UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

OFFICE OF
PREVENTION, PESTICIDES, AND
TOXIC SUBSTANCES

April 02, 2007

MEMORANDUM

SUBJECT: Transmittal of Meeting Minutes of the FIFRA Scientific Advisory Panel Meeting Held January 9 - 12, 2007 on the Review of Worker Exposure Assessment Methods.

TO: Anne Lindsay, Acting Director
Office of Pesticide Programs

FROM: Myrta R. Christian, Designated Federal Official
FIFRA Scientific Advisory Panel
Office of Science Coordination and Policy

THRU: Steven Knott, Executive Secretary
FIFRA Scientific Advisory Panel
Office of Science Coordination and Policy

Clifford J. Gabriel, Ph.D., Director
Office of Science Coordination and Policy

Attached, please find the meeting minutes of the FIFRA Scientific Advisory Panel open meeting held in Arlington, Virginia on January 9 - 12, 2007. This report addresses a set of scientific issues being considered by the Environmental Protection Agency pertaining to the Review of Worker Exposure Assessment Methods.

Attachment

cc:

James B. Gulliford
James J. Jones
William Jordan
Margie Fehrenbach
Janet Andersen
Steven Bradbury
William Diamond
Debbie Edwards
Richard Keigwin
Oscar Morales
Tina Levine
Jack Housenger
Lois Rossi
Frank Sanders
Betty Shackleford
Enesta Jones
Douglas Parsons
Vanessa Vu (SAB)
Jeff Evans
Jeff Dawson
Cassi Walls
David J. Miller
Mathew Crowley
OPP Docket

FIFRA Scientific Advisory Panel Members

Steven Heeringa, Ph.D. (FIFRA SAP Chair)
John R. Bucher, Ph.D., D.A.B.T.
Janice Elaine Chambers, Ph.D., D.A.B.T.
Stuart Handwerger, Ph.D.
Kenneth J. Portier, Ph.D.

FQPA Science Review Board Members

Henry T. Appleton, Ph.D.
Dana B. Barr, Ph.D.
Brian Curwin, Ph.D.
Paul Y. Hamey, M.Sc.
Cynthia J. Hines, M.S., CIH
Brian J. Hughes, Ph.D., M.P.H., D.A.B.T.
Dallas E. Johnson, Ph.D.
David Kim, Ph.D.
Andrew J. Landers, Ph.D.
Chensheng Lu, Ph.D.
Peter D.M. Macdonald, D.Phil., P.Stat.
William J. Popendorf, Ph.D.
Mark G. Robson, Ph.D., M.P.H., A.T.S.

SAP Minutes No. 2007-03

A Set of Scientific Issues Being Considered by the Environmental Protection Agency Regarding:

REVIEW OF WORKER EXPOSURE ASSESSMENT METHODS

JANUARY 9 - 12, 2007
FIFRA Scientific Advisory Panel Meeting,
held at the
Environmental Protection Agency Conference Center
Arlington, Virginia

NOTICE

These meeting minutes have been written as part of the activities of the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), Scientific Advisory Panel (SAP). The meeting minutes represent the views and recommendations of the FIFRA SAP, not the United States Environmental Protection Agency (Agency). The content of the meeting minutes does not represent information approved or disseminated by the Agency. The meeting minutes have not been reviewed for approval by the Agency and, hence, the contents of these meeting minutes do not necessarily represent the views and policies of the Agency, nor of other agencies in the Executive Branch of the Federal government, nor does mention of trade names or commercial products constitute a recommendation for use.

The FIFRA SAP is a Federal advisory committee operating in accordance with the Federal Advisory Committee Act and established under the provisions of FIFRA as amended by the Food Quality Protection Act (FQPA) of 1996. The FIFRA SAP provides advice, information, and recommendations to the Agency Administrator on pesticides and pesticide-related issues regarding the impact of regulatory actions on health and the environment. The Panel serves as the primary scientific peer review mechanism of the EPA, Office of Pesticide Programs (OPP), and is structured to provide balanced expert assessment of pesticide and pesticide-related matters facing the Agency. Food Quality Protection Act Science Review Board members serve the FIFRA SAP on an *ad hoc* basis to assist in reviews conducted by the FIFRA SAP. Further information about FIFRA SAP reports and activities can be obtained from its website at http://www.epa.gov/scipoly/sap/ or the OPP Docket at (703) 305-5805. Interested persons are invited to contact Myrta R. Christian, SAP Designated Federal Official, via e-mail at christian.myrta@epa.gov.

In preparing the meeting minutes, the Panel carefully considered all information provided and presented by EPA, Health Canada, California EPA, the Agricultural Handler Exposure Task Force (AHETF) and the Antimicrobial Exposure Assessment Task Force II (AEATF II), as well as information presented by public commenters. This document addresses the information provided and presented by these groups within the structure of the charge.

TABLE OF CONTENT

SAP Minutes No. 2007-03

A Set of Scientific Issues Being Considered by the Environmental Protection Agency Regarding:

REVIEW OF WORKER EXPOSURE ASSESSMENT METHODS

JANUARY 9 - 12, 2007
FIFRA Scientific Advisory Panel Meeting,
held at the
Environmental Protection Agency Conference Center
Arlington, Virginia

Steven G. Heeringa, Ph.D.
FIFRA SAP Chair
FIFRA Scientific Advisory Panel
Date: April 02, 2007

Myrta R. Christian, M.S
Designated Federal Official
FIFRA Scientific Advisory Panel
Date: April 02, 2007

Federal Insecticide, Fungicide, and Rodenticide Act
Scientific Advisory Panel Meeting
January 9 - 12, 2007

REVIEW OF WORKER EXPOSURE ASSESSMENT METHODS

PARTICIPANTS

FIFRA SAP Chair

Steven G. Heeringa, Ph.D., Research Scientist & Director for Statistical Design, University of Michigan, Institute for Social Research, Ann Arbor, MI

Designated Federal Official

Myrta R. Christian, M.S., FIFRA Scientific Advisory Panel, Office of Science Coordination and Policy, EPA

FIFRA Scientific Advisory Panel Members

John R. Bucher, Ph.D., D.A.B.T., Deputy Director, Environmental Toxicology Program, NIEHS, Research Triangle Park, NC

Janice E. Chambers, Ph.D., **D.A.B.T.**, William L. Giles Distinguished Professor & Director, Center for Environmental Health Sciences, College of Veterinary Medicine, Mississippi State University, Mississippi State, MS

Stuart Handwerger, M.D., Professor of Pediatrics, University of Cincinnati Children's Hospital Medical Center, Cincinnati, OH

Kenneth M. Portier, Ph.D., Program Director, Statistics, American Cancer Society, Statistics and Evaluation Center, Atlanta, GA

FQPA Science Review Board Members

Henry T. Appleton, Ph.D., National Toxicologist, U.S. Forest Service, Arlington, VA

Dana B. Barr, Ph.D., Chief, Pesticide Laboratory, Centers for Disease Control and Prevention, National Center for Environmental Health, Atlanta, GA

Brian Curwin, Ph.D., Associate Research Fellow, National Institute of Occupational Safety and Health, Cincinnati, OH

Paul Y. Hamey, M.Sc., Principal Scientist, Human Exposure Assessment, Pesticides Safety Directorate, York, United Kingdom

Cynthia J. Hines, M.S., C.I.H., Senior Research Industrial Hygienist, National Institute of Occupational Safety and Health, Cincinnati, OH

Brian J. Hughes, Ph.D., M.P.H., D.A.B.T., Toxicologist, Michigan Department of Agriculture Pesticide and Plant Pest Management Division, Lansing, MI

Dallas E. Johnson, Ph.D., Professor Emeritus, Department of Statistics, Kansas State University, Manhattan, KS

David Kim, Ph.D., Research Fellow, Harvard University, Department of Environmental Health, Harvard School of Public Health, Boston, MA

Andrew J. Landers, Ph.D., Director, Application Technology Group and Associate Professor Cornell University, Department of Entomology, Geneva, NY

Chensheng Lu, Ph.D., Assistant Professor, Emory University, Department of Environmental and Occupational Health, Rollins School of Public Health, Atlanta, GA

Peter D.M. Macdonald, D.Phil., P.Stat., Professor of Mathematics and Statistics, McMaster University, Department of Mathematics and Statistics, Hamilton, Ontario, Canada

William J. Popendorf, Ph.D., Professor, Utah State University, Department of Biology, Logan, UT

Mark G. Robson, Ph.D., M.P.H., A.T.S., Director of the New Jersey Agricultural Experiment Station and Professor of Entomology, Rutgers University, New Brunswick, NJ

INTRODUCTION

The Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA), Scientific Advisory Panel (SAP) has completed its review of the Worker Exposure Assessment Methods. Advance notice of the meeting was published in the *Federal Register* on October 27, 2006. The review was conducted in an open Panel meeting held in Arlington, Virginia, from January 9 – 12, 2007. Dr. Steven G. Heeringa chaired the meeting. Myrta R. Christian served as the Designated Federal Official.

The FIFRA SAP met to consider and review the Worker Exposure Assessment Methods. The Agency issued its first occupational exposure testing guidelines in the early 1980s. These guidelines were intended to standardize the methodology used to conduct the studies necessary to allow the Agency to determine the potential exposures and consequent risks associated with the activities surrounding the use of pesticides. These activities include handling pesticides (i.e. mixing, loading and applying) as well as working in treated sites following pesticide applications (e.g., harvesting, thinning, weeding; servicing cooling towers). In the early 1990s, two databases--the Pesticide Handlers Exposure Database (PHED) and the Chemical Manufacturers Association (CMA) database were constructed in order to estimate exposures resulting from mixing/loading/applying pesticides. The data assembled for use in these databases were taken from published literature as well as from industry studies submitted to the Agency. These databases have been used as the main sources for estimating occupational exposures to workers handling pesticides for both registration and reregistration actions. Since the early 1980s, the Agency has been using a scenario-based approach in its assessments for estimating exposures for occupational pesticide handlers (e.g., mixers, loaders, and applicators). This approach is consistent with the Agency's guidelines for exposure assessment which can be found on the EPA website at http://cfpub.epa.gov/ncea/cfm/recordisplay.cfm?deid=15263.

Over the years since the issuance of the exposure guidelines, scientific issues have been raised about the accuracy of exposure estimates based on data developed using these methods. In addition, recent protocols for the generation of new agricultural pesticide handler exposure data are being generated by a pesticide industry task force and were reviewed by the Agency's Human Studies Review Board (HSRB) (see http://www.epa.gov/osa/hsrb/files/june2006finaldraftreport82806.pdf for further information). The board raised questions concerning the scientific merits of the proposed protocols.

Given the scientific issues that have been raised regarding occupational pesticide exposure estimates and study protocols, including the recent comments from the HSRB, at this time EPA asked the FIFRA Scientific Advisory Panel (SAP) to evaluate, in detail, issues associated with certain methodologies used to generate exposure studies and the procedures used to develop exposure estimates. As part of the background for the SAP meeting, the Agency developed a case study that details the procedures and data the Agency uses to evaluate 6 exposure scenarios that are common in agriculture. These data can be found in the existing Pesticide Handlers Exposure Database.

The Charge to the FIFRA SAP focused on the following issues: data needs (e.g., availability of data in the Pesticide Handlers Exposure Database); sample collection methods (e.g., whole-body

dosimetry, handwashing, facial/neck wipes, and biological monitoring [BM]); unit exposure (e.g., relating the amount of exposure to the amount of chemical active ingredient handled); and sample size issues (e.g., inter-/intra-worker variability and representativeness).

The agenda for this SAP meeting included presentations from the Health Effects Division (HED) and Antimicrobial Division (AD) in the OPP. In addition, presentations were provided by Health Canada, Pest Management Regulatory Agency; California EPA, Department of Pesticide Regulation; Agricultural Handler Exposure Task Force (AHETF) and the Antimicrobial Exposure Assessment Task Force II (AEATF II).

Dr. Tina Levine (Director, HED, OPP) offered opening remarks at the meeting.

PUBLIC COMMENTERS

Oral statements were presented as follows:

Rebeckah Adcock, on behalf of the Pesticide Policy Coalition, American Farm Bureau Federation
Larry Olsen, Ph.D., on his own behalf
Andrew Moore, on behalf of the National Agricultural Aviation Association
Pamela Roa, Ph.D., on behalf of Farmworker Justice

Written statements were provided by:

Richard Fenske, Ph.D., MPH. University of Washington's School of Public Health and Community Medicine

SUMMARY OF PANEL DISCUSSION AND RECOMMENDATIONS

Measuring dermal exposures is not simple, but is rather one of the most complex tasks in the field of exposure assessment. The Panel was frequently complimentary toward the efforts and presentations made by the Agency and both of the task forces in dealing with these important and complex issues.

Because terminology is so important within the topic of dermal hazards from chemicals, a glossary of terms was prepared and is included as Appendix A to this Panel Report to help readers understand the particular definition of "exposure" and its relation to "dose" used in this context and the distinction among the terms “bias,” “uncertainty,” and “variability.” In exposure and risk assessment, bias and uncertainty need to be minimized; conversely, variability needs not to be minimized but examined. Therefore, when evaluating dermal exposure data, EPA should consider (1) the validity of the sampling and analytical methods in terms of bias, (2) the magnitude of the uncertainty inherent within the methods, and (3) the ability of the resulting information to illuminate the variability. These issues are integral to virtually all of the Panel's responses.

1) Data Needs

The Panel agreed with the Agency’s concern about the limitations of the existing PHED exposure database. Furthermore, they concluded that additional data could significantly improve the Agency’s ability to assess worker exposure. They listed eight limitations within PHED including its inconsistent data quality; a patch-work of methods, some with high uncertainty and data censoring; a high level of "clustering," and an inadequate number of samples and diversity within some scenarios. Panelists also expanded on the following three broad weaknesses within PHED.

The inclusion within PHED of studies where either not all parts of the body were monitored or a substantial number of exposures were undetectable do not allow the results to yield accurate exposure statistics of interest for regulatory assessments. New Agricultural Handlers Exposure Database (AHED) data and software may be able to correct these weaknesses. The AHED study design will also include more reliable exposure assessment methods (especially of the hands; see also Charge #2) and newer ("modern") pesticide application equipment and techniques (see also Charge #4). The probable uncertainty in the calculated exposure values creates an unrecognized weakness. A rationale is presented in Appendix B to expand the current concept of “data grading” based only on the analytical method to include the “probable uncertainty” of the calculated exposure level. Such a broader grading scheme could help users interpret exposure values better as well as create a direct means for the Agency to demonstrate both the weakness of the existing PHED and the improvements that should result from a new database.

The Panel viewed the selection criteria proposed by AHETF and AEATF to be reasonable for generating exposure data for using in exposure assessments, with the following caveats. The monitoring duration requirement may be too stringent. Some provision to allow the inclusion of data from settings where only short-term uses are the norm may need to be added. The criteria to use biomonitoring data only if primate dermal absorption and pharmacokinetic data exist for the

chemical may also be too restrictive. The Panel suggested that "extrapolation" factors appropriately estimated from rat and porcine models to humans should be allowed. Better justification is needed to either include or exclude air sampling from the protocol, and the criteria for sampling an "inspirable" aerosol needs to encompass large droplets or particles. Further, the array and location of patches (if they are used) should be standardized.

Finally, the Panel noted (here and in our response to Charge #6) the need to generate a database that documents the frequency and the duration with which handlers in general are exposed to pesticides in the scenarios being considered, as a complement to the database of the intensity of exposures of participants in the studies being contemplated herein.

2) Passive Dosimetry

The response to this charge is divided into four parts that address bias, a correction factor for handwashing, a correction for breakthrough of dosimeters, and complementary uses of biomonitoring and dosimetry.

The Panel concluded that although a bias may exist, no bias between dermal exposure monitoring and biological monitoring could be detected in large part because of the statistical uncertainty inherent in the exposure data (see also Appendix B). The uncertainty for patch dosimeters can be a factor of 4× to 7× due to the calculation needed to scale up from deposition onto a patch of *circa* 40 cm^2 to a body-part area of *circa* 1000 cm^2 and the potential variability in the spatial pattern of dermal deposition onto any given body-part. In comparison, the probable uncertainty when using whole-body dosimeters that cover *circa* 90% of the body is likely to be no more than about 1.5×. A similar analysis in our response to Charge #3 will show that the probable uncertainty of a dermal exposure value derived from biomonitoring could range from ±20× to ±100×.

The Panel was slightly more equivocal about a need to correct handwashing for its efficiency at recovering pesticides from the skin. Existing data clearly indicate that adsorption of certain pesticides can occur within a matter of minutes after the exposure, that hand wiping underestimates dermal exposure more than does hand washing, and that recovery efficiency is really not a constant. A first-order kinetic model of adsorption that depends upon the pesticide's $K_{O/W}$ or octanol-water partition coefficient was suggested. Limited results using this model indicate that the interval from initiating exposure until washing the hands can be important when measuring pesticide handler exposure over the planned 4 to 8-hour day. However, others on the Panel pointed out that the accuracies of either modeling or experimental data could be confounded in field conditions by the effects of repetitive (multiple) rinsing or washing that can change the skin's absorption rate, either enhancing or decreasing recovery from the skin. Also wearing chemical protective gloves will decrease the importance of hand exposures categorically. Biomonitoring presents an interesting but often impractical approach to correcting handwash data due to its inclusion of other routes of exposure and the impact of the additional burden on the participant. Overall, the Panel recommended that use of a hand washing technique should be accepted in AHETF or AEATF studies if it is supported by either laboratory data and/or a model that predicts and can correct for its efficiency over the sampling time for the pesticide being studied.

