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

OFFICE OF CHEMICAL SAFETY
AND POLLUTION PREVENTION

June 5, 2012

MEMORANDUM

SUBJECT: Transmittal of Meeting Minutes of the FIFRA Scientific Advisory Panel Meeting held March 6 - 7, 2012 on Methods for Efficacy Testing of Bed Bug Pesticide Products

TO: Steven Bradbury, Ph.D.
Director
Office of Pesticide Programs

FROM: Joseph E. Bailey, Designated Federal Official
FIFRA Scientific Advisory Panel
Office of Science Coordination and Policy

THRU: Laura Bailey, Executive Secretary
FIFRA Scientific Advisory Panel
Office of Science Coordination and Policy

Frank Sanders, Director
Office of Science Coordination and Policy

Attached, please find the meeting minutes of the FIFRA Scientific Advisory Panel open meeting held in Arlington, VA on March 6 - 7, 2012. This report addresses a set of scientific issues associated with Methods for Efficacy Testing of Bed Bug Pesticide Products.

Enclosure:

cc:

Jim Jones
Louise Wise
Vicki Dellarco
William Jordan
Margie Fehrenbach
Keith Matthews
Donald Brady
Jack Housenger
Joan Harrigan-Farrelly
Lois Rossi
Jay Ellenberger
Karen Whitby
Richard Keigwin
Kevin Sweeney
Dale Kemery
Douglas Parsons
Vanessa Vu
OPP Docket

FIFRA Scientific Advisory Panel Members

Janice Chambers, Ph.D., DABT, Fellow ATS
Stephen Klaine, Ph.D.
Kenneth Portier, Ph.D.
Martha Sandy, Ph.D., M.P.H.

FQPA Science Review Board Members

Mark Feldlaufer, Ph.D.
Jody L. Gangloff-Kaufmann, Ph.D.
Edward E. Gbur, Jr., Ph.D.
Matthew Kramer, Ph.D.
Alvaro Romero, Ph.D., DVM
Michael Trevor Siva-Jothy, D. Phil.
Rajeev Vaidyanathan, Ph.D.

SAP Minutes No. 2012-03

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

Methods for Efficacy Testing of Bed Bug Pesticide Products

March 6 - 7, 2012
FIFRA Scientific Advisory Panel Meeting
Held at the
Environmental Protection Agency Conference Center
Arlington, VA

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 Environmental Protection Agency, Office of Pesticide Programs (OPP), and is structured to provide balanced expert assessment of pesticide and pesticide-related matters facing the Agency. FQPA 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 Joseph E. Bailey, SAP Designated Federal Official, via e-mail at bailey.joseph@epa.gov.

In preparing these meeting minutes, the Panel carefully considered all information provided and presented by EPA, as well as information presented by public commenters.

TABLE OF CONTENTS

SAP Minutes No. 2012-03

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

Methods for Efficacy Testing of Bed Bug Pesticide Products

March 6 - 7, 2012
FIFRA Scientific Advisory Panel Meeting
Held at the
Environmental Protection Agency Conference Center
Arlington, VA

Martha Sandy

Martha Sandy, Ph.D. M.P.H.
FIFRA SAP Session Chair
FIFRA Scientific Advisory Panel
Date: JUN 05 2012

Joseph E Bailey

Joseph E. Bailey
Designated Federal Official
FIFRA Scientific Advisory Panel
Date: JUN 05 2012

Federal Insecticide Fungicide and Rodenticide Act
Scientific Advisory Panel Meeting
March 6 - 7, 2012

PARTICIPANTS

FIFRA SAP Chair

Kenneth Portier, Ph.D., Program Director, Statistics, American Cancer Society, National Home Office, Atlanta, GA

FIFRA SAP Session Chair

Martha Sandy, Ph.D., M.P.H., Senior Toxicologist and Chief, Cancer Toxicology and Epidemiology Section, Reproductive and Cancer Hazard Assessment Branch, Office of Environmental Health Hazard Assessment, California Environmental Protection Agency, Oakland, CA

Designated Federal Official

Joseph E. Bailey, FIFRA Scientific Advisory Panel, Office of Science Coordination and Policy, EPA

FIFRA Scientific Advisory Panel Members

Janice Chambers, Ph.D., DABT, Fellow ATS, Director, Center for Environmental Health Sciences, College of Veterinary Medicine, Mississippi State University, Mississippi State, MS

Stephen Klaine, Ph.D., Professor, Department of Biological Sciences, Director, Institute of Environmental Toxicology, Clemson University, Pendleton, SC

FQPA Science Review Board Members

Mark F. Feldlaufer, Ph.D.
Invasive Insect Biocontrol & Behavior Laboratory
USDA/Agricultural Research Service
Beltsville, MD

Jody L. Gangloff-Kaufmann, Ph.D.
New York State IPM Program
Faculty Fellow, Atkinson Center for a Sustainable Future
Cornell University
Babylon, NY

Edward E. Gbur, Jr., Ph.D.
Agricultural Statistics Laboratory
Arkansas Agricultural Experiment Station
Dale Bumpers College of Agricultural, Food and Life Sciences
University of Arkansas
Fayetteville, AR

Matthew Kramer, Ph.D.
USDA/Agricultural Research Service
Beltsville, MD

Alvaro Romero, Ph.D., DVM
Department of Entomology, Plant Pathology, and Weed Science (EPPWS)
New Mexico State University
Las Cruces, NM

Michael Trevor Siva-Jothy, D. Phil.
Department of Animal and Plant Science
University of Sheffield
Sheffield S10 2TN UK

Rajeev Vaidyanathan, Ph.D.
Associate Director, Vector Biology and Zoonotic Diseases
SRI International
Center for Advanced Drug Research
Harrisonburg, VA

INTRODUCTION

The Federal Insecticide, Fungicide and Rodenticide Act Scientific Advisory Panel (FIFRA SAP) has completed its review of the Methods for Efficacy Testing of Bed Bug Pesticide Products. Advance notice of the meeting was published in the Federal Register on January 11, 2012. The review was conducted in an open panel meeting held in Arlington, VA, on March 6 - 7, 2012. Dr. Martha Sandy chaired the meeting. Joseph E. Bailey served as the Designated Federal Official.

EPA-registered pesticide products are an important part of bed bug management programs. In the past five years pesticide users have reported control failures, which they claim is due to the lack of bed bug susceptibility to pesticide products that have been applied. Subsequently, resistance to pyrethroid insecticides was documented although it has not been directly correlated to field application failures. In response, EPA evaluated the database for registered products and concluded that there is a need to standardize approaches to bed bug product testing methods to ensure the quality and validity of the efficacy data for these types of products. In order to accomplish this, the Agency developed a draft product performance guideline for bed bug pesticide products. The Agency sought advice and recommendations from the Scientific Advisory Panel (SAP) on scientific issues associated with the proposed Draft Product Performance Guidelines "Laboratory Testing Methods for Bed Bug Pesticide Products."

PUBLIC COMMENTS

Oral Statements were presented as follows:

James Ballard, Ph.D., BCE, representing self and Ballard Pest Management Consulting, LLC
Robert Rosenberg and James Fredericks, representing National Pest Management Association
Timothy Drake, Ph.D., representing Association of Structural Pest Control Regulatory Officials
Kate Shenk, representing Responsible Industry for a Sound Environment
Robin G. Todd, Ph.D., BCE, and Reginald Coler, Ph.D., BCE, representing ICR, Inc.

Written Statements were provided by:

Anonymous commenters
Robert Rosenberg, National Pest Management Association
John F. Wright, Product and Regulatory Associates
Dini M. Miller, Ph.D., Virginia Tech
Philip G. Koehler, Ph.D., University of Florida
Joseph Latino, Allergy Technologies, LLC
Derrick Lastinger, Association of Structural Pest Control Regulatory Officials
Robin G. Todd, Ph.D., BCE, ICR, Inc.
Reginald Coler, ICR, Inc.
Aaron Hobbs, Responsible Industry for a Sound Environment
Carol Somody, Syngenta Crop Protection, LLC
Sibylle Rahlenbeck, Ph.D., and Birgit Habedank, Ph.D., German Environment Agency
James Ballard, Ph.D., BCE, Ballard Pest Management Consulting, LLC
Kelly K. Naujock, Amvac Chemical Corporation (received post public comment period)

SUMMARY OF PANEL DISCUSSION AND RECOMMENDATIONS

Charge Issue 1(a). Appropriateness of the test methods to evaluate the efficacy of bed bug products with regard to kill and knockdown; repellency; pesticide resistance; and discriminating dose. In general, the Panel noted that some methods in the draft guidelines are not standardized (*i.e.*, testing methods could differ) and therefore, it is not known how appropriate or repeatable they may be for the particular type of evaluation proposed. The Panel recommended that rearing systems for bed bugs need to be standardized and that the blood source is important. They proposed reducing the number of strains used in each test and the number of replicates while maintaining scientific rigor. Other considerations noted included when to consider morbidity the same as mortality and how to maintain 'resistant' strains under laboratory conditions (*i.e.*, either to continuously challenge the colony or replace it if resistance is lost).

Charge Issue 1(b). Are there additional or alternative laboratory test methods beyond those discussed in the draft guidelines for testing the efficacy of bed bug pesticide products? The Panel noted that bioassays for some types of insecticides were not included in the draft guidelines, including tests for efficacy of fumigants and volatile compounds, slow-release insecticide strips, metabolic insecticides and insect growth regulators.

Charge Issue 2 (a). The appropriateness of suggested environmental conditions (such as temperature and humidity). The Panel recommended that a temperature range or variation of ± 1° C be added to the test temperature and that the range of acceptable humidity be narrowed. The proposed photoperiodic regime of 16L:8D is acceptable.

Charge Issue 2(b)(i). Should testing be done only with one species, the common bed bug* Cimex lectularius *and is it appropriate to use unfed adult bed bugs with a sex ratio of 1:1 to test adulticide products? The Panel recommended that *Cimex lectularius* be used as the test organism, and only males generally used for testing. All bed bugs used for testing should be fed to repletion 7 days prior to testing and be at least 3 weeks old.

Charge Issue 2(b)(ii). The advantages and disadvantages of including a susceptible strain in every test and whether the Harlan strain is the preferred susceptible strain. The Panel concurred that the Harlan (pyrethroid-susceptible) strain is the preferred susceptible strain. The advantage of using a susceptible strain is that it acts as a control (for bed bug strains) and is also necessary to calculate resistance ratios. In addition, a susceptible strain allows one to compare results across studies and over time, and to provide a baseline efficacy of a product.

Charge Issue 2(b)((iii). The advantages and disadvantages of including a resistant strain in every test and how an investigator should select a resistant strain or strains. The Panel pointed out that maintaining the status of pyrethroid resistance in resistant strains is difficult in the laboratory due to their reversion to 'susceptible' unless challenged and selected with insecticide. The advantage of using pyrethroid resistant strains for insecticide testing is that they represent

actual field populations across the United States, and they serve as a benchmark for comparing results from different studies or changes over time.

Charge Issue 2(b)(iv). The adequacy of the use of field collected bed bug strains from urban areas in three regions of the U.S. to represent the variability in susceptibility of bed bug populations for conducting efficacy evaluations. The Panel concluded that the number of field-collected strains from different regions as recommended in the draft guidelines can be reduced.

Charge Issue 2(b)(v). The recommended approach for grouping and evaluating the data from the tested bed bug strains to assess efficacy of bed bug products. The Panel discussed grouping and data evaluation in their statistical discussions about the individual study designs associated with each specific test protocol.

Charge Issue 2(c). The advantages and disadvantages of including a positive control in every test. The Panel agreed that positive controls should be used in tests and when testing a new pyrethroid insecticide and it is assumed that a pyrethroid-resistant strain of bed bugs will be utilized. In this case, challenging the pyrethoid-resistant strain with deltamethrin (a pyrethroid that many field strains exhibit resistance to) would act as a positive control, insure that the strain of bed bugs is pyrethroid resistant, and would allow for the comparison of efficacy between the new compound and deltamethrin. The Panel pointed out that for insecticides with modes of action different from pyrethroids, there may not be suitable bed bug strains to use as controls. Controls (positive and negative) add an additional burden in terms of numbers of bed bugs and replicates needed.

General Study Design and Statistical Analysis Discussions

In general, the Panel recommended the use of "concentration-response" in place of "dose-response" throughout the document and in concentration-response studies, there should be at least one concentration above and one concentration below the key concentration parameter (either the LC_{50} or LC_{90}) to ensure accurate estimation. In time-to-effect studies, there should be at least one time measurement above and one measurement below the key time parameter (either the LT_{50} or LT_{90}). Specifying a level of confidence is important for all label claims. Negative (untreated) and positive (known pesticide) control treatments should be required as part of concentration-response studies, where possible. An appropriate statistical model should be selected for estimation and testing. Goodness-of-fit should be assessed and the model revised until the goodness-of-fit criteria are satisfied. The Panel recommended that EPA specify which analysis models should be considered and specify levels of acceptable control group mortality in order for a study to be deemed acceptable. The Panel recommended analyses based on generalized linear models (rather than attempting a transformation in order to use a linear model; *e.g.*, a 'probit model'), and conduct statistical testing within the model (*e.g.*, using contrasts).

Charge Issue 3. Resistance ratio determination, characterization of bed bug strain susceptibility and discriminating dose selection.

Charge Issue 3(a). Which insecticides should be chosen as standards for other insecticidal modes of action? The Panel's experience is that the only pesticide class to which resistance is seen in wild bed bug populations is pyrethroids and that deltamethrin is the standard positive control for this group.

Charge Issue 3(b). Usefulness of comparing the resistance ratios of bed bug field populations to the resistance ratio for a corresponding dose of deltamethrin when evaluating pyrethroid insecticides. The Panel concluded that this comparison is a useful exercise because it benchmarks the activity of the evaluated product against an agreed standard (deltamethrin). By using resistance ratios (RR) (with Harlan as the susceptible strain), the tests are also benchmarked against a standardized strain baseline.

Charge Issue 3(c). Is a resistance ratio of 100x adequate to screen field strains for resistance against insecticides and if not, what other approaches are recommended for detecting resistance in bed bug populations? The Panel's consensus was that a resistance ratio of 100x was appropriate, given that a utilitarian resistance ratio should be biologically realistic and practical.

Charge Issue 3(d). What type(s) of data analysis and statistical testing would be most appropriate for these data sets? The Panel recommended using model-based inference when analyzing (and evaluating) data. The link function chosen for use in a generalized linear model (GLM) is less important for estimates near the LC_{50} (or other similar measures), but requires careful selection and justification to ensure good LC_{90} estimates. The Panel noted that the draft guidelines suggest taking logs of concentration to produce a straight-line relationship with the probit model. They recommended against that suggestion because taking the log of the concentration may, or may not, produce a straight-line relationship. The recommendation was made to include a standard error on resistance ratio, calculated using either a Taylor's series expansion or a linear model with contrasts if concentration is log-transformed. Including sufficient information in the description of the testing protocol was emphasized as an important step in ensuring that EPA clearly understands how experiments were conducted. The standard format given in Appendix A is provided for use as a guideline.

Charge Issue 3(e). The discriminating dose selection for products, including whether the LD90 value of the least susceptible field population is the best value to use as the basis for discriminating dose selection. The Panel's discussion concluded that the charge should read "most susceptible field population" rather than "least susceptible." The Panel concluded that the LC_{90} was acceptable to use but they recommended that the protocol for selecting the discriminating dose take into account the uncertainty associated with the estimated LC_{90} values for susceptible field populations. Use of doses that are much larger or smaller than the LC_{90} will result in high estimate uncertainty and an inaccurate discriminating dose.

Charge Issue 4(a). Do the exposure times provide sufficient data to measure the efficacy of a bed bug pesticide or should other exposure times or testing scenarios be used?
The Panel pointed out that the draft guidelines needed to clarify whether mortality observations are to be made during insecticide exposure and/or post-exposure. The single dose exposure time is sufficient for section (j) on forced exposure residual surface treatments. It was also noted that in section (l) on testing impregnated materials, exposure times in the 'tunnel test' may not be true exposure times because of behavioral variation in avoidance among bed bug strains.

Charge Issue 4(b). Should percent mortality values or lethal time values be used as endpoints to assess the success of product applications and what are the advantages and disadvantages of each one? The Panel concurred that percent mortality is the preferred endpoint, although both parameters can yield important information.

Charge Issue 5(a). Will the experimental unit described in the draft guidelines provide sufficient data to show how effectively and how quickly a bed bug product knocks down and kills bed bugs, when applied as a residual treatment to surfaces? Overall, the Panel concluded that this particular test is large and complicated as proposed, and they pointed out that providing surfaces and habitats in testing units that mimic the natural habitat, human habitat, will provide the most accurate test results. This is especially important because field testing is not included in these draft guidelines. The Panel recommended testing three treatment surfaces, rather than five specified in the draft guidelines and use of 2 or 3 bed bug strains (including a resistant and a field collected).

