LoC scenarios possible. In order to assess their level of discrimination between the various

scenarios and identify the most commonly selected number of categories, the number of LoC categories is not pre-determined and rather left to each participant to decide; the maximum number of categories is 10. Each participant will do the sorting exercise twice with an intervening interval of ~2-4 weeks to evaluate the reliability of the participants' placement of cards in Trial 1A versus Trial 1B. NTP will use Phase 1 input to set the number (X levels) of LoC categories for Phase 2.

The goals of Phase 2 are to select labels for each of the X-level LoC categories, determine the self-reported confidence and reliability of card assignments to the categories, and evaluate factors contributing to classification of cards into categories. Phase 2 will have two trials, C and D, and participants in Phase 2 will conduct a closed card sort, in which the number of LoC categories is fixed at X as established in Phase 1. This second group of 100 participants will sort 40 cards into the X-level LoC categories determined from Phase 1. Each of the 100 participants from Phase 2 will do the sorting exercise twice with an intervening interval of ~2-4 weeks to evaluate the reliability of the participants' placement of cards in Trial 2C versus Trial 2D. The additional 10 cards will be designed to test discrimination between juxtapositioned categories. The group for Phase 2 will include 60 new participants and 40 repeat participants. The subset of repeat participants from Phase 1 (4) will include equal proportions of experts, randomly selected, who placed cards into (1) the designated number (X-level) of categories for Phase 2, (2) fewer categories, and (3) more categories. For example, if we decide to have 4 LoC categories, we would use equal proportions of repeat participants who used <4, 4, and >4 categories in Phase 1.

In Phase 2, Trial C, NTP will also ask participants to propose suitable names for each new X-level LoC category; the consensus category labels will be used in Trial D. NTP will not define the lexicon used by participants for labeling the categories. NTP will consider all proposed labels and work with its risk communication advisors to narrow the list of proposed names toward a set of category labels. If needed, NTP will convene an external stakeholder panel in a public forum to obtain input on the label set. In addition to sorting the LoC cards into X-levels, participants in Phase 2, Trials C and D will be asked to identify and rank order the scenario factor(s) (hazard category, human evidence, animal evidence,

mechanistic/other evidence, exposure description, and population group for concern) that most influenced their selection of the LoC category and rate their confidence in that category selection using a 7-point scale (1 = “not confident” to 7 = “highly confident”). NTP will use the Phase 2, Trials C and D responses to determine which LoC factors most contributed to participants placing LoC scenarios into upper, lower, and in-between categories and their rating-of-confidence opinions. We will also examine the Trial D responses based on expert sector (e.g., federal and state governments, academia, industry, non-profit organizations).

For Phase 1, the analysis will include calculating the mean, median, and modal number of LoC categories in Phase 1, Trials A and B to determine the number of LoC categories (X-levels) to use in Phase 2. To assess reliability of the number of LoC categories that participants propose, paired t-tests (or Wilcoxon signed ranks tests or kappa statistics for agreement, if nonparametric methods are warranted) will be used. The specific test selected will depend on the range and distribution of the number of categories selected by the 100 participants. The number of categories will be judged reliable if there is no significant change between Trial A and Trial B at significance level 0.05. If the number of categories selected in Phase 1 is not judged reliable, the Trial B result will be used for Phase 2, because it is assumed that the participants are more experienced with sorting on their second trial than on their first.

Responses of participants in Phase 2, Trial C will be used to identify suitable labels for the X-level LoC categories, based on majority nominations. The reliability of card assignments to LoC categories in Trials C and D will be assessed several ways, including (1) determining, for each card, the percentages of participants placing it within each of the LoC categories within each Trial, (2) for each card, evaluating a participant’s agreement of category placement between Trials C and D using kappa statistics and McNemar’s chi-square test, and (3) judging the concordance of LoC cards’ rank orderings between Trials C and D to determine the consistency of categorization of LoC scenarios, using Kendall’s tau correlation coefficients or a contingency table measure of association, depending on the number of X-level categories selected in Phase 1. Analysis of variance and step-wise regression analysis will be conducted to determine which LoC scenario factors (e.g., hazard category, human evidence, animal

evidence, mechanistic/other evidence, exposure description, and population group for concern) most contributed to the participants' sorting individual scenarios into the new X-level categories. We will also use analysis of variance to analyze LoC rating-of-confidence opinions based on expert sector (e.g., federal and state governments, academia, industry, non-profit organizations).