The Panel concluded that generally there is no need to correct whole-body dosimeters worn under outer garments for the breakthrough of residues. Patch dosimeters should have an impervious backing to prevent breakthrough, but they may have an unacceptably large probable uncertainty to meet the proposed uses of AHED or Biocide Handlers Exposure Database (BHED). The use of a whole-body dosimeter placed directly against the skin is recommended, but no sure means of detecting its breakthrough from liquid saturation was identified.

The Panel also concluded that biological monitoring can be complementary to dermal exposure monitoring to detect if not estimate the amount of breakthrough from a whole body dosimeter (WBD). Therefore, the Panel generally supports an Agency proposal that biomonitoring *may* be included in any sampling plan, but recommended that it not be required because of its potential to bias participant selection. Furthermore, the detection of breakthrough should not be grounds to discard the sample, again to avoid selection bias.

3) Passive Dosimetry versus Biomonitoring

The relationship between passive dosimetry and biomonitoring depends in part upon what form of dosimetry is being used and upon the question being asked. Within this caveat, the Panel concluded that passive dosimetry can generate data that can be used to predict worker exposure for a wide variety of scenarios and activities.

The above conclusion was supported by two analyses. A basic analysis of the pathways taken by the pesticide from exposure to excretion was used to show that concurrent whole-body dosimetry and biomonitoring will interfere with each other, precluding any expected correlation between their data. The second analysis was used to show that the "probable uncertainty" in dermal exposure values calculated from either biomonitoring or patch dosimetry are sufficiently large to weaken correlations using data from independent studies. Thus, the agreement in the data presented to the Panel is about as good as can be expected and is sufficient to support the Agency's conclusion that a passive dosimetry-based approach can generate data that can be used to develop relatively predictive estimates of worker exposure for a wide variety of scenarios and activities.

4**) Normalization of Exposure by Amount of Active Ingredient Handled (AaiH)**

Most Panel members agreed that the data shown to the Panel did not consistently support a linear relationship between exposure and AaiH. A linear relationship seems intuitively logical, but a physical rationale should be developed to support that hypothesis (or other hypotheses) in all scenarios. Several good reasons were given as to why a linear relationship might exist but not be detectable within the PHED data. Some arguments were presented to accept and/or explain an apparent non-linear relationship between AaiH and exposure. And some factors other than AaiH were suggested that might be better correlated with exposure than AaiH.

For dermal exposure to be proportional to AaiH implies that a consistently small fraction of the amount of pesticide handled is deposited onto the handler's skin. The Panel proffered three reasons why so few of the exposure examples from PHED were found to be proportional to AaiH. The "ecological fallacy" is the mistaken assumption that all members of a group have the

same characteristics as the group at large. The “engineering fallacy” is the mistaken assumption that all work practices and equipment within a scenario are in fact the same. Either fallacy might cause proportionality to apply within a closely defined portion of a scenario (such as within a particular cluster), but fail to apply across the whole scenario. Alternatively, the uncertainty within the data might just be too large to permit a linear proportionality to be detected, even if it exists.

Several Panel members pointed out that strict adherence to proportionality is not dictated. If a nonlinear relationship exists between AaiH and dermal exposures to pesticides, then the log-log regression coefficient that gives the best fit of the relationship should be used, whether it is equal to 1 or not. And some other Panelists took the position that at least in certain scenarios, no correlation at all should be expected between AaiH and exposure, and went on to suggest other factors that might be better predictors of exposure such as the number of times a handler contacts a contaminated surface or pesticide residues on surfaces that are at steady-state or saturated, regardless of the AaiH.

Some suggested that the variability in field conditions will make it difficult for a well-designed observational study such as that proposed to illuminate a clear predictive relationship. An array of short documents that put forth a plausible argument based on physical mechanisms applicable to each scenario might comprise an alternative way to justify any of the above assumptions and could also help in designing "purposive non-random sampling" to support any of these predictive correlates of exposure. The cluster-based sampling strategy is being optimized to test linearity (see Charge #6). In scenarios where correlation with a variable other than AaiH is hypothesized, the degree of cross-correlation between these variables and AaiH needs to be considered.

5) Within-worker and Between-worker Variability

The Panel agreed that exposure data collected from observational studies has the potential to address all three potential sources of variation identified in the background documents: within-handler, between-handler, and between-study.

Much of the Panel felt that within-handler variability should be de-emphasized because repeated measurements on any one handler are likely to demonstrate low within-handler correlations and that between-handler variability can predict the distribution of long-term cumulative exposures better than within-handler variability. The Panel also recognized that repeated measurements may introduce a selection bias within participants and the narrow time window for pesticide application can make obtaining repeated measurement data infeasible for some scenarios. Also a decision to include repeated measures could raise ethical or other issues at a Human Studies Review Board.

At the same time, some on the Panel were concerned that EPA might be missing an opportunity to obtain at least some limited repeated measurements. One suggestion was to analyse within-handler data currently within PHED to provide some further quantitative evidence to support a later decision. The Panel was somewhat split among members who suggested that repeated measurements require a large increase in both effort and analytical costs and those that felt that adding limited repeated passive dosimetry measurements on the same or next day would provide

the most information for the least additional investment. The Panel recommended that EPA and the AHETF evaluate whether within- and between-worker variability might be evaluated for selected scenarios where application frequency and logistics are favourable. Statistical techniques were also suggested using statistical methods that incorporate cost information to see how to get the most precision for the least total cost.

6) Sample Size: Number of Sites and Subjects per Scenario/Activity

While the appropriate number of Monitored Units is integral to the goals set for any database, the goals set by the Agency and by the AHETF may be in conflict. The goal of the Agency is to "adequately represent the range of exposure of people who engage in a particular scenario and activity." The goals of AHETF are to be able to estimate exposure after normalizing by the amount of active ingredient handled [AaiH] within a proposed factor [*K*] of 3 and to be able to "distinguish" between complete proportionality and complete independence of exposure and AaiH.

To meet the accuracy goal, the number of sampling units will depend in part on the *K* factor. The Panel felt that a default value of $K = 3$ seems reasonable, although it need not be a fixed value and might best be varied among scenarios. The number of sampling units under the cluster-based sampling plan is also quite sensitive to the specified geometric standard deviation [GSD] and intra-class correlation coefficient [ICC]. Therefore, EPA and the AHETF should consider building in one or two "check points" after a certain amount of new data is collected to evaluate assumptions about the GSD and ICC and refine the scenario sample sizes. In order to understand the robustness of the cluster design to the assumption of log-normality and its impact on calculated sample size, the Panel also recommended that sample size simulations be performed using an alternative skewed distribution for concentration values, such as a Gamma distribution.

Overall, the Panel believes that the recommendation to have 5 handlers per cluster with approximately 5 clusters for each scenario/activity seems reasonable at this point in time. However, because the variation in ICCs observed to date comes from sparse data, the Panel again recommends building in one or two "check points" to consider adjustments in the numbers of clusters and monitoring units (MUs) within clusters to be sampled. At the same time, clear guidelines are needed on how to add new clusters that guard against the potential to "parse" the target population into more clusters just to limit sample size requirements. More clarity was desired about how a cluster is defined, whether that definition needs to be scenario dependent, and the notion that "geographic differences" are important for establishing clusters.

The Panel questioned the adverse effect that was being imposed on sample selection by the secondary goal proposed by AHETF to be able to elucidate a potential predictive relationship between exposure and AaiH. Is it feasible to span the desired range of AaiH without biasing the selection of application equipment and/or work practices? Will the "bias toward conditions that might yield higher exposures" conflict with the assumption that the purposive sample of the MUs approximates a probability sample from the target population? The conclusion of a detailed critique of the AHETF sampling plan contained within Appendix C is that the lack of a database that documents the distribution of real-world tasks, activities, or pesticide usage may make it

very difficult to judge or compensate for the biases introduced by the sampling selection design being proposed (see also the last comment in our response to Charge #1). Further clarification is needed about this sampling bias and how it might affect the distribution of exposure values and subsequent uses of the database.

The AEATF Study Plan deals with a very different situation and appears much more amenable to experimental control. The proposal to take 15 monitoring units initially seems adequate to give an overview, and yet in this case it should be feasible to increase the sample size for any scenario at a future date if more observations are needed. The Panel also suggested that the AEATF undertake a pilot study to compare the results of studies conducted at one simulated site versus three field locations.

PANEL DISCUSSION AND RECOMMENDATIONS

Agency Charge

1) Data Needs

EPA believes that many studies within our current database have limitations. In some cases, the Agency is lacking data to address modern pesticide application equipment and techniques. EPA believes that additional data could significantly improve our ability to estimate and better characterize the range of worker exposure with greater certainty.

Please comment on these limitations and EPA's conclusion that additional data could improve significantly the Agency's ability to assess worker exposure. Also, please comment on the selection criteria proposed by the AHETF and AEATF in their respective submissions for evaluating the extent to which existing data would meet EPA's exposure assessment needs.

Panel Response

The Panel agreed with the Agency's concern about the limitations of the existing PHED exposure database and added a number of concerns of their own. Furthermore, the Panel also concluded that additional data could improve significantly the Agency's ability to assess worker exposure. The limitations of the current PHED database are summarized below.

- Inadequate QA/QC in many of the available data and the limitations of the current data grading criteria to adequately depict the uncertainty within the results (see Appx. B).
- Inclusion of data based on sampling methodologies (*e.g.*, patch dosimeters and outer whole-body dosimeters) that yield exposure estimates with significantly higher levels of uncertainty than data based on inner whole-body dosimeters.
- Inclusion of data with high amounts of data censoring coupled with treating all censored data as LOD/2 and not informing users of the database about the extent of censoring.

- Inclusion of some data based on incomplete dermal sampling (*i.e.*, not from the entire body) that requires a method (algorithm) that combines data from different individuals to assemble a complete dermal exposure value.
- The existence and unknown effect of high levels of "clustering" (a lot of data from one study) within the data that comprise many scenarios (see also Charge #6).
- The short sampling period of some of the data in selected scenarios resulting in more uncertainty when scaling up to full day exposures.
- An inadequate number of measurements in many scenarios resulting in less reliable interpretations and extrapolations.
- A potential lack of “representativeness” due to the absence of modern work practices and equipment in the exposure database (and the possible inclusion of some older work practices that may now be less common or no longer used).
- Lack of diversity of test conditions within some scenarios due to a combination of limited numbers, clustering, the age of the data, and evolving technologies.

The Panel agrees that these limitations decrease our confidence that PHED can reliably estimate exposures for pesticide handlers in all handling scenarios. The ability of such a database to be able to estimate exposure has become an essential part of a regulatory risk assessment that ensures there is a sufficient margin between the likely exposure and the toxicological endpoint of concern. Thus, there is a need to be able do this consistently with a degree of confidence that protects the health of handlers while permitting products that present acceptable risks into the market for the benefit of growers, industry, and consumers. While PHED has served its initial purpose, its goals have evolved. In addition to the above limitations, Panelists elaborated on the following three broad weaknesses within PHED.

The first weakness is the structure of the database and the algorithms necessitated by that structure. These reflect the fact that many of the original data came from studies where either not all parts of the body were monitored or exposures were undetectable. As a consequence of the structure, extrapolation algorithms hidden within the software are often necessary to utilize the data to estimate exposure. While the structure and algorithms may have been expedient at the time, analyses of these data do not represent individual exposures or yield accurate exposure statistics or confidence limits for those statistics. This is true for both the central tendency (arithmetic mean) and higher exposure values (*e.g.*, 95^{th} percentiles), both of which may be of interest for regulatory assessments. The AHED software represents an opportunity to correct this deficiency and to calculate estimates of the mean and higher percentiles and their associated confidence intervals.

The data from the existing PHED presented on p. 33 of the Agency's *Review* is a good (although perhaps inadvertent and not fully explained) example of the weakness of relying on the current data and algorithm used to accommodate censored data. As a result of PHED's use of incomplete and/or censored data from mixed sources using various methods, it predicts that hand exposure without gloves (0.0095 mg/lb a.i.) is equal to or slightly less than exposure with gloves (0.0097 mg/lb a.i). Such a result is completely counter to a logical expectation of reality. Continued reliance on such outcomes may not only lead to incorrect decisions for handler

protection but also weakens the policy of the Agency to require the use of appropriate protective gloves.

Several Panelists encouraged the Agency to adopt more sophisticated statistical methods of dealing with censored data (*e.g.*, MLE) that would yield less statistically biased estimates of the distribution of exposures. In addition (and at the very least) an indication of the degree of censoring should be included in the output of the exposure assessment database. Indeed, there is a strong body of scientific opinion, with much agreement at the international level, that the characterization of both the variability and uncertainty in exposure assessments should be transparent.

The second weakness relates to the actual data. It is clear from the materials submitted to the Panel that the limitations of PHED could severely hamper the Agency's ability to adequately assess pesticide handler exposures. In addition to problems with conducting a reliable exposure assessment using PHED for those scenarios where the data quality (QA/QC) or the number of monitoring units is low, much of the study methodology within PHED is dated. The studies are a minimum of 12 years old, while some are as much as 30 years old. Study designs and the sampling and analytical methodologies for measuring pesticide exposure have improved over the years, resulting in exposure to a greater proportion of the body being measured, vastly improved limits of detection for pesticides in sample media, less censored data (*i.e.*, non-detectable data), and improved overall quality assurance of exposure assessment data. As evidence of the continuing evolution of methods, the Panel discussed relatively recent studies that show the rapid dermal adsorption of some active ingredients and the analyses of probable uncertainty to be described in response to Charge #2, that suggest even some current data within AHED that are based on hand washing should be interpreted with caution.

The applicability of older data to newer ("modern") pesticide application equipment and techniques presents some interesting dilemmas. Clearly, exposure cannot be assessed using PHED in scenarios with new technologies, work practices, or product formulations that are not reflected within PHED. This fact justifies some new studies. But if much old equipment is still in use (a likely possibility given farming traditions), then the old PHED data (despite their limitations) are applicable, have value, and should not be discarded or entirely abandoned. It should be possible to test for a statistical difference between new and old data. One would then need to decide whether such a difference was due either to better study methods or to safer technologies. If the latter were true, the Agency would have to decide whether, for instance, the lower exposure data that might result from the utilization of newer equipment and techniques would not be applicable to the older but still functioning equipment and techniques. Such an outcome might necessitate creating, for example, a new use category and possibly restricting the use of certain high-toxicity chemicals to those newer equipment and techniques that limit exposures.[a] Underlying these dilemmas is the lack of a database describing the actual distribution of equipment and techniques among current users (see further comments on this point in the last portion of this response). The evolution of equipment and work practices within a scenario will also affect our response to Charge #4.

[a] A similar but more vexing problem would result if the data showed that newer equipment and modern techniques actually resulted in higher exposures.

The third weakness pertains to the unrecognized impact that the assumptions, statistical variations, and calculations implicit within each dermal exposure assessment method have on the uncertainty in the resulting exposure values. Assigning a "Data Grade" to each study is a good concept. However, expanding the concept of grading to the end result (the measured exposure level) will help users to better interpret the resulting exposure values as well as create a direct means for the Agency to demonstrate both the weakness of the existing PHED and the improvements that should result from a new database. The only grading criterion that is currently integrated into PHED is the quality of the analytical method (defined within PHED by a combination of "mean % Lab recovery" and the "coefficient of variation (CV) for Lab recovery" or the "mean % Field recovery" or "mean % Storage [recovery] Stability.") While the percent recovery contributes to the uncertainty of the resulting exposure data, Appendix B explains why the CV is the more important of these two parameters. Parts of the Panel's response to Charges #2 and #3 will expand on this same theme to explain why calculations implicit within the sample collection method (*e.g.*, PD versus BM) have a much larger effect on the uncertainty of the resulting exposure value than the CV of the analytic method and should therefore be integrated into the Data Grade. The same principle (with less quantitative measures than CV) could also be applied to other data attributes such as the lack of complete body monitoring, the lack of complete urine collected (if biomonitoring data were included), and the impact of the simplistic method currently used to account for undetectable samples (due to censoring). (If censoring is not integrated into the grading criteria, at the very least the fraction of undetectable samples should be an output to users of PHED and AHED.) These other measures of data quality are likely to be better indicators of the poor quality of the existing PHED data than the current Data Grade based only on the analytical method and could be used by the Agency to help justify the need for new exposure data whose uncertainty can be improved by orders of magnitude.

In 2003 the International Life Sciences Institute, Risk Science Institute, convened an international workshop to consider how to conduct probabilistic assessments of worker exposure to agricultural pesticides. This workshop brought together exposure assessors, modelers, toxicologists, and statisticians. After considering case studies developed for that workshop, it became apparent to participants that the PHED contained so much unexplained variation (likely due to the limitations in the data and mixed study protocols) that the objective could not be achieved reliably. Consequently, it was concluded that more robust representative data are required to attempt to fulfill the objective.