Charge Issue 5(b). Is there is a single surface type that could be used as a standard or representative surface for testing product residual activity in lieu of testing multiple surfaces as recommended in the draft guidelines? The Panel recommended using the following three surface types, depending on how the product will be used by the end-user: 1) dry wall painted with a satin latex paint (which satisfies the criteria of impervious and nonporous), 2) varnished wood paneling (which is solid and textured) and 3) cotton cloth with stiff cardboard backing (which is porous and absorptive).

Charge Issue 5(c). Modifications or additional tests that could improve residual surface treatment testing. The Panel concluded that this test is important but costly, and recommended using 4 x 4 inch square surface material rather than 6 x 6 and using stiff cardboard as backing for the cotton cloth test rather than glass. The Panel recommended counting dead bed bugs at 1, 24, 48 and 72 hours, if necessary. Further, the Panel indicated that adding stimuli that mimic bed bug habitat, such as host or harborage cues, could ensure more accurate results, but could also raise some questions (see Charge 6(b)).

Charge Issue 5(d). What type(s) of data analysis and statistical testing would be most appropriate for these data sets? The Panel discussed the extent to which the termination rule ("Observations terminated at 100% mortality in product treatment or 10% mortality in control treatment") would result in insufficient mortality to estimate the parameters of interest (LT_{50}, LT_{90}, KT_{50} or KT_{90}). The Panel recommended that guidance be provided on what the researcher

is to do if this happens. They pointed out that whether bed bug strains are considered a fixed effect or a random effect in the analysis model determines the type of test statistics used to estimate the bed bug effect. In the proposed draft guidelines, the Panel indicated that some factors could be confounded with day of week or time of day if the experiment was poorly designed and they stressed that it is important that the study design not confound the most important effect (*e.g.*, bed bug strain) with the time factor. They recommended that study implementation properties such as restrictions on randomization, potential experimental related covariates and correlations related to measurements taken over time be noted in the study design summary so that they can be taken into account in the analysis model. Examples of analysis methods to consider are provided, including Kaplan-Meier approaches (survival models) and generalized linear models and/or generalized linear mixed models (survival trend modeling), assuming a binomial family, and a probit or logit link function. Since this testing is for a residual surface treatment, some objective measurement of the amount of pesticide remaining available and active on the treated surface over time is needed.

Charge Issue 6(a). Will the experimental unit described in the draft guidelines provide sufficient data to measure the duration and extent to which a pesticide product's residues repel bed bugs? The Panel concluded that a more appropriate descriptive term to use for this test is "avoid" rather than "repel" and that this test is more suitable as a subset under the category of "forced exposure (no choice) residual surface" treatment because it does not meet the draft guideline's definition of "repellent." The Panel concluded that "duration" of avoidance is difficult to measure and that the proposed experimental set up using Petri dishes makes it difficult to determine if a product is repellent or attractive. An alternative T-maze experimental unit was suggested along with a recommendation to conduct the experiment under bed bug dark period using a red light.

Charge Issue 6(b). Is conditioning of experimental bed bug harborages necessary? While harborages conditioned with fecal dropping would be more attractive to bed bugs, conditioning the harborages adds considerable time/effort to the experiment and raises the question of how to quantify (amount and 'shelf-life') the effect of bed bug aggregating factors contained in the fecal droppings.

Charge Issue 6(c). Should individual responses and/or group responses be used to determine whether bed bugs are repelled by pesticide product residues in harborages? The Panel concluded that these experiments should be conducted with both individuals and groups, because each represents an infestation type normally seen in natural conditions. However, the Panel recognized that using group evaluations complicate the statistical analysis because bed bugs in groups do not behave independently.

Charge Issue 6(d). What type(s) of data analysis and statistical testing would be most appropriate for these data sets? Many Panel members did not recommend use of the study set up as proposed and suggested the consideration of alternative experimental layouts, such as a T-maze. They recommended that the study design be modified to include only one type of tent in each container and use a previously determined fixed, shorter time on test. Each container

provides one replication for one product treatment. Different containers and bed bugs are required for each of the product treatments for each replication. This produces a simpler protocol but one that requires more bed bugs.

Charge Issue 7(a). Is the experimental unit adequate for evaluating mortality and blood feeding inhibition following exposure to impregnated materials? The Panel suggested modifications to make this experimental unit simpler and less expensive to run and easier to analyze statistically. To that end, the Panel recommended eliminating Treatment 1 and adding a positive control treatment, as well as redesigning the treatment "tunnel" apparatus using heat and CO_2 as attractants. The Panel also recommended using a set number of adult male bed bugs (maintained for 7 days after feeding to satiation) residing in a vessel with harborage that would be placed into the test system farthest away from the attractants. The Panel recommended that the test be run under a normal L:D circadian cycle regime with mortality and repellence examined only in the dark phase of the regime.

Charge Issue 7(b). Is use of an artificial membrane system to simulate an animal host and provide blood for questing bed bugs adequate, or is an animal host necessary? The Panel recommended redesign of the testing apparatus eliminating the need for a membrane feeding device or live animals.

Charge Issue 7(c). Adequacy of the assessment period. The Panel settled on the assessment period of 24 hours as the best standard and that it would capture any circadian cycle differences among wild bed bug populations.

Charge Issue 7(d). Modifications or additional tests that would improve pesticide impregnated product testing. The Panel's recommendations to modify this test are discussed in the response to Charge Issue 7(a).

Charge Issue 7(e). Adequacy of the experimental unit to provide an experimental design and adequate data to evaluate repellency. The Panel's discussions, as elaborated in parts (a) through (e) of this Charge, recommended major modifications to the proposed experimental unit that are expected to make the test simpler, easier to run, and generate data easier to analyze.

Charge Issue 7(f). What type(s) of data analysis and statistical testing would be most appropriate for these data sets? With the Panel's recommended redesign of the experimental unit, the subsequent data analysis will be straightforward and follow the general guidance outlined in the Panel responses to Charge Issues 2 and 6. While the assessment period of 24 hours was deemed adequate, it was unclear whether single or multiple observation times are most appropriate, with the latter requiring data analysis that accommodates correlations induced by repeatedly observing the same bed bugs. Such correlations have been discussed in association with other tests. Each trial will use a group of bed bugs in which individual bed bugs are likely to act non-independently, hence, the analysis model will need to account for this "over-dispersion" effect in its estimation and testing.

Charge Issue 8(a). Adequacy of the experimental unit for testing indoor foggers and misters. The Panel concluded that the proposed experimental unit was not sufficient for testing fogger products because pesticide particles emitted from foggers settle and therefore, would not adequately penetrate the proposed PVC pipe design.

Charge Issue 8(b). Modifications or additional tests that could be recommended to improve indoor fogger testing. The Panel concluded that the following two questions required answering in the guidance document: 1) does the formulation kill bed bugs, and 2) does the delivery method penetrate harborages where bed bugs spend most of their time? An initial test was suggested using small scale droplets and direct exposure to determine dispersion efficacy of the fogger. A subsequent test was suggested using a mesh cage (in order to allow better particle movement) containing harborage such as folded filter paper, small pieces of wood, cardboard and fabric to resemble conditions and habitat of bed bug infestations. The Panel recommended using one susceptible strain, one resistant field strain and one resistant lab strain.

Charge Issue 8(c). What type(s)of data analysis and statistical testing would be most appropriate for these data sets? Due to the need for fogger type pesticides to penetrate into cracks and crevices where bed bugs are often found, the Panel recommended that this test be conducted using fed bed bugs that remain in harborages. An alternate design was suggested that uses an objective measurement of the amount of pesticide penetration in various kinds of harborages (without involving bed bugs), related to a direct application test of the insecticide on bed bugs at the measured levels. Inclusion of a validation test (using bed bugs) to make sure the two components "mesh" is necessary.

If the draft guideline design is used as proposed, and not the alternate design suggested by the Panel, the Panel recommended that each run of the Peet-Grady chamber be considered an experimental design block. Each such run should include as many of the levels of the treatment factors of interest as possible. Special handling of repeated measures data is necessary for this study design. The Panel suggested and described two methods that could be used for analysis of resulting data. One method is based on generalized linear models (or generalized linear mixed models) and the other is based on use of survival models (see Charge Issue 5(d)).

Charge Issue 9(a). Adequacy of the experimental unit for testing ovicidal products. The Panel concluded that Experimental unit 1 is more appropriate for an oviposition repellence trial and not for an ovicidal trial and recommended deleting it as an option. The Panel recommended a specific protocol to generate eggs and collect them from a cohort of females within a 2-day window. The eggs should be tested at the same time in their development and within the first 2 days of oviposition. The timeline to record egg mortality can be shortened to 10 to 14 days and the blood source for feeding the females should be consistent.

Charge Issue 9(b). Modifications or additional tests that could be recommended to improve ovicidal product testing. The Panel recommended using fed and mated females that are not older than eight weeks old to produce eggs for testing. Older females will produce smaller eggs which introduce surface area to volume ratio increases and an increase in pesticide susceptibility.

To achieve repeatability of experiments, egg-laden papers should be placed onto treatment papers rather than individual eggs.

Charge Issue 9(c). What type(s) of data analysis and statistical testing would be most appropriate for these data sets? For the ovicidal test, the Panel recommended the use of all eggs from one female as an experimental design blocking factor and the use of at least 20 female bed bugs.

Charge Issue 10). Please provide comments on the overall clarity, accuracy and completeness of the draft guidelines: "Laboratory Testing Methods for Bed Bug Pesticide Products". Please provide any additional comments that highlight any areas of the draft guidelines that may need to be clarified and note any relevant topics that may be missing. Please include references to any published literature that could help improve the completeness and clarity of the draft guidelines. The Panel offered a number of specific editorial changes and corrections/suggestions to the definitions section of the draft guidelines. In general, the Panel stated the importance of the draft guidelines being clearly organized to ensure that pesticide registrants readily know which tests are required to meet their desired label claims. The Panel also emphasized the importance of keeping the guidelines as simple and as clear as possible while preserving biological relevance. Bed bug husbandry was noted as an important aspect in the laboratory setting and should be emphasized early in the guidelines. Relevant references are provided. Additional references provide more information on selection of appropriate diagnostic/discriminating doses.

The Panel reiterated that any restrictions on randomization of any of the testing components can affect both the experimental design and data analysis. The Panel suggested that such restrictions be documented clearly in the protocol submitted with a manufacturer's application so that when the results are evaluated, it is clear how the experiment was actually run and, that the restrictions are correctly accounted for in the statistical analysis. The Panel encouraged the employment of a statistician by the registrant to support study design and data analysis and by EPA for evaluation of the submitted data.

Finally, the Panel stressed the importance of managing the development of bed bug resistance to insecticides and the need to consider resistance management in further development of the guideline.

DETAILED PANEL DELIBERATIONS AND RESPONSE TO CHARGE

Charge 1 - Laboratory test methods. The draft guidelines describe laboratory test methods for evaluating the efficacy of a variety of bed bug pesticide products. Please discuss:

(a) Whether, given the objectives and the types of products being evaluated, the test methods are appropriate to evaluate the efficacy of bed bug products with regard to kill and knockdown; repellency; pesticide resistance; and discriminating dose.

The Panel pointed out that several methods in the draft guidelines are not standardized and there are no publications that support how appropriate or repeatable these protocols may be for the particular type of evaluation proposed, largely because they do not take into account the biology and behavior of bed bugs. As an example, the protocol for "specific guidance for laboratory studies for pesticide impregnated material products" (p. 17) proposes to record the position of the bed bugs after feeding. However, after feeding occurs, bed bugs typically stay away from hosts and display harborage-seeking behavior. Consequently, in this particular protocol bed bugs might move out of the host area and abandon the tunnel apparatus.

The Panel emphasized that rearing bed buds in a laboratory setting is difficult and producing the number of bed bugs required for these tests may prove challenging. They suggested standardizing bed bug rearing systems because all systems currently used may not be interchangeable. Today, bed bug feeding membrane systems (using blood from different vertebrates) are used as well as live vertebrates such as chickens, rabbits, and humans in some cases. The Panel pointed out that blood source is important and bed bugs may become accustomed to particular food sources and react differently to different hosts. Also, blood storage has limitations.

The Panel noted that it is important to consider mortality versus morbidity and to include measures of morbidity over time. Morbidity that lasts over 72 - 96 hours is not usually something from which insects recover (moribundity), but registrants must understand the possibility of early morbidity and that subsequent recovery may occur. The question that needs to be answered is at what point is long term morbidity (or moribundity) considered the same as mortality?

Resistant field-collected bed bugs that are raised in the laboratory for a period of time will exhibit a decrease in resistance levels over time (docketed public comments from Miller; Romero personal communication). The question is then raised concerning how long a resistant field-collected strain should be maintained in the lab before it loses resistance? In addition, there are no criteria to select strains for resistance. Another question that is then raised is whether the resistant strain should be periodically treated with that pesticide to which it is already resistant in order to maintain that resistance (i.e., continuously challenge the colony) or replace it if resistance is lost?

The very beginning of the draft guidelines should more clearly define and outline that the intrinsic insecticide efficacy must be established before the insecticide delivery method (*i.e.*, fogger, residual spray, etc.) is tested. Further, it is important that the laboratory methods mimic the natural settings where bed bugs typically infest as closely as possible.

It is understood that the draft guidelines intend to capture the diversity of field bed bugs. However, it is unlikely to find populations of bed bugs with significantly different resistance profiles. The guess is that all field populations will have a background of resistance to pyrethroids, the most commonly used insecticide group for bed bug control in the United States in the last 10 years. The use of 5 strains, as recommended in the draft guidelines, increases the number of bed bugs required for each test enormously and, as noted previously, there are serious limitations in rearing systems for bed bugs. Many field bed bugs do not adapt to artificial feeding systems and if they do, it might take at least one year to produce the number of bed bugs required in the bioassays. Thus, the Panel proposed reconsidering the number of strains to be used and the number of replicates proposed in each test without affecting the test's scientific rigor.

(b) Whether there are additional or alternative laboratory test methods beyond those discussed in the draft guidelines for testing the efficacy of bed bug pesticide products.

The Panel pointed out that the draft guidelines do not include research bioassays for other types of insecticides that may be used for bed bug control, such as fumigants, slow-release insecticidal strips (*e.g.*, DDVP), slow-acting insecticides such as the metabolic insecticide chlorfenapyr, and Insect Growth Regulators (IGR).

IGR tests for resistant and susceptible strains could include direct sprays, residual (no-choice), and vapor or volatile component exposure tests. Variables that can be measured for IGR efficacy include incomplete eclosion, uneven cuticle, prolapsed gut, supernumerary nymphs, failure to develop to adult, and effects on egg laying and egg hatch. IGRs can be coupled with other toxicants and a rigorous test would include these formulations. Discussions of bioassays for IGRs are presented in Naylor et al., 2008 and Bajomi et al., 2011.

The Panel pointed out that there are no convenient tests to study the efficacy of volatile pesticides. Some tests that have been proposed earlier include testing DDVP in a type of gas-exchange chamber (Sun and Johnson 1963), testing the volatile effects of the IGR methoprene on German cockroaches (Atkinson et al., 1992), and other tests using fumigants such as sulfuryl fluoride (Vikane) and plant essential oils (Yang et al., 2004). Fumigants, which are often 100% effective, could be tested using a Peet-Grady chamber with deep harborages to determine LC_{50} or LT_{50}. Finally, research on aggregation, sex pheromones and attractants is missing. A test could be conducted for products that could potentially contain these elements and that could consider whether bed bugs orient to (such as follow a plume – considering the deterrent effect of turbulence) certain substances without competing stimuli and with competing stimuli.

Charge 2 - Standardized test system elements and conditions (Section (h) p. 10). The draft guidelines describe standard elements and conditions that are recommended for each type of efficacy evaluation. Please discuss:

(a) The appropriateness of suggested environmental conditions (such as temperature and humidity).

This charge issue discusses the recommended materials, procedures and environmental conditions for bed bug efficacy tests. Included under this topic are (1) test substance; (2) test organisms; (3) bed bug rearing, handling and maintenance; (4) testing conditions; and (5) disposition of test bed bugs. While not a direct charge question, it was noted that under "Test materials" [p.6; section (c)(1)(ii)] "Efficacy should be tested using end-use formulation as registered or as proposed for use." At least one Panel member inquired about formulations that contained one or more active ingredients, and was interested in knowing if the efficacy of individual components of the formulated product was considered or important.

Regarding appropriateness of environmental conditions, the draft guidelines state that "The temperature during the test should be kept at 25°C, with a relative humidity of 50-80%, and a photoperiod of 16:8. The temperature during the test should be kept as constant as possible because changes can affect the performance of the product treatments." The Panel agreed that maintaining constant temperature is critical for both reproducibility and for comparing of bioassays. While the suggested temperature of 25°C was acceptable, the Panel suggested adding a range or variation (*i.e.*, ± 1°C) to the draft guidelines. With regard to the humidity, the Panel concurred that a 30% range (50-80%) was quite wide, and believed a smaller range could be achieved by either humidity-controlled incubators or assay dishes being held above constant humidity salt solutions. Finally, the literature indicates that the photoperiodic regimes employed during bed bug research have varied from 12-16 hours of light to total darkness, but the Panel agreed that the photoperiodic regime specified in the draft guidelines (16L:8D) will ensure reproducibility.

