**Review of AHETF Sampling Design Documentation**
**Work Assignment 4-06 of Battelle Contract EP-W-04-021**

**Warren Strauss and Louise Ryan**
**May 22, 2008**

## Executive Summary:

We have reviewed the materials provided by EPA related to a multi-year pesticide handler worker exposure monitoring program. We acknowledge the complexity of the task and the formidable challenges of designing a cost effective and logistically feasible monitoring program. The Agricultural Handlers Exposure Task Force (AHETF) should be commended for it's efforts to address these challenges through a variety of creative strategies. We concur that a multistage sampling strategy is the most appropriate approach, though we have concerns about limitations in planned sample size. We also have concerns that over-reliance on purposive sampling strategies will lead to bias that will be difficult if not impossible to characterize with an acceptable degree of accuracy. We provide some suggestions on how the design could be improved, as well as some suggestions of analytical strategies that might be employed to improve prediction for specific groups of workers.

## 1. Introduction and Background

Based on a request from EPA's Office of Pesticide Programs, we conducted a review of various documents prepared by the Agricultural Handlers Exposure Task Force (AHETF) associated with the sampling design of studies intended to assess occupational exposures encountered by workers who mix, load and/or apply agricultural pesticide chemicals. This review was conducted under Work Assignment 4-06 of Battelle Contract EP-W-04-021, in which EPA identified various different sections of text in the AHETF documentation that were to be reviewed (Appendix A). The review focused on the following documents:

- ***AHETF Volume IV: AHETF Revised Governing Document for a Multi-Year Pesticide Handler Worker Exposure Monitoring Program.*** This report provides general guidance regarding the scope of the AHETF occupational exposure design, which will be applied to approximately 33 independent field studies used to quantify potential exposures that will be experienced by agricultural workers across a range of different job functions, equipment types, and product formulations. This document provides the foundation for more specific study designs that are being prepared by the AHETF.
- ***AHETF Volume I: General Information and Scenario Sampling Design.*** This document provides the more specific design for a series of field studies intended to characterize the range of exposures experienced by agricultural workers who apply pesticides using closed-cab airblast equipment. This document stratifies the closed-cab airblast exposure sample by crop type (nut crops, citrus, pome fruits, stone fruits, trellis crops) and geographic area (state) with more specific protocol information for the field experiments to be conducted within each stratum provided in separate documentation.
- ***AHETF Volume II: AHE55 – Protocol and Supporting Documents.*** This document defines the protocol for five field studies to be conducted on citrus crops in Florida.
- ***AHETF Volume II: AHE56 – Protocol and Supporting Documents.*** This document defines the protocol for five field studies to be conducted on pecan crops in Georgia.

- ***AHETF Volume VII: Reference Materials Cited in Other Volumes.*** This document provides information on AHETF's consultation with outside experts on the design for the closed-cab airblast studies.

The AHETF, in consultation with EPA/OPP, has adopted a purposive diversity sampling approach for the design of the field study experiments that are meant to characterize the range of potential occupational exposures for workers who mix, load and/or apply agricultural pesticide chemicals. Our understanding prior to reviewing these documents was that the purposive diversity sampling approach would incorporate sound statistical design principles (including representative sampling) whenever possible/practical, and would diverge from these principles with purposive selection only when necessary. We agreed with this strategy prior to the review, and our comments are partially based on our opinions about whether this has been achieved within the General Guidance Document and the more specific design information for the Closed-Cab Airblast Field Studies. Our review, while focused on specific subsections of the documentation listed above, was also intended to address two primary questions:

1. Is the overarching sampling design guidance proposed by AHETF (documented in Volume IV) reasonable, sufficient, and based on sound statistical principles?
2. Is the more specific design for the closed-cab airblast scenario (documented in Volumes I, II, III, and VII) also reasonable, sufficient, and based on sound statistical principles?

We have structured our review to address the above two primary questions, followed by some additional input regarding minor comments on the subsections of the text that were part of our review.

## 2. Comments on Overarching Design (Volume IV)

The AHETF Revised Governing Document (Volume IV) provides a set of design principles that will be followed by the industrial consortia when designing and implementing field studies that characterize potential occupational exposures for workers across a range of different job functions, equipment types, and product formulations. Among the material reviewed, the following principles are most germane to the adequacy of design:

***2.1 Scope of a Scenario:*** The AHEFT Revised Governing Document defines 33 different scenarios that represent a combination of job function (e.g. mixing/loading, pesticide application, seed treatment), equipment type (e.g. enclosed vs. open cab or cockpit), and product formulation (e.g. liquid, granule, solid, etc.). These 33 combinations were based on expert consultation, with the expectation that each combination may have a different distribution of occupational exposure (both dermal and inhalation) that would require separate studies to be completed. While it is outside our expertise in determining whether the 33 identified scenarios adequately cover the range of different occupational exposure profiles across agricultural industry, we agree with AHEFT's strategy of partitioning into well defined and separate exposure study scenarios.