The Panel agreed with the Agency's conclusion with regard to the limitations of PHED and that additional worker exposure data could vastly improve their ability to assess worker exposure, provided that the studies to collect additional data are designed in such a way as to address the shortcomings of PHED. It is important that the new studies be representative of the diverse nature of handlers and use settings, minimize bias, monitor a significant proportion of the handler's working day, use methods that decrease the uncertainty associated with unavoidable extrapolations, and use a sufficient number of monitoring units to allow some meaningful separation of the effects of uncertainty, natural variations, and bias within the results. Furthermore, it appears from the supporting documentation presented to the Panel that the proposed generation of exposure data from AHETF and AEATF is designed to address the limitations of PHED and the needs of the Agency.

A recent Human Studies Review Board review of the proposal to collect new data questioned the need for new human exposure studies, citing that the Agency did not clearly demonstrate the need for new data. This Panel is clearly of the opinion that additional worker exposure data collected on human volunteers under field conditions and label requirements on chemicals that have been approved by the Agency are necessary.

AHETF and AEATF study selection criteria

The Panel has also been asked to comment on the selection criteria proposed by AHETF and AEATF for using existing exposure data to meet the exposure assessment needs of the Agency. While the criteria outlined in documents submitted to the Panel by the AHETF and AEATF appear to be reasonable for including existing data into regulatory exposure assessments, the Panel expressed concerns regarding the selection criteria for future data including those that follow and those regarding "clustering" and "sample selection" discussed in our response to Charge #6.

The monitoring duration requirement may be too stringent. For instance, the requirement to monitor for at least one-half day is unlikely to capture all high intensity and short-term dermal exposures. Some provision to allow the inclusion of data from settings where only short-term uses are the norm may need to be added to the criteria.

The criteria for the use of biomonitoring data may also be too restrictive. The AEATF states in their criteria document that biomonitoring data would only be acceptable if primate dermal absorption and pharmacokinetic data exist for the chemical being monitored. The point is made that extrapolation parameters must be available for the study to be selected by the AEATF. However, there is a paucity of studies that have "extrapolation" factors to humans, and most are estimated from rat and porcine models (*e.g.*, McDougal, 2002 and Williams, 1996). The requirement that only primate dermal absorption data be included in this database may be too stringent. There are mathematical ways of treating rodent and porcine data to make them more applicable to human dermal absorption, namely Fick's law of diffusion and adjusting for the thickness of the skin. At the same time, the use of pharmacokinetic and dermal absorption data to back-calculate from urinary excretion to dermal exposure that can be used in the generic database is likely to introduce a very large amount of uncertainty into the generic exposure assessment, as will be discussed further in Charge #3.

The criteria for air sampling are not consistent. The AHETF states that inhalation data are not required. While it is recognized that dermal exposure generally accounts for the majority of the total pesticide exposure in most scenarios, there may be circumstances when inhalation exposure may contribute significantly to the total pesticide exposure. In contrast, the AEATF is requiring inhalation exposure data if the pesticide is volatile, could be volatilized due to environmental conditions, or if the application method produces "inspirable aerosols." Some objective means should be sought to assure that such circumstances are not overlooked. Perhaps the respiratory data in PHED can be used to provide justification either for omitting or including air sampling (although particular care may be needed to sort the real from the default censored values). The criterion for "inspirable" needs to be clearly defined (as well as the corresponding sample

collection criteria) to deal with the fact that most sprays in fact will contain a wide range of droplet sizes from smaller than 1 μm to over 100 μm droplets, not all of which are "inspirable." It should also be remembered that although large droplets or particles may not be respired into the lung, they may be deposited in the nasal region or mouth where they are available for absorption.

To the degree that patch dosimeters are allowed, the Agency should standardize the array of patches that must be used in the assessment of dermal exposures to pesticides, in particular the location of each patch. The current PHED includes some data that relied on the skills and observations of the researcher to determine the areas of the body to monitor, which introduced the biases of the researcher and a lack of consistency across studies. And finally, some means should be incorporated within the AHED to allow estimating dermal exposures for people of different shapes and sizes from either patch or whole-body dosimeter data.

Other databases

Neither the current PHED nor the proposed AHED or BHED include data that document the distribution of tasks, activities, or pesticide use information within any given exposure scenario. The variance observed within an exposure database is the combined result of statistical uncertainty and imprecision within the assessment methods and natural variability within an exposure scenario. Better interpretation of the observed variability in exposure and the adequacy of an exposure database require (1) measuring the intensity of exposure with consistent methods (which is the only one of the three that should happen within the new proposals), (2) examining the frequency of exposure, and (3) identifying the duration of exposure. The Panel is unaware of any database that contains the latter two important descriptors of exposure scenarios. Although we are not suggesting that a requirement to generate such a database be placed on the task forces, such a descriptive database would greatly increase confidence in any exposure database that will serve as the basis for conducting adequate and reliable risk assessments (see also Appendix C and our response to Charge #6).

Agency Charge

2) Passive Dosimetry

The common approach for conducting dermal exposure monitoring studies relies on the use of whole-body dosimetry, handwashing, and facial/neck wipes. In some cases, biological monitoring is also used as an alternative method. Exposure estimates in Agency risk assessments, however, typically rely on "to the skin" measurements (*i.e.*, potential dose) coupled with dermal absorption data or dermal toxicity studies in order to calculate risks. The Agency believes that these methods are complementary and that they can provide appropriate estimates for exposure assessment but that the results directly relate to the reliability of the inputs used. Please comment on the Agency's conclusion regarding passive dosimetry and biological monitoring, including whether a systematic bias exists in either approach.

Based on the information presented, the Agency has particular concerns over three specific aspects of how these studies are conducted including (1) the possible need to correct for the efficiency of the handwashing technique; (2) compensating for absorption of residues through the skin during sample collection periods; and (3) the breakthrough of residues under whole-body dosimeter garments. Please comment on the need to systematically account for residue losses due to these potential method biases. If there is a need, please describe how these corrections should be accomplished in a way that could reduce uncertainties in the resulting exposure estimates.

Panel Response

This response is divided into four parts. The first three parts each address a "numbered" portion of the charge. A discussion of the complementary nature of biological monitoring [BM] and passive dosimetry [PD] comprises the last part.

1) Is there a systematic bias in either approach?

Bias between dermal exposure monitoring and biological monitoring is not detectable within the data presented to the Panel. Statistical uncertainty is at the crux of this response. Other less comprehensive evidence in the published literature suggests that a bias may exist, but if a bias does exist within PHED, it is much smaller than the uncertainty of the two methods. The level of uncertainty in the estimates of dermal exposure using passive dosimetry will be discussed below; the level of uncertainty in the estimates of dermal exposure using biological monitoring [BM] will be discussed in our response to Charge #3.

Any form of dosimeter will be affected by imprecision within the method used to calculate exposure. The uncertainty in exposure measurements derived from patch dosimeters is due, for example, to the effect of scaling up from deposition onto a patch of *circa* 40 cm^2 to a body-part area of *circa* 1000 cm^2 and the variability of dermal deposition onto any given body part. The interaction of these two factors as defined by Equation 2.1 has the same effect on the probable uncertainty in the resulting exposure value as spike recovery had in

Equation B.1 in Appendix B. (The variability in absorption and metabolism will have a similar effect on the exposure value calculated from urinary excretion, as will be described in the response to Charge #3).

$$\text{Magnitude of Probable Uncertainty} = 1+\left[\frac{\text{CV of Dermal Deposition}}{100\times(\text{dosimeterarea / body area})}\right] \qquad \text{Eqn. 2.1}$$

The variability in the spatial deposition onto the skin of a given handler is generally unknown.[b] Placing two patch dosimeters on each portion of the body only covers between 2% and 4% of each body part, resulting in a scaling factor of 25-50. The use of patch dosimeters and gloves (a "whole body-part dosimeter") means that dosimeters cover about 6% percent of the whole body. Figure 2-1 allows one to see how even a small amount of variability in the spatial pattern of dermal deposition will cause a large probable error when using patch dosimeters.

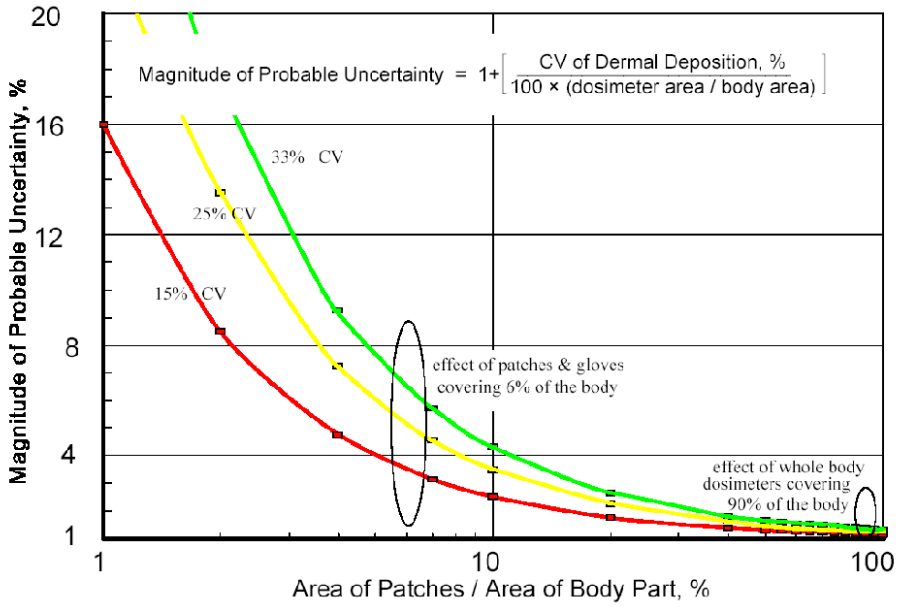

Figure 2-1. The interaction of the CV in dermal deposition and the fraction of the body covered by dosimeters upon the magnitude of the Probable Uncertainty

In comparison, whole body dosimeters cover circa 90% of the body, and even a large variability in the uncovered area (face and neck) can only have a relatively small effect on the probable uncertainty in the resulting dermal exposure value. Any time the overall efficiency of such dosimeters is at least 70% or better, the analyses presented in Figure 2-1 indicate that the uncertainty in the estimates of dermal exposure using whole-body dosimeters is quite a bit smaller than the uncertainty using either patch dosimeters or biomonitoring.

[b] While it is possible to measure that variability by cutting a WBD into small pieces (about the size of a patch) and analyzing each piece separately, such a study is more amenable to research than to a new data requirement.

The implications of these levels of uncertainty will be discussed more fully in Charge #3, but the simple answer to the question herein is that statistical uncertainty precludes the detection of any bias between PD and BM data within the data presented to the Panel. The rationale to believe that a bias may exist in either approach to assess dermal exposure is discussed in the next two subsections of this Response.

2) Is there a need to correct handwashing for its efficiency or to compensate for the absorption of residues through the skin during sample collection periods?

Overall, the Panel believes that if a correction method for adsorption and/or absorption can be derived that can significantly decrease the uncertainty of the resulting measurement at a reasonable cost and within approved human studies guidelines, then it should probably be applied. However, to answer this second question, one has to realize that the skin is more than a filter that just slows absorption. Skin comprises a complex system of layers onto which pesticides may reside, adhere, diffuse, or even be metabolized. Pesticide reaching the skin may first reside on the skin from which it can be washed with relative ease (very analogous to a "dislodgable pesticide residue" on foliage). A pesticide may also be temporarily adsorbed onto the top layer of the skin (the *stratum corneum*) before being absorbed through the skin and into the body. The degree of adsorption and the rate of absorption either far enough into or through the *stratum corneum* where the pesticide is not removable with a weak solvent determine the efficiency of handwashing as an assessment method. Added to these physiologic mechanisms is the varying degree of importance of hand exposure within a scenario and its mitigation by the use of chemical protective gloves.

Adsorption would be expected to be pesticide- (and possibly formulation-) dependent and to be related to the ability of the particular pesticide to adhere to the *stratum corneum*. The fraction adsorbed would be expected to vary based upon the time the pesticide remains on the skin prior to washing, the amount of pesticide that was deposited onto the skin (and perhaps the rate of deposition), and both intra- and inter-personal variability. The efficiency of handwashing would depend upon the ability of the solvent (*e.g.*, water, alcohol, detergent) to remove the chemical from the skin or to promote its absorption into the skin which may vary based upon the physical and chemical properties of the pesticide and the handwashing protocols used.

Existing data clearly indicate that adsorption and absorption of certain pesticides can occur within a matter of minutes after the exposure has occurred. For example, data presented in Fenske and Lu (1994) show that several handwashing solvents recover less than 50% of chlorpyrifos from the skin immediately after exposure and recover only about 20% from the hands one hour after exposure. The handwashing efficiency data summarized in Table 1 of a review by Brouwer *et al.*, 2000 cited in the EPA *Review* range from 23 to 96%. The result ranges from a negligible bias to a four-fold underestimation of the amount of pesticide measured on the skin.

In contrast to hand washing, much of the hand wipe data presented (such as the 10% mean recovery with a CV of 33% for azinphos-methyl in Table 3-4 of the *Review* (from Fenske *et al.*, 1999) and the many chemicals with *circa* 50% mean recovery with a similar CV in Table

2 of the review by Brouwer *et al.* (2000) indicates that hand wipes may be both more biased and more variable than hand washes. The combined impact that lower levels of recovery efficiency and higher variability has upon the probable uncertainty of the resulting measured hand exposure as depicted in Figure 2-1, led some Panelists to recommend that hand wiping not be used in deference to either hand washing or gloves.

The duration of the monitoring interval (p. 38 of the *Review*) appears to be a factor affecting both glove dosimeters and handwashing. An analysis was conducted by one Panelist of the dermal recovery via hand washing data cited in the *Review*'s Table 3-7 for captan and Table 3-8 for chlorpyrifos. In both cases, the efficiency of recovering a single ("spike") deposit onto the hands appears to follow the same exponential model of the form shown in Equation 2.2 that results if one assumes that the rate of adsorption is a constant fraction of the dose to the hands.

$$\%\ \text{Recovery from spike to hands} = [\%\ \text{from immediate wash}] \times 2^{(-\text{time} / \text{HalfLife})} \qquad \text{Eqn. 2.2}$$

Using the above data in this model yields half-lives for the recovery by handwashing of slightly over 2 hours for captan and about 1 hour for chlorpyrifos. Morever, the HalfLife for these two chemicals [in hours] is approximately "5.1/log $K_{O/W}$" with a coefficient of variance of only about ±5%.[c] This type of model was then used to generate a predictive equation for wash recovery following an assumed uniform rate of dermal exposure over the work period, as shown in Equation 2.3.

$$\%\ \text{Recovery from constant exposure} = \frac{\text{Deposition Rate}}{100} \times \frac{\text{HalfLife}}{\ln(2)} \times [1 - 2^{(-time / HalfLife)}] \qquad \text{Eqn. 2.3}$$

Equation 2.3 indicates that if exposure is constant with time, recovery via washing will reach a "steady state" in two to three half-lives. These facts suggest that the interval from initiating exposure until washing the hands can be important if monitoring from the planned minimum of 4 hours to a nominal maximum of an 8-hour day, as proposed by AHETF. Equation 2.3 predicts that waiting this long to wash the hands of handlers of these particular chemicals would result in recoveries that would underestimate exposure by factors of 3× to 5×. Thus, the use of a hand washing technique in AHETF studies should be accepted if it is supported either by a predictive model or by animal or human laboratory recovery data for retention times extending up to the maximum sampling time for the pesticide being studied (see also the Panel's comment regarding primate dermal absorption data as part of the AHETF and AEATF study selection criteria near the end of our response to Charge #1).

Others on the Panel pointed out a weakness to both the modeling and experimental approaches. The accuracy of a correction using either approach could be confounded by the effects of repetitive (multiple) rinsing or washing of the skin of the hands during a work shift for data collection, to attend to personal hygiene needs, or for other purposes. This repetitive rinsing or washing raises the possibility of changing an individual's absorption rate due to alterations in the physiological nature of the outer layers of their skin (*i.e.*, desaponification or protein binding), co-solvent effects, or other mechanisms. Some of these alterations can

[c] Despite such a good fit and the fact that several Panelists noted that $K_{O/W}$ is a part of other models for dermal absorption rate (although here it is only modeling adsorption), an analysis of dermal recovery for only two insecticides each at only two points in time (nominally zero and one hour) is not sufficient to validate this model for all other chemicals.

be expected to enhance dermal penetration rates, causing handwash data to under-estimate exposure, while other alterations could increase the recovery from the skin, potentially causing handwash data to over-estimate exposure. The uncertainty associated with hand washing represents a major topic for research. The above reasons might lead the AHETF to favor the use of cotton glove dosimeters to assess hand exposures in future AHED monitoring studies; however, they are not without their own limitations. [d]

An alternative approach to using one of the methods described above to correct handwash data would be to quantify the amount of absorbed dose based upon excreted metabolites and pharmacokinetic information and add this to the passive dosimetry estimates. Unfortunately, because biomonitoring provides data that are independent of the route of exposure, other routes of exposure unrelated to handling would be included which (if not accounted for) might overestimate the total dermal dose due to the handling task. Also, the biomonitoring approach is not applicable to pesticides that do not have a reliable biomarker or where sufficient pharmacokinetic information is lacking. The burden to the participant becomes larger if they are requested to provide 24-hour urine samples over a period of days, which might bias participant selection in some way.