B.3. Methods to Maximize Response Rates and Deal with Nonresponse

Since we are targeting stakeholder groups from sectors that have previously used LoC conclusions from NTP literature-based evaluations, are familiar with NTP and its research and analysis activities, and/or are familiar with the conduct, analysis, interpretation, and use of data from environmental health hazard assessments, we hope to receive a high response rate rate (~80%). We are taking several actions to maximize the response rate. Pre-survey targeted advertisement has been shown to spur interest and improve response rates (Dal Grande et al. 2016; Dillman DA et al. 2014)). In addition, we plan to send out email reminders to participants approximately one week and two weeks after they have been granted access to each trial. In anticipation of a possible 20% non-response rate, we plan to recruit an additional 25 participants from each sector (total 125), with the goal of obtaining 20 responses from each sector (total 100). Among proposed participants will be scientists who previously served on an NTP advisory group, including representatives of academia, industry, and state government, or are from a federal agency represented on the NTP Executive Committee.¹ We plan to recruit for Phases 1 and 2, and will add the list of non-selected registrants from Phase 1 to the pool of experts who enroll for Phase 2.

We believe that using targeted advertisement for recruitment of participants will enable us to readily communicate information about this project to potentially interested experts. We believe that the pool of toxicologists, epidemiologists, and individuals engaged in risk assessment, who would meet the eligibility requirements of education and work experience, is more than sufficient to provide a pool of candidates to fill our need for approximately 160 participants (~32 from each of the five stakeholder sectors). While we hope to obtain similar numbers of scientists from each of the three specialties across

¹The NTP Executive Committee has representatives from nine HHS and non HHS federal research and health regulatory agencies.

the five sectors, it is not required. To ensure broad and diverse recruitment for this study, we will send recruitment information to professional societies such as the Society of Toxicology (membership 6,500), International Society of Environmental Epidemiology (membership 800), American Public Health Association (membership 25,000) and Association of Occupational and Environmental Medicine (250 members); contacts at all 50 U.S. state environmental health departments (80); current and former NTP advisory boards; academic and schools of public health contacts (182) including the Association of Schools and Programs of Public Health; various environmental non-government organizations (43) such as Environmental Defense Fund and Environmental Working Group; industry groups (130) including the American Chemistry Council; other federal health regulatory and research agency contacts (144); and NTP listserv that include subscribers from various stakeholder sectors (4282).

B.4. Test of Procedures or Methods to be Undertaken

Our study concept was formally reviewed by the NTP Project Review Committee on October 30, 2014. This project was mentioned as a research activity to the NTP Board of Scientific Counselors on December 14, 2014, and was presented in more detail to the NTP Board of Scientific Counselors at its meeting at NIEHS on June 16, 2015, and will be shared with agencies represented on the NTP Executive Committee through agency points of contact. Additionally, the concept presented at the June meeting was available to the public for comment. There was enthusiastic support for this project from the NTP Board of Scientific Counselors and public comment at the June 16 meeting.

We have convened an internal NTP advisory group for feedback on our proposed study plans, LoC card scenarios, and the web-based instrument for the LoC card-sorting exercises. We will continue to consult this group during the study's execution. This outreach provides opportunities for input from the NIEHS intramural community, NTP advisory boards, and externally by federal and academic experts.

B.5. Individuals Consulted on Statistical Aspects and Individuals Collecting and/or Analyzing Data

Dr. Mary Wolfe (919-541-7539 and wolfe@niehs.nih.gov), Deputy Division Director for Policy, Division of NTP, NIEHS, and Dr. Kristina Thayer (919-541-5021 and thayer@niehs.nih.gov), Deputy Division Director for Analysis and Director, Office of Health Assessment and Translation (OHAT), Division of NTP, NIEHS, are co-investigators on this study. Kyla Taylor (919-316-4707 and kyla.taylor@nih.gov), a health scientist in OHAT, is associate investigator. The study plan was developed with external input from Dr. David Budescu (718-817-3786 and budescu@fordham.edu), Anne Anastasi Professor of Psychometrics and Quantitative Psychology, Department of Psychology, Fordham University; Dr. Thomas Wallsten (301-405-3562 and tswallst@umd.edu), Professor emeritus, Department of Psychology, University of Maryland-College; and Dr. Barbara Forsyth, private consultant. Dr. Budescu, Dr. Wallsten, and Dr. Grace Kissling (919-541-1756 and grace.kissling@nih.gov), a biostatistician in the NIEHS Biostatistics and Computational Branch, have provided input on the statistical approach and they will participate in the data analysis.

References

- Dal Grande E, Chittleborough CR, Campostrini S, Dollard M, Taylor AW. 2016. Pre-survey text messages (sms) improve participation rate in an australian mobile telephone survey: An experimental study. *PLoS One* 11:e0150231.
- Dillman DA, Smyth JD, LM. C. 2014. Chapter 2. Internet, phone, mail, and mixed-mode surveys; the tailored design method, 4th edition. John Wiley & Sons, Inc.
- Rooney AA, Boyles AL, Wolfe MS, Bucher JR, Thayer KA. 2014. Systematic review and evidence integration for literature-based environmental health science assessments. *Environmental health perspectives* 122:711-718.