(b) The proposed test organism. In particular, please discuss:

(i) Whether the testing should be done only with one species, the common bed bug* Cimex lectularius. *Whether it is appropriate to use unfed adult bed bugs with a sex ratio of 1:1 to test adulticide products.

The Panel agreed that testing should be done with the common bed bug *Cimex lectularius*. It was mentioned by one Panel member that should the tropical bed bug, *C. hemipterus,* become established in parts of the United States, a mechanism should be in place for adding this species to an existing label in a timely manner. Another Panel member mentioned that there is increasing evidence that *C. hemipterus* is moving into the temperate zone, underscoring this need. The Panel had several recommendations with regard to the feeding status and sex ratio of bed bugs used for efficacy testing. Adult bed bugs used for testing should be standardized as much as possible with respect to age, mating status and feeding status. All

adult bugs used for testing should be fed to repletion seven days prior to testing, and should be at least three weeks old. Given the difficulty of laboratory rearing (particularly with field-collected bed bugs) and the cost, the Panel agreed that insecticide testing should generally be conducted with adult males only, thereby preserving the breeding females for future colony maintenance (in bioassays where fewer numbers of bed bugs are required (e.g., testing impregnated materials) separate groups of males or females could be used. The differences between male and female bed bug responses to the exposure to most pesticides are not well known, but are likely to be minimal as is the case with pyrethroid pesticides.

(ii) The advantages and disadvantages of including a susceptible strain in every test and whether the Harlan strain is the preferred susceptible strain.

The advantage of using a susceptible strain when testing pyrethroid insecticides is that it acts as a control (for bed bug strains); using a susceptible strain is also necessary to calculate resistance ratios (Charge 3), particularly when dealing with pyrethroid-based insecticides since widespread resistance to pyrethroids has been documented in field populations of bed bugs. A susceptible strain also allows comparison of results across studies and provides a baseline efficacy of a product. The Harlan strain has been kept in the laboratory for over 40 years, is widely available, and relatively easy to maintain. The Panel agreed that the Harlan strain would be the preferred susceptible strain for insecticide testing (while noting behavioral differences with field strains that might affect results in some assays). It was noted that other susceptible strains do exist.

(iii) The advantages and disadvantages of including a resistant strain in every test and how an investigator should select a resistant strain or strains.

As mentioned in the previous paragraph, resistance to pyrethroid insecticides has been widely documented in field populations of bed bugs. The Panel thought it was important to note that until the time insecticide resistance is documented for other chemical classes (*e.g.*, neonicotinoids), the use of the term "resistant strain" should be synonymous with "pyrethroid-resistant strain." The advantage of using pyrethroid-resistant strains for insecticide testing lies in the fact that they may represent "real" field populations across the United States and they serve as a benchmark for comparing results from different studies or changes over time. The selection of resistant strains is covered in Charge 3 and involves comparing the susceptibility of field strains to laboratory (Harlan) strains of bed bugs to a particular insecticide. The Panel agreed that the primary disadvantage of using laboratory-reared resistant strains for insecticide testing is that their resistance to a particular insecticide decreases (lose resistance, reverting to susceptible) unless they are subject to insecticide pressure (exposure). This adds another variable that needs to be continually evaluated.

(iv) The adequacy of the use of field collected bed bug strains from urban areas in three regions of the U.S. to represent the variability in susceptibility of bed bug populations for conducting efficacy evaluations.

The Panel concluded that using strains from three urban regions of the United States would represent a huge and possibly unrealistic undertaking, given the need to collect and subsequently maintain field populations of bed bugs. Since the life history of field-collected bed bugs would be unknown, they would have to be maintained in a laboratory for a period of time in an attempt to normalize or standardize their mating and feeding status. Since most, if not all, field populations will demonstrate resistance to pyrethroids, one Panel member thought a single field strain would suffice for testing. Another Panel member thought this component could be completely omitted from the draft guidelines, but highly resistant strains would still need to be included in testing. There was consensus among the Panel that three field-collected populations is excessive.

One Panel member suggested that companies set up an organized exchange system, so each company would only have to maintain one colony of a field strain. Another suggestion by a Panel member was that field strains be tested at lower sample sizes (essentially for validation but not for concentration-determinations).

(v) The recommended approach for grouping and evaluating the data from the tested bed bug strains to assess efficacy of bed bug products.

The Panel addressed grouping and data evaluation in their statistical discussions about the individual study designs associated with each particular test protocol.

(c) The advantages and disadvantages of including a positive control in every test.

The Panel agreed that positive controls should be used in tests. When testing a new pyrethroid insecticide, it is assumed that a pyrethroid-resistant strain of bed bugs will be utilized. In this case, challenging the pyrethoid-resistant strain with deltamethrin (a pyrethroid that many field strains exhibit resistance to) would act as a positive control, insure that the strain of bed bugs is pyrethroid resistant, and would allow for the comparison of efficacy between the new compound and deltamethrin. For other insecticides with different modes of actions, there may not be suitable positive controls (assuming a pyrethroid-resistant strain of bed bugs is being used as the test organism). A disadvantage of using controls (positive and negative) in every test is that it adds an additional burden from the standpoint of bed bug numbers, replicates, etc.

General Study Design and Data Analysis Discussions

As part of the general discussion on standardized test system elements and conditions associated with Charge 2, the Panel provided a broad overview of issues related to general study design and statistical analyses of resulting data that may apply to most of the individual protocols presented in the draft guidelines. EPA requests and expects that data used for estimation of key

toxicity parameters and establishment of differences across concentration or time be derived from standardized test systems and be scientifically sound. What happens if a laboratory study is unable to replicate what happens in the real world? Full reality is sacrificed to gain greater control, greater standardization, and, hence, greater ability to establish concentration-response relationships and other required results. For these draft guidelines, efficacy needs to be defined quantitatively based on controlled experimental conditions. Additional research is needed to verify that experimentally-derived efficacy assessments correlate with real world efficacy, that is, field studies are necessary to validate results found in laboratory studies.

The Panel noted that it is not unusual for two experienced statisticians to have different opinions about what constitutes the best design or on exactly how to analyze a well-defined experiment. Each individual may analyze the resulting data in a slightly different manner, but if the study objectives are well defined, their study conclusions should not differ. Because study designs have to accommodate specific situations and available resources, it is not easy to specify study objectives such that there is an obvious unique study design. However, it is possible for experienced statisticians or scientists to assess the likelihood of success of a properly documented study design.

In general, the Panel recommended using the terminology "concentration-response" in place of "dose-response" throughout the draft guidelines. The Panel also believed that "efficacy", defined in the draft guidance as "the test dose that demonstrates bed bug kill or repellency as claimed on the product label" (p. 6), does not, from a statistical point of view, provide sufficient information to promote the use of good experimental designs.

STUDY DESIGN

Concentration Test Ranges - Confidence in the estimated LC_{50} is highest when this value is in the middle of the range of tested concentrations. An optimal study would place the middle concentration of the range of five nearest the estimated LC_{50} and the lowest and highest concentrations placed to produce 10% and 90% mortality, respectively. This may not result in the logarithmic scale of concentrations yielding a straight line relationship as stated in the draft guidelines (p. 11); indeed, other functions of concentration (*e.g.*, square root) may better yield a straight-line relationship. Studies where the concentration range is fully below or above the LC_{50} will result in highly uncertain estimates for this value because it requires extrapolation beyond the range of the available data. The problem will be much worse if the concentrations used are all below the LC_{50} and the LC_{90} is the key parameter. If this should occur, the LC_{90} estimate will be extremely uncertain. Researchers should explore the use of two-stage bioassays where the first stage utilizes three or four concentrations to roughly establish the concentration-response curve. This information can then be used to identify the concentrations used in the second stage that ensure that data is taken above and below the LC_{50} and to identify at least one concentration close to the expected LC_{90} (preferably bracketing it). It is important to have controls used in both stages to measure whether the test environment changes significantly between the two stages. If such an environmental change does occur, it may be taken into account in the final data analysis.

Researchers have a large number of available study designs, some of which yield better estimates than others. Choices must be made not only about which concentrations are tested, but also exactly how bed bugs are grouped and/or randomized to treatments and how treatments are implemented over time. For example, researchers have the option of testing all strains at one time, or of testing some strains at one time and others at another time. The latter is a poor design because it confounds strain and time (*i.e.*, any differences can be the result of differences in strain or time effects or some combination of both). There is likely to be an optimal study design that produces the smallest confidence interval (CI) for the parameter of interest (*e.g.*, LC_{90}) for a given sample size, that is, a design which most efficiently uses available bed bug and time resources. Unfortunately, there is no one design that is optimal for all situations. Clearly, researchers need to think critically about study design and, in many cases, it will be beneficial to consult with a statistician to determine what the best design is for their specific situation. The Panel has provided suggestions throughout these meeting minutes for improving the design of the experiments outlined in the draft guidelines. The same concepts used for estimating the LC_{50} or LC_{90} apply for studies aimed at estimating the LT_{50} or LT_{90}.

Level of Confidence - The discussion in the draft guidelines on representative sampling (p. 7) states "The sample size should be large enough to be likely to yield a definitive answer to the research question being addressed, and its size should be justified statistically, taking into account the specific characteristics of the proposed research and necessary accuracy and precision of the results." The draft guidelines, however, do not specify a target precision for any of the estimates (LC, LT and KD at the key 50 and 90 percentiles), nor do they specify the level of confidence at which statistical comparisons must be made. Statements like, but not limited to, the following, are necessary for establishing the size and effectiveness of study designs and should be expected:

- For studies where the goal is estimation of a key concentration parameter (LC_{50}, LC_{90}, KC_{50}, or KC_{90}) or time to death parameter (LT_{50} or LT_{90}), the sample size should be sufficient to ensure that expected 95% confidence intervals for the key parameters are no wider than x units, where x is a length measured in the same units as concentration.

- For studies where the goal is comparison of mortality (adult bed bugs or bed bug eggs) among levels of a factor (application concentrations, surface types, bed bug strains, etc.), the size of the study should be sufficient to ensure Type I error rate of $\alpha = 0.05$ and Type II error rate $\beta = 0.2$ (equivalently a power of 0.8) for a specified significant difference of y units in mortality rate.

The targets x and y should be specified in the study designs. The Panel suggested that EPA may wish to recommend acceptable values for x and y to ensure comparability among registrants in level of study effort. The Panel also suggested that EPA recommend use of typical Type I ($\alpha = 0.05$) and Type II ($\beta = 0.2$) error rates unless the registrant offers strong justification for alternative values.

Setting these targets will force study designers to take into account all of the factors that can affect uncertainty, such as underlying variability in susceptibility from bed bug to bed bug; correlated responses among bed bugs placed in the same experimental unit (grouped units); and differential responses due to developmental stage, gender, geography, and time of year. It also forces analysts to scrutinize their models for excessive lack of fit that result in increased confidence interval widths and for missing terms (especially random effects) that would result in decreased confidence interval widths (*e.g.*, the data are over-dispersed but this is ignored in the model) or weaker statistical comparisons.

Determining appropriate study size is not straightforward when the analysis incorporates a non-linear response function with multiple factors, some of which may be random effects. There are a number of statistical software packages available that can be used for sample size and power calculations, several of which deal exclusively with sample size issues. Also, there are general procedures that can be used for design situations where the objectives may not be those typically covered in "canned" programs. For example, if the objective is to find the minimum sample size necessary to estimate differences in LT_{50} values for two different surfaces at a specified confidence level from a proposed design, use the best available "guesstimates" of the design parameter values (*e.g.*, variability among bed bug strains) and the minimum difference between LT_{50} values required for the two surfaces. Choose a sample size (*e.g.*, n = 5) and, using the software that will be used eventually in the data analysis, generate a typical data file. Then use the software to analyze the generated data set to obtain an expected standard error and confidence interval width for the difference in LT_{50} values for that sample size (*e.g.*, n = 5). Compare these estimates of precision to the required guideline specifications. Adjust the sample size as needed and rerun the above steps until a sample size is obtained that can be expected to satisfy the specifications. Gbur et al. (2012, Chapter 7) provide additional details and worked out examples. The procedures presented there can be adapted to other statistical software packages. As a last resort, use a simpler model and analysis (such as a *t*-test of proportions, where proportions have been probit or arcsine transformed) to determine sample size with the assumption that confidence levels and/or Type II error rates of associated statistical tests will be smaller when done with the more complex model. It should be emphasized that such simplifications should be a last resort and the need for simplification should be justified.

Negative and Positive Control Treatments - While negative and positive controls are discussed in the draft guidelines (p. 7), their use is not consistently emphasized. For example, for the studies to determine the resistance ratio, the draft guideline emphasizes the use of a deltamethrin positive control for pyrethroid insecticides testing (p. 11, section (i)(2)(i)(E - "Positive Control")). The untreated control is hidden under section (B), "Number of Treatments." The requirement of an untreated control should have its own section. The deltamethrin (or some other standard pesticide) positive control should always be used if it makes sense to do so (Ballard docketed public comment). Having these two treatments as required parts of the protocol allows comparison of bed bug population survival among laboratories, among strains and over time, thereby helping to place a specific study into a broader context. It should be used for quality control within laboratories to ensure lack of population drift and changes in susceptibility. Also, since laboratory strains (including very resistant strains) lose resistance over

time unless they are constantly challenged, using a known pesticide during studies demonstrates that presumed "resistant strains" are still resistant (and "susceptible strains" still susceptible).

DATA ANALYSIS

Use of Appropriate Models - A number of the methods in the draft guidelines reference use of analysis of variance (ANOVA). ANOVA is a way of summarizing data variability and for allocating that variability to various components of the underlying analysis model; the usual assumption is that residuals are normally distributed and have the same variance for each treatment. All data analysis should explicitly describe the model being used and demonstrate whether or not the assumptions underlying the model have been met (Bailar and Piegorsch, 1997). For example, comparisons of mortality depend on the proportion of bed bugs that died. Proportions are typically related to binomial models. If normal (or Gaussian) models are assumed, these proportions have to be transformed in some way. One approach taught in basic statistics courses is to base comparisons on the arcsine square root transformation which would allow for 0% or 100% dead. Another approach is to base the analysis on transformations using probit or logit analysis, which set 0% or 100% dead to $-\infty$ or $+\infty$, respectively. Software is available for a broader class of models, known as generalized linear models (GLM) (McCullagh and Nelder, 1989; Gbur et al., 2012), which fit models directly to non-normal (*e.g.*, binomial) data using a probit or logit link function. A GLM approach to analysis of these data brings with it a consistent and well documented set of estimation and testing tools and does not rely on the ad hoc transformation of the dependent variable to probits for an analysis based on normal distributions. GLMs do not involve transforming the response variable, thereby allowing the data to remain on the original scale of measurement. In a GLM, the model for the mean is on a different scale, referred to as the link scale (or model scale). This is the scale on which the mean of the response variable is related to other study factors (treatment and design). GLMs have also been expanded to include Bayesian approaches to estimation and testing. Both linear models and GLMs allow the data analysis to incorporate factors into the model that account for structural aspects of the study design (so called fixed effects, such as temporal blocking and multiple strains), other restrictions on randomization and hierarchical levels of sampling (so called random effects).

Another important distinction when doing statistical modeling is how tests (*e.g.,* of differences) will be performed. The preferred method, with the most power (and sensitivity), comes from model-based inference, *e.g.*, using contrasts. This allows one to "borrow strength" from all the observations, not just those involved in the contrast, when estimating some parameters (*e.g.*, variances). Perhaps more importantly, it takes into account correlation in the parameter estimates while removing variation due to redundant parameters (*e.g.*, the intercept in a contrast of two means). The alternative (using out-of-model comparisons) estimates means and their variances from separate statistical models. Comparisons of means are made by subtracting one estimate from another and the standard error of this difference is obtained by taking the square root of the sum of the two variances.

The probit analysis, as used in the draft guidelines, falls into the class of linear models, but is very simplistic, not taking advantage of model-based inference and not incorporating design and experimental effects that capture the way the study was actually run.

Bed bugs placed in a treatment experimental unit as a group are likely to display similar responses since they are encountering the same environment as well as possibly being influenced by chemical or behavioral signals from others in the group. This creates correlations that must be taken into account in the model used for analysis. GLMs can be structured to take this effect into account in the estimation of parameters and confidence intervals, and when performing statistical tests. In addition, these models can be specified to take into account covariates or other independent variables that can affect variability and thus impact test conclusions and parameter estimates. Over-dispersion (more variation in the data than would be expected under the assumed distribution, typically binomial or Poisson) can result from a number of factors, including an incorrectly specified probability distribution for the response and an incorrect model for the mean (or transformed mean). Failure to correct for over-dispersion in the analysis can result in under-estimated standard errors and inflated test statistics that have large Type I misspecified error rates. Gbur et al., (2012) discuss this problem and offer solutions, particularly in the case where one is dealing with proportions. Morgan (1996) contains essays that also discuss many of these statistical issues using examples from toxicology.