Further data are needed for better quantification of chemical absorption through human skin *in vivo*. Currently, the Agency uses the "% absorbed" (or absorption factor) to quantify the amount of dermal exposure that is taken up into the body. This "% absorbed" is based on empirical observations for a narrowly defined exposure scenario (see McDougal, 2002). However, the amount of absorption is dependent upon the intensity of the exposure; for larger exposures, the "% absorbed" is less. The reason for this is that only the layer of chemical in direct contact with the skin is available for uptake while the entire exposure is included in the calculation of fraction absorbed. Further studies should be conducted to estimate the percent absorbed for a range of exposure levels.

3) <u>Is there a need to correct dosimetry for the breakthrough of residues under whole-body dosimeter garments?</u>

Breakthrough on a dosimeter could occur either slowly, due to the migration of a collected pesticide residue due to a mechanism like particle filtration or permeation or rapidly, due to the saturation of the dosimeter with a liquid.

Early patch dosimeters had an impervious backing to prevent breakthrough; however, as shown via Figure 2-1, the variability in exposures over individual parts of a handler's body is likely to be sufficiently high in application settings to make the probable uncertainty of dermal exposures calculated from patches unacceptably large. The same analyses depicted in Figure 2-1 showed that the probable uncertainty associated with whole-body dosimeters was much more acceptable.

[d] The limitations of cotton glove dosimeters include the following: a) they are not a surrogate of human skin and in particular have very different adsorbent and absorbent properties than human skin, b) some pesticides cannot be analytically recovered well from cotton, c) breakthrough can happen with gloves, d) it may be difficult for applicators to do some tasks while wearing the gloves, and e) wearing cotton gloves may modify the handler's behavior.

Some researchers used the handler's outer clothing as a dosimeter, in which case breakthrough could occur both from liquid saturation and as a result of the penetration of dry pesticide filtering through the clothing due to air movements. Various penetration factors were suggested ranging from perhaps 5% to 20% with a nominal default value of 10%, implying a default efficiency of 90% with a CV of ±100%. While the correction for this level of breakthrough is slight, the probable uncertainty in the resulting calculated exposure is about ±2×, as would be shown in Figure 2-1 if a CV of ±100% were included.

A whole-body dosimeter placed directly against the skin would minimize the slow penetration due to air movement. However, it would still be susceptible to breakthrough from liquid saturation. Unfortunately, one cannot put an impervious backing on a whole-body dosimeter to protect against liquid breakthrough without creating a heat-stress hazard to the wearer. Moreover, no obvious quantitative criterion for liquid saturation has been identified, and a qualitative criterion such as "visible saturation of the outer clothing or exposed whole-body dosimeter" is probably not reliable. Patch dosimeters placed strategically underneath the whole body dosimeter are unlikely to detect saturation, should it occur, because saturation is unlikely to be uniform, as discussed in part 1 above. Detecting and responding to saturation via biomonitoring is discussed further in part 4 below.

4) Is biological monitoring complementary to dermal exposure monitoring (*i.e.*, dosimetry and handwashing)?

Biomonitoring is possibly one of the few viable approaches available to at least detect if not estimate the amount of breakthrough from a whole body dosimeter, especially breakthrough due to saturation. In theory, the Panel supports the proposal (*Review* p. 61) that "biomonitoring be included in any sampling plan" as a validation that the passive dosimetry collected virtually all of the handler's exposure. However, the Panel also foresees that requiring concurrent biomonitoring could severely restrict study participants to those with no prior (or near-term subsequent) exposure to the chemical, introducing a potentially serious bias in the results. Thus, the Panel recommends that concurrent biomonitoring not be required.

Should biomonitoring be performed and if it were to detect breakthrough, the preferred response is to add the additional dose estimated using a PBPK or another exposure-excretion model to the exposure calculated from passive dosimetry. Discarding all samples associated with a breakthrough greater than some to-be-defined threshold (*e.g.*, 30%) would prevent the inclusion of data suspected to be low, but excluding saturated samples that may not recur in a replacement assessment would bias the data downward. If biomonitoring were to detect breakthough but a correction cannot be estimated, the occurrence of saturation should be reported both within the data entered and to a user of the database's output (much like an earlier recommendation in response to Charge #1 to report the number of undetectable samples).-

If biomonitoring were used as a complement to dermal exposure monitoring, then caution should be exercised in the use of creatinine. Traditionally, creatinine has been used either to correct for dilution in a spot urine sample or to check for completeness of a 24-hr urine

sample. An individual's creatinine excretion rate can vary by age, race/ethnicity, physical condition, and creatinine should only be used to correct for urine dilution for metabolites excreted in a manner similar to creatinine. Thus, creatinine may not always yield an appropriate correction for urine dilution. The use of a PBPK model and two timed-interval urine samples as an alternative to creatinine will be discussed in our response to Charge #3.

Agency Charge

3) Passive Dosimetry vs. Biomonitoring

EPA believes that a comparison of exposure estimates derived from data collected through biomonitoring with data collected through passive dosimetry is the most appropriate way to assess the predictive nature of a passive dosimetry-based approach for estimating worker exposure. Please comment on the strengths and limitations of this kind of comparison for judging the potential utility of passive dosimetry data in conducting exposure assessments.

EPA has conducted such a comparison using available data and believes that the comparison shows sufficient concordance of estimates based on biomonitoring data and passive dosimetry data to support the conclusion that a passive dosimetry-based approach can generate data that can be used to develop relatively predictive estimates of worker exposure for a wide variety of scenarios and activities. Please comment on the adequacy of the analysis to support EPA's conclusion.

Panel Response

A basic schematic flow chart is presented in Figure 3-1 as a visual aid to this response, to depict the relationships between dermal exposure and the chemical at its various steps en route to excretion. This figure tries to depict these relationships for three passive dosimetry [PD] options: skin washing (hands or face/neck), patch dosimeters, and whole-body dosimeters. Each option is represented by a column of boxes and arrows. The top box in each column represents the work activity that results in a dermal exposure. The width of each arrow varies (although not to scale) to represent the amount of the chemical passing from step to step. Thus, the width of the first arrow is the same in each column because it represents the same exposure. In general the width of the arrow decreases as the chemical passes through or around clothing and/or dosimeters, through the skin, and through the metabolic pathways of the body. The amount of urinary excretion potentially measurable via biological monitoring [BM] is at the bottom of each column.

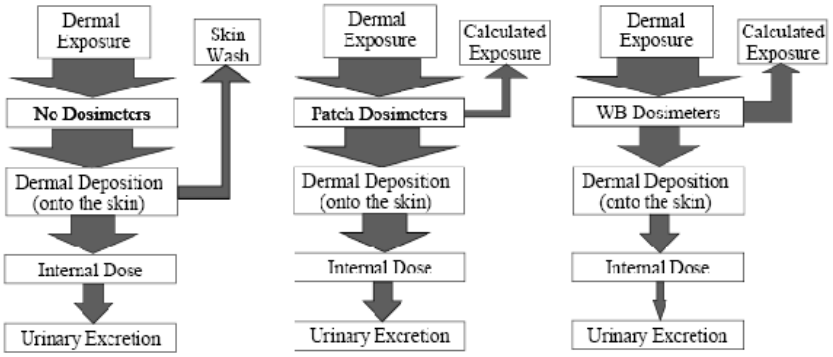

Figure 3-1. Depiction of the decreasing mass of chemical passing through various levels of dosimetry, the skin, and the body to urinary excretion that could be used to extrapolate back to dermal exposure

The first column depicts washing only (typically of the hands and possibly wiping the face and neck, although wiping has been shown to recover less pesticide than washing as discussed in Charge #2). The Panel's response to Charge #2 also discussed the calculations necessary to scale-up the analytically measured amount of chemical removed via hand washing to estimate the dermal exposure. Although hand washing is typically used in conjunction with either patch or whole body dosimeters, the small change in the width of the arrow in the first column of Figure 3-1 that reaches the skin reflects the expectation that washing protected hands will not greatly reduce the total amount of pesticide absorbed and eventually excreted. The second column depicts patch dosimetry in which only a small portion of the pesticide that reaches the skin is captured, chemically analyzed, then mathematically scaled-up to estimate the dermal exposure. Because patch dosimeters cover only a fraction of the skin, they do not greatly reduce the amount of chemical that reaches the skin, the fraction of that dermal deposition that is absorbed to become an internal dose, or the fraction of that amount that is excreted. The third column depicts whole body dosimetry in which (ideally) all of the chemical that would have reached the skin is captured (except typically that reaching the face, neck, and hands), chemically analyzed, and reported as dermal exposure. A good whole body dosimeter should greatly reduce the dermal exposure that reaches the skin, the absorbed dose, and the amount excreted (probably reduced by more than the arrow width shown).

The depictions in Figure 3-1 illuminate two relationships between passive dosimetry and biomonitoring. First, the sequence of events presented in Figure 3-1 causes some forms of dosimetry to reduce the amount of chemical that can be deposited onto the skin, absorbed, and excreted. This interaction between methods would interfere with any expected correlation between the exposure (as represented by PD data) and biomonitoring (as represented by BM data). Second, the large uncertainty in the calculated dermal exposure characteristic of biomonitoring and patch dosimetry will increase the scatter of both variables and decrease the ability of a statistical regression to detect an existing correlation, especially over a relatively small range of exposure levels. A good portion of the Panel's response to Charge #2 discussed the uncertainty associated with passive dosimetry, especially washing and patch dosimetry. A good portion of the rest of the Panel's response to this question will elaborate on the uncertainty of dermal exposure predicted by biomonitoring.

Several Panelists agreed with the Agency (*Review* p. 41-42) that biomonitoring may well allow a more reliable prediction of toxicological risk from the mass of a particular chemical that reaches the body's internal tissues (including the target organ for adverse health effects). However, the goal of the database is to predict exposure of any pesticide used within a scenario. The longer chain of events and the effect of multiple sources of uncertainty depicted in Figure 3-1 between exposure and excretion caused the majority of the Panel to believe that probable uncertainty of dermal exposure values estimated from biomonitoring is greater than similar dermal exposure values estimated from passive dosimetry. In other words, the dermal exposure estimated from the use of good dermal dosimetry (represented by whole body dosimeters and either head patches or a head-neck wipe and either hand wash adjusted for adsorption or gloves) is subject to fewer assumptions and less uncertainty than is the dermal exposure estimated by a back-calculation from urinary excretion. The latter is subject to the uncertainty of inter-personal variations in the rate of dermal absorption, the rate of metabolism and excretion, and the existing body burden from recent prior exposures. Any

variance in these rates or percentages either among participants within a study or scenario or within a participant due to heat stress and/or their work rate (or inter-species differences if applicable) will have a magnified effect on the uncertainty in the calculated dermal exposure. The interaction of the variability from both of these calculations can be interpreted by Equation 3.1 in a manner analogous to the interaction between CV for Lab Recovery and % Lab.Recovery in Equation B.1 and between the variability in dermal deposition and percent of body covered by dosimeters in Equation 2.1. [e]

$$\text{Magnitude of Probable Uncertainty} = 1 + \left[\frac{\text{CV of the \% excreted}}{\text{mean \% of the dermal exposure excreted}} \right] \qquad \text{Eqn. 3.1}$$

The range of mean % of the dermal deposition that would be excreted in the examples cited in the comparison by the AHETF (slide #16 by John Ross and Graham Chester "Comparison of Human Dosimetry and Biomonitoring Data") is 0.18% to 8.9% (the product of "Human Dermal Abs. (%)" and "Excretion Fraction (%)"). The Agency agrees (*Review* p. 42) that the "rate of urinary excretion can vary considerably among individuals for many reasons." A CV of ±10% is probably minimal; the chlorpyrifos urinary data on p. 46 of the *Review* has a mean of 1.3% of the dermal dose excreted but a CV of 65%. As shown in Figure 3-2, the resulting probable uncertainties in a calculated dermal deposition from such urinary excretion data is easily more than one order of magnitude and could approach two orders of magnitude.

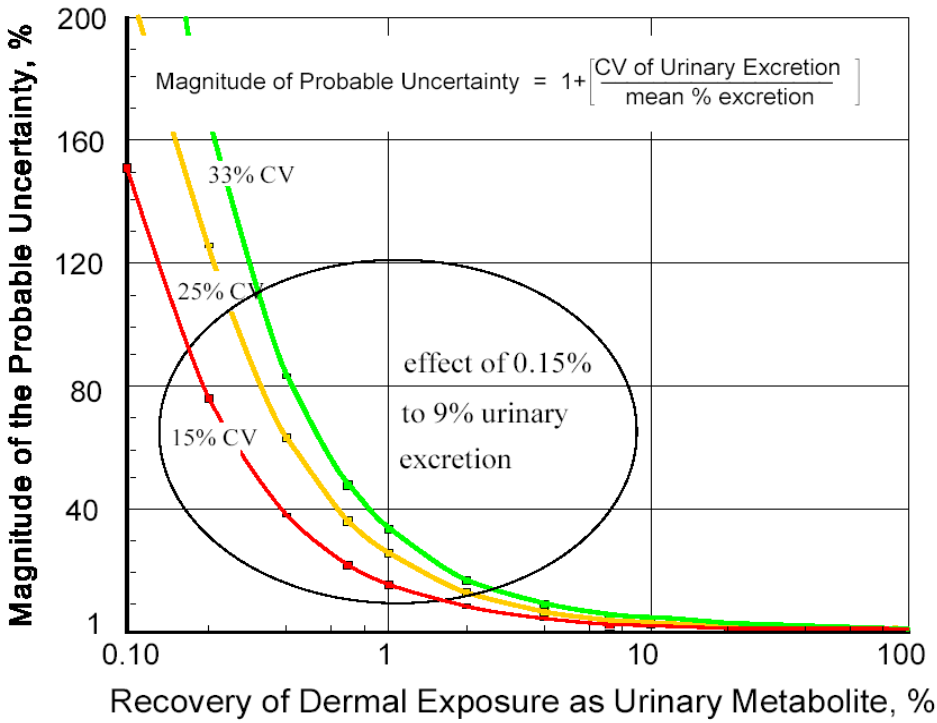

Figure 3-2. The interaction of the CV of the fraction of the dermal exposure or actual dermal deposition that could be recovered in the urine upon the magnitude of the Probable Uncertainty

[e] Technically, Equation 3.1 only approximates the result of a propagation of error analysis of the two steps of absorption and excretion as only one step, but the result is not far off because the fraction of the dermal exposure that is absorbed is generally much lower than the fraction of the absorbed dose that is excreted.

The combination of statistical uncertainties and the potential effect of interferences probably contribute to the ±10× dispersion from the 1:1 regression line for individuals with concurrent monitoring in slide #24 by John Ross and Graham Chester and an even wider range of dispersion in their slide #26. The fact that their slide #22 shows a dispersion of only about ±3× from the 1:1 regression line is probably the result of exposures being based on PD and BM values from independent studies and group averaging. Although some Panelists would have preferred to see the effect of adding all the data (including that rejected for well defined reasons) into the regressions, the above analysis of uncertainties indicates that the agreement in the data presented to the Panel is about as good as can be expected and is sufficient to support the conclusion that a passive dosimetry-based approach can generate data that can be used to develop relatively predictive estimates of worker exposure for a wide variety of scenarios and activities. Another recently published analysis of the well-characterized herbicide 2,4-D by Durkin *et al*. (2004) sponsored by the U.S. Forest Service and EPA is offered as evidence of further agreement between these analytical methods.

The large uncertainties associated with biomonitoring are much, much larger than any of the uncertainties in Figure B-3 associated with acceptable "% Lab. Recovery" values defined by the PHED Grading Criteria discussed in Charge #1 and Appendix B. They are also much larger than the uncertainties in Figure 2-1 associated with whole-body dosimetry, however, generally smaller than the other uncertainties in Figure 2-1 associated with patch and glove dosimeters. For biomonitoring to yield an estimate of exposure with a probable uncertainty close to the *ca*. ±3× dispersion noted above for PD versus BM data, the urinary recovery must be in the range of 20 to 40% of the dermal exposure (values more easily visualized in Figure 2-1 than in Figure 3-2). Thus, while the grouped results of the two methods are in good agreement, the analyses presented in Figures 2-1 and 3-2 indicate that, in general, biomonitoring data do not create as certain a measure of dermal exposure as do passive dosimetry data and that whole body dosimeters are strongly recommended over patch dosimeters.

However from a broader perspective, other divisions of EPA have recently adopted physiologically-based pharmacokinetic (PBPK) models for risk assessment (*e.g.*, for methylene chloride). PBPK models can be used to quantify the relationship between exposure and the absorption, distribution, metabolism, and elimination of a chemical based on two urinary voids within a known interval (rather than requiring full 24-hour excretions to be collected). Such models can also be used to estimate the dose to the target-tissue, for which there is no alternative approach other than conducting invasive animal studies. In fact, statements on page 35 and 42 of the *Review* indicate that the Agency views biological monitoring as a good method "to quantify absorbed dose." "Such an estimate of absorbed dose, which avoids potential confounding from assumptions of dermal penetration or inhalation retention, may be more useful in assessing risk than route-specific doses estimated from passive dosimetry." The technical development of the simplified pharmacokinetic model is ongoing within the EPA Office of Research and Development, and the current model development for chlorpyrifos is relatively mature at this point in time. Therefore, even though we do not currently have validated PBPK models for performing reverse-dosimetry using biomonitoring data, the Panel believes that such models will become available in the future.