***2.2 Target Population:*** The results of the planned field studies will be used to characterize the potential risk of occupational exposure for current and future populations of workers who handle pesticides. Given the fact that future populations of workers cannot be studied, the goal of the AHETF exposure studies is to characterize those factors that influence worker exposure in the current population of workers and assume that these factors will have similar influence in future populations of workers. Obviously, future populations of workers may be subjected to different equipment types, protective measures, use patterns, etc. than the current population of pesticide handlers. To this extent, it is important that AHETF collect as much information as possible that will allow future users of the data to make reasonable interpretation of the data in light of these future conditions. We believe that AHEFT has ensured that the appropriate factors that might influence exposure are being recorded as part of the data collection protocol in these field studies, and assume that the full study data will be available to allow for appropriate interpretation of the study results in light of changing worker conditions over time.

***2.3 Type of Study:*** To the extent feasible[1], the AHETF studies are intended to be observational in nature (in contrast to a controlled experiment). Based on work conducted by Cochran (1977) the series of observational exposure studies being planned by AHETF would be classified as "analytical" studies rather than "descriptive" studies. A "descriptive" study has the objective of describing a population of interest in terms of measurable characteristics (e.g. NHANES), whereas an "analytical" study has the objective of identifying differences between subpopulations of the population of interest. As described above, AHETF has partitioned the population of pesticide handling jobs into 33 different scenarios based on the combination of features such as job function, equipment types, and product formulation. Each scenario represents a different subpopulation which may have a separate and distinct distribution of exposure.

The actual exposure profile of an individual worker may represent a combination of these scenarios in real-world application. For example, a single worker may conduct mixing and loading using an open pour system, followed by application of the pesticide using open cab airblast equipment. To understand the total potential exposure for this worker, the results of the AHETF planned studies could be analytically combined to form an estimate. In order to do this within a descriptive study, AHETF would identify a sufficient sample size of workers who perform both tasks jointly as part of the study and conduct exposure assessment over the combined activities. Obviously, it would be difficult and resource inefficient to plan a study that covered all possible combinations that a worker might do in a typical day. AHETF has employed other clever design strategies (Normalization and use of a Generic Active Ingredient, both discussed below) that will allow users of the study data to analytically extend the results to a much broader range of pesticides and agricultural chemicals beyond those that are specifically used in these planned field studies.

***2.4 Representative Sampling and Constructing a Sampling Frame:*** Often, researchers consider descriptive studies as requiring the concept of representative sampling, and falsely conclude that analytical studies do not have this requirement. This fallacy is based on the fact that many analytical studies are designed and controlled experiments, where the experimental units are assumed to have identical characteristics prior to being studied. However, in the case of the AHETF planned studies, the requirement for representative sampling cannot be relaxed, because we wish to draw inferences on the distribution of exposure experienced within defined subpopulations of pesticide workers.

The AHETF correctly states that worker exposure is a complex combination of factors, some of which are defined within the scenario (job type, equipment and product formulation) and some that are not (worker behavior, training, clothing type, use of protective equipment, weather,

[1] Note that in some cases, the study may require that a participant handle more (or less) pesticide than they would normally use to ensure proper representation across the range of anticipated exposure for the study. This strategy will likely be necessary due to the small sample sizes planned within each scenario, and the possibility that the sampling strategy will not identify workers that would naturally handle the proper amount of pesticide to satisfy the sampling design.

getting out of an enclosed cab to adjust equipment, etc.). In order to draw inferences on the current/future population of workers, it is important that the sampling design provides adequate coverage across the range of these factors. It is also important that these factors be recorded, as discussed in 2.2 above.

Statisticians utilize the concept of a sampling frame to enumerate the members of the population of interest, such that a sample can be drawn from that population. The goal in representative sampling is that every member of the population has a known and positive probability of being included in the sample before the sample is drawn. The most simple and statistically efficient (in terms of precision) sampling method is the simple random sample – where each member of the population is listed as part of the sampling frame and has an equal chance of entering the sample.

Clearly, it will be difficult (if not impossible) to define a nationally representative sampling frame that would support simple random sampling for any of the 33 scenarios that is within the scope of the AHETF exposure studies. There is no list of all pesticide handlers who use open-cab airblast equipment to apply pesticides to trellis crops - nor is there likely to be a national list that defines the population of pesticide handlers that are defined within any of the other 32 AHETF defined scenarios.

***2.5 Multi-Stage Clustered Designs:*** When the simple random sample is not feasible, sampling statisticians often either limit the sampling frame to a smaller population, or employ more complex sampling methods such as multi-stage clustered designs. Limiting the sampling frame to a smaller population could be as simple as confining the study to a single state, such as California. The result of this action would be that the study would be representative of California, but may not be representative of the Nation as a whole.

To exemplify a multi-stage design, we consider the Open-Cab Airblast Scenario Design that is summarized in the Volume I documentation. Page 29 (of 52) in Volume I provides 2006 Summary Statistics from USDA on the number of bearing acres for each of 5 crop types (nut crops, citrus, pome fruits, stone fruits, and trellis crops) across States (and is reproduced in Section 3 below). Assuming that AHETF desires equal representation across the 5 crop types as described in the narrative, the first stage of sampling in a multi-stage design could be used to select a sample of States (e.g., 2-3) within each crop type to include in the study – where the probability of selection is proportional to the number of bearing acres. Assuming that these USDA statistics are also available by County (within State), a second stage of sampling could be used to select a small number of Counties within each of the States selected in the first stage of the design – where again, the probability of County selection is proportional to the number of bearing acres for the crop type being considered. By following this clustered design concept, we can reduce the listing requirements from the entire Nation, to a sample of a few counties in 2-3 States. Identifying the pool of Open-Cab Airblast pesticide handlers within these few counties should be much more feasible.