In GLM model-based inference, a common link function (probit, logit, or other) and model structure is used to compare treatment and design factors from within the model. For all commonly used link functions, when these models incorporate a concentration (or time-to-event) factor, they tend to provide poor estimates and weak comparisons at very high concentration (time) values. For the typical link function used (probit, logit, log-probit, log-logit or complementary log-log) the LC_{50} or LT_{50} estimates will be roughly the same. The differences will be in the estimates of the tail percentiles (in this case, interest is primarily in the upper percentiles such as the LC_{90} or LT_{90}), especially if there is a wide range of concentrations.

Software to analyze GLMs or generalized linear mixed models (GLMM) is available in all professional statistical analysis packages. The free software environment, "The Comprehensive R Archive Network" (http://cran.r-project.org), has functions to allow specifying generalized linear (mixed) models and accomplishing associated estimation and testing. Faraway (2006) contains a discussion about the analysis of generalized linear models using R.

Goodness-of-Fit - Assessing goodness-of-fit for linear models (of which probit, logit and log-log models are members where the proportions have been transformed but they are not models in the GLM framework) is not straightforward. Goodness-of-fit is typically assessed by examination of the residuals and standardized residuals. Residuals of the fit to transformed data are useful to determine whether the linear model fits the transformed data, against alternative models that use the same transformation but look for systematic deviations from randomness (*e.g.*, a significant quadratic component to the fitted line). Throne et al. (1995) demonstrates an approach to assessing goodness-of-fit across different transformations that is relatively easy to

use and at the same time is a powerful tool for getting the concentration-responses modeled correctly. McCullagh and Nelder (1989) and Gbur et al. (2012) also provide discussions of statistical tools for the assessment of model adequacy. Assessing goodness-of-fit for GLMs is similar to that for linear models, but requires an understanding of what a "residual" means, for example, for binomial data.

Specify Which Analysis Models Should Be Considered - A large number of models are available for use in statistical analysis. There is value in EPA limiting the models considered to those traditionally used in the analysis of mortality data. In addition to the probit model, other linear models such as logit, log-log, complementary-log-log, log-logit and log-probit are used to analyze insect bioassay experimental data. These transformations are all feasible for use in the computational systems typically used by researchers/biologists performing these bioassays, and they are all amenable to goodness-of-fit assessment and model comparisons. Researchers should justify why a particular model is used for a specific set of data using graphical assessments, formal goodness-of-fit assessments, and comparisons of the Akaike Information Criterion (AIC) or its small sample corrected version (AICc) (Burnham and Anderson, 1998).

However, the Panel recommended the use of GLMs where binomial data are modeled directly, instead of a linear model where the proportion has been transformed to try to satisfy linear model assumptions, In general, when comparative analyses have been conducted using linear models and GLMs on the same data, results from GLMs have both more accurate parameter estimates and more accurate standard errors, which translates into more accurate *p* values. A major difference is that "over-dispersion" must be addressed when using GLMs; in linear models the over-dispersion becomes part of the residual error. For a discussion on these and other differences in these two model frameworks refer to Gbur et al., 2012.

Generalized Linear Models (GLM) - A GLM allows for many different distributions of the dependent variable, including binomial counts. In addition, this class of models has been generalized in various ways, and can account for random effects (GLMMs) and time series correlation (induced by repeatedly assessing the state of each bed bug over time). For normally distributed data, the time series correlation is modeled through the residual (and is independent from the mean). However, for binomial counts, where the variance (the sampling error) is a function of the mean, modeling the time series correlation through the sampling error creates what is known as a marginal model. These models are typically not difficult to estimate but questions exist about the soundness of their theoretical underpinnings (Gbur et al., 2012). The alternative is to use a conditional model, where the time series correlation is modeled as part of the linear predictor (*i.e.*, through the mean). However, modeling the time series correlation this way can result in estimation problems and produce unrealistic covariance parameter estimates. In any case, the time series correlation cannot be ignored.

An additional problem with use of GLMs (especially GLMMs) for binomial data is that all sources of variation need to be specified, otherwise standard errors are too small. One way to handle the over-dispersion that results from missing sources of variation (and other things that

might contribute to over-dispersion) is to include a random effect for each binomial sample. In a GLM (without any random effects), one can instead include an over-dispersion parameter.

In addition to modeling the time dependent correlation, time also needs to be modeled as a decreasing trend (trends might differ among levels of a factor). There are many ways to model the time trend. A common way is to fit it with a polynomial (usually 2nd order), which may require rescaling of the time variable (*e.g.*, using orthogonal polynomials) to decrease the correlation between the estimates of the linear and quadratic components. Alternatively, time can be scaled as a series of decreasing numbers that follow a known function (*e.g.*, exponential decay function), which is very interpretable if the data fit the function. For example, if the time points are (0, 1, 2, 4, 6), a rescaling is (10, 5, 2.5, 0.625, 0.156), where half the remaining bed bugs die each hour. A third alternative, if one has difficulty fitting a smooth function, is to consider time as discrete levels (which is equivalent to fitting a high degree polynomial).

Survival Models - Survival models may fit the data better, especially if 100% die after some time interval. This kind of model has different assumptions than a GLM (and there are various types of models within this class), but, for a reasonable number of subjects (here individual bed bugs), necessary sample sizes should easily be met. Given a sufficient sample size, these models are fairly robust, and like GLMs (and GLMMs), available in several statistical software packages (*e.g.*, SAS and R). The survival models allow one to test main effects (*e.g.*, differences between strains) in a similar way to a GLM. One thing that differs in this class of models is that they assume observations continue until all subjects (bed bugs) die; otherwise, the data are (right) censored. In contrast, GLMs have difficulty when all subjects (bed bugs) die in a group because the link functions used (*e.g.*, probit, logit) never allow for 100% dead. A drawback of these models is that random effects (such as blocking factors) are incompletely implemented; some models do allow for clustered data.

Acceptable Control Group Mortality Limits - The draft guidelines recommend use of Abbott's Formula to adjust treatment data for control group mortality. It does not place limits on the amount of control group mortality that is allowed other than to end data collection once control group mortality reaches 10% in certain situations. High control group mortality suggests poor study implementation or exceptionally "weak" study populations. Strong study results require very low control group mortality. The Panel recommended EPA specify the level of acceptable control group mortality for a study to be deemed acceptable as evidence (as well as including confidence intervals on the ratio).

Charge 3 - Specific guidance for laboratory studies for resistant ratio determination, characterization of bed bug strain susceptibility, and discriminating dose selection (Section (i) pp.11). The draft guidelines propose a method to collect the data necessary to calculate a resistance ratio. Resistance ratios should be calculated for bed bug populations collected from the field in three regions in the U.S. The lethal dose values to be used in these calculations are to be derived from probit analysis. For pyrethroid insecticides, deltamethrin is proposed as the laboratory standard. From the collected data and resistance ratio values, the procedure recommends an approach to select a discriminating dose for a product. Please discuss:

(a) Which insecticides should be chosen as standards for other insecticidal modes of action?

The range of pesticide classes and their respective modes of action in relation to bed bugs is limited [neonicotinoids, hydroprene, and a pyrrole (chlorfenapyr)], and the empirical evidence for resistance to any of these (apart from pyrethroids) is non-existent. The Panel believed that several issues should steer the guidance towards a simpler protocol. The charge question asked for 'benchmark' compounds to be identified under discrete modes of action. The Panel thought that this was (a) ambiguous, since this depends on how you define 'mode of action'; (b) difficult to provide robust guidance on since real-world efficacy of certain types of pesticides (*e.g.*, desiccants) requires that they are used in conjunction with other groups of compounds (*i.e.*, attractants and pyrethroids); and (c) that commercial formulations labeled as single types (*e.g.*, neonicotinoids, such as imidacloprid) often contain low concentrations of synergists from other groups (*e.g.*, pyrethroids). Because of all of these concerns, the Panel regarded this charge question as unresolvable as put.

However, the Panel arrived at a simple, practical solution to the core issue. The Panel's experience is that the only pesticide class to which resistance is seen in wild bed bug populations is pyrethroids. Deltamethrin is the standard for this group (and is flagged as such throughout the document). Given that 'natural' resistance is largely restricted to this group of pesticides, and that there is already an accepted standard, the discussion ended with the conclusion that the draft guideline focus on deltamethrin as the standard positive control and that is contextualized by pyrethroid resistance being the only known form of natural resistance. Should the situation change, EPA would need to amend the guidelines.

(b) Whether it is useful to compare the resistance ratios of bed bug field populations to the resistance ratio for a corresponding dose of deltamethrin when evaluating pyrethroid insecticides.

The Panel believed that this comparison is a useful exercise because it benchmarks the activity of the evaluated product against an agreed standard (deltamethrin). By using resistance ratios (RR) with Harlan as the susceptible strain, the tests are also benchmarked against a standardized strain baseline. However, the Panel believed that the protocol in the draft guidelines (p. 11, section 2(A)) should be replaced with a simpler, less expensive option. Instead of a 500 ml beaker (there are only 600 ml beakers now), one should use glass Petri dishes of a standard diameter (6 cm, or 10 cm). Moreover, the filter paper does not need to be taped to

prevent bed bugs from crawling underneath. If the paper is initially saturated with the test solution, both sides of the paper are treated. Finally, it is unlikely that 50-100 µl will saturate a piece of filter paper of the size necessary to cover the bottom of a 500 ml beaker. About 175-200 µl is necessary to saturate a filter paper disk 4.7 cm in diameter.

(c) Whether a resistance ratio of 100x is adequate to screen field strains for resistance against insecticides and if not, what other approaches are recommended for detecting resistance in bed bug populations.

The Panel's consensus was that a resistance ratio (RR) of 100x is appropriate. This outcome arose from a discussion about RRs set by other bodies (*e.g.*, World Health Organization (WHO) recommends 2x) and the RRs that scientists observe in the field (approximately 1000x). Given that a utilitarian RR has to be biologically realistic and practical, 100x was deemed a logical and practical ratio to adopt. [It was noted that the definition does not agree with the WHO, where the 'discriminating dose is double the LC_{99} (WHOPES, 1998) or twice the minimum concentration that kills 99.9% of mosquitoes of the susceptible strain (WHOPES, 2009)]

(d) What type(s) of data analysis and statistical testing would be most appropriate for these data sets?

There is a difference between what the draft guidelines refer to as a "probit model" and what the Panel is recommending (GLM model with a probit [or logit, etc.] link). The first is a transformation on the dependent variable so that a model based on normality can be used (similar to an arcsine square root transformation). The Panel recommended a fundamentally different class of models where the dependent variable is modeled as binomial counts rather than as samples from a normal distribution. This has important consequences on the characteristics of residuals, and that the data and the model are on different scales (*i.e.*, in a GLM, the model is on the probit or logit scale but the data scale remains binomial counts).

Regardless of the choice of link function (or transformation on the proportions if linear models are used), estimates are best made at or near LC_{50} (or similar measures) for several reasons. Competing models (*e.g.*, using logit rather than probit) produce a very similar relationship between the expected value of the dependent variable and the (transformed) concentration near LC_{50}, but the relationships diverge as one approaches 0% and 100% bed bugs affected. Thus, for estimates near LC_{50}, especially if the range of concentrations is small, the choice of link function or transformation on the proportions is less important. However, the link choice or transformation on the proportion starts to matter the further one departs from it (*i.e.*, at LC_{90}, the choice of transformation or link function makes a difference). Also, since an LC_{50} is in the center of the range of proportions, there are presumably data on both sides (lower proportions and higher proportions) that lend strength to estimating the LC_{50}. At higher concentrations, it becomes more difficult to bracket the target concentration (*e.g.*, LC_{90}).

When fitting a wide range of concentrations, or just high concentrations (*e.g.*, near LC_{90}), the link or transformation matters. There will be situations (*e.g.*, where the dose response curve is not symmetric about the LC_{50}) where the log-log, complementary log-log, log-logit or log-probit link (or transformation) will provide a noticeably better fit to the overall curve. Any of these links (or transformations) should be considered acceptable, but the one chosen for a specific situation or set of data should be justified by model comparisons (see discussion in Charge 2).

Neither the logit nor probit links, nor transformations (mentioned above) make much sense near their limits (0% and 100% bed bugs affected), since these are asymptotes that can never actually be reached without tweaking the models, which then has to be justified in some way. The Panel believed that estimating an LC_{90} is still acceptable as long as there are sufficient observations in that region and one has chosen the link or transformation that gives a good fit to the data. One potential problem when picking a specific value of concentration is that relationships may change if a different value is picked, so strain A may appear more resistant than strain B at LC_{50}, but the reverse is true at LC_{90} (this implies that slopes and intercepts for the two strains differ). For very resistant strains, it may not be possible to even estimate an LC_{50} but, at that point, it probably is not necessary to determine exactly how much more resistant this strain is than a susceptible one.

The draft guidelines suggest taking logs of concentration to produce a straight-line relationship with the probit model. The Panel recommended against that suggestion because taking the log of the concentration may, or may not, produce a straight-line relationship. The Panel could think of no theoretical reason that taking logs has to be the right transformation. Experience has often shown that the square root transformation on concentration will get the points to align better on a straight line and also allows one to include a concentration of zero, and the Panel encouraged applicants to explore a variety of transformations on concentration. Note that, since these are transformations on the independent variable, the usual measures of goodness-of-fit are applicable (*e.g.*, which transformation yields the lowest AIC or smallest residual sum of squares).

The Panel recommended including a standard error on resistance ratio estimate. Standard errors of ratios can be calculated using either a Taylor's series approximation to the variance of the ratio (Stuart and Ord, 1998) or by using a linear model with contrasts. In the case where concentration is log-transformed and a linear model is used for the analysis, the resistance ratio becomes a linear contrast and the standard error of that contrast can be obtained directly from the fitted linear model (see below). Since the LC_{50} values are estimated from sample data, they are random variables and have associated sampling distributions and standard error terms. The resistance ratio is, therefore, also a random variable, but one that will not have good statistical properties (ratios of random variables do not have good statistical properties). However, it is critical that precision targets (acceptable uncertainty limits) be placed on the estimates of the LC_{50} (or similar measures) as well as ratios involving them. Thus, the Panel recommended that estimates of the resistance ratio be accompanied by standard error estimates (draft guidelines p. 12, section (i)(3)(iii)).

An approximation of the variance of a ratio of two random quantities (using a first or a second order Taylor's series expansion) can be found at www.stat.cmu.edu/~hseltman/files/ratio.pdf. The formulas given there become a little simpler for the intended use if there is no covariance between the two estimates, so use of the formula will differ depending on how estimates of the LC_{50} are calculated. Estimating the third moment will be difficult (necessary for the second order approximation); hence, the first order approximation, involving just means and variances, would typically be used. This formula can also be found in Stuart and Ord (1998). The square root of this estimated variance is then the standard error of the ratio.

An alternate formulation for this ratio, and one that will have better statistical properties, is to use a linear function of the uncertainties in the two LC_{50} values used in the estimate, but on the log scale. That is, instead of using concentration for field strain/concentration for susceptible strain or log(concentration for field strain)/log(concentration for susceptible strain), one should use log(concentration for field strain/concentration for susceptible strain), since that can be estimated as log(concentration for field strain) - log(concentration for susceptible strain). If this is done using model-based inference, it becomes a linear contrast, which is easily done using off-the-shelf statistical software and will output a standard error of this difference (the square root of the linear function of the uncertainties). This approach also requires rescaling how large this ratio has to be to claim a strain is resistant. For example, if previously a ratio of $(\text{concentration}_f/\text{concentration}_s) > 100$, where f = field and s = susceptible, was needed, now the criterion is $\log(\text{concentration}_f/\text{concentration}_s) > \log(100) = 4.6$ (this example is in natural logs). This presupposes that working on the log scale makes sense, that is, if one takes logs of concentration, the logged concentrations have a straight line relationship with the dependent variable. As stated above, taking logs may not produce a straight line relationship.

The Panel recommended that EPA include sufficient information in the description of the protocol to be followed so that it is clear to EPA how the experiments were conducted. The standard format given in Appendix A can be used as a guideline. It is not clear in the draft guidelines exactly what the experimental unit is, what a replicate is, etc. Study design characteristics must carefully be defined if a proper analysis of the resulting data is to be performed. If EPA is not clear about how the study was designed and conducted (especially any restrictions on randomization), it will be unable to evaluate the submission. One Panel member suggested that EPA consider recommending guidance-related studies to be summarized in a standard format such as the one given in Appendix A. This format would allow a researcher to describe the experiment in such a way that others can quickly understand exactly what was done and what data were collected. Table 1 illustrates the use of this template as a summary for the proposed resistance ratio study. It should be noted that the design and treatment factors, as well as the randomization, is described in the form.