Agency Charge

4**) Normalization of Exposure by Amount of Active Ingredient Handled (AaiH)**

The normalization of exposure by AaiH - the unit exposure - has, since the mid-1980s, been the principle relationship underlying the use of exposure data in the Agency's pesticide handler exposure assessments. It is based on the assumption that the two variables are proportional. That is, if one doubles the amount of pesticide they handled or applied, the resultant exposure will be doubled as well.

The Agency is unsure whether the results of our exploratory work showing that proportionality between exposure and AaiH is reasonable in some but not all cases, is a function of limitations of the data within PHED or whether this relationship is in fact not a reasonable assumption for all scenarios. It may be the case that an additional ancillary variable (e.g., boom length, # of tank mixes, or # de-couplings in a closed loading system), in addition to or in place of AaiH, may improve the predictive capabilities of our exposure model.

Though it is recognized that neither the studies in our current database nor the proposed studies by the AHETF were designed for the primary purpose of examining proportionality between exposure and AaiH or to determine the extent to which other parameters influence exposure, compared with our current database, the Agency believes that the proposed AHETF studies will generate data that will reinforce the assumption of proportionality between exposure and AaiH or, alternatively, inform the applicability of another variable as a more appropriate predictor of exposure.

Based on the themes presented on this topic including its historical precedent, its application in risk assessment and subsequent risk management decisions, the Agency's exploratory work using the six PHED scenarios in the case study, and the study design and objectives of the AHETF, please comment on the assumption of proportionality between exposure and AaiH, as a default. Also, please provide comments on whether the proposed AHETF study design is adequate to evaluate proportionality between exposure and AaiH? What other parameters should AHETF study designs measure in order to improve the prediction capabilities of our exposure model?

Panel Response

Most Panel members agreed that the data shown did not consistently support a linear relationship between exposure and AaiH. A linear relationship between AaiH and exposure seems intuitively logical, but a physical rationale should be developed to support that hypothesis (or others) in all scenarios. Several good reasons were given why a linear relationship might exist but not be detectable within the PHED data. Some arguments were presented to accept and/or explain an apparent non-linear relationship between AaiH and exposure. And some arguments were presented that suggest factors other than AaiH that might be better predictors of exposure. The following paragraphs elaborate on these various perspectives.

For exposure within a scenario to be proportional to AaiH implies that a consistently small fraction of the amount of pesticide that workers handle is deposited onto their skin. It seems logical to expect exposure to be proportional to AaiH under certain circumstances (such as open cab airblast applications using similar application equipment moving at similar speeds and in similar wind conditions). However, a description of the physical mechanism that dilutes the *a.i.* handled into a small but constant fraction of the AaiH that actually reaches the handler in each scenario has not been made. Developing a written array of hypotheses based on physical mechanisms applicable to each scenario would seem like a good place to start.

In theory, such an argument is merely an extension of the finding that absorbed pesticide dose increases proportionally with exposure, as depicted in Figure 4-1. Data obtained for three pesticides [(atrazine (Lu, *et al*., 1997), diazinon (Lu, *et al*., 2006), and chlorpyrifos (Lu, *et al*., 2007)] clearly demonstrated the proportionality between exposure and dose. By extending the exposure-dose continuum to the left in Figure 4-1, the Agency is extrapolating the proportionality as seen in the exposure-dose relationship to the one involving AaiH and exposure.

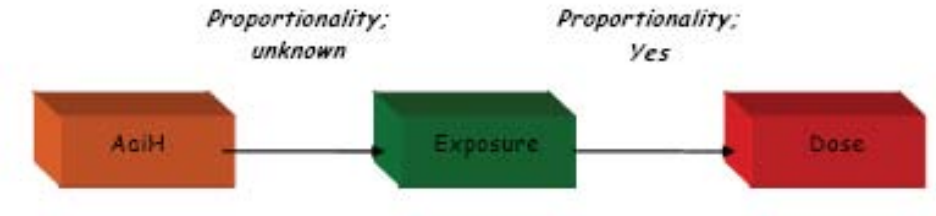

Figure 4-1. The proposed continuum of the amount of active ingredient (pesticide) handled [AaiH] to the exposure and to the absorbed dose

In practice, according to the examples from the PHED database presented to the Panel, the assumption that exposure is proportional to AaiH is only valid in a few scenarios but is more often invalid. Three reasons were proffered by the Panel to explain why such a proportionality was not found in the PHED data: an "ecological fallacy," an "engineering fallacy," and the statistical uncertainty intrinsic to the experimental data.

The "ecological fallacy" is the mistaken assumption that all members of a group have the same characteristics as the group at large. The analyses provided by EPA contained several examples of a proportional relationship being observed within a scenario but not when examining each study within that scenario separately. Such a pattern could be caused by some of methodological or study design differences described in Charges #1 and #3 or by unique conditions such as clean-up and repair activities encountered or assessed in some but not all studies. If unit exposures are going to be applied in a risk assessment/management context, it then should hold across most if not all studies and scenarios.

The reverse may also occur, where proportionality applies within a closely defined portion of a scenario (such as within a particular cluster); however, the proportionality fails to apply across a wider range of work practices or equipment within that same scenario. Such a finding may constitute what another Panel member termed an "engineering fallacy." For example, ground boom sprayers may range from small or almost "antique" machines to state-of-the-art modern

large self-propelled machines with induction bowls, clean water supplies for washing gloves before removal, glove lockers, automatic folding booms, and with the operator and the equipment controls positioned in closed air-conditioned cabs. If the unit exposures are derived from "antique" machines, we might expect the value to be over-protective in the case of modern equipment. However, if the unit exposures are derived from modern equipment, they would not be adequately protective of the smaller, less advanced equipment. And if the unit exposures are derived from a mix of antique and modern equipment, they are unlikely to support a proportionality assumption across the full range of AaiH. Therefore, it is important to consider the specific circumstances used to generate the data in some detail.

Yet another Panel member pointed out that the failure of the regressions of PHED data presented to the Panel to support a strictly linear proportionality between exposure and AaiH should not necessarily lead to rejecting the existence of such a linear hypothesis. This view is justified by the results of the analyses of the probable uncertainty presented as part of our response to Charges #2 and #3. These analyses indicate that the probable uncertainty in exposure values derived from patch dosimetry or biomonitoring data were *circa* ± one order of magnitude. Thus, rather than viewing each individual pair of data as a point in a scatter plot to be tested for correlation, the data would be better portrayed as an array of vertical lines ("lines" because the probable uncertainty of the AaiH values are virtually zero or perhaps ±1% at most). Given the imprecision of the Y values for regression, the necessary range of X values (AaiH) must be at least 100× to yield a truly significant correlation, as proposed by the AHETF study design and addressed further by the Panel in Charge #6.

In contrast to the above justifications not to reject proportionality, several arguments were presented to accept a nonlinear relationship between exposure and AaiH.

Several Panel members questioned the need for strict adherence to proportionality. For instance, in PBPK modeling (and in models for many other fields), nonlinear scaling laws are determined empirically by estimating the regression coefficients from a log-log regression analysis, e.g., cardiac output is proportional to body weight to the 0.75 power. Similar nonlinear relationships might apply to using AaiH to predict dermal exposures to pesticides. To apply this viewpoint, the log-log regression coefficients that give the best fit of the relationship between AaiH and exposure should be used, whether or not they are equal to 1. In a scenario where an increase in the amount of active ingredient handled is expected to result in an increased exposure, whether the ratio is 1:1 could depend upon the amount of technology used when handling pesticides as well as other factors that one may or may not be able to control.

Physical mechanisms may also predict or be used to explain nonlinear relationships. For instance, it may be notable (or given the probable uncertainty in the exposure values previously described herein, the following observation may just be serendipitous) that in virtually all of the "other-than-hand" examples presented in the background document (p. 90-97), the slope of both outer dosimeters was less than 1 while the slope for the inner dosimeters was greater than 1. One can rationalize this bracketing of 1 (unity) if one assumes that the AaiH per unit of time are approximately equal within each of the scenarios but that the transfer of pesticide from the outer clothing to the inner dosimeters is delayed in time (as has been observed in some experimental

field studies). In this case, the longer duration scenarios (which also handled more a.i.) would appear to have more than a linearly proportional dose, *i.e.*, a power greater than unity.

And finally, some Panelists took the position that at least in certain scenarios, no correlation should be expected between AaiH and exposure, and went on to suggest other factors that might be better predictors of exposure in these situations.

In some scenarios, discrete events might contribute the predominant fraction of a handler's total exposure. One such situation is where exposure results mainly through contact with a contaminated surface or with pesticide residues on surfaces that are at steady-state or saturated, regardless of the AaiH. For example, a mechanical transfer device used for loading liquids might limit exposure to concentrated residue left on the dry-break coupling which is independent of the amount transferred (but of course is affected by the concentration of active ingredient in the formulation or tank mix). Another example might be a situation where the user is protected in a closed cab when making ground boom applications so that exposure mainly occurs when handling contaminated boom / nozzles and the outside of the cab (door handle), *e.g.*, (Kline *et al*., 2003). The majority of a mixer/loader's exposure might occur each time they handle a bag of a solid formulation (independent of the volume of that bag or the amount (fraction) of that bag they actually dispensed), each time they open a can or jug of liquid formulation (independent of the volume dispensed), or at the moment that the concentrate is added to the water diluents (independent of the volume or mass involved). Under all these scenarios, exposure may be proportional to the number of discrete events but may not be proportional to the AaiH.

The previously mentioned short documents that put forth a plausible argument based on physical mechanisms might also identify other potential predictors of exposure. For instance, a time-based rate of the activity (*e.g.*, speed or rate of application such as AaiH/hour) might be a predictive factor when the applicator is moving away from the immediate, relatively localized point of application such as a horizontal boom (*cf.*, an airblast sprayer), an airplane, or possibly even if on foot (as in a hand wand application). The swath of a ground boom might be another correlate because it determines the distance from the applicator from which part of that AaiH is released. A combination of application rate and ventilation rate (and not necessarily the amount applied) might underlie the ambient dilution of an aerosol spray used indoors.

Looking forward, the Agency's desire to investigate the proportionality between exposure and the AaiH in future studies is warranted. The AHETF also has recognized the need to verify the "accepted" assumption that AaiH has a 1:1 relationship with exposure. To test this hypothesis, studies should be conducted to evaluate the relationship of exposures under variable AaiHs while controlling for confounding variables. If one is able to spread out the amounts of ingredient handled within each cluster, then one should be able to estimate the relationship (whether linear or not) between exposure and the amount of active ingredient handled for each scenario, and use this information to more accurately estimate exposure for a given scenario/activity. If the cluster-based sampling strategy is used, then the AHETF's analysis clearly shows the desirability of obtaining the maximum within-cluster range in the amount of a.i. handled. Achieving a range in the AaiH as high as 100 fold will have the effect of reducing the needed sample size as compared to a 10-fold difference. However, given the AHETF's proposed threshold of 5 lbs a.i., the 100-fold difference may be achievable for some pesticides (*e.g.*, certain herbicides applied by

ground boom to row crops) but not for others. To the degree that the AHETF plans to assess the correlation of exposure to other variables such as the number of acres or the duration of mixing or application or the number of events such as contact with contaminated surfaces or tank mixes, the degree of potential cross-correlation between these variables and AaiH should also be considered. Such cross-correlations may preclude a strictly observational study from illuminating all other significant predictors. In the meanwhile, the Agency is encouraged to develop an array of short documents that put forth a plausible argument based on physical mechanisms that would justify either using a default assumption of a linear relationship between exposure and AaiH and could help in designing "purposive non-random sampling" to support other predictive correlates of exposure.

In conclusion, Panel members agree that a great many factors are associated with field conditions, *e.g.*, application techniques, equipment types, meteorological conditions, formulations of pesticide. A well-designed observational study such as that proposed by the AHETF may illuminate the relationship between exposure and AaiH, but a controlled experimental study beyond the scope of the studies envisioned herein may be a better way to ascertain this relationship. Given this dilemma, the Agency may wish to consider establishing its own criteria for the strength of evidence needed either to accept or to depart from the existing default assumption of direct proportionality. In addition, the AHETF may wish to consider a more strategic allocation of their resources by focusing their efforts on fewer exposure scenarios but employing sufficient monitoring units to establish a lack of proportionality with greater certainty.

Agency Charge

5) Within-worker and Between-worker Variability

The proposed AHETF study design does not include true worker replicates and is not intended to examine the issue of variability within workers. The AHETF notes that to appropriately investigate this issue would require significantly more sampling and resources. They propose, however, that their single-day exposure distribution results can be used to evaluate longer term multiple day exposures by placing reasonable limits on expected intra-class correlation coefficients (ICC): they indicate that, from their own research and review of the literature, the ICC is likely to be between 0.3 and 0.5 over relatively short periods of time (*e.g.*, seasonal), and likely to be even lower over longer periods of time.

Please comment on the AHETF's approach to estimating the number of samples (MU) needed to determine within worker variability and their conclusion on the importance of measuring such variability in their proposed studies.

Panel Response

The Panel agreed that exposure data collected from observational studies has the potential to address all three potential sources of variation identified in the background documents: within-handler, between-handler, and between-study. Within-handler variability is defined as the variation among different measurements on the same individual doing the same or a similar task under similar environmental or other conditions (or what is referred to as "repeated measurements" in the background document). Between-handler variability is defined as the variation among different individuals doing the same or a similar task, possibly under the same but typically under differing environmental or other conditions. Between-study variability is defined as the variation among different individuals doing the same or a similar task under different environmental or other conditions but at either different locations (separated by miles rather than meters) or times (separated by days rather than hours). Often between-study variability is confounded with between-handler variability since researchers may have to go to different locations and/or different times to find handlers doing the same task. It may also be confounded with within-handler variability if the same handler is measured at multiple time points or locations. A necessary and basic component of any quantitative risk assessment is a good measure of the variability expected from independent handlers doing the same or a similar task under similar conditions. In some assessment scenarios, the variability term required may be the sum of between-individual and between-study variability.

The term "repeated measurements" may have a different meaning for different researchers. In statistics, a repeated measurement would occur if one unique handler were to do the same task multiple times and his/her exposure were measured separately for each repetition of the task. Other researchers use the phrase repeated measurements to refer to more than one measurement of a task, regardless of whether the measurements are on one or many individual handlers. Still other researchers may think of repeated measurements as different tasks measured individually on one handler. The concern in all of these definitions is that measurements cannot be considered truly independent. The issue is further confounded if the less defined term

"replicates" is used, although for many, replicate measurements imply independence of responses.

The important issue for design and analysis of exposure studies is the potential for exposure measurements to be correlated. The responses from different handlers doing the same or similar task are often assumed to be independent (uncorrelated), that is, the exposure of one handler is not expected to affect or be affected by the exposure of another handler. Within-handler measurements on the other hand are typically assumed to be correlated. The "simple sample variance" computed from all applicable exposure measurements will be an underestimate of the true risk-related variability unless the expected correlation of measurements in the data is taken into account in the estimation methodology. If the database consists of only uncorrelated measurements, as would be the case where within-handler data were specifically excluded, then the simple variance would be an acceptable estimator of variability for the risk assessment.

The AHETF proposal argues, fairly strongly, that the within handler source of variation is unimportant and/or too expensive to measure given the objective of the resulting data to support benchmark or minimal adequacy requirements for Tier I and Tier II risk assessments. The proposal also suggests that measurements be taken in studies (clusters of measurements) that are linked to specific locations and times. This design can also result in significant but moderately-sized correlations among within-study measurements. The concern with within-study correlation is that handlers doing the same or similar tasks at one site and time may produce similar exposure values because the measurements are taken under common environmental or other conditions. The measure of similarity used to quantify within-study variability is the "intra-class correlation due to clustering" (referred to herein as the ICC, *cf.*, the intra-class correlation coefficient in the charge) and the range of interclass correlations for measurements taken over short periods of time was reported in the background documents to be between 0.3 and 0.5.

The true model for MU exposures (Equation 5.1 below) is a modification of Equation (1) in the Procedures for Determining the Required Number of Clusters and Monitoring Units per Cluster to Achieve Benchmark Adequacy (AHETFb, 2006).

$$Log\left[\frac{E_{ijk}}{H_{ijk}}\right] = LogQ_{ijk} = LogGM_{Q} + C_{i} + W_{ij} + R_{ijk} \qquad \text{Eqn. 5.1}$$

where

E_{ijk} = the exposure obtained for MU j in cluster i in repeated measure k
H_{ijk} = the amount of *a.i.* handled for MU j in cluster i in repeated measure k
Q_{ijk} = the exposure for MU j in cluster i in repeated measure k normalized by amount of *a.i.* handled
GM_Q = the population geometric mean for normalized exposure
C_i = a random effect for cluster (study or condition) i
W_{ij} = a random effect of MU j in cluster i
R_{ijk} = a random effect of MU j in cluster i for repeated measure k

In this model, all three random effects C_i, W_{ij} and R_{ijk} are assumed to be normally distributed with means of zero and variances of V_C, V_W and V_R, respectively. Note that the distribution of random effects could alternatively be parameterized using a total variance term [V], an intra-

class correlation due to clusters term [ICC], and a within-handler correlation term [R_{ww}]. The formulation reported in Equation 5.1 is to clarify potential confusion that might exist about the definitions of the ICC and the R_{ww} terms.