The term clustered design assumes that the results of an observational study performed on members of the population (growers and/or individual workers) that end up in the sample would have common characteristics because they are selected from within the same county or state. These workers could have similar training, similar background, or experience similar weather patterns that would make the results of their exposure monitoring more similar than results between workers/growers that are from different Counties or States.

Usually, when multi-stage designs are employed, the design calls for multiple study participants to be sampled from within the same cluster. This usually leads to a loss of statistical efficiency in terms of characterizing the distribution of exposure, but will also lead to large increases in operational efficiency that make these designs worthwhile pursuing.

The AHETF has incorporated concepts of a multi-stage clustered design in developing guidance for the Governing Document. The sample size calculations used for the scenario-specific study designs assume a two-stage design that properly adjusts for the potential loss of efficiency attributable to clustering.

***2.6 Random vs. Purposive Sampling and the Use of Stratification.*** The AHETF Governing Documentation makes a significant investment in defining different types of sampling approaches that include random sampling, purposive sampling, diversity sampling, and the use of stratification in both Chapter 9 and Appendix B. These definitions have been customized for the AHETF application, and are not necessarily consistent with conventional sampling theory. The Governing Document clearly states the divergence from conventional sampling theory, for example, using terms like selection rather than sampling.

As stated in the introduction, our understanding prior to reviewing these documents was that the purposive diversity sampling approach would incorporate sound statistical design principles (including representative sampling) whenever possible/practical, and would diverge from these principles with purposive selection only when necessary. We agreed with this strategy prior to the review, and are not convinced that this objective has been reached based on our review of the Volume IV Governing Document and the more specific information in Volume I which describes the more specific design associated with the Open Cab Airblast Scenario Field Studies. In particular, we focus on the following cited text from Section 9.2.2 of the Volume IV Governing Document:

> *"Diversity selection is an attempt to make a small set of MUs more useful for regulatory purposes when it is treated as an approximate (two-stage) random sample. Often, some characteristics that are likely to influence exposure are known or can at least be reasonably hypothesized. Diversity selection is any procedure that improves the chance that different MUs differ with respect to such characteristics. If these differing handler-day conditions used for the MUs are associated with exposure, then a diversity pseudo-sampling will tend to be more variable with respect to exposure than would be a same-sized representative sample. As a result, a diversity sample should tend to have more extreme exposures (both higher and lower) and fewer exposures in the middle. Thus, a diversity selection sample should tend to predict central tendencies of a future generic exposure distribution reasonably well but will tend to under predict lower percentiles and over-predict upper percentiles."*

Sample size issues aside, a sampling purist would argue that there is no compelling reason why AHETF should not adopt a more formal multi-stage probability-based sample design, and employ stratification to ensure representation across the range of normalized active-ingredient handled that is currently being used to define the currently planned diversity sample. As discussed previously, stratification could also be used to ensure that there is desired representation in the sample among other factors such as crop type.

Assuming that a formal multi-stage sampling approach is used to select a group of geographic areas to perform the study (i.e., counties within states) for a particular scenario, AHETF could develop a list of all possible pesticide handlers within those selected counties that are eligible for study within that scenario. With some reasonable effort, AHETF could then make contact with all growers/workers that are eligible to participate in the sample of geographic areas selected, and ask questions about their use of pesticide products to aid the process of stratification (i.e., determine how much active ingredient they use on a typical work day within the sampling period).

Assuming all (or a high proportion of) eligible members of the sample provide this information, the highest-stage sampling frame (workers within geographic areas selected within the lower stages of the design) can be properly stratified and a sample can be selected for the field study. Furthermore, sampling weights can be constructed which will allow proper statistical analysis of the study data which will avoid the biases that are associated with the diversity sample as characterized in Figure B4 in Appendix B of the Governing Document.

Although the planned sample size for the actual scenario-specific study is quite small – and may include only one worker who participates in the field study from each constructed stratum – the essential feature is that the sample of counties identified in the earlier stages collectively contain enough workers so that each stratum contains at least one eligible participant who is willing to enroll in the study.

The approach outlined in our comments above is not significantly different from what is outlined in the Governing Documentation from an operational efficiency perspective. This approach will also help improve the quality and acceptance of the study data, while allowing both the AHETF and EPA avoid criticism regarding scientific validity of the sampling design due to the potential for selection bias of unknown magnitude associated with purposive sampling.

***2.7 Use of a Generic Active Ingredient and Measures of Normalized Exposure:*** There are two key principles that allow AHETF to leverage a small number of dosimetry field studies conducted with a select few target chemicals to provide data across a much larger class of agricultural chemicals are described in Appendix B (Sections B2 and B3). The first principle (Generic Active Ingredient Principle) suggests that "*if all other conditions are the same, the magnitude of exposure does not depend on the particular active ingredient used*." This allows AHETF to identify a small number of surrogate chemicals to predict worker exposure across all pesticides that are consistent with a particular scenario (job type, equipment type and product formulation). The second principle (Normalized Exposure) suggests that the exposure results from each study should be normalized based on the amount of active ingredient handled by the worker during the workday. This would allow the results to be extended to other agricultural chemicals on the same normalized scale (amount of active ingredient handled). These two principles are extremely important to the validity of the design, but unfortunately are outside the expertise of the reviewers. These principles should be carefully reviewed by experts in human exposure assessment.