Table 1. Completed study design protocol summary for proposed resistance ratio study. Entries are based on a combination of the original EPA draft guideline and the Panel's recommendations for changes and are purely for illustrative purposes.

Design Components	
Product Being Tested:	Chemical X [active pyrethroid]
Experimental Unit:	Beakers with white filter paper taped to bottom [350]
Measurement Unit:	20 Bed bugs of one strain per beaker
Repeated Measurement:	None
Termination Rule:	After 24 hours observation
Exposure Period:	24 hours continuous exposure
Response:	Mortality, Knock-down
Design Factors:	
Strains:	5 [susceptible, resistant, 3 geographic wild strains]
Concentrations:	0.0001%, 0.001%, 0.01%, 0.1%, 1.0% active
Negative Control:	Diluent only
Positive Control:	deltamethrin
Randomization Factors:	
	50 beakers per treatment
	10 beakers per strain for each treatment
Analysis Approach	
Exclusions:	Exclude resistant strain - 0 mortality
Model Type:	GLM: link(probit); family (binomial)
Repeated Measures Factor:	None
Fixed Effects:	Bed bug strain, treatment (concentration), chemical
Random Effects:	beakers within strain by treatment
Parameters of Interest:	LC_{50}, LC_{90}, KC_{50}, KC_{90}
Goodness of Fit:	
AIC for best model:	xxx
Non-linear response:	quadratic not significant ($P<0.05$) for example

(e) The discriminating dose selection for products, including whether the LD90 value of the least susceptible field population is the best value to use as the basis for discriminating dose selection.

The draft guidelines stated that the discriminating dose be selected based on the LC_{90} for the least susceptible field population tested (p. 12, section (i)(2)(i)(I)). The Panel's discussion concluded that this should read "most susceptible" rather than "least susceptible." While agreeing that the LC_{90} is acceptable to use, they pointed out that the rule does not, however, take into account the relative uncertainty in the LC_{90} estimates. Consider the estimates given in Table 2 below. While the minimum LC_{90} at 0.05 came from field population 1, there is a 5% chance that field population 2 or field population 3 could see 90% kill with roughly half this dose. While concluding that the LC90 was acceptable to use, the Panel recommended that the protocol for selecting the discriminating dose take into account the uncertainty associated with the estimated LC_{90} values. This will reinforce the importance of getting the set of doses used in the study correct since use of doses that are "far" from the LC_{90} will result in a high estimate of uncertainty. This will result in a very wide confidence interval, which should alert the applicant (and EPA) that there was a problem estimating the LC_{90}. While the Panel was not asked for guidance about how wide an allowable confidence interval could be (since it was the Panel's recommendation to include a measure of uncertainty for estimates like LC_{90}), in much of biological research acceptable coefficients of variation range from 10% to 30% (the latter for behavioral studies, the former for physiological studies). Thus, Table 2 illustrates that when one takes into consideration the confidence intervals on the estimates, determining which is the most susceptible population may not be straight-forward. More data are needed to better separate these populations.

Table 2. Example estimates and confidence intervals illustrating the importance of uncertainty in selection of the strain and dose on which the discriminating dose is computed.

Field Population	LC_{90}	Lower 95% Bound	Upper 95% Bound
1	0.05	0.04	0.07
2	0.06	0.03	0.1
3	0.07	0.03	0.08

Charge 4 - Exposure times to product treatments [Sections (j) (p. 13) and (l) (p.17)]. For residual surface and impregnated material testing, the draft guideline proposes two possible approaches to exposing bed bugs to pesticide product applications in the laboratory: 1) single dose with a fixed exposure time of 24 hours followed by mortality assessments through 96 hours unless all bed bugs die or control mortality exceeds 10%; and 2) single dose with an exposure time that is continuous until all bed bugs die or control mortality exceeds 10%. The single dose is generally the lowest label recommended dose for the product. Please discuss:

(a) Whether the exposure times provide sufficient data to measure the efficacy of a bed bug pesticide or whether other exposure times or testing scenarios should be used.

Charge issue 4 addresses two components (exposure times/methods and mortality endpoints) that are also embedded in subsequent Charge issues associated with draft guidelines sections (j) and (l). One exposure scenario that the draft guidelines propose is a "fixed exposure period" of 24 hours following the transfer of test bed bugs to an untreated surface. It was not clear whether mortality assessments (at the lower time intervals of 15 minutes; 30 minutes; 1, 2, 4, 24 hours) were to be made during the exposure period or at time intervals (15 minutes; 30 minutes; 1, 2, 4, 24, 48, 72, and 96 hours) 24 hours post-exposure and after bed bugs had been transferred to an untreated surface. The other exposure scenario proposed in the draft guidelines is for a "continuous exposure period" wherein the bed bugs are never removed from the treated surface, and mortality assessment is recorded as a lethal time to 100% mortality.

With regard to section (j), the Panel believed that the single dose exposure time was sufficient. An important caveat, however, was that under field conditions, exposure times for foraging bed bugs (those leaving a harborage in search of a host, or those returning to a harborage after feeding) would be relatively short – minutes rather than hours. Assuming a harborage was 'untreated', shorter exposure periods (< 24 hours) might be necessary to better mimic field conditions. It was also noted that in section (l), exposure times in the 'tunnel test' may not be true exposure times because of behavioral variations in avoidance among bed bug strains.

(b) Whether percent mortality values or lethal time values should be used as endpoints to assess the success of product applications and the advantages and disadvantages of each one.

The Panel concluded that while both parameters yield important information about a product, 'percent mortality values' using fixed exposure times is the best option. The advantage is that this endpoint is more realistic; the disadvantage is that for an insecticide in which mortality is observed outside the reference times (*e.g.*, chlorfenapyr because of mode of action), modifications in observations for knockdown and mortality may need to be adjusted accordingly. The advantage of including 'lethal time values' as an endpoint in addition to 'percent mortality' is that it supplies additional information. The obvious disadvantage is that it requires another battery of tests.

Charge 5 - Specific guidance for laboratory studies for forced exposure (no-choice) residual surface treatment tests (Section (j) pp.13). The methods described for testing residual surface treatments recommend five types of surfaces: unpainted plywood; linoleum tile; concrete board; cotton sheet; and medium pile carpet. Please discuss:

(a) Whether the experimental unit described in the draft guidelines will provide sufficient data to show how effectively and how quickly a bed bug product knocks down and kills bed bugs, when applied as a residual treatment to surfaces.

When testing the efficacy of insecticide residuals, which may be among the most important pesticide tests because of the difficulty of directly treating bed bugs in harborages, providing surfaces and habitats that mimic their natural habitat (*i.e.*, human habitat) may provide the most accurate results. This is especially important because field testing is not included in these draft guidelines. For this and the other charge issues, it is important to emphasize that these are recommendations. If a manufacturer is interested in marketing a product for bed bug control in apartments or dormitories, for example, instead of hotels, those specifications and instructions for the end-user should inform the testing organization's choice of surface testing material.

With regard to the specific testing surface, the Panel recommended not using unpainted plywood, linoleum tile, and medium pile carpet if the goal of this test protocol is to quantify the effect of an agent on a surface that mimics a natural bed bug harborage. Instead, the Panel recommended using the following three treatment surfaces: 1) dry wall painted with a satin latex paint, 2) varnished wood paneling and 3) cotton cloth with stiff cardboard backing. Regarding the number of surfaces tested, the Panel concluded that 5 treated surfaces, 5 untreated surfaces with 5 replicates each, with 10 bed bugs per unit, weekly over 3 months (12 different test runs) is cumbersome and unnecessary. This would require a total of 6,000 bed bug adults per strain and the Panel recommended reducing this to 3 surfaces (treated and untreated = 6 total) and 2-3 strains (including a resistant and a field-collected strain). As several public commenters suggested, the LT_{50}, or length of time of exposure needed to kill 50% of bed bugs, should be the measure used in this test (p. 14 in the draft guidelines under section (j)(2)(E)(2), Continuous exposure period). Continuous exposure tests mimic treated harborages, and should be included in the draft guidelines. Timed, limited exposure tests mimic the wandering activity of bed bugs through the environment as they encounter sprayed surfaces.

In general, the Panel thought that this test, as proposed, is very large and complicated. No one particular research question is being asked and the test must be teased apart into its simplest components with the following questions considered:

1. Does a residue on X surface kill bed bugs? At what dose? (continuous exposure test).
2. Does a residue kill bed bugs that cross the surface? (limited exposure test).
3. Does a residue of an effective pesticide maintain toxicity and, and if so, for how long? (toxicity through time test)

(b) Whether there is a single surface type that could be used as a standard or representative surface for testing product residual activity in lieu of testing multiple surfaces as recommended in the draft guidelines.

As stated above, the Panel recommended using three surface types, depending on the expectation of how the product will be used by the end-user. Bed bugs do not inhabit carpets, although they may be found near the carpet edge next to a solid item or wall, and concrete is not a standard surface inside a home and is unlikely to be populated by bed bugs. However, varnished and unvarnished wood surfaces and other nonporous surfaces such as painted walls and wallpaper are relevant textured surfaces. Paints available for interior home use include flat, eggshell, satin, semi-gloss and gloss. Therefore, considering these factors, the Panel recommended dry wall painted with a satin latex paint (which satisfies the criteria of impervious and nonporous), varnished wood paneling (which is solid and textured) and cotton cloth with stiff cardboard backing (which is porous and absorptive) as the test surfaces to be used.

(c) Whether there are modifications or additional tests that could be recommended to improve residual surface treatment testing.

The test for residual efficacy is very important, but costly. To help reduce the cost of materials, the Panel recommended using a 4 x 4 inch square surface rather than a 6 x 6 inch square. Presumably this would not affect the experimental unit of the test or the outcomes. Further, a suitably sized Petri dish that fits on a test surface could be inverted onto the surface for testing and would not constitute a change in protocol. To further reduce costs, the Panel suggested using stiff cardboard for the "backing" material for the cotton cloth which would be less expensive and safer than glass.

It is worth reiterating that these tests should be performed with adults only, preferably all males. The Panel recommended that it would be helpful to clarify this point throughout the draft guidelines or at the beginning when defining the assay design.

If the goal of this test protocol is to quantify residual efficacy, fixed exposure period and recording of mortality seems excessive and the Panel recommended counting dead bed bugs at 1, 24, 48, and 72 hours, if necessary.

The Panel also thought it might be useful to elicit some natural behaviors to mimic the bed bug/host habitat. This will depend on what question the testing organization is trying to answer. Adding stimuli, such as host or harborage cues, may ensure more accurate results for the treatment of areas where bed bugs would be wandering or hiding. However, as pointed out in Charge 6(b), conditioning of experimental harborages (*e.g.*, use of bed bug fecal droppings) can be difficult to quantify and raises questions.

(d) What type(s) of data analysis and statistical testing would be most appropriate for these data sets?

A summary of this study design and its analysis approach is provided in Table 3. In addition to overall estimates of parameters of interest (LT_{50}, LT_{90}, KT_{50} and KT_{90}), this study design allows estimation of 95% confidence intervals and testing to determine answers to the following questions:

- Do parameters of interest differ among panel types?
- Do parameters of interest differ among week?
- Does weekly change in parameters of interest demonstrate an expected pattern (increase in LT and KT percentiles with increased weathering)?
- Is bed bug strain variability significantly greater than zero (bed bug effect if assumed random) or are there significant bed bug strain differences in parameters of interest?

A statistical issue discussed by the Panel is the extent to which the termination rule ("Observations terminated at 100% mortality in product treatment or 10% mortality in control treatment") would result in insufficient mortality to estimate the parameters of interest (LT_{50}, LT_{90}, KT_{50} or KT_{90}). Guidance should be provided on what the researcher should do if this happens. Most of the Panel members believed that high control mortality is highly suggestive of a failed experiment. In this case, the study would need to be repeated. It was not clear whether failed experiments would necessarily need to be reported to EPA. Too many failed experiments by one particular laboratory would suggest that the facility could be providing unreliable findings.

The Panel noted that some parameters in the current draft guidelines are left unspecified. For instance, the summary in Table 3 above assumes that all panels of one bed bug strain are started on the same day of the week, however, this could be done other ways. How exactly will observation times be assigned to days of the week? Further, other tests specify the sex ratio of adult bed bugs to be assigned to each panel/treatment/week. Do sex ratios need to be specified for this study? If not, why not?

The Panel also discussed the issue of whether bed bug strains should be a fixed or random effect in the analysis model. If bed bug strains are assumed to be a sample from the population of all wild strains of bed bugs in the United States, then bed bug strains becomes a "random effect" in the analysis model. If researchers are interested in determining whether there are differences among the specific bed bug strains used, then bed bug strain is a "fixed effect" in the model. This decision determines the type of model and test statistics used to examine the bed bug effect. A determination that bed bugs are "fixed effects" implies comparison of differences among strain means whereas "random effects" implies testing whether the associated variance is

Table 3. Summary of the proposed forced exposure (no-choice) residual surface treatment test. Entries are based on a combination of the original EPA draft guidelines and the Panel's recommendations for changes and are purely for illustrative purposes.

Design Components	
Product Being Tested:	Chemical X at minimum label rate - dilution from concentrate [alternate: Ready-to-use application]
Experimental Unit:	"panels" each 6" by 6" surface area [3000]
Measurement Unit:	10 Bed bugs of one strain per panel
Repeated Measurement:	Time [9] prespecified [15 min, 30 min, 60 min, etc.]
Termination Rule:	100% mortality in product treatment or 10% mortality in control
Exposure Period:	24 hours fixed exposure [alternate - to termination of study]
Response:	Mortality, Knock-down - at each observation time
Design Factors:	
Strains:	5 [susceptible, resistant, 3 geographic wild strains]
Concentrations:	minimum label rate
Negative Control:	Diluent only
Positive Control:	None
Surface type:	5 [plywood, linoleum tile, med pile carpet, concrete, cotton sheet
Week of test:	12 weeks [1,2,3,...]
Covariates:	volume, weight [alternate amount delivered to panel]
Randomization Factors:	
	600 panels per surface type
	300 to negative control, 300 to Product treated
	60 to week within surface type and treatment
	5 to bed bug strain within week, surface type, treatment
	All panels of one bed bug type started on same day of week.
Analysis Approach	
Exclusions:	Exclude resistant strain - 0 mortality
Model Type:	GLM: link(probit); family (binomial)
Repeated Measures Factor:	Time of observation [12]
Fixed Effects:	Surface type [5], Week [12], BB Strain [5], interactions [?]
Random Effects:	panels w/l surface, week and strain
Parameters of Interest:	LT_{50}, LT_{90}, KT_{50}, KT_{90}
Goodness of Fit:	
AIC for best model:	xxx
Non-linear response:	quadratic not significant (P<0.05) for example

greater than zero. Since estimating a variance (*e.g.*, bed bugs strains as a random effect) requires more than just a few units (*i.e.*, strains) (a statistical rule-of-thumb is that 30 units are necessary to estimate a variance well), most testing will consider bed bug strain as fixed. The downside to this approach is that the experiment's inference space is limited to just the strains used, rather than to the wider North American bed bug population, where the product will actually be applied. This reinforces the need to include the most resistant bed bug strains available when conducting tests.

The choice of using a 24-hour fixed exposure time or continuous exposure for the full length of the study does not impact the structure of the study design or of the underlying analysis model. This choice could have large practical considerations in that with continuous exposure, living bed bugs would have to be removed at the end of the week leaving less time for setting up the next week's test. This could result in restrictions on randomization, confounding of some factors and/or missing data. Similarly, use of low doses may result in lower exposures and longer times to death requiring longer observation times. The goal is to have adequate mortality in a reasonable time. Low exposure and/or slow kill treatments will require modification of some design components such as what is, or can be, observed in one week.

If the analysis plan specifies that one is interested in only making inferences at specific observation times (*i.e.*, only interested in comparing mortality at, for example, 24 hours), then there are 9 separate analyses performed, each with one response (no repeated measures). There is still an issue with using 9 separate analyses, suggesting that some control of the experiment-wise error rate (multiple comparison error) might have to be used.

The negative control treatment is important in establishing how long it takes before there is no difference between treated and untreated panels (residual pest control). While treated and untreated responses can be compared each week, a criterion for deciding when there is no longer sufficient chemical to control the pest should be specified.

The Panel noted that there are two different time dimensions to this study. Mortality and knock down are observed for exposed bed bugs at 9 prescribed times following placement of the bed bugs on each panel. The other time dimension is age of panel (week) which has to do with residual control. In a sense, for each panel one is able to establish a time response curve (form defined by the model) and to examine how this time response curve changes from week-to-week, strain-to-strain and surface type-to-surface type. Since a different set of panels are tested each week, the week factor is not a repeated measure. However, since there is an expectation of a residual pesticide effect, the time response curves (or more properly, the parameters that describe the time response curve) for each week may be correlated.