It was pointed out that regardless of the source of the data, within-worker variation will always be confounded with errors in the monitoring technique and the chemical analysis (referred to as "probable uncertainty" within responses to Charges #1, 2, and 3). That is, one can never really measure the true residual error.

The Panel felt that the AHETF arguments to de-emphasize within-handler variability in section 5.3 (AHETFa, 2006, Technical Summary Document) are clear and compelling. In particular, the AHETF argues that:

1. The combination of the probable uncertainty inherent in exposure measurements and the typical influence of uncontrolled environmental factors on the measured exposure would result in repeated measurements that would be expected to demonstrate low correlation or R_{ww} values between 0.2 and 0.4 (page 19 of AHETFa, 2006). The argument for low R_{ww} values is derived from limited published literature and not on an analysis of relevant PHED data.
2. The between-handler data which will populate the AHED database is expected to support Tier I and Tier II risk assessments that focus on cumulative exposures over long time periods. The distribution of individual long-term cumulative exposures will be best described by the between-handler distribution regardless of whether the R_{ww} is 0 or 1.
3. The between-handler data distributions could be used to simulate both within-handler and between-handler variability in any probabilistic (Monte Carlo) risk assessment by specifying and drawing from a distribution of R_{ww} values such R_{ww} is between 0 and 1.

In addition, the Panel noted that expecting to conduct repeated measurement on each handler would constrain the eligibility of handlers to participate, thus introducing a selection bias.

One Panel member estimated the within-worker correlation coefficient using the repeated measures data presented in Figure 5-1 of the EPA *Review* document. Variance components were estimated using a one-way random effects model with 10 individuals having from 2 to 6 repeated measurements for a total of 39 observations. The within-worker variance component was estimated as 0.38, and the between-worker variance component as 2.5, resulting in an estimated within-subject correlation coefficient of 0.9. This estimate was significantly greater than 0, the value assumed by AHETF. This result supports the conclusion that little additional information would be gained from repeated sampling of one individual.

While there was little interest among Panel members in increasing dramatically the total number of measurements taken per scenario by requiring repeated measurements for every handler in every scenario, there seemed to be some concern that EPA might be missing an opportunity by not pressing or investing in some limited repeated measurements.

Current literature on within-handler correlation is small and problematic. The Panel suggested that an analysis of the within-handler data currently within PHED could serve as a starting point for understanding within-handler correlation. For instance, it would be prudent to attempt to estimate R_{ww} using PHED data (as mentioned above using Figure 5-1 of the EPA *Review*

document) for different exposure scenarios. The analysis would be purely exploratory, providing some evidence to support the assumption that the within-worker correlation is relatively small and/or bounded. Additional data might be necessary to provide evidence-based justification for limiting the range of R_{ww}, something that might be needed if indeed the AHETF approach is used to incorporate within-handler variability in a future assessment. Any probabilistic risk assessments will want to incorporate both within-handler and between-handler variability. The AHETF approach is always going to be weaker than an approach that is based on estimates that are backed by actual data.

Those Panel members with in-the-field experience performing the types of studies being considered suggested that there are significant challenges to getting good data and that requiring repeated measurements can result in large increases in both effort and analytical costs. Others felt that adding limited repeated measurements should be relatively cheap, especially when compared to the cost of starting a new study (new location and time) and or recruiting new handlers. Repeating measurements on the same or next day would provide the most information for the least additional investment. The need for information on the costs associated with the various aspects of sampling was pointed out a couple of times in the discussion. Cost information would help to better inform the decision on repeated measurements. There are statistical techniques that can be used to adjust sampling efforts among the three variance components to essentially get the most precision for the least total cost.

There was also recognition among Panel members that the time window for pesticide application is often narrow in agricultural situations and that the number of tasks per worker per pesticide per year may be few. These constraints substantially limit both the number of sampling opportunities and the number of eligible workers, especially within the same geographic region (since all applicators are usually working within the same time window), making repeated measurement studies infeasible for some scenarios. Given these pressures, the Panel recommends that unless repeated measurement data are specifically allowed and properly handled within AHED, in cases where a participant withdraws from a study that a new worker be recruited rather than using a previously sampled worker.

The discussion on cost and timing did not reduce some Panelists' interest in seeing repeated measurements performed in at least a couple of important scenarios. Although a repeated measurements sampling strategy may not be possible for all exposure scenarios, the Panel recommended that EPA and the AHETF determine whether within- and between-worker variability might be evaluated for selected scenarios where application frequency and logistics are favorable. As an example of a scenario where repeated measurements might be important, one Panel member suggested orchard programs where repeated chemical applications are often performed every 7 to 10 days by the same handler. One Panel member noted that when EPA and other researchers use these data to examine potential predictive determinants of exposure, the best data for identifying predictive exposure factors is to have measurements taken under different levels of the suspected factor on the same individual (*i.e.*, to use the individual as the block or its own control). Thus, some limited repeated measurements on handlers in the database could result in more powerful identification of predictive exposure factors.

Repeated measurements data need not be balanced for subsequent data analysis when using modern mixed model software. Although analyses by such mixed model statistics can estimate within handler variability, between-handler variability, correlations, overall average exposure, and associated confidence intervals, such software (or this ability) may not be compatible with the type of software used for PHED or being contemplated for AHED/BHED. Thus, the presence of just a few workers with repeats in the database might raise practical data management and statistical analysis issues. A number of questions will need to be addressed in the AHED interface to ensure that exposure estimates and uncertainties are valid. These questions include: Should the repeat measurements be considered independent? Should a mixed model be used to estimate overall average exposure and associated confidence intervals? Or should the repeats be removed (or automatically averaged) and simple data analysis techniques be used?

One Panel member argued strongly that repeat exposure studies on the same and different handlers are needed to identify the biological differences between and within handlers that are important in interpreting biomonitoring data (were such BM data to be collected). Differences in metabolism, body weight, age, gender, BMI, and ethnicity can account for much of the between-individual variability in biomonitoring data. But this variability will need to be gauged against the within-individual variability estimated using measurements from task replicates on the same handler. Repeated exposure studies will also support an understanding of the relationship between passive dosimetry and biomonitoring results. That Panel member suggested that based on the estimated R_{ww}, only a few repeated samples; say 2 or 3 would be sufficient to provide this understanding.

The Panel discussed extensively those factors relating to repeated measures that could give rise to ethical or other issues at a Human Studies Review Board review. Risk would be increased if a handler were asked to do something they would not normally do, use pesticides that they would not normally be handling, or use amounts of pesticide that they might not usually use. Some opportunities for repeated measurements may not add risk, but further justification to HSRB would be needed if the handler were asked to do the same task multiple times if it were something that might be typical for the scenario but that with timing and or amounts would not be typical in that handler's normal work assignment. Another issue was the use of scripting to achieve repeated measures. Scripting takes the handler outside of his/her usual mode of work and hence has the potential to change risks. The Panel suggests that these issues be carefully addressed by the AHETF. Although unrelated to repeated measures, some of the measurement techniques such as the use of whole-body dosimeters, create a burden on the handler not typically encountered in a set of tasks. This additional burden might be of concern to the Human Studies Review Board.

Agency Charge

6) Sample Size: Number of Sites and Subjects per Scenario/Activity

The Agency's goal is to ensure that monitoring studies rely on sample sizes that adequately represent the range of exposure of people who engage in a particular handler scenario and activity. It is also recognized that occupational monitoring studies are costly and have many logistical obstacles. The Agency is also concerned about limiting the numbers of participants in these types of studies in accordance with the ethical requirements described in Subpart K (40CFR26) and the recent criteria outlined by the Agency's Human Studies Review Board. The Agency's current guidelines recommend 15 monitoring units for each scenario. In addition, the AHETF has provided a rationale for the number of samples in their study design.

Please comment on the uncertainties associated with the Agency's and AHETF's recommended number of monitoring units. In your comments, please include any recommendations you may have regarding specific statistical analyses that may assist the Agency in developing better understanding of these uncertainties and characterizing them in a complete and transparent manner in Agency assessments based on these data.

Panel Response

To design a monitoring program that may be used for a variety of regulatory purposes by various organizations challenges the developers to anticipate all possible future applications while keeping costs in line. The Panel agrees with the background criteria given in the first paragraph of the charge above. The Panel also agrees that the current PHED database needs to be updated and modernized. Our response to this charge is divided into four main parts. The first part comments on the uncertainties associated with the Agency's and AHETF's recommended number of monitoring units. The second and third parts comment on the planned clustering within the study design. And the fourth part comments on sample selection and bias within the study design. Several comments made during discussion regarding statistical analyses are in Appendix C or were covered herein in our response to Charge #4.

Recommended Number of Sampling Units

The appropriate number of Monitored Units is integral to the goals of database users. The Panel noted that the benchmark objectives for data adequacy as established by the AHETF, listed on slide 17 in the presentation by Larry Holden entitled "Summary of Statistical Issues for the AHETF Monitoring Program: Sampling Methods and Sample Sizes," may not support the goal of the Agency, stated in the above Charge, to "adequately represent the range of exposure of people who engage in a particular scenario and activity." The AHETF's primary and secondary "benchmark objectives" for data adequacy were to meet for all exposure scenarios a degree of data accuracy to within K-fold when exposures were normalized by amount of active ingredient (*a.i.*) handled, with *K* proposed to be 3; and for some scenarios, users of the data should be able

to "distinguish" between complete proportionality and complete independence of exposure and amount of *a.i.* handled, respectively.

The first benchmark objective is to estimate the parameters of the distribution of dermal exposure to an adequate level of precision. The criterion chosen (that the upper 95% confidence bound for the parameter be no more than *K* times the parameter and the 95% lower confidence bound be no less than the parameter divided by *K*) makes sense under the lognormal assumption. A closely related criterion, giving similar results, is to require that the upper 95% confidence limit be no more than K^2 times the lower 95% confidence limit (K above times K below equals K^2); this might be easier to communicate and has the advantage of not requiring the true parameter value to be explicitly in the formula.

The Panel also discussed the need to think strategically about the allocation of resources and to establish sampling priorities for scenarios. We were told that regulatory personnel have not had difficulty in specifying what, for them, would be an acceptable value of *K*. A default value of *K* = 3 seems reasonable, although it need not be a fixed value. Scenarios with higher exposure might warrant allocation of more MUs. Alternatively, the *K* value could be larger for a scenario with a larger than typical MOE, permitting an acceptably smaller sample size. In addition, if EPA has a strong need to estimate exposure levels in the upper tail of the exposure distribution (the 90^{th} percentile for example), more samples will be required than suggested in the background documents.

A similar cautionary note would apply if the EPA uses the 95^{th} percentile for risk assessment in the future, in which case the AHETF's analyses indicate that under cluster-based sampling the sample size required to estimate the 95^{th} percentile within a 3-fold accuracy is quite sensitive to the specified GSD and ICC. If the actual GSD or ICC is greater than anticipated, then much larger sample sizes will be needed to achieve the desired accuracy. EPA and the AHETF should consider building in one or two "check points" after a certain amount of new data is collected to evaluate assumptions about the GSD and ICC so that any needed refinements to scenario sample sizes can be made. The examples shown to the Panel assumed that the most extreme upper percentile of exposure that anyone would want to estimate was the 95^{th}, in which case K = a 3-fold relative accuracy can likely be achieved with 5 clusters and five monitoring units per cluster. This means 25 monitoring units per scenario. However, this sample size will be inadequate if at a future time it is necessary to estimate the 99.9^{th} percentile. One Panelist noted that it appears that 11 or 12 clusters would be needed to achieve K = a 3-fold accuracy for the 99.9^{th} percentile. Thus, the total number of monitoring units would be more than doubled to 55 or 60.

One Panelist gave three reasons why it may be advisable to have at least 50 monitoring units per scenario: to estimate upper percentiles of exposure, to make effective comparisons among scenarios, and the possibility of measuring within-worker variation. This Panelist thinks that the value of the database will be greater in the future if costs are controlled now by making a thoughtful choice of scenarios in which to sample heavily rather than by using small samples in all scenarios. If at a future date a sample size is found to be inadequate for regulatory purposes, it will be impossible to return and get more observations that are consistent with the original sample. It will be much easier to do a complete study of new scenarios as they are needed.

If the data will be used to compare scenarios (*e.g.*, to compare different application methods with the same pesticide), then the design needs to be considered more as a stratified sample and there have to be enough observations within each stratum to make the test powerful enough to be worthwhile. If the sample size meets the first benchmark objective, it may also be good enough for this objective, but it would be worth checking this out.

The Panel also noted that all of the sample size calculations presented are dependent upon the assumption of lognormal distributions, and any attempt to estimate extreme upper percentiles from a small sample will be an extrapolation into the tail of the assumed distribution. The larger the sample, the more robust will be the ability to validate that assumption and support the conclusions. To understand the robustness of the cluster design (see below) to the assumption of lognormal concentration values, the Panel recommended that sample size simulations be performed using an alternative skewed distribution for concentration values, such as a Gamma distribution.

Clustering within a Scenario

The discussion in the Procedures for Determining the Required Number of Clusters and Monitoring Units per Cluster to Achieve Benchmark Adequacy (AHETFb, 2006) was relatively straightforward, clear, proper, and representative of good statistical thinking. The Panel compliments Dr. Holden on creating a clear conceptual model for the sampling process and following it through to the particulars of the sampling design. The cluster sampling design proposed by the AHETF makes good sense, as there are cost savings in sampling a number of pesticide handlers in a single field operation.

The Panel notes that if one is going to fix (set) the total number of monitoring units, then it is generally better to have more clusters within each scenario/activity and fewer numbers of handlers within each cluster than it is to have fewer clusters and more handlers per cluster. The usual practice in survey design when there is an intra-class correlation within clusters is to consider the costs of getting to a cluster relative to the costs of sampling monitoring units within a cluster. The optimal cluster size and the number of clusters can then be chosen to minimize the variance of a parameter estimate subject to constraints on the total cost. For this study, AHETF has determined from experience that the intra-class correlation is modest in size and that it is often practical to monitor five pesticide handlers within a cluster.

The Panel has no problems with the values of geometric standard deviation and intra-class correlation used in their examples. However, the variation in intra-class correlations observed to date comes from sparse data and variability in monitoring methods and can't be attributed to specific scenarios. The Panel would expect that as more monitoring data are collected, it will become evident that some scenarios may have very different intra-class correlations from others, and adjustments in the numbers of clusters and monitoring units within clusters that are sampled may need to be considered (see prior comment on "check points").

The Panel also noted that for a variety of reasons not all clusters are likely to come in neat units of five MUs. Similarly, it may be very difficult to assess each scenario with only five clusters for each scenario. The science will be greatly served by not requiring five clusters for each

handler task. The AHETF might explore simulations where the average cluster size is five but with some variability in cluster size to assess the robustness of results with respect to cluster size.

Given that analyses presented by the AHETF indicated that greater sample size efficiency was generally achieved by increasing the number of clusters rather than increasing the number of monitoring units per cluster, clearer guidelines are needed for cluster selection so that the addition of new clusters will achieve some desired degree of dispersion on the variable of interest and to guard against "parsing" the population of interest into more clusters just to limit sample size requirements without improving stratification or representation.

To summarize the above comments, the Panel believes that the Agency's and AHETF's recommendation to have 5 handlers per cluster with approximately 5 clusters for each scenario/activity is reasonable at this point in time. The most important issue for the Agency to consider is what value of *K* for a *K*-fold accuracy is appropriate and reasonable.

Cluster Selection

The Panel noted that the reports being reviewed had information about what data will be collected, but there was not much information about how the data will be collected. It is important to consider such questions as:

- How will the clusters within each scenario/activity be selected?
- How will the monitoring units within a cluster be selected?
- Will AHETF have control over the amount of ingredient handled?
- When will the data be collected?

With respect to the first two bullets above, the Panel suggests that the definition of a cluster and how clusters will be selected be clarified and tightened (see also the above comment on "parsing"). For example, clarify whether clusters are defined by crop, state, county within state, crop-state combination, geographic region, *etc*. Would the cluster definition need to be scenario dependent? Also, the Panel suggests that the AHETF provide additional evidence to support the notion that "geographic differences" are important for establishing clusters.

Sample Selection and Bias

With respect to the third bullet above, there was some concern expressed that the sample selection was being adversely affected by the secondary benchmark objective proposed by AHETF to be able to elucidate predictive relationships with exposure such a proportionality with AaiH. Their "Technical Summary Document" (p. 17, 21, and especially p. 40-41) describes their plan to use "purposive non-random sampling to achieve a diversity of major factors likely to influence exposure" which they list as "the amount of active ingredient handled, number of unique workers, and number of geographic locations." These concerns start with feasibility but extend to bias and the ultimate value of the database.

With regard to feasibility, their page 41 is not clear with respect to what aspect of handling will (or can) be varied to achieve their example range of 5 to 2000 pounds of *a.i.* in "a period of time

that is representative of a full workday," *viz.*, from at least 4 to a maximum of 8 hours. Can the same type of application equipment span this range? Will work practices be the same across this range? What predictive artifacts are introduced by attempting to maximize the range of AaiH? And is manipulating the conditions they select to maximize A*aiH* worth the loss of representativeness within the resulting data? Throughout these efforts to maximize variability, the Panel recommends that currently approved practices be used and that maximum amount under currently approved limits not be exceeded.