- That being said, we do have several specific questions related to these to principles:
- Is the generic active ingredient principle apply to mixtures of chemicals as well?
- When it comes to implementation, will MUs always be working with an assigned surrogate chemical? Or will it sometimes be acceptable to measure exposure for the chemical(s) that growers are already using, if their chemicals are part of the approved surrogate list?

***2.8 Adequacy of Sample Size:*** AHETF developed Benchmark Objectives of the AHETF Monitoring Program, and conducted power studies consistent with a two-stage design and an assumed log-normal distribution of worker exposure. The data quality objective suggests that for any statistic of interest (geometric mean, 95$^{th}$ percentile, etc.), $Max(\hat{\theta}/T_{2.5}, T_{97.5}/\hat{\theta}) \leq 3$, where $\hat{\theta}$ is the estimate of the statistic generated from the study data, and $T_{2.5}$ and $T_{97.5}$ are the lower and upper bounds of an appropriately constructed 95% confidence interval around the statistic. Using this data quality objective, if the observed geometric mean exposure value for inhalation were 100 μg of active ingredient inhaled per day, we would expect that the 95% confidence interval for the geometric mean to be no wider than (33.33 μg ai/day, 300 μg ai/day). This is a relatively wide confidence interval – but is likely influenced by the high cost associated with conducting the dosimetry field studies. AHETF also provides a secondary benchmark objective used to verify the assumption that exposure is proportional to the amount of active ingredient handled. The general model to assess this assumption is well conceived, and consistent with the assumed two-stage sampling approach.

The benchmark sample size calculations adjust for the potential effects of positive correlation among MUs that are observed within the same geographic area. However, the sample size calculations do not make any attempt to ensure that the statistics generated from the field study are unbiased. To do so, the sample size calculations would require the assumption of a more formal probability-based design and an adjustment for unequal probabilities of selection (associated with the planned diversity-based stratification scheme).

Figure B4 provides a nice summary of the potential biases that could be introduced in the diversity sample. In this figure, the simple random sample has wider confidence intervals for each of the sample statistics – but is unbiased. The diversity sample has more narrow confidence intervals, but is severely biased for the upper 95$^{th}$ percentile (to the point in which the generated 95% confidence interval is far removed from the true value). Our intuition suggests that if the statistics from the diversity sample were properly adjusted for the selection probabilities (weights), the diversity sample estimates would be unbiased – but would have wider confidence intervals (perhaps slightly wider than the simple random sample). Thus, properly accounting for the unequal probabilities of selection associated with a diversity-based stratified sampling design would (1) lead to unbiased estimates, and (2) require larger sample sizes to meet the same data quality objectives. Treating the data from the study as if they were elements of a simple random sample that had equal selection probability will lead to biased results, and should be avoided.

We note a telling few sentences on page 49 of Volume I: *".... This suggests that the estimates of upper percentiles will tend to be overestimated (and lower percentiles underestimated) in the resulting monitoring data. With the small samples sizes used in this scenario, however, such estimation bias is probably trivial relative to ordinary uncertainties due to sampling, whether random or purposive".* In our opinion, this is a dangerous admission and that points to a basic design flaw in the proposed monitoring program.

***2.9 Flexibility in Scenario Design:*** Appendix C of the Governing Document (Volume IV) provides several sections (C.8 through C.14, on pages 138-152) dedicated to optimizing the design in terms of meeting the data quality objectives and optimizing operational efficiency. AHETF concludes that a two-stage design with 5 geographic sites and 5 Monitoring Units per site achieve these objectives. Our interpretation is that there exists a family of designs that achieve similar performance (e.g. 4 sites with 8 MUs per site) and could very well match better with the different strata that need to be covered for a particular AHETF scenario. We would recommend that study planners take all factors into consideration (including power and operational efficiency) when optimizing the design – allowing greater flexibility in choosing an appropriate design (including so-called unbalanced designs) to meet the research objectives for each scenario.

***2.10 Sample size limitations:*** Page 15 talks about the challenges of estimating worker exposures when so many factors (the text uses the term parameters) can influence exposure in any given setting. The strategy taken to address this limitation is to establish representative scenarios, generating data for workers wearing the minimum recommended PPE, using data for one chemical/product as a surrogate for another and assembling generic databases that contain surrogate data applicable to many products. These are all good ideas. We wonder whether there might be some additional statistical strategies that could be used to sort out some of these factor effects. Because the current proposed strategy will analyze data one scenario at a time, it will not be possible to use statistical strategies such as regression modeling to adjust for the impact of varying factors on worker exposure. We believe that there could be added value to the concept of synthesizing information from multiple different scenarios into one large data base. "Scenario" could then be considered as one of the factors that influence exposure using statistical techniques such as multiple linear regression. Or, even better, characteristics of each scenario could be constructed as covariates to help predict exposure for specific types of workers. Potentially, it might also be possible to include scenario-specific random effects to help gain precision for specific subgroups. This strategy of borrowing strength has wide precedent, for example in the context of small area estimation and so on. Working with a synthesized database and applying some statistical modeling could potentially overcome some of the sample size limitations associated with analyzing the small number (25) of MUs to be assessed for each separate scenario.