Deciding how to assign all of the required observations over a week takes careful evaluation. It is clear that in the current draft guidelines something will have to be confounded with day of the week or time of day. The design objective is to establish the priority among the three main effects (bed bug strain comparisons, surface comparisons, treatment-control differences) and to have a study design that does not confound the most important main effect

with day of week or time of day. The design rule is to confound day of week or time of day with the comparison that is of least importance. Therefore, if surface type difference is the least important comparison, one would make all measurements on a particular surface on the same day each week. In this case, day of week is confounded with the surface type effect.

Equally important are other restrictions on randomization, potential experimental related covariates (*e.g.*, amount of spray used on each panel), and correlations related to measurements taken over time (*e.g.*, repeated measurements or measurements that are taken in "batches"). These properties of study implementation should be noted in the study design summary so that they can be properly accounted for in the analysis model. An example of another "restriction on randomization" might be a situation where the registrant decides to study one panel type at a time, confounding panel type with study time.

The Panel discussed how time-to-failure or the survival trend might be analyzed with these data. Kaplan-Meier approaches (Miller, 1981) which model survival as a step function might typically be used in simpler experiments but are inadequate analysis models in that it is more difficult to accommodate random effects such as blocking or other randomization restrictions using this approach. Reasonable analysis models include GLMs and/or GLMMs assuming a binomial family and a probit or logit link function (Gbur et al., 2012). In these models, dose and time to death would typically be modeled with trend functions (*e.g.*, linear in the logit or probit), with bed bug strain and surface type as discrete factor levels. Week of study (to examine residual effect) could be looked at as 9 discrete factor levels or it might be modeled with a trend function (*e.g.*, an exponential decay between two thresholds). Blocking factors arising from restrictions on randomization would also be modeled as discrete factor levels. Even events such as missing values (skipped observations), lost bed bugs (failure to follow up) and early termination can be handled in these models. The KT and LT estimates and their standard errors can be computed for a combination of conditions and compared using appropriate contrasts.

If the same panel is used each week there could be issues with using continuous exposure instead of the 24-hour exposure. For example, as the residual pesticide decreases, the dose delivered to the bed bugs is less, resulting in less mortality. With continuous exposure, a stopping rule is needed that specifies exposure until a specific fraction of bed bugs on test is dead.

It is recognized that the draft guidelines specifically address study protocols that directly address bed bug mortality. The issue of pesticide residual suggests that, along with mortality studies, there needs to be some objective measure of the amount of pesticide remaining available and active on a treated surface over time. Previous studies will have established a concentration-response relationship between the active ingredient and bed bug mortality (direct contact studies). If direct measurement of remaining pesticide indicates that x amount is available on the surface, one should be able to use the concentration-response curve to estimate the expected number of bed bugs accumulating a toxic dose over time. This situation changes the mortality study from one with the primary objective of modeling of the concentration-mortality trends over

time to a study with the objective of validation of the expected mortality. This change in objective has the potential to reduce needed study resources and costs, but at the same time, it increases the risk of a failed study. The study could fail to validate that the concentration-response curve actually works in situations where the bed bug is not forced to be in continuous and direct contact with the pesticide.

The Panel also discussed potential design changes to the proposed forced exposure laboratory study. For example, instead of having a separate treated panel for each week, the same panel would be tested each week for the life of the study. In this case, the current week's panel treatment would be a combination of pesticide residual and "conditioning" of the panel through use in previous weeks by other bed bugs. This change in design would need to be reflected in both the design protocol summary and in the factors used in the analysis model. Therefore, for this example, this change has the effect of making "week of study" a second repeated measurement factor resulting in a more complex analysis model.

Some Panel members wondered if enough is known about bed bug biology and pesticide mechanisms of action to make the analysis model used to describe survival one that is more physiologically driven. After discussion, the Panel decided that there did not seem to be adequate knowledge at this time to incorporate this approach into the modeling.

Charge 6 - Specific guidance for laboratory studies to determine if bed bugs are repelled by, or attracted to, pesticide product residues (Section (k) p. 15). For pesticide product treatments to be efficacious, bed bugs must contact pesticide residues. Testing methods should determine whether or not pesticide residues alter bed bug behavior, either by repellence from or attraction to pesticide treated areas. Please discuss:

(a) Whether the experimental unit described in the draft guidelines will provide sufficient data to measure the duration and extent to which a pesticide product's residues repel bed bugs.

The Panel concluded that Section (k) (p. 15) of the draft guidelines should be a subset of Section (j) (pp 13), "forced exposure (no-choice) residual surface," because Section (k) does not meet the draft guideline's definition of a "repellent" [p.4, (b)(1)] which disrupts "host-seeking behavior...in the presence of a host." The Panel concluded that a more appropriate term to use in Section (k) is "avoidance". Under Section (j) one presumably has identified a pesticide that is efficacious against bed bugs (determined via 'no-choice' tests) and would now want to determine if the bed bugs will "avoid" a surface treated with that pesticide.

The Panel concluded that "duration" of avoidance would be difficult to measure and laboratory findings are difficult to translate into field situations; therefore, it may be moot. For instance, assume a pesticide-treated surface caused avoidance for 24 hours. A hungry bed bug might avoid a treated surface for 24 hours, subsequently contact that surface and acquire a lethal dose. Alternatively, if a hungry bed bug avoided a surface for longer time frames (not determined in these experiments), it might move away from the area to another area.

The experimental set-up, as described in the draft guidelines, consists of a covered Petri dish with at least two harborages, one of them treated with a relatively high concentration of a pesticide. The Panel expressed concern that, even if the test pesticide is only slightly volatile, the small confined air space would become saturated with the compound and the olfactory senses of the bed bugs would not be able to differentiate between the two harborages. Further, there are three choices (harborage A, harborage B, not in a harborage) complicating the analysis. Unless the Petri dishes were kept under constant surveillance, it will not be clear when an individual bed bug occupied its final location. Since bed bugs aggregate (except possibly a group of adult males), bed bugs would not be expected to exhibit a non-random distribution when given a choice of two identical harborages, so lack of independence in tests using groups is another complication. These factors make it difficult to determine if a product is repellent or attractive using the proposed set-up.

An alternative that should be considered is modifying the commonly used T-maze (sometimes the arms appear more like a "V"), which is a tool often used by researchers interested in preference for both vertebrates and invertebrates. The arms (top of the "T") should separate the stimuli sufficiently such that the olfactory system of the bed bug is not overwhelmed by odors emanating from the product in the test arm. Bed bugs are quite mobile (distance of a few meters does not discourage host seeking), so arms that would sufficiently separate the two stimuli should not be an issue. Traps could be placed at the end of the arms, so that once a choice was made the bed bug could not return to explore the other arm (this assumes that the repellency or attractiveness of the product is apparent to a bed bug before it physically contacts the stimulus). If the product is neutral, then the bed bugs should arbitrarily choose one of the two arms. A product that is attractive when far, but repels when close, would be indicated by a final location in the T-maze in the product arm but not trapped at the end. Whether or not a T-maze is used, confining a group of bed bugs to a small Petri dish to determine the attractiveness or repellency of a product is probably not a good idea. It should be noted, however, that the T-maze methodology has not been used specifically to evaluate bed bug responses to insecticides and would need to be standardized.

Finally, the guidelines do not mention the time of the day to run the bioassays. The Panel suggested conducting these experiments under the dark period of their light-dark cycle. Observation of bed bugs in darkness can be performed with a red light.

(b) Whether conditioning of experimental bed bug harborages is necessary.

Conditioning of harborages with bed bug fecal droppings adds considerable time and effort to these experiments and also raises numerous questions. The Panel pointed out that harborages conditioned with fecal droppings would be more attractive to bed bugs (Romero et al., 2009; Levinson and Bar Ilan, 1971; Parashar et al., 2003), but thought it would be quite difficult to quantify the amount and shelf-life of bed bug aggregating factors contained in the fecal droppings or the timeframe conditioned harborages remained attractive. Therefore, the Panel called into question the use of conditioning in the guidelines.

Results with pyrethroid-treated harborages have shown that aggregating factors in harborages overcome the repellent effect that pyrethroids are known to have in bed bugs. However, it is not known if the same would occur with other insecticides. Thus, the Panel proposed that assays with conditioned harborages be only included when the product has a non-pyrethroid active ingredient. On the other hand, this bioassay with bed bug harborages should be carefully standardized because bed bug responses to insecticide-treated harborages might well depend on the amount of aggregating factors present, and this would depend on how long the papers have been pre-exposed to bed bugs.

(c) Whether individual responses and/or group responses should be used to determine whether bed bugs are repelled by pesticide product residues in harborages.

The decision of whether to use individual bed bug responses or group responses is both a biological and a statistical issue. Affecting this decision are bed bug behavioral issues that are either unknown (*i.e.*, trail pheromones) or that might complicate the interpretation of the results (*e.g.*, individual bed bug movement not independent). The Panel concluded that these experiments should be conducted with individuals and groups, because both represent infestations normally seen in natural conditions. However, group evaluations complicate the statistical analysis.

(d) What type(s) of data analysis and statistical testing would be most appropriate for these data sets?

The repellency study design is summarized in Table 4 using the template provided in Appendix A. The Panel pointed out that if the set-up proposed in the draft guidelines is used (something many Panel members did not recommend), the analysis will be complicated and possibly will not yield useful data. An example of how non-independence of insect action affects the analysis is discussed in Domingue et al. (2010), where experiments were run with two harborages per Petri dish. The statistics used in this paper are complicated due to the non-ideal design of this assay system, and probably not the type of analysis a pesticide registration submitter would want to use.

There was confusion among some Panel members about whether or not the draft guidelines are considering three different experiments utilizing subsets of the product treatments defined by "experimental units" 1 to 3 under section (i) (A), each of which addresses one of the objectives listed in the Data Analysis section (iii). In any case, it is clear that the "treatment" with one bed bug per container is designed to provide the bed bug the opportunity to choose to harbor under the untreated or treated "tent". Observed at 24 hours, it is possible for the bed bug to be in one of the "tents" or to be outside both. This results in a multinomial response (with either 3 or 4 levels) for the "test" that occurs in each container. This can be simplified to a binomial if the response of interest is established as "found under the product treated tent" or "not found under the product treated tent." Analyses of multinomial responses are more

Table 4. Summary of the proposed study to assess if bed bugs are repelled by or attracted to pesticide product residues. Entries are based on a combination of the original EPA draft guidelines and the Panel's recommendations for changes and are purely for illustrative purposes.

Design Components	
Product Being Tested:	Chemical X at minimum label rate - dilution from conc [alternate: Ready-to-use application]
Experimental Unit:	"containers" [200]
Measurement Unit:	1 [alternate 10] Bed bugs of one strain by sex per container
Repeated Measurement:	None
Termination Rule:	None
Exposure Period:	H hours fixed exposure [specify from 1 to 24h]
Response:	Mortality, Knock-down, Location - at end of exposure [in Ctl, in Product treated, neither]
Design Factors:	
Strains:	5 [susceptible, resistant, 3 geographic wild strains]
Concentrations:	minimum label rate
Negative Control:	Diluent only [conditioned or unconditioned filter paper tent]
Positive Control:	None
Treatment	Product [applied to conditioned or unconditioned filter paper in tent]
Covariates	None
Randomization Factors:	
	Within strain - 20 bed bugs each of male and female tested
	Within strain by sex - 10 containers randomly assigned
	Within strain, sex, container - Product treat. randomly to L or R
Analysis Approach	
Exclusions:	None
Model Type:	GLM: link (probit); family (binomial for Mortality, KD; multinomial with 3 (or 5) categories)
Repeated Measures Factor:	None
Fixed Effects:	BB Strain [5] {none if BB assumed random effect}
Random Effects:	Container within sex. {BB strain if assumed not fixed}
	Probability of death at 24h, knockdown at 24h, in Product treated at 24 h
Parameters of Interest:	
Goodness of Fit:	
AIC for best model:	Not applicable
Non-linear response:	Not applicable

complex than analyses for a binomial response, especially if one is also interested in testing for the other main effects and strain differences.

If a group of 10 bed bugs is used instead of individual bed bugs, the result at the 24 hour observation would be a "composite value" consisting of a set of proportions {p1, p2, p3}, where p1 is the fraction of the 10 bed bugs that are found under the control tent, p2 is the fraction found under the product treated tent, and p3 is the fraction found under neither tent. The sum of the three proportions is 1, causing a dependency in these values that is characteristic of a multinomial response. If conditioning is included in the protocol (experiment #3) and one experiment is considered instead of three separate experiments, then there are five possible states for the "composite value" (under product treated and conditioned tent, product treated and unconditioned tent, untreated and unconditioned tent, untreated but conditioned tent and "none of these") whose observed proportions sum to 1. In addition, as pointed out by some Panel members, the group of 10 bed bugs is not likely to respond as 10 independent individuals but rather having some degree of correlation resulting in the quite likely observation of all bed bugs being at one of the 5 "nodes" of the compositional space. That is, for one "test" you might find all 10 bed bugs under the same one tent (say the untreated, unconditioned tent) in which case the corresponding composite value set of proportions would be something like {0,0,1,0,0}. This would result in reduced response heterogeneity, meaning that estimates will be less precise and differences more difficult to demonstrate statistically than if bed bugs had behaved independently. In extreme cases, it may not be possible to obtain estimates of all parameters without a very large number of replications ("containers") and bed bugs.

While approaches to analyze compositional responses are available (Aitchison, 1986) this is not a typical statistical analysis and it would likely require expert statistical help to accomplish. Treating the data as a multinomial logit in a generalized linear model context is difficult to accomplish as well. In either analysis, interpretation of the results is complicated. For this reason, a number of Panel members suggested that this design should be avoided due to the high likelihood of an improper analysis leading to incorrect conclusions. If multiple observation times are used in the protocol instead of only a single time, then a repeated measurement is included and the model and analysis must be modified to accommodate this. Repeated measures in the multinomial context require careful analysis and can be complicated.

The Panel recommended that the study design be modified to include only one type of tent in each container and decide on a fixed, shorter time on test. Each container would provide one replication for one product treatment. Different containers and bed bugs would be required for each of the product treatments for each replication. The inability of the recommended study protocol to provide the data needed to answer the question about product attractiveness or repellency to bed bugs led the Panel to suggest the simpler test protocol, but the simpler protocol would potentially require more bed bugs than the original proposal. With the simpler protocol, after a suitable orientation period, the container would be observed to determine if the bed bug was willing to go into the tent. If encouragement (light for example) were used to drive the bed bug to the tent, then the test would be one of repellency – the desire of the bed bug to harbor is countered by something in the product that drives the bed bug to stay out of the tent. The

randomization for this design is straightforward with bed bugs and tents assigned at random to "containers".

For the simplified protocol, each set of containers results in a set of binary responses that represent answers to the following questions. Did the bed bug die within the test period? Was the bed bug knocked down at the end of the test period? Did the bed bug go into the tent during the observation period? These binary response data can easily be analyzed with the methods previously described. This new study protocol uses more bed bugs and more containers than the proposed protocol but is simpler to implement and simpler to analyze.

Mortality and knockdown are binary responses and are not subject to the complications described above for multinomial data. They would be analyzed as described previously in this report.

The Panel also suggested considering alternative experimental layouts. One Panel member pointed out that if a layout similar to a T-maze were used, the analysis would be straightforward. Individual bed bugs can be tested and a strong effect (repellency or attractiveness) should be apparent in less than 100 trials (using a test comparing an observed proportion to 0.5). Testing bed bugs in groups should also be possible since there are no host attractants (heat and CO_2) nor harborages (so aggregating pheromones should not be an important issue for small numbers of bed bugs if the testing time is not too long). Pilot tests should be conducted with a blank stimulus at each end to verify this.

Charge 7 - Specific guidance for laboratory studies for testing pesticide impregnated material products (Section (l) p. 17). The draft guideline proposes no-choice and choice tests for treated materials. Both are to be conducted in the presence of a non-human host or artificial membrane system to provide a source of blood. The proposed tests evaluate mortality, blood feeding inhibition, and preference for treated versus untreated surfaces. Please discuss:

(a) Whether the experimental unit is adequate for evaluating mortality and blood feeding inhibition following exposure to impregnated materials.

The Panel members had difficulty understanding the proposed protocol for testing impregnated materials and their discussion resulted in recommendations for major modifications to the proposed study design. The recommended changes simplify the study and produce a design similar to that summarized in Charge 6, only in this design "containers" become "tunnel arenas" and the repellency outcome is "still in vessel/harborage", "in tunnel" or "in trap". The Panel's discussions quickly focused on altering the experimental unit to be logistically simpler and consequently less expensive to run and easier to analyze statistically. The Panel recommended that Treatment 1, in which the tunnel had half its length covered with the test product material and the remaining half untreated, be eliminated since it does not provide an exposure scenario that is substantially different from Treatment 2. The test would then consist of exposing bed bugs to Treatments 2 and 3 and a positive control using deltamethrin-treated fabric/textile included as Treatment 4.