The AHETF Technical Summary Document refers to sampling such that there is a "bias toward conditions that might yield higher exposures." This bias is in conflict with the statement that users "must also assume that the purposive sample of the MUs approximates some type of probability sample from the target population." Further clarification is needed on this sampling bias and how it might affect exposure distributions and subsequent uses of the database. A detailed critique of the AHETF sampling plan is contained in Appendix C. The conclusion of that critique is that the underlying distribution of pesticide use conditions must be considered either during sample selection or during data analysis. The impact of this bias is further complicated by the lack of a database that documents the distribution of tasks, activities, or pesticide use information within any given exposure scenario from which to judge or compensate for the biases introduced by the sampling selection design being proposed (see also the last comment in our response to Charge #1).

Recommendations Regarding AEATF

The AEATF Study Plan is dealing with a very different situation and is much more amenable to experimental control. In particular, it should be feasible to increase the sample size for any scenario at a future date if more observations are needed. The proposal to take 15 monitoring units initially is adequate to give an overview. For probabilistic assessments and determination of exposure at extreme upper percentiles, 15 units will not be enough.

The Panel was also asked whether the AEATF study should be conducted at one simulated site or three field locations. The simplest way to answer this is to try both options a few times in a pilot study and compare the results. Perhaps three field sites should be treated as blocks or strata rather than clusters. In summary, the most important difference between the two studies is the possibility of increasing the sample size at a future date.

REFERENCES

Brouwer DH, Boeniger MF, Van Hemmen J. 2000. Hand Wash and Manual Skin Wipes. Ann. Occup. Hyg. 44(7):501-510.

Durkin P, Hertzberg R, Diamond G. 2004. Application of PBPK Model for 2,4-D to Estimates of Risk in Backpack Applicators. Environmental Toxicology and Pharmacology 16:73-91.

Kline AA, Landers AJ, Hedge A, Lemley AT, Obendorf SK, Dokuchayeva T. 2003. Pesticide exposure levels on surfaces within sprayer cabs. Applied Engineering in Agriculture 19(4):397-403.

Lu C, Anderson LC, Morgan MS, Fenske RA. 1997. Correspondence of salivary and plasma concentrations of atrazine in rats under variable salivary flow rate and plasma concentration. J. Toxicol. Environ. Health 52:317-329.

Lu C, Rodríguez T, Funez A, Irish R, Fenske RA. 2006. The assessment of occupational exposure to diazinon in Nicaraguan plantation workers using saliva biomonitoring. Ann. New York Acad. Sci. 1076:355-365.

Lu C, Rodriguez T, Thetkathuek T, Funez A, Pearson M. 2007. Using Salivary Biomarker in Exposure and Risk Assessments for Organophosphorus Pesticides: Possibilities and Pitfalls. (submitted).

McDougal JN and Boeniger MF. 2002. Methods for assessing risks of dermal exposures in the workplace. Critical Reviews in Toxicology 32:291-327.

Williams PL, Thompson D, Qiao G, Monteiro-Riviere N, Riviere JE. 1996. The use of mechanistically defined chemical mixtures (MDCM) to assess mixture component effects on the percutaneous absorption and cutaneous disposition of topically exposed chemicals. Toxicology and Applied Pharmacology 141:487-496.

APPENDICES

Appendix A: *Definitions and Abbreviations*

This glossary of terms was prepared not only because terminology is so important within the topic of dermal hazards from chemicals but more specifically to help assure that the responses of the Panel are both internally consistent and properly interpreted. Most of the terms in this glossary come from the *Review of Worker Exposure Assessment Methods* document prepared for the Panel by the U.S. Environmental Protection Agency, Health Canada, and the California Environmental Protection Agency (referred to herein as the "*Review*").

Absorption Factor = A measure of the flux or amount of chemical that crosses a biological boundary such as the skin (% of the total exposure that is absorbed). p.38

AHED = Agricultural Handlers Exposure Database to be developed by the AHETF.

BHED = Biocide Handlers Exposure Database to be developed by the AEATF.

bias = One form of a bias is a consistent or overall difference between the result and the true value being estimated, sometimes called a systematic bias. Results from a known bias (such as sample recovery efficiency or representative measurements) can be adjusted for by a simple calculation. A bias exists for virtually all of the dosimetry and biomonitoring methods discussed in this report.

Another form of bias is in sample selection in which some members of the population are more likely to be included than others. While the existence of sample selection bias can be identified, its effect and methods to adjust the results are not always known.

biological monitoring = [BM or biomonitoring] "is usually employed to quantify the absorbed dose (also referred to as body burden)." The Agency accepts monitoring data based on the collection of biological media such as urine or blood. Biomarker data can also be used for predicting exposures. In addition to biological sampling media, the Agency also requires that additional supporting pharmacokinetics and/or pharmacodynamic information be submitted that can be used to develop exposure and/or risk estimates. p. 35. For the purposes of this discussion, biomonitoring was restricted to urinary excretion.

By definition, dose estimates resulting from biological monitoring "integrate exposure across all routes. Such an estimate of absorbed dose, which avoids potential confounding from assumptions of dermal penetration or inhalation retention, may be more useful in assessing risk than route-specific doses estimated from passive dosimetry." p.41-42

BM = See biological monitoring.

CV = coefficient of variation calculated as the standard deviation divided by the group's mean (sometimes also called a coefficient of variance).

Dose = the amount of chemical absorbed from exposure to a pesticide in a given scenario. p. 21

Dermal exposure = defined by U.S. EPA (1996b) as the process of pesticide residue deposition onto the skin, as well as the measurement of the deposited residue. p.38

Exposure = Amount deposited on the surface of the skin that is available for dermal absorption or amount that is inhaled, also referred to as potential dose. p. 21 See also Dermal exposure.

Factor-set = a large number or combination of factors that characterize a specific setting, *e.g.*, climatic conditions, combinations of equipment, task times, *etc*. As used in Appendix C, each unique combination of factors is denoted by the symbol S_i.

inner dosimeter =a dosimeter worn under the handler's work clothing and held in some way against the skin.

LOAEL = Dose level in a toxicity study, the lowest dose level where an adverse effect occurred in the study (mg pesticide active ingredient/kg body weight/day). p.38

MOE = Margin of exposure, value used by the Agency to represent risk or how close a chemical exposure is to being a concern (unitless). p.38

MU = Monitoring Unit or one person whose exposure is assessed via measurements such as biomonitoring or passive dosimetry.

NOAEL = Dose level in a toxicity study, where no observed adverse effects occurred in the study (mg pesticide active ingredient/kg body weight/day). p.38

outer dosimeter = a dosimeter worn outside of regular work clothing (in some cases, the regular work clothing has comprised the outer dosimeter).

passive dosimetry = [PD] "employs some sort of physical monitor that traps residues from the surface of the skin (*i.e.*, they absorb or remove, such as a dosimeter or a wash) to determine dermal exposure (*i.e.*, also referred to as a potential dose). Passive dosimetry methods (*e.g.*, patches, gloves, dosimeters or washes). p. 35 For the purposes of this discussion, passive dosimetry [PD] included hands or face/neck washing, patch dosimeters, and whole-body dosimeters.

Patches = Various forms of absorbent pads (usually made of gauze but sometimes of alpha-cellulose) placed on the body at fixed locations.

PD = see Passive Dosimetry.

PHED = Pesticide Handlers Exposure Database consists of dermal and inhalation exposure measurements compiled by EPA beginning in the late 1980s from a wide variety of scenario specific pesticide handler exposure studies.

Probable uncertainty = The expected diversity of the result based on an analysis using propagation of error theory. Its magnitude is based on the precision of the individual measurements used to calculate a result and the form of the calculations being made. For linear calculations, probable uncertainty may be characterized by a standard deviation, coefficient of variation, or geometric standard deviation after adjusting for the multiplication or division factor used to achieve the result, such as after adjusting for sample recovery from a dosimeter or in urine.

variability = The diversity of the population being studied. The true variability of a population is only known if repeated measures of whatever is being studied can be made with an accurate and completely precise method. Variability is usually characterized in terms of the observed variation or geometric standard deviation of a random sample of such measurements.

WBD = A whole-body dosimeter preferably composed of full-length cotton underwear, cf., a WBD composed of coveralls and worn as outer clothing. The latter suffers the combined problem of not retaining all of the pesticide deposited onto them and passing through them some fraction of the pesticide that was deposited.

Appendix B: Expanding the Concept of Grading Data

The current data grade (p. 31 of the *Review*) is based (in part) on two relevant criteria: the "% Lab recovery" and the "CV for Lab recovery" which is also a percent. The current "Grading" evaluation of PHED data is based on these two criteria independently, *i.e.*, a lower letter-grade is assigned if the data exceed either criterion. The effect of this interpretation is shown by the small shaded rectangles in Figure B-1 below. In fact, these two criteria are two characteristics of the same data that interact and can be combined within the concept of "data grade" to indicate the magnitude of the probable uncertainty within the resulting exposure value, as calculated by the equation below. The "magnitude of uncertainty" is shown as a multiplying factor herein but could be viewed as a percent by subtracting one and multiplying by 100 (as a percent, it would be analogous to a CV, but interpreting values that exceed several hundred percent was deemed easier as a multiplier such as $\pm 4\times$ than as the equivalent $\pm 500\%$).

$$\textit{Magnitude of Probable Uncertainty} = 1 + \left[\frac{\textit{CV of Lab Recovery [\%]}}{\textit{low \% Lab. Recovery}} \right] \qquad \text{Eqn. B.1}$$

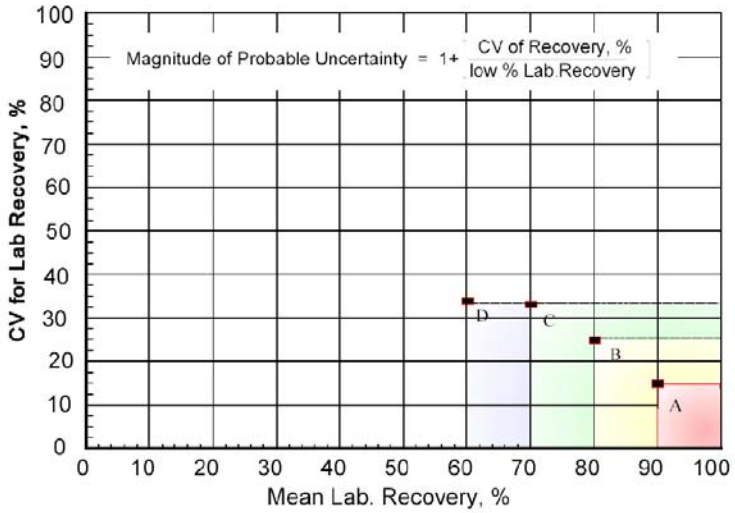

Figure B-1. A narrow view of graded data quality

Each of these two criteria can contribute to the magnitude of the uncertainty within the resulting calculated exposure; however, only together do they define the magnitude of that statistically probable uncertainty. Figure B-2 shows an array of possible values of these two criteria that have the same magnitude of probable uncertainty represented by the relatively large shaded triangles comprising each grade. One can discern several points from this figure. One is the broader range of possible criteria that result in the same quantitative effect on the probable uncertainty as an indicator of data quality. The Agency could consider broadening their criteria for "Data Grade" by using the above equation without decreasing the precision of their resulting data.

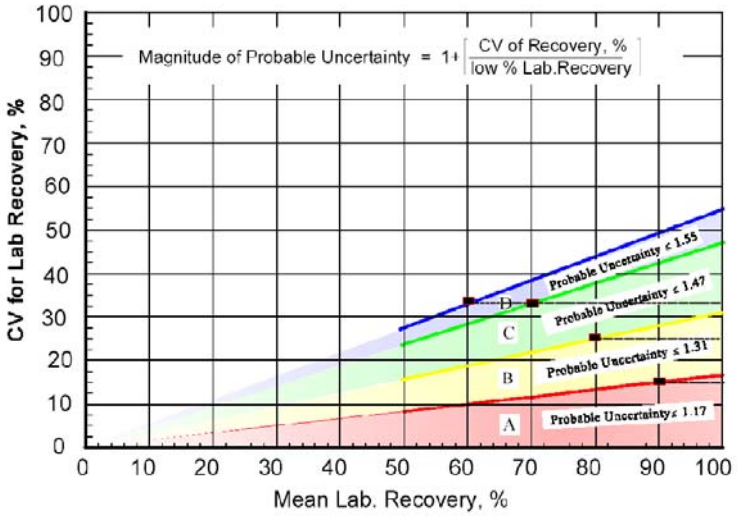

Figure B-2. A broader view of graded data quality

Second, one can see in Figure B-2 that the CV of the Recovery has a greater effect on the Grade than the mean Lab. Recovery. In fact, one can see that a sufficiently precise estimate of a poor sample recovery can yield a smaller probable uncertainty in exposure than an imprecise estimate of a good sample recovery. An extension of this observation is that the reliability of the result (as characterized by the magnitude of the probable uncertainty) cannot be defined by only specifying a mean recovery (any value on the X-axis in either figure) without also specifying a CV.

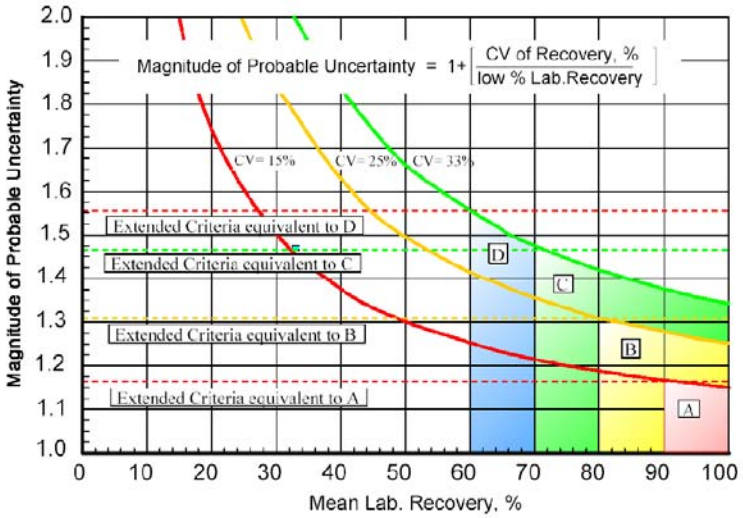

Figure B-3. The impact of the CV of the mean Lab. Recovery on the magnitude of Probable Uncertainty

This observation is perhaps more visible in Figure B-3 in which the Y-axis is the probable uncertainty and the data's CV is a parameter. The Agency's current definition of Grade is still depicted as the shaded zones in the lower right, and the suggested extended definition is shown as horizontal dashed lines at constant values of probable uncertainty. Notice that one cannot predict the variability of the result by only specifying a value on the X-axis (a % Lab. Recovery)

without also specifying a CV. Thus, the Agency should consider adding CV limits to their Field Recovery and Storage Stability criteria.

A third observation is that the magnitude of the probable uncertainties shown in these figures (even using the Agency's worst current Data Grade of "D") are all much smaller than the range of variabilities in the data discussed within their *Review* (such as a 3× range in dermal absorption or differences between passive dosimetry and biomonitoring). The magnitude of these probable uncertainties will also be seen to be much smaller than other uncertainties when the same principle behind Equation B-1 is applied to passive dosimetry data in Charge #2 and to urinary biomonitoring data in Charge #3. The same principle (with less specific measures of CV) could also be applied to other data attributes such as the lack of complete body monitoring in much of the PHED dosimetry data, the variability within the calculated results implicit within each form of passive dosimetry, and the impact of the simplistic method currently used to account for undetectable samples (due to censoring). Any one of these quantitative attributes of the data within PHED probably has a greater impact on the variability within the results than does sample recovery but are currently not part of the grading criteria. Furthermore, incorporating these other measures of data quality into the “data grade” is likely to illuminate the poor quality of much of the existing PHED data and the potential for the new AHED data to improve that quality by orders of magnitude.

Appendix C: A Critique of the AHETF Study Design

The Panel's understanding of risk assessment is that the exposure value used in the risk equation is expected to be "representative" of the average exposure that would be experienced by the population potentially exposed to the chemical. For probabilistic risk assessments, individual exposure values are drawn from a distribution of exposures that are expected to describe the distribution of long term average exposures for individuals in the population potentially exposed to an active ingredient. The Agency should look at the proposed sampling design through the lens of its "representativeness."

The first assumption made (by AHETF) is that a surrogate cluster sampling model that assumes underlying random selection can be used to estimate sample sizes even though the proposed sampling methodology does not advocate random sampling for clusters. The second assumption (by AHETF) relates to the normality of variance components in the nested-effects linear model on log-normalized exposure. The real concern herein is with the acceptability of using the surrogate random sampling model.

The discussion in Sections 5.1 and 5.2 of the AHETF Technical Summary background document (AHETFa, 2006) is excellent in that it provides a good framework for thinking about sampling for exposure assessment. The following will use a slight modification of their conceptual model to illustrate a technical concern with the sampling protocol that is being proposed.