***2.11 Biomonitoring:*** We concur with the rationale given on page 75 for passive dosimetry rather than biomonitoring. However, it seems a shame not to build in at least some biomonitoring as well. Most of the test chemicals being considered have established metabolytes that can be easily measured in urine or less easily measured in blood. Perhaps a subset of participants could be identified for additional biomonitoring.

***2.12 Conclusions on Review of Governing Document:*** The AHETF Governing document has incorporated many design features that are attractive, including

- The global design and implementation of 33 independent scenario-specific studies that are observational in nature, and whose results can be analytically combined to assess potential current and future worker exposures as a function of job type, equipment type, product formulation, and a variety of other factors;
- Employing the concepts of a Generic Active Ingredient and Normalization of Results to allow extension of the dosimetry studies to a wide array of pesticides and other agricultural chemicals;
- Utilizing a multi-stage design and accounting for potential effects of clustering to achieve data quality objectives while also maximizing implementation efficiency; and
- Ensuring that the sample collected spans across a range of values for the amount of active ingredient handled – which is believed to be one of the most important factors that will affect worker exposure levels.

The chief concern that we had with the Governing Document is that by abandoning probability-based representative sampling, AHETF has introduced the potential for biases of unknown magnitude to influence the results of this important series of studies. AHETF acknowledges that their recommended purposive diversity selection method will lead to biased results, particularly with respect to lower and upper percentile estimates that will be constructed from study data.
In our review of the governing document, we did not see any mention of the principle discussed with EPA – that AHETF study planners should adopt elements of probability-based sampling whenever feasible, and use purposive methods of selection as a last resort. Instead, it appeared that AHETF has abandoned the notion of using probability-based sampling at all stages of the design – opting for purposive selection instead. The basis for this decision appears to be operational efficiency – however, no careful analysis has been provided for how much more a

probability-based design would cost. These conclusions are based partially on our review of the Closed Cab Airblast Design summarized in Volume I, where we believe that a multi-stage probability-based design would be feasible and not much more costly to implement than the purposive diversity selection. This analysis is provided in the following section of our review.

Given (1) the high profile nature of the information being collected (which will support national risk assessments for current and future generations of workers), and (2) that crafting study designs that minimize the potential for bias are not significantly less resource efficient to implement – we think that AHETF should provide more careful analysis of the potential use of probability-based designs within this research program.

### 3. Comments on the Design of the Closed-Cab Airblast Scenario (Volume I)

Our comments in Section 2 above provide a comprehensive review of the Governing Document with respect to sampling design issues. Our comments suggest that (1) purposive sampling, as discussed in these documents, will lead to estimates that suffer from selection biases of unknown magnitude, and (2) that a probability-based design could be achieved with similar operational efficiency and perhaps only slightly increased cost.

Our comments on the design of the Closed-Cab Airblast Scenario (Volume I) are intended to illustrate how a probability-based design could be accomplished with a similar level of operational efficiency. The design approach outlined below is primarily based on information that was contained in Table 2 of the document on page 29 of 52 (which we reproduce below). We assume that the 2006 USDA reported number of bearing acres for each crop-type can be used as a surrogate measure for the amount of pesticide that is applied by workers using Closed-Cab Airblast Equipment in a manner consistent with this scenario.

AHETF decided to stratify the design by the 5 identified crop types, purposively selecting a single (and unique) state to perform the field studies for each crop-type strata, with the expectation that 5 MUs would be selected (across a range of active ingredient amount handled) in each state. AHETF selected the following states in the first stage of selection:

| Strata | State |
|---|---|
| 1. Citrus | Florida |
| 2. Trellis Crops | California |
| 3. Nut Crops | Georgia |
| 4. Pome Fruit | Washington |
| 5. Stone Fruit | Michigan |

**Table 2. Acreages for the Five Crop Types by State (NASS)***

| Crop Type | State | Bearing Acres, 2006 | Percent of Total US Acres for Crop Type |
|---|---|---|---|
| Nut Crops | CA | 797,667 | 53% |
| | TX | 180,719 | 12% |
| | GA | 128,550 | 9% |
| | OK | 85,740 | 6% |
| | NM | 37,763 | 3% |
| | OR | 28,300 | 2% |
| | HI | 17,800 | 1% |
| | 35 Others | 109,905 | 7% |
| Citrus | FL | 576,400 | 66% |
| | CA | 250,500 | 29% |
| | TX | 27,300 | 3% |
| | AZ | 23,500 | 3% |
| Pome Fruits | WA | 180,000 | 40% |
| | NY | 46,400 | 10% |
| | CA | 42,000 | 9% |
| | MI | 41,300 | 9% |
| | PA | 22,900 | 5% |
| | OR | 21,600 | 5% |
| | VA | 15,000 | 3% |
| | 28 Others | 76,100 | 17% |
| Stone Fruits | CA | 217,000 | 53% |
| | MI | 41,100 | 10% |
| | WA | 36,350 | 9% |
| | OR | 14,250 | 3% |
| | SC | 14,000 | 3% |
| | GA | 11,500 | 3% |
| | TX | 5,800 | 1% |
| | PA | 5,000 | 1% |
| | 22 Others | 66,800 | 16% |
| Trellis Crops | CA | 806,920 | 77% |
| | WA | 90,032 | 9% |
| | OR | 32,596 | 3% |
| | NY | 31,700 | 3% |
| | MI | 30,900 | 3% |
| | PA | 12,100 | 1% |
| | NJ | 7,600 | 1% |
| | CA | 7,000 | 1% |
| | 10 Others | 28,317 | 2% |

***Note: This table was reproduced from AHETF Volume I – General Information and Scenario Sampling Plan***

Upon inspection of the 5 strata, we noticed that approximately 49% of the bearing acreage, across all crops, was found in the State of California – and that California had the highest number of bearing acres for 3 of the 5 crops (Nut Crops, Stone Fruits and Trellis Crops), and the 2nd highest acreage for one of the crops (Citrus). Yet, AHETF's design only included sampling in California for one crop type (Trellis Crops).