The Panel suggested that the 'tunnel' should be manufactured from 4 mm polycarbonate twin wall greenhouse sheeting which should be stapled and the joints sealed with silicone sealant. The 'gutter' (rather than tunnel) should be 5 cm deep, 20 cm wide and 100 cm long with solid ends, but no top. The top edges of the wall should be treated with fluon and the bottom lined with fabric or textile. The attractant end of the tunnel should have a 36° C heat-mat and a CO_2 source that generates approximately 1200 ppm CO_2 at source (no more than 1200 ppm should be generated since higher levels become repellent). This 'attractant' or simulated host cues should be contained within 20 cm of the attractant-end wall. There should be a pitfall trap immediately in front of the attractant so that bed bugs moving towards it are prevented from moving forward or backward once they have entered the trap's boundary. A fluon-coated pitfall trap would be appropriate.

At the beginning of each trial, a set number of adult male bed bugs (maintained for 7 days after feeding to satiation) residing in a vessel with harborage (*e.g.*, a Petri dish with accordion-folded paper) should be placed into the apparatus at the end farthest away from the attractant in such a way that they can easily exit and re-enter the refuge. Over the life of the test, the bed bugs would have a choice between staying where they are or of going through the tunnel to the attractant. A conclusion that the product material has repellent properties would result if (1) bed bugs in the control trial choose to cross the untreated tunnel to the attractant and (2) bed bugs in the product treated trial choose to not cross the tunnel.

All bed bugs found in the pitfall traps are assumed to have traveled the tunnel at least once. Bed bugs might also die while crossing or after contacting the impregnated material and then returning to the vessel/harborage. Mortality is recorded for all bed bugs that are dead in the "arena" regardless of where they are found dead. The Panel recommended that the test run for 24 hours under a normal L:D circadian cycle regime with mortality and repellence examined only in the dark phase of the regime.

(b) Whether the use of an artificial membrane system to simulate an animal host and provide blood for questing bed bugs is adequate, or whether an animal host is necessary.

The Panel provided a number of recommendations to simplify the experimental unit in the response above, including the use of simulated host cues (heat and CO_2) as an attractant. Therefore, there is no need for a membrane-feeding device or for live animals. Calculation of bed bug blood feeding inhibition is also eliminated.

(c) Whether the assessment period is adequate.

The Panel concluded that the assessment period of 24 hours presents some issues but the time period was settled on as the best standard. In wild bed bugs, foraging is driven by strong circadian cycles and it is likely that different populations may have slightly different cycles. Running the apparatus for 24 hours will capture activity regardless of the inherent cyclicity of

the population. The assessment period represents a simple logistic set-up for the practitioner and provides enough time to ensure the disposition of bed bugs is accurately captured by the assay.

(d) Modifications or additional tests that would improve pesticide impregnated product testing.

Modifications are discussed under part 7(a) above.

(e) Adequacy of the experimental unit to provide an experimental design and adequate data to evaluate repellency.

The new design as proposed by the Panel is less expensive, simpler, more easily run and generates data that are easier to analyse. The revised design (presented herein) is a better fit to purpose.

(f) What type(s) of data analysis and statistical testing would be most appropriate for these data sets?

As noted previously, the Panel had difficulty understanding the proposed protocol for testing impregnated materials and, as a result, difficulty summarizing the design using the format provided in Appendix A. However, the following comments and suggestions were offered:

Experimental Unit - The protocol must still specify exactly how many "tunnels" will be used and whether tunnel reuse will be allowed. At least one Panel member suggested that it is traditional to use a new tunnel for each test (no reuse) since the cost of a tunnel is quite low. While conditioning of the tunnel was discussed, the Panel came to no conclusion on this issue.

Positive control - The Panel was unable to formulate a positive control treatment for this test since there are no known bed bug repellents.

Feeding Inhibition - Since the bed bugs will no longer be feeding, calculation of blood feeding inhibition is eliminated.

Observation times - The Panel was unclear on whether mortality and knockdown should be assessed at one study time (8 or 24 hours), or at multiple observation times. Multiple observation times result in repeated measures data and the analysis needs to accommodate the expected correlations induced by repeatedly observing the same bed bugs. See comments made for other Charge issues about correlations induced by repeatedly observing the same bed bugs.

Independence of Action - Since each trial uses a group of bed bugs and it is established that the behavior of bed bugs in groups is correlated, the analysis model will need to account for this effect ("over-dispersion") in its estimation and testing. Note that in this case, the effective sample size is smaller than the actual number of bed bugs used. This needs to be taken into

consideration when determining the statistical power of the experiment. The amount of over-dispersion differs by sex and possibly life stage. Domingue et al. (2010) found that male bed bugs in a group display near independent movement, so if only males are used in testing, "over-dispersion" may not be an issue. In the presence of over-dispersion, when the magnitude of the correlation parameter can be specified, an effective sample size can be computed (Schabenberger and Pierce, 2002).

Data analysis - As now envisioned, analysis of this experiment should be straightforward and follow the general guidance outlined in the Panel's responses to Charges 2 and 6.

Charge 8 - Specific guidance for laboratory studies of indoor fogger products (Section (m) pp.19). The methods described for indoor fogger testing recommend use of an experimental unit that includes a 216 cubic feet or larger Peet-Grady chamber. The test container for bed bugs has 20 holes (1/16" diameter) in it to simulate a crack and crevice treatment and a bed bug refuge. Please discuss:

(a) Whether the experimental unit is adequate for testing indoor foggers and misters.

The Panel concluded that the proposed experimental unit (a PVC pipe, capped and fitted with twenty 1/16 inch holes to allow pesticides in and prevent bed bug escape) would not be an adequate test chamber for foggers and misters. Bed bugs usually remain hidden most of the time in cracks and crevices. However, when infestations are heavy, bed bugs can be seen "exposed" or wandering. When insecticides are needed to eliminate a bed bug infestation, results are better if the spray, dust, etc., is placed directly into areas where pests are likely to be hiding.

Foggers deliver pesticide particles that initially float upward through the air and then settle. They are known to have low penetration into cracks and crevices. Therefore, fogger insecticides will not reach inside the PVC pipe due to low penetration, even if the pipe has holes.

(b) Modifications or additional tests that could be recommended to improve indoor fogger testing.

There are two different questions to be answered about the efficacy of foggers: 1) does the formulation kill bed bugs and 2) does the delivery method adequately place the pesticide into areas where bed bugs spend most of their time?

The Panel recommended initially testing the effectiveness of the fogger insecticide by exposing bed bugs directly to insecticide droplets at a small scale. In this assay bed bugs are not in harborages and do not have the option to seek a harborage. If the fogger insecticide is effective, then a second assay should be used that mimics natural infestations where bed bugs are more typically found in harborages.

The Panel concluded that the test container or unit should be different. A mesh cage is commonly used in efficacy trials for aerosol/fogger formulations for the control of mosquitoes (Bonds et al., 2010, for example). A mesh cage, specifically made of cloth mesh, such as commonly available and inexpensive window screening, would better allow the movement of droplets of aerosolized product inside the cage. To replicate a harborage inside such a cage, folded pieces of filter paper, cardboard, fabric, or small pieces of wood can be added to more closely resemble the conditions and habitat of a household bed bug infestation.

This type of cage will probably not elicit unnatural behaviors of bed bugs. The representation of residual pesticides on the mesh cage, if bed bugs wandered out of the harborage and encountered the mesh, will simulate bed bug contact with similar fabrics and porous surfaces within a typical household. Bed bugs that are hungry are more likely to encounter residual pesticides outside of their harborage, whereas bed bugs that are satiated will remain in the harborage. The feeding status of bed bugs used in this test should be considered, based on the question to be answered. For example: Does the pesticide get inside harborages? If so, use satiated bed bugs with realistic harborage. Does the pesticide leave sufficient residues in the right places to kill bed bugs wandering outside the harborage? If so, use unfed bed bugs with a realistic harborage, possibly with a stimulus to seek a host (heat source).

The use of a Peet-Grady chamber is standard in household insecticide bioassays, as well as many agricultural pest bioassays. It is adequate for use in the testing of indoor total release fogger products. However, it may be inadequate to consider caged bed bugs left outside the chamber as controls. A different untreated chamber or a chamber treated with water or diluent should be used.

The Panel recommended the use of one susceptible and one field strain with known resistance characteristics for this test and one resistant laboratory strain.

The more rigorous these fogger tests are to demonstrate efficacy, the better. The use of foggers is commonly thought to be ineffective for bed bug management. Anecdotally, it is believed that foggers are widely misused and overused for the control of bed bugs.

Regarding the inclusion of aerosol products in this type of test, one Panel member stated that "handheld aerosols" should be grouped with other tests, such as direct application testing of products. For example, some insecticide labels have directions that instruct the user to direct sprays into limited cracks and crevices or onto specific surfaces. Aerosols are not intended for use as foggers or space sprays.

(c) What type(s) of data analysis and statistical testing would be most appropriate for these data sets?

The Panel discussed major implementation issues in sections (a) and (b) of this charge that impacted the final study protocol. For this reason, this study design has not been summarized using the template presented in Appendix A.

The use of fed versus unfed bed bugs impacts the extent to which they stay in harborages. Since the goal of using a fogger is to kill all bed bugs (in harborages as well as those exposed), the Panel suggested using the more difficult target, *i.e.*, fed bed bugs in harborages.

The Panel suggested an alternate design that considers an objective measurement of the amount of pesticide penetration in various kinds of harborages (without involving bed bugs), then relate that to direct application tests of the insecticide on bed bugs at the measured levels. The extent to which bed bugs must be directly exposed to pesticides in the fog depends on how well the fog can penetrate into cracks and crevices. It was not clear whether the goal of the experiment was to test pesticide-in-fog contact mortality/knockdown or the extent to which the pesticide-in-fog gets deep enough into cracks to provide sufficient exposure to kill bed bugs in their harborages. Some of the Panel members thought that if the objective was related to penetration into cracks, then objective measurement of the amount of pesticide penetration in various kinds of harborages might provide as much information as running trial after trial in Peet-Grady chambers trying to kill bed bugs in cracks.

If the alternate design is followed, a validation test, using live bed bugs in a Peet-Grady chamber, should be included, but a large number of runs using this chamber appears to be unnecessary. However, if there are discrepancies between the validation test and the previous results, then an investigation into why that occurred is required and needs to be continued until resolved. The applicant needs to demonstrate that the validation (live bed bugs in a Peet-Grady chamber) yields results that were predicted from the direct application tests.

Indoor fogger products by their nature produce air and surface residual concentrations that vary in different areas of the room. The same is expected in the Peet-Grady chamber. Therefore, it is extremely important that design factors other than bed bug strain are considered. A short list of questions associated with the design factors discussed by the Panel include the following:

- Where are cages located on the wall (*i.e.*, are they in open wall spaces and/or corners between two walls and/or corners between a wall and the floor)?
- How high from the floor are cages located?

The appropriate negative controls would come from the same set-up in a duplicate Peet-Grady chamber, but with no fogger released. The bed bugs would be followed over time, like those exposed to the fogger. Death of controls would indicate that there is some general problem in maintaining (untreated) bed bugs and the Peet-Grady chamber tests will need to be repeated.

Bed bug Strains: It is best if all bed bug strains are included in each run of the chamber. Using only a single strain per chamber per run completely confounds the strain effect with the run and its associated variation in product delivery and any other changes in environmental conditions that may change from run-to-run. If more than four bed bug strains are required, as currently specified by the guidelines, then an incomplete block design of some form can be used to assign

a subset of strains to location-limited chamber characteristics such as walls, corners, etc. Each chamber run then forms a block with respect to bed bug strains. An example of such a design, outlined below in Figure 1, would allow four strains to be used in each run and require five runs to ensure each strain is tested four times. This is not the only possible design that could be used, and in this situation it would be best if a statistician familiar with advanced experimental designs was consulted. Cochran and Cox (1992) is a good reference for more advanced designs.

Repeated Measures: The observation times, in this case, time following fogger treatment, form a repeated measure and must be handled as such in subsequent analyses. The same important characteristics of repeated measurements discussed for designs in the previous charge questions are important here. Since the same bed bug (in a group with its cage-mates) will be followed over time, results from one time period to the next will be correlated. As a bed bug is either alive or dead each time it is checked (presumably removed if dead), the data are binomial in nature. The Panel suggested two kinds of analyses, one based on GLMs or GLMMs if random effects are included in the design, and one based on survival models (the models are different and have a different set of assumptions). Without data, the Panel could not make a recommendation as to which would be more suitable. However, survival models are unable to handle models with many random effects as are likely to occur in this experimental set-up.

Figure 1. Example of an incomplete block design for the Peet-Grady chamber.

Design:

Experimental unit: one cage with 10 bed bugs per cage

Factors:

treatments - 2 (treated, untreated control);
strains of bed bugs - 5 (coded A-E);
heights on wall - 3 (low, medium, high);
locations on wall - 2 (open space, corner);
walls - 4 (north, east, south, west);
cage orientations at each wall location - 2 (horizontal, vertical)

Block: one run in a Peet-Grady chamber

Number of blocks: 5 (coded 1-5)

Using a balanced incomplete block design, each run of the chamber can accommodate only one of the two treatments, hence there would need to be 10 runs - 5 runs with the untreated control, and 5 runs with the treatment. For each run, only 4 of the 5 bed bug strains can be accommodated since there are only 4 walls to the room. In five runs, each bed bug strain is replicated only 4 times, and each pair of bed bugs (e.g., A & B) only occurs together in the same run 3 times. One example of an allocation of bed bug strains to walls for the treatment runs is given below. On any wall, sufficient cages would be needed to ensure that cages were placed at all 3 heights, 2 locations and 2 cage orientations, requiring 12 cages per wall or 48 (12x4) cages per run.

	Bed bug strain				
Run of chamber	A	B	C	D	E
1	north	east	south	west	
2	east	south	west		north
3	south	west		north	east
4	west		north	east	south
5		north	east	south	west

Table entry is the wall of the chamber to which each bed bug strain is assigned for a run; blank means that bed bug strain is not used in that particular run.

Charge 9 - Specific guidance for laboratory studies for testing ovicidal products (Section (o) p. 22). Efficacious ovicidal products are likely to improve the effectiveness of bed bug management programs. The draft guidelines describe methods for evaluating product efficacy against eggs from direct application and contact with residual surface applications. Please discuss:

(a) Adequacy of the experimental unit for testing ovicidal products.

The Panel thought that Experimental Unit 1 is more appropriate as an oviposition repellence trial, not an ovicidal trial. Oviposition repellence is a different question and is unlikely to elicit candidates for repellence. There is an unequal probability of oviposition occurring on both strips of the filter paper called for in the draft guidelines, regardless of treatment. The female may lay eggs directly in the Petri dish. For these reasons, the Panel recommended deleting Experimental Unit 1.

The Panel recommended blood feeding a cohort of females and collecting their eggs within a 2-day window. This is done by feeding freshly eclosed females every week and mating them every 4 weeks, immediately after they feed. After the second mating, 4 weeks after eclosion, isolate 20 females in a 5 cm high by 3 cm in diameter sample tube with a fabric top and filter paper strips (2 cm x 0.75 cm) inside. Each chamber will generate 300 eggs per week if the females are fed weekly and mated once per month.

Eggs need to be tested at the same time in their development cycle and within the first 2 days of oviposition is most practical. By synchronizing 2-day cohorts of eggs, the testing laboratory will guarantee synchronized embryonic development and hatching. The timeline to record egg mortality can be shortened to 10 to 14 days. The testing laboratory should consider that the host blood source will influence female fecundity and egg viability; thus, all trials need to be fed on the same blood source. The Panel did not advocate one blood source over another, but the laboratory must be consistent in its choice.

(b) Modifications or additional tests that could be recommended to improve ovicidal product testing.

The Panel indicated that females must not be older than 8 weeks old, have mated, and be actively ovipositing over that period. The female must be fed and mated before testing. If the female rearing regime described above is followed, the female will deposit viable eggs after mating and with regular weekly blood meals. Egg phenotype changes with age of the female; eggs from older females are significantly smaller than eggs from younger females. Because surface area to volume ratio increases as egg size decreases, susceptibility to pesticides will also increase. The Panel recommended using eggs from females that are not older than 8 weeks old. Females older than 8 weeks can be returned to the colony for routine maintenance. To achieve repeatability of experiments, the testing laboratory should add egg-laden papers onto treatment papers rather than individual eggs.

(c) What type(s) of data analysis and statistical testing would be most appropriate for these data sets?

The ovicidal product laboratory study design utilizes direct application of pesticide product on bed bug eggs. The Panel discussed a number of aspects of this design and concluded that Experimental Unit 1 (a newly molted gravid female bed bug allowed 7 days to lay eggs on conditioned filter paper in a Petri dish) was not particularly feasible for a number of reasons. Hence, the statistical discussion will assume Experimental Unit 2 is used in the study design. An example of the summary for this study is given in Table 5.