The goal of the AHED dataset is the estimation of the true exposure, E for a specific handler task. To collect these data, AHETF proposes a cluster sampling or hierarchical sampling design in which clusters (or studies) are essentially examinations of handlers performing the handler task of interest at specific locations and times. As mentioned in the background documentation, there exist a very large number of potential studies.

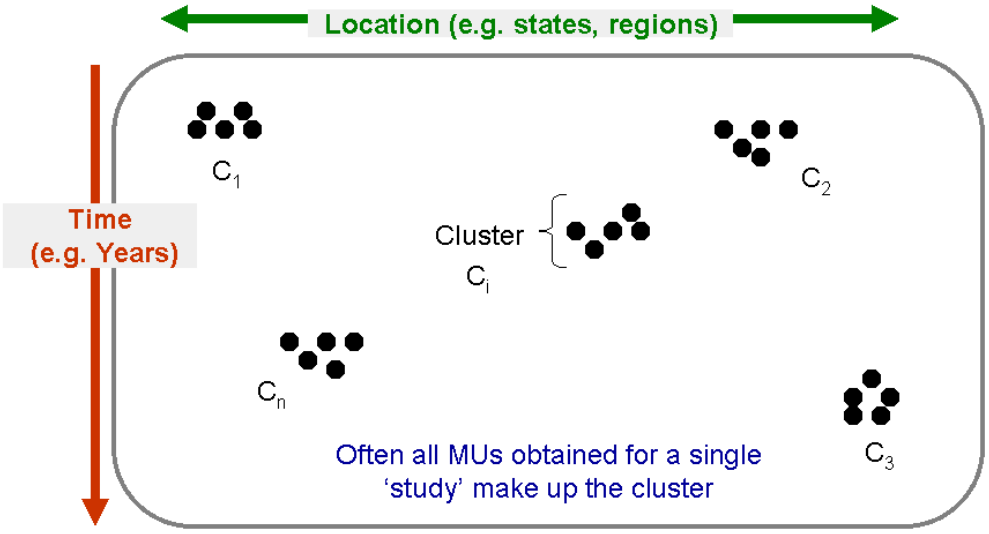

Figure C-1. AHETF proposed cluster sampling design

Conceptually each C_i is characterized by specific settings for a large number of factors, *e.g.*, climatic conditions, equipment combinations, task times, *etc*. Denote each unique combination of factors by S_i for factor-set. (Note: in the discussion before the Panel, the word "scenario" was used for "factor-set," but this caused some confusion and has been changed here. The word scenario as used by AHETF and EPA applies to a handler task.) In theory, if one knew all of the conditions that affected exposure one could compute a true average exposure concentration E_i for each factor-set. Each cluster/study is essentially a replicate of some set of factors. Since many of the factors that impact exposure are continuous, theoretically there are an infinite number of factor-sets and hence there are an infinite number of potential studies.

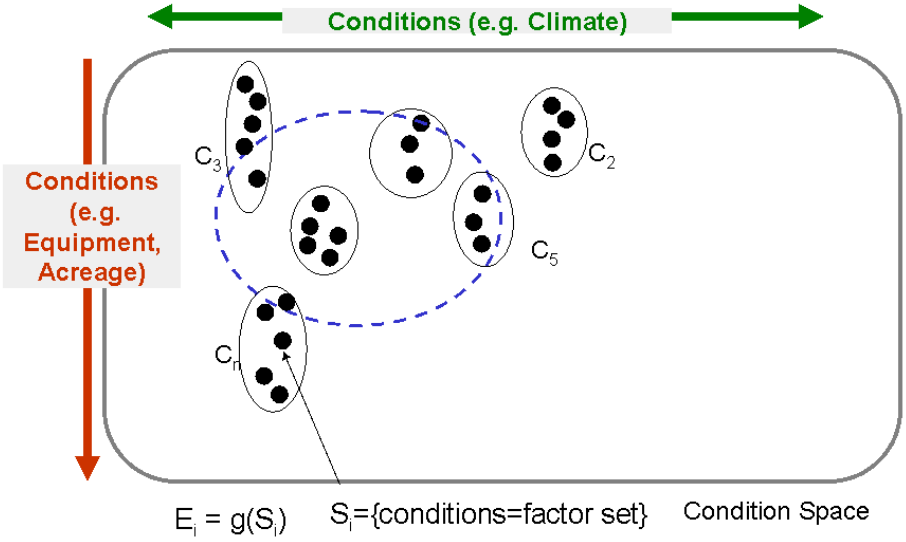

Figure C-2. AHETF study design conceptualized as a set of factor conditions

One way to think of the true exposure for a handler task within a cluster would be as the average of exposures for the handler task across all possible factor-sets.

$$E = \sum_{i=1}^{\infty} E_i = \sum_{i=1}^{\infty} g(S_i) \qquad \text{Eqn. C.1}$$

But this equation assumes that each possible factor-set has an equal probability (or frequency) of occurrence, and we know this is not true. Pest application tasks have certain climatic conditions that define when they must or can be performed. Some types of equipment are more common than others. For example, the factors-set contained within the blue circle in Figure C-2 might represent the more common conditions. If one knew the relative frequency or probability of each factor-set, denoted as P_i, then the true exposure could be estimated as a weighted average.

$$E = \sum_{i=1}^{\infty} E_i P_i = \sum_{i=1}^{\infty} g(S_i) P_i = \int_{s \in S} g(s)\,dF(s) \qquad \text{Eqn. C.2}$$

Since the true condition space is continuous, the most mathematically appropriate way to describe the true exposure is given in the integral part of Equation C.2. The term dF(s) essentially describes the relative probability for each possible factor-set, s, in the total factor space S.

What does the surrogate random sampling model mean in terms of clusters C and factor-set S? With a random selection of clusters, one essentially selects factor-sets at random for inclusion in the study, and their proportion within the sample will be their relative frequency within the population of interest. This also means that averaging the study-specific average exposures should produce an unbiased estimate of the true handler task exposure. That is, Equation C.3 is an unbiased estimate of what we need for the risk assessment.

$$\hat{E} = \frac{1}{n}\left[\hat{E}_1 + \hat{E}_2 + \hat{E}_3 + \cdots + \hat{E}_n\right] \qquad \text{Eqn. C.3}$$

From Figure C-2, simple random sampling would result in relatively more of the sampling effort placed in the blue circle than would be placed outside of it.

Now consider the "Diversity" sampling approach as proposed by AHETF. The approach does not include randomness. Through thoughtful considerations of location and time, it is possible that a large number of factor-sets will be examined. In fact, the background document to the Panel seemed to indicate that the locations and times would be selected to ensure that different conditions would be represented. The problem with the Diversity approach is that the relative frequency of scenarios will not necessarily be considered in the selection of the scenarios. Hence, it is possible (if not likely) that only one of a really common (high relative frequency) factor-set will be included in the sample set at the same time as one really rare (low relative frequency) factor-set is included. The dots and clusters in Figure C-3 depict an array of sampled conditions chosen for their diversity that are not representative of the relative frequency of factor-sets. The relatively common factor conditions found within the blue circle has small representation compared to the uncommon conditions outside the blue circle. When the sample average of estimated exposures is computed, estimated exposure for the rare scenario is weighted equally with the estimated exposure for the common scenario, and as a result the sample average will be a biased estimate of the true exposure.

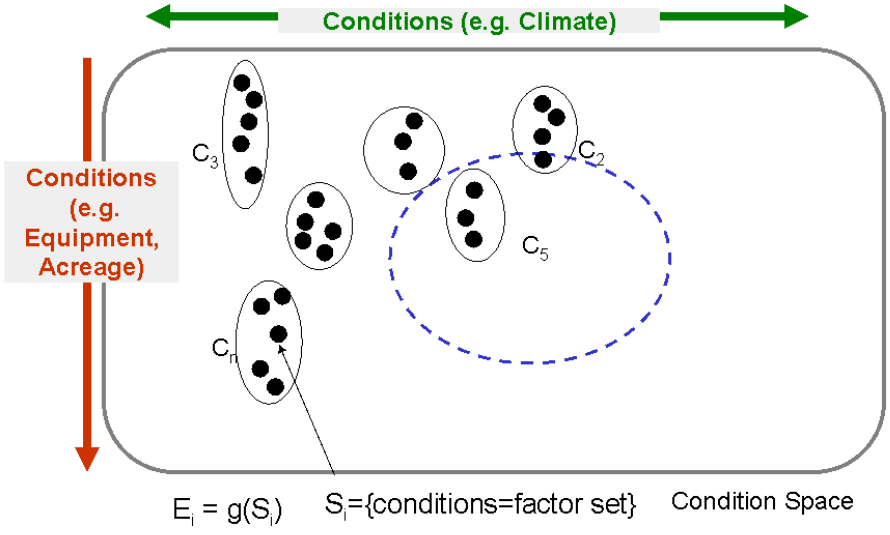

Figure C-3. Non random sampling can result in samples that do not properly represent the distribution of conditions

The issue is even worse if one wants to estimate a distributional parameter of the true exposure like the standard deviation or an upper percentile. The non-probability sample will not produce a faithful estimate of the population distribution. Worse, one cannot even predict the direction of the bias. The study design would produce overestimates if the over-represented rare factor-sets produce high exposures, or the design would produce underestimates if over-represented rare factor-sets produce low exposures.

This then is the basis for the statement made by the AHETF statistician that "non-random sampling means that statistical methods alone are insufficient for generalizing to the target population." Most statisticians and many risk assessors are aware of this problem. This problem is not new. Almost every environmental dataset has this problem. The question is whether we want to support the creation of another environmental dataset with this problem.

AHETF acknowledges the above problems in the background document and points out that rarely are the true relative frequencies of the factor-sets known. At the same time, it is not possible to create a simple random sample that is guaranteed to appropriately represent a specific scenario. The goal of the "Diversity" sampling approach proposed for populating AHED is to "achieve a diversity of major factors that are likely to influence exposure" and to attempt to "capture the major aspects of" the actual distribution of exposures. In essence, AHETF will attempt to identify specific C_i to sample that are "representative" of the whole set of possible conditions such that the distribution of exposures from the Diversity sample is approximately equal to the distribution of exposures appropriately weighted for all factor-sets.

Statisticians have heard this kind of sampling proposal many times but have never seen a true success. It is actually impossible to purposively define a sample that produces a distribution of exposures that duplicates the true population distribution when one has no knowledge of the true population distribution to start with. Rare events are seldom given proper consideration and common events are often under represented. Selecting to get true representation does not work. Randomness in selection must be used somewhere in the design to even approach an unbiased estimate.

So, is this really a hopeless situation? Not necessarily. EPA and AHETF have at this point the opportunity to rethink these issues and possibly come up with some new approaches that might get them closer to their stated goals. While more creative thinking may come up with a number of feasible alternatives, consider the following approach.

1. Create a list of all of the factors that are known to impact exposure levels within a specific scenario. The list may be long but is not infinite.
2. Rank-order the factors by their expected magnitude of impact on exposure variation. A Delphi approach might be used with a Panel of expert risk assessors to accomplish this ranking.
3. Select the top two to four factors, and identify for each factor two to three categories or levels.
4. Create the set of all possible combinations of factor levels. Consider these combinations as strata of the population of interest. In a sense, these become the factor-sets of interest, S_i*.
5. Next assign a weight, w_i, to each factor-set, S_i* that approximates its relative frequency in the population. Sampling theory tells us that these weights don't have to be exact for us to gain

large improvements in estimator precision. Here again, the use of a Delphi approach with a Panel of agricultural experts could help.

6. Selection of studies (two options).

Option 1 (see Figure C-4): Select at-random studies and/or MUs for each factor-set such that the relative number of exposure estimates obtained for the factor-set equals its weight. The population exposure estimate is the <u>average</u> of the estimated exposures for the MUs.

Option 2: (see Figure C-5): Select at random a fixed number of studies or MUs for each factor set and assign each the factor set weight, w_i. The population exposure estimate is the <u>weighted average</u> of the estimated exposures for the MUs.

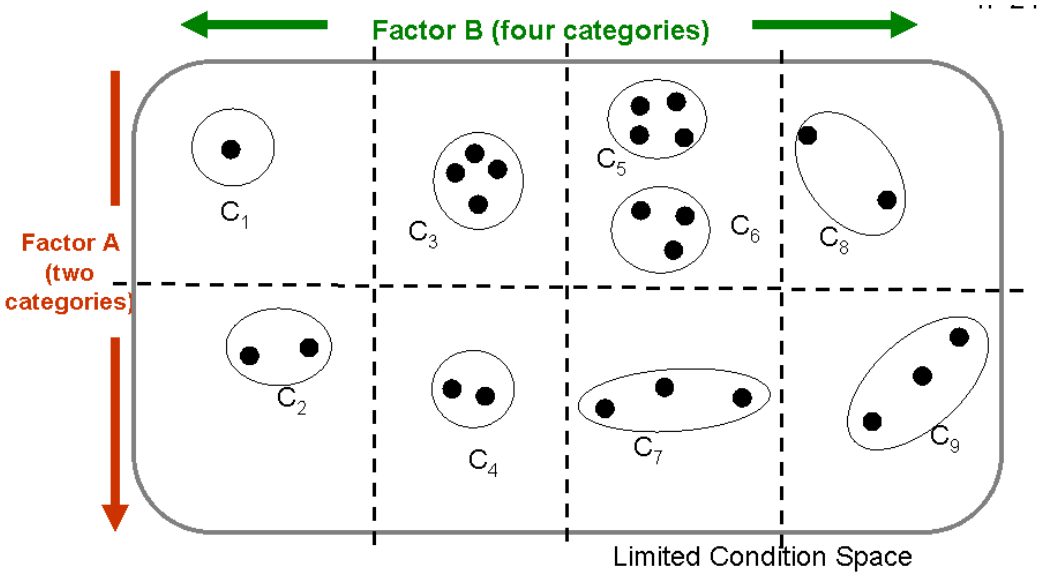

Figure C-4. Allocation of studies and MUs to factor-set strata according to their relative weights.

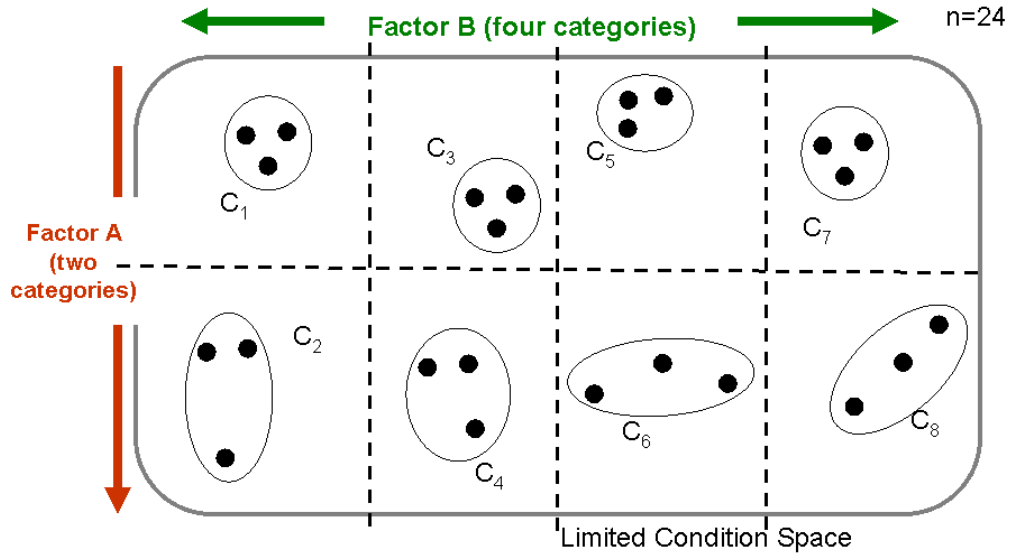

Figure C-5. Uniform allocation of studies and MUs to each factor-set strata.

This approach incorporates both "representation" and "randomness" into the creation of the database and should result in an average exposure estimate that is less biased than the "average" that would be obtained from the Diversity sampling protocol.

The above approach is quite similar to what AHETF is actually proposing. The major difference is that in the approach outlined here an attempt will have been made first to map out the possible condition space in a rough categorized way, to assign relative importance to each category, and finally to sample according to that relative importance.

A number of issues with this approach remain to be addressed by the Agency and AHETF statisticians. For example, if a uniform sampling plan (Option 2) were used, the sample weights would be used to estimate simple statistics such as the mean or standard deviation. Much more complex is the use of these weights to estimate the upper percentiles for exposure, to estimate the exposure distribution, or to test the exposure distribution for a specific distributional form.

It was pointed out that many seemingly unrelated variables are correlated in pesticide application studies (*e.g.*, the number of acres sprayed, type of PPE, use of a tractor with a cab, *etc*.). It is also acknowledged that sufficient information on condition factors would be needed to understand how factors co-vary. A two-dimensional stratification based on Factor A and B would be much less effective if Factors A and B were highly correlated. It would be better to use two factors that are relatively uncorrelated in the definition of the strata. In this case, each factor could be seen as a surrogate or representative of a large set of correlated factors (kind of like a principal component).

Finally, the interface between the User and the AHED dataset should reflect not only the data contained within the database but also the sampling design used to collect that data. Hence, if a uniform sampling design were used, weighted estimates and correspondingly appropriate tests for distributions should be presented as a result of a User query.