We offer the following scheme for a probability-based design approach, as an alternative for consideration:

1. To ensure diversity of geographic location within crop-type, select 2 states per crop-type to include in the sampling design
   a. Assuming 5 MUs per strata (which may need to be adjusted based on sample size requirements), one state will have 3 MUs, and the other will have 2 MUs.
2. Within each stratum (crop-type) sampling of states will occur with probability proportional to the number of bearing acres
   a. If a state within a stratum represents 50% or more of the proportion of bearing acres (e.g. CA for Nut Crops, Stone Fruits and Trellis Crops, and FL for Citrus), it is selected with certainty to become the 3 MU State for that stratum.
   b. Similarly, if a state within a stratum represents between 30% and 49% of the proportion of bearing acres (e.g. WA for Pome Fruits), it is selected with certainty to become the 2 MU State for that stratum.
   c. All other states are selected with probability proportional to the number of bearing acres
3. The probability of selection of each State is recorded for use in developing appropriate sampling weights at later stages of the design.

The following table provides an example of how the design would appear for a single hypothetical realization:

| Strata | State Selected | Number of MUs | Probability of Selection |
|---|---|---|---|
| 1. Citrus | Florida | 3 | 1.00 |
| | California | 2 | 0.83 |
| 2. Trellis Crops | California | 3 | 1.00 |
| | Oregon | 2 | 0.13 |
| 3. Nut Crops | California | 3 | 1.00 |
| | Texas | 2 | 0.26 |
| 4. Pome Fruit | Michigan | 3 | 0.15 |
| | Washington | 2 | 1.00 |
| 5. Stone Fruit | California | 3 | 1.00 |
| | Michigan | 2 | 0.21 |

In this realization of the design, California is represented in 4 of the 5 crop-type strata with 11 MUs (of 25) being assigned. Recall that approximately 49% of the bearing acres across crop types were located in California, and 44% of our sample (11/25) is located in California. This hypothetical probability-based design includes 6 states in total (California, Florida, Oregon, Texas, Michigan and Washington) – compared to 5 states in the purposive diversity sample. We assume that operational efficiency will be similar between the two designs (as a matter of fact, the probability-based design might be more efficient due to the larger number of samples planned in the single state of California). The costs may be marginally higher of course due to the spreading of MUs over states.

In the next stage of the design, counties from within the above selected states could be sampled (again proportional to size according to the number of bearing acres). AHETF could sample one or more counties for each MU to be eventually studied, in order to identify a large enough pool of workers to cover the range of active ingredient handled that will be used as a stratification variable in a later stage of the design. For example, the pool of pesticide handlers using Closed-Cab Airblast equipment might be sufficiently large in each possible county identified such that only 2-3 counties will need to be enumerated per MU to identify a sufficient pool of growers/workers. Identification of a sufficient pool of pesticide handlers for the rotary wing enclosed cockpit Aerial application scenario might require a much larger sample of counties (or perhaps all counties in some states) because there are fewer workers using this configuration of equipment.

The probability of selection for the counties in this second stage would again be recorded, such that sampling weights for the MUs can be properly constructed.

Once the counties have been selected, AHETF would then contact the within-county growers of the appropriate crop type, and ask them for information regarding the number of person-days of closed-cab airblast application that will occur on their crops within the specified date range selected for the scenario study, as a function of the number of pounds of active ingredient that will be used each day. Larger growers may provide a distribution across several workers, pieces of equipment, and levels of active ingredient –whereas a smaller grower might only provide information for a single day of pesticide application. Note that the AHETF researchers may need to be trained on how to collect this information, so that sufficient data is recorded to complete the design. These data can be combined across the counties of both states within each crop-type to develop a stratified sampling frame for the selection of individual workers for the field study – while ensuring that the sample is diversified across the range of active ingredient handled.

Once this stratification is completed, methods for probability-based selection can be used to select growers, and then workers within growers to complete the design.

Weights associated with the probability of selection can be generated for each MU in the study sample, allowing researchers to develop unbiased estimates for the distribution of current worker exposure.

We note that the process for developing the list of eligible growers could be adapted for the purpose of constructing the list of growers in the chosen second-phase sample, as described above. Importantly, however, we note that willingness to participate should not be considered as an eligibility criterion. Eligibility for inclusion on the list should be based purely on characteristics of the grower (sufficient acreage, at least one worker with experience making closed cab airblast applications etc). Unwllingness to participate should be considered as part of non-response and analyzed as such.

## 4. Minor Comments Associated with the Review of Each Document

The following subsections provide minor comments on sections of the documents that were not captured in the previous comments.