The strongest experimental design is one in which multiple female bed bugs are used and where the eggs from one female are divided into two groups using random assignment, and then one group is randomly assigned to the product treatment and the other group assigned to the untreated control. In practice, two Petri dishes would be prepared for each female, one containing product treated paper and the other containing untreated paper. Roughly half of the eggs from one female would be placed in one Petri dish and half in the other with assignment at random. This restriction on randomization conceptually defines each female as a "complete block". The replicate in this case is the female bed bug (or more accurately the set of eggs from one female bed bug). The experiment then becomes a randomized complete block design with the measurement unit being the egg and the response being egg viability at specified observation times. Note that one benefit of this design is that not all eggs have to be available at the same time, hence, as a female bed bug begins laying eggs a new pair of Petri dishes come on test and are then observed independently of all other Petri dishes.

The Panel recommended that limits be placed on when and for how long eggs would be collected from a female bed bug. Typically, each female would lay about 4 eggs over a two to three day period.

Since the replicate in the above experiment is the "block" (all eggs from one female bed bug), the statistical power for comparisons depends on the number of female bed bugs as well as on the number of eggs. Power calculations can be carried out using the methods described in the statistical issues section of Charge 2 by using various combinations of females and eggs per female.

Table 5. Summary of design and analysis for testing the efficacy of bed bug ovicidal products. Entries are based on a combination of the original EPA draft guidelines and the Panel's recommendations for changes and are purely for illustrative purposes.

Design Components	
Product Being Tested:	Product X at label rate (containing pesticide y)
Experimental Unit:	group of eggs (half of eggs from one female) [40] - female BB as block [20]
Measurement Unit:	Egg
Repeated Measurement:	Every 24 hr up to 30 days
Termination Rule:	Fixed end of observation
Exposure Period:	Observation time (Contact on treated filter paper, time since direct application)
Response:	Viability at observation times
Design Factors:	
Strains:	Five (resistant, susceptible, 3 wild strains)
Concentrations:	minimum label rate
Negative Control:	Untreated
Positive Control:	None
Randomization Factors:	
	All eggs from one female randomly split two equal groups
	1 Group from a female randomly assigned to Treatment, other to negative control
Analysis Approach	
Exclusions:	None
Model Type:	Generalized linear mixed model
Repeated Measures Factor:	Observation time
Fixed Effects:	Treatment (Product treated/untreated)
Random Effects:	Block (female bed bug ID)
Parameters of Interest:	Probability not viable at specific time or LT_{10}, LT_{50} (times to viable)
Goodness of Fit:	
AIC for best model:	
Non-linear response:	

The experimental design based on blocks (all eggs from one female bed bug) would be straightforward to analyze. Each egg produces a binary response (viable/non-viable) at each observation time. If one observation time is used, the analysis model is a generalized linear mixed model with treatment group as a fixed effect and female (block) as a random effect. Because the response is binomial, it is possible, and in fact, preferable, to include a treatment × female interaction term in the model. A logit or probit link function would typically be used. If all observation times are included in one analysis, observation time is a repeated measurement. Some Panel members recommended a single observation time which would simplify the analysis.

It is possible to run this experiment by disregarding the specific source of the eggs (disregard blocks). In this case, egg-to-egg variability is combined with female-to-female variability for testing purposes, but would likely lead to over-dispersion issues that must be addressed when fitting the model.

Charge 10 - Please provide comments on the overall clarity, accuracy and completeness of the draft guidelines: "Laboratory Testing Methods for Bed Bug Pesticide Products". Please provide any additional comments that highlight any areas of the draft guidelines that may need to be clarified and note any relevant topics that may be missing. Please include references to any published literature that could help improve the completeness and clarity of the draft guidelines.

The Panel suggested a number of editorial comments that would either correct certain statements in the draft guidelines or generally improve them for their intended purpose. Two broader based comments were: 1) the title lacks the word "efficacy" and it should be included and 2) on p. 3 in the "Introduction (2) Organization of the Guideline", it states "provide general guidance applicable to protocol development for bed bug efficacy testing conducted in the laboratory and field". This draft guideline only proposes laboratory tests, not field tests, and the statement should be revised to reflect that.

Specific comments were offered by the Panel regarding clarification of the definitions of terminology (pp 4-5) used in the draft guidelines. Those comments for the individual terms are as follows:

- "Crossing" (p. 4, (b)(2)(i) - It would be helpful to add "...or over a treated surface to another untreated surface." There might be a situation encountered with bed bugs, where they must cross a treated surface to reach a host (untreated surface). This should also include resting on a surface.

- "Host" (p. 4, (b)(2)(ii) - The definition for "host" is included under a category of "events" and it is not an event. Host should be defined above or below, not as a part of this section.

- "Questing" (p. 4, (b)(2)(iii) - Questing is generally associated with (hard) tick behavior. "Host-seeking" is a term generally applied to hungry bed bugs seeking a blood meal.

- "Mortality" (p. 4, (b)(4) - What if a bed bug is not dead, but unlikely to recover (host-seek, feed, mate, etc.)? Would a registrant simply wait (time) until death? In other words, can a statement be made that "the LT_{90} is 7 days, though no feeding will occur on days 1-6"?

- "Pesticide Resistance" (p. 5, (b)(13) - "Expected level of control" would be clarified if it stated "expected level of mortality".

- "Resistant Bed Bug Population" (p. 5, (b)(15) - Use of the word "normal" here should be changed to or coupled with the definition of "susceptible" or "baseline", as is done in the definition of "discriminating dose" (p.5, (b)(17).

- "Discriminating Dose" (p.5, (b)(17) - Definition does not agree with definition on p.12 (I) due to possible typographical error on p.12. Also, the original definition does not agree with the publications listed below in which a discriminating dose is used as a measure of insecticide resistance. Suggest redefining as "a discriminating dose is used to measure insecticide resistance, and is usually a multiple (*e.g.*, 2X or 100X) of the minimum dose required to kill 90% (or 95% or 99%) of a susceptible bed bug population."

Additional comments provided by the Panel are discussed in each of the topic areas below.

Scientific Design of Research (p. 6)

Regarding "Test materials" (p. 6, 1(ii) - "Efficacy should be tested using the end formulation as registered...", the Panel asked about formulations that contain more than one active ingredient. For instance, there is a registered product that contains both imidacloprid (21%) and beta cyfluthrin (10.5%). Would there be interest in knowing the efficacy of the individual pesticides? A question was also raised about products with several "inerts". Would there be interest in knowing the efficacy of the active ingredient and/or whether the inert has a synergistic action? Therefore, it seems using the "end formulation" can raise important questions.

The Panel thought that the unit of measure (fl ounces/cubic foot) under "Dose determination" (p. 6, 1(iii)) is confusing. This may be a useful concentration under field conditions, but is not particularly applicable in the laboratory. The LD_{50} (the dose) is specifically difficult to define in these tests because there is no way to determine the dose that a bed bug has picked up on its tarsi or its abdomen through movement or activity. The unit of measure should be concentration per area or mg/ai and then time.

Of the $6 - 10 billion per year global pesticide market, less than 2% is dedicated for vector control and far less than that for bed bug control. There are only four main groups of insecticides with different modes of action. If the efficacy testing guidelines are very complex, registrants will be discouraged from bringing new products to market for bed bug control. The Panel sought to offer suggestions and recommendations to simplify and clarify the test protocols, as much as possible, while preserving the biological relevance. The Panel believed it is important to have products that are safe, effective and help manage resistance development in the process.

The Panel believed it is important that registrants and their testing laboratories conduct good research. The guidelines should be as helpful to that end as possible and should therefore, be clearly arranged to help the registrant know what tests they need to conduct (define objectives of each test) by their desired label claims. Providing "if-then" statements may be helpful, such as the following examples:

- "If you want to know whether your product has an effective residual on various surfaces, then conduct a study as described in the guideline for residual surface treatments with the following parameters (age, sex, mating, feeding status, strains.)"
- "If you want to show that your product kills bed bugs within their harborage, and your product is an impregnated fabric, then conduct a study as described in the guideline for testing pesticide impregnated material products".

The Panel suggested a format for design and reporting analysis, experimental unit, and blocking that the registrant could complete to certify that they have thought about the parameters of the test design. They also pointed out that it should be clear to the user of the guidelines that the intrinsic insecticidal activity (efficacy) must be determined before application or delivery method efficacy is tested. This would save time and money for the registrant.

Biology and Ecology

The Panel emphasized the importance of bed bug husbandry. Many groups and laboratories have their own rearing methods and the draft guidelines should provide guidance early on about key aspects of bed bug husbandry.

The Panel pointed out the lack of references regarding bed bug behavior and ecology. To be meaningful, the tests in the draft guidelines need a biological context and the Panel was concerned that the applied community may not be aware of the pure research that is available. Additional references that could provide information regarding this issue were noted (Usinger, 1966; Reinhardt and Siva-Jothy, 2007; and Siva-Jothy, 2006). The Panel included references that could provide additional information that may be useful in selecting appropriate diagnostic/discriminating doses (Halliday and Burnham, 1990; WHOPES, 2009).

Statistics

As pointed out earlier, any restriction on the randomization of any of the components used in carrying out an experiment (including allocation of individual bed bugs, cages, fabrics, allocation of cages to a wall, batches of sprayed panels, date of experiment, etc.) affects both the experimental design and its analysis. Restrictions on randomization are a normal feature of experiments and the Panel suggested that they be documented clearly in the protocol submitted with a manufacturer's application so that when the results are evaluated it is clear how the experiment was actually run and that these restrictions are correctly accounted for in the statistical analysis. In general, restrictions on randomization create blocking variables (*e.g.*, repeating an entire experiment five times, each done two weeks apart), which the Panel suggested treating as random effects. The Panel encouraged that a statistician be employed at the design and analysis (manufacturer submitting claim) stage and another at the evaluation (EPA) stage to verify that the analysis has been conducted following the way the experiment was run and that the design and statistical model match.

Resistance Concerns

The Panel emphasized that EPA consider the development of pesticide resistance when further developing the guidelines. Since the development of resistance is a general concern for bed bug control, as well as for other arthropod pests, the Panel encouraged EPA to consider resistance development as the guidelines progress and during evaluation of efficacy testing data. Repeated applications of a single chemical that does not eliminate all the bed bugs in the treated area encourages the development of resistance to that chemical (or class of chemicals). Mixtures of two or more active chemicals from different classes are less likely to lead to resistance. The development of resistance is a problem not only because it makes bed bugs more difficult to kill, it also forces use of concentrations that may harm humans or pets if exposed to the chemicals repeatedly over many years, and potentially cause environmental damage as treated furniture, carpets, etc. eventually find their way to landfills.

At the same time, the Panel was mindful that EPA is trying to encourage new product development to increase ways of killing bed bugs, and that will not occur with complicated and onerous guidelines. The Panel asked that EPA consider whether use of the candidate product is likely to lead to resistance quickly and suggest ways that would decrease the chance of resistance developing. Perhaps simply mentioning in the guidelines that EPA does consider how rapidly bed bugs could develop resistance to the product might encourage manufacturers to consider that in their product formulation.

Miscellaneous Comment: Several Panel members commented that the "lubricant" used for confining bed bugs in some of the proposed test, when necessary, should be inert, and standardized to limit any unintended variables. In addition, the recommendation in the draft guidelines that bed bugs should be "tightly secured to prevent bed bug escape" may prove more difficult than assumed, depending on the testing surface.

REFERENCES

Aitchison, J. 1986. The Statistical Analysis of Compositional Data. Chapman and Hall, London. 416 p.

Atkinson, T.H., P.G. Koehler, R.S. Patterson. 1992. Volatile effects of insect growth regulators against the German cockroach (Dictyoptera: Blattellidae). J Med Entomol. 29(2): 364-367

Bailar, A. J. and W. W. Piegorsch. 1997. Chapter 7 - Dose Response Modeling in Statistics for Environmental Biology and Toxicology, Chapman & Hall, London.

Bajomi, D., J. Szilagyi, R. Naylor, L. Takacs. 2011. S-Methoprene formulations laboratory tests for efficacy against bed bugs. Proceedings of the Seventh Annual ICUP. 249-256.

Bonds, J.A.S, M. Greer, J. Coughlin, V. Patel. 2010. Caged mosquito bioassay: Studies on cage exposure pathways, effects of mesh on pesticide filtration, and mosquito containment American Mosquito Control Association, 26(1):50–56.

Burnham, K.P. and Anderson, D.R. 1998. Model Selection and Inference: A Practical information-theoretic Approach. Springer-Verlag, N.Y. 353 pp.

Cochran, W. G. and G. M. Cox. 1992. Experimental Designs (Second edition), John Wiley & Sons, Inc., NY.

Domingue, M.J., M. Kramer, M. F. Feldlaufer. 2010. Sexual dimorphism of bed bug (*Cimex lectularis*) attraction and aggregation responses to cuticular extracts from nymph exuviae. Physiol Entomol. 35(3):203-213.

Faraway, J.J. 2006. Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models, Chapman and Hall/CRC, Boca Raton, FL.

Gbur, E. E., W. W. Stroup, K.S. McCarter, S. Durham, L. J. Young, M. Christman, M. West, M. Kramer. 2012. Analysis of Generalized Linear Mixed models in the Agricultural and Natural Resources Sciences, American Society of Agronomy, Madison, WI.

Halliday, W.R. and Burnham, K.P. 1990. Choosing the optimal diagnostic dose for monitoring insecticide resistance. J Econ Entomol. 83(4): 1151-1159.

Levinson, H.Z. and A.R. Bar Ilan. 1971. Assembling and alerting scents produced by the bedbug *Cimex lectularius*. Experientia. 27:102-103.

McCullagh, P. and J. A. Nelder. 1989. Generalized Linear Models (Second edition), Chapman and Hall, NY.

Miller, R. G., Jr. 1981. Survival Analysis, John Wiley & Sons, Inc., NY.

Morgan, B. J. T. (editor). 1996. Statistics in Toxicology, Oxford University Press, NY.

Naylor, R., D. Bajomi, C. Boase. 2008. Efficacy of (S)-methoprene against *Cimex lectularius* (Hemiptera: Cimicidae). Proceedings of the Sixth Annual ICUP, Hungary. Available at: http://www.icup.org.uk/reports%5CICUP863.pdf. 115-121

Parashar, B. D., K. Ganesan, D. Sukumaran, Y.V.S. Rao, V. Veer, and S. Prakash. 2003.Aggregation activity induced by excreta extracts of *Cimex hemipterus* (Hemiptera: Cimicidae). Entomon 28: 215-222.

Reinhardt, K. and M.T. Siva-Jothy. 2007. The biology of bed bugs (Cimicidae). *Annu Rev Entomol.* 52, 351–374.

Romero, A., M.F. Potter and K.F. Haynes. 2009. Behavioral responses of the bedbug to insecticide residues. *J Med Entomol.* 46(1): 51-57.

Schabenberger, O. and F.J. Pierce. 2002. Contemporary Statistical Models for the Plant and Soil Sciences. CRC Press, NY, NY.

Siva-Jothy, M. T. 2006. Trauma, disease and collateral damage: conflict in cimcids. Phil. Trans. R. Soc. Lond. B 361:269–275.

Stuart, A and Ord, K., 1998. Kendall's Advanced Theory of Statistics Sixth Edition Volume I Distribution Theory. Arnold, London, p. 351.

Sun, Y.P. and E.R. Johnson. 1963. A new bioassay technique, with special reference to the specific bioassay of DDVP insecticide. J Econ Entomol. 56(5): 635-641

Throne, J. E., D. K. Weaver, J. E. Baker. 1995. Probit analysis: Assessing goodness-of-fit based on backtransformation and residuals, J Econ Entomol, 88(5), 1513-1516.

Usinger, R., 1966. Monograph of Cimicidae (Hemiptera-Heteroptera). The Thomas Say Foundation, Vol. 7, Entomological Society of America, College Park, Maryland, USA.

WHOPES. 2009. Guidelines for efficacy testing on insecticides for indoor and outdoor ground-applied space spray applications. World Health organization, Geneva, Switzerland. Available at: http://whqlibdoc.who.int/hq/2009/WHO_HTM_NTD_WHOPES_2009.2_eng.pdf

Yang, Y-C., H-S Lee, J.M. Clark, Y-J Ahn. 2004. Insecticidal activity of plant essential oils against *Pediculus humanus capitis* (Anoplura: Pediculidae). J Med Entomol. 41(4): 699-704

Appendix A Proposed study protocol summary template.

Design Components	
Product Being Tested:	
Experimental Unit:	
Measurement Unit:	
Repeated Measurement:	
Termination Rule:	
Exposure Period:	
Response:	
Design Factors:	
Strains:	
Concentrations:	
Negative Control:	
Positive Control:	
Randomization Factors:	
Analysis Approach	
Exclusions:	
Model Type:	
Repeated Measures Factor:	
Fixed Effects:	
Random Effects:	
Parameters of Interest:	
Goodness of Fit:	
AIC for best model:	
Non-linear response:	