### *4.1 AHETF Volume IV: Revised Governing Document*

- It seems awkward to have the first sentence of the executive summary refer to the glossary of terms. It would read much better if there were a few sentences of framing first. It would also help to provide brief context on why the new database (AHED® or the Agricultural Handlers Exposure Database) is needed? What are the limitations of the old one (PHED - Pesticide Handlers Exposure Database)?
- We have concerns about the statement "Thus, every MU will provide an estimate of a single handler-day of exposure to that pesticide." It does not seem sound to consider this realization of a single individual's exposure to be an estimate. It is really more of a realization.
- Last paragraph on the bottom of page 6 is very hard to follow. Certainly it becomes clearer on reading the details. But it should be made clearer in the executive summary as well.
- Page 46 of 153 – Section 9.2.2 on Diversity Selection: If EPA permits AHETF to utilize purposive diversity sampling as described in the governing document (rather than a more traditional probability-based sampling approach), simulation studies should be conducted to assess the potential degree of bias that might affect the statistics generated using the study data. It is possible that methods could be developed to make adjustments to these statistics to adjust for potential bias as a result of such studies.
- Page 47 of 153 – 3rd paragraph: This is an example of where the governing document failed to suggest that traditional probability-based sampling approaches are preferable over purposive selection. Here, the governing document could have recommended achieving diversity in the sample using a stratified sampling approach – but instead it states "*the most straightforward approach is to purposively select MUs that appear to be sufficiently different with respect to the characteristic(s) of interest.*"
- Page 48 of 153 – last sentence: Does AHETF know (apriori) the distribution of NF that will be used across all pesticide handlers? It seems that there is strong potential for errors that could affect the design.
- Page 50 of 153 – Section 9.3.1 suggests a reference sampling model that is based on a two-stage design with probability based sampling (and equal weights among the sites/MUs selected at each stage) – however, the design approaches pursued don't follow this model. AHETF should conduct some study of sample size requirements for designs that violate the assumption of equal selection probabilities for sites and MUs, as the purposive diversity sampling approaches will certainly violate these assumptions.
- Page 119 of 153 – Figure B4: Would like to see the corresponding "probability boxes" for a stratified sample (similar to the manner in which the diversity-sample is constructed), where the sample statistics are based on a weighted analysis.
- Page 121 of 153: 4th paragraph "*The number of strata must be at least as large as the number of units that will be selected*" Don't agree with this sentence – there are many design strategies in which multiple units are selected from within the same strata
- Page 121 of 153: 4th paragraph "*If there are more strata than units to be selected (Figure B7), then a subset of the strata should be selected first. This could be either purposively (to increase diversity or reduce costs) or randomly (to reduce intentional selection bias).*" Alternatively – the strata could be collapsed or combined, such that there are equal numbers of strata and units to be selected. The second sentence fails to again

prioritize random (or probability-based) sampling over purposive/convenience sampling, which is problematic.

- Page 121 of 153 – last two sentences: "*There is no attempt to sample or to weight results in proportion to stratum size. In fact, stratum sizes in the future 'generic' HD population are unknown*." While it is true that the stratum sizes for future 'generic' HD populations are unknown – the failure to use proper weighting leads to biased statistics for the current population. It also provides no basis for comparing future distributions to put the current results into proper perspective. For example – if new generations of pesticides allow workers to use a smaller amount of active ingredient with similar results, the current study (as planned) does not provide adequate information on the distribution of active ingredient used across HD contained in the sample.
- Page 122 of 153: Conceptually, Figure B6 is preferred over Figure B7 – because in B6, all members of the target population have a chance of being included in the study, whereas in B7, members of the target population is some strata could be excluded based on some selection criteria other than chance.
- Page 129 of 153: Section C2. We suspect that the reference model is specified on the natural-log scale (ln) – and not on the $Log_{10}$ scale. The notation should be more clearly specified to avoid confusion (and the natural log is used in later sections of Appendix C).
- Page 139 of 153: There would be value in reproducing Table 4 with different assumptions for the values of the GSD (2, 4, 6) and ICC (0.3 and 0.6). These statistics should also be regenerated to include designs that do not have sites and MUs selected with equal probabilities.
- Page 141 of 153 – Last sentence in Section C.9: "*Over this entire range, it appears that a configuration of $N_c$=5 and $N_m$=5 is the most reasonable choice*." A configuration of $N_c$=4 and $N_m$=8 has almost identical performance and cost, while yielding 7 additional samples. We think that these calculations are informative, but that AHETF should really think about optimizing the study design for each scenario, based on the unique characteristics (e.g. crop types, etc.).
- Pages 137 – 152 Sections C7 – C14: These analyses should also be conducted in a manner that properly accounts for the unequal probabilities of selection for sites and MUs

### *4.2 AHETF Volume I: General Information and Scenario Sampling Design*

- In Table 1 (page 14 of 52), it is not entirely clear whether the 5 MUs are from the same or different growers. It emerges later (page 41 of 52) that no more than 2 MUs can be from the same grower. It our opinion, it would be preferable if no more than 1 MU came from the same grower, since this would cut down on the intraclass correlation, thereby improving precision.
- Page 14 of 52: 6th Step in the Table: "*Stratify the practical range of amount of active ingredient handled (AaiH) into 5 levels*" – Can AHETF determine the fraction of worker Handler Days that would naturally fall into these strata?
- Page 17 of 52: The existing PHED could be used to verify some of the assumptions that were integrated into the current study design (such as values for the GSD and ICC).
- Pages 17-21 of 52: Section 4.1 – 4.3 appears to be copied straight out of the Revised Governing Document, and provided little insight into the specific design for the Closed Cab Airblast scenario studies.
- Discussion about the "generic database" is confusing and needs more clarification
- Page 18 of 52: discussion about the "smallest total sample sizes" corresponding to the setting of one MU per site is not so clearly expressed. It would help to say a little more and talk about the precision issue that is implicit in these statements.

- Section 4.3 (on pages 20 and 21 of 52) is confusing. We disagree with the statements that "statistical theory can only be used …" for the two settings described (random, representative sampling or randomization). Statistical theory is used to analyze observational data all the time when these scenarios don't apply. We can see that the authors of the document are trying to address these more general settings using the concept of "reference model". However, this section needs to perhaps be more carefully worded.
- The issue of clustering is not well described (page 21 of 52). If the MUs are from the same grower within a site, then the intraclass correlation would be expected to be much higher than if the MUs were from different growers.
- On page 22 of 52, the term AaiH is introduced, I believe for the first time in the document, but not defined. Later in the document, it is defined. Make sure that acronyms are defined the first time they are used.
- On page 25 of 52, there is a list of factors identified by experts as influencing potential exposure. Some of these need clarification (e.g. canopy density).
- Will growers be provided with the surrogate active ingredients to be used in the application of interest? Would it ever be appropriate to do measurements based on what the grower is planning to spray anyway?
- We disagree (page 50, Volume 1) that the regression coefficient ($\beta$) should equal 1 under proportional sampling. This should be approximately true, but given random sampling errors and so on, there could be realizations where a misleading conclusion (i.e. independence) is drawn when the value of $\beta$ is found to be different than 1.

### *4.3 AHETF Volume II: AHE55 – Protocol and Supporting Documents*

- Page 24-25 of 43: "*Screening of the growers (in the order of the random list) continues until the pool of eligible growers (and/or commercial applicators) contains at least 10 workers who may potentially volunteer for the study, and at least 2 workers are available for each of the AaiH strata*" Assuming that a probability-based design would be under consideration – the most meaningful difference in operational efficiency is that we would constrain the geographic coverage of the location to a select number of counties (in a second stage of location selection) to ensure similar goals (at least 10 workers, and 2 workers available for each of the AaiH strata) and conduct full enumeration among the growers within these counties.

### *4.4 AHETF Volume II: AHE56 – Protocol and Supporting Documents*

- Page 25 of 44: "*Screening of the growers (in the order of the random list) continues until the pool of eligible growers (and/or commercial applicators) contains at least 10 workers who may potentially volunteer for the study, and at least 2 workers are available for each of the AaiH strata*" Assuming that a probability-based design would be under consideration – the most meaningful difference in operational efficiency is that we would constrain the geographic coverage of the location to a select number of counties (in a second stage of location selection) to ensure similar goals (at least 10 workers, and 2 workers available for each of the AaiH strata) and conduct full enumeration among the growers within these counties.

### *4.5 AHETF Volume VII: Reference Materials Cited in Other Volumes*

- No Comments

## Appendix A
## Guidance from EPA regarding Specific Documentation to be Reviewed

AHETF Documents for Battelle Review

Note: it will be best to review in the order outlined below

1. AHETF Volume IV: AHETF Revised Governing Document for a Multi-Year Pesticide Handler Worker Exposure Monitoring Program

This document describes the general design that will be employed for each of the 33 scenarios in the AHETF program. It should be noted that only one scenario (closed-cab airblast application) is under current consideration.

Battelle should focus on the following sections:

- Glossary of Terms (pages 98 – 105 of 153)
- Executive Summary & Sections 1 – 1.3 (pages 5 – 13 of 153)
- Sections 7 – 7.1 (pages 35 – 38 of 153)
- Section 7.3 (page 41 of 153)
- Sections 9 – 11 (pages 43 – 64 of 153)
- Sections 12 – 12.3 (pages 70 – 74 of 153)
- Appendices B & C – (pages 111 – 153 of 153)

2. AHETF Volume I: General Information and Scenario Sampling Design

Volume I describes the closed-cab airblast application scenario, and explains the proposed sampling design and how AHETF made the purposive choices of the crops and growing areas (sites) to monitor.

Battelle should review this document in its entirety (52 pages), although Sections 3 & 7 do not require extensive review.

3. AHETF Volume II: AHE55 – Protocol and Supporting Documents

Volume II defines the protocol for one of the five field studies associated with the closed-cab airblast application scenario: citrus in Florida.

Battelle should focus on the following sections:

- Sections 3 – 6.3 (pages 25 – 30 of 107)
- Sections 7.8 – 7.9 (pages 33 – 34 of 107)

4. AHETF Volume III: AHE56 – Protocol and Supporting Documents

Volume III defines the protocol for a second of the five field studies associated with the closed-cab airblast application scenario: pecans in Georgia. It is closely similar to Volume II, and introduces no new ideas relevant to Battelle's review.

Battelle should focus on the following sections:

- Sections 3 – 6.3 (pages 26 – 30 of 108)
- Sections 7.8 – 7.9 (pages 34 – 35 of 108)

5. AHETF Volume VII: Reference Materials Cited in Other Volumes

This document documents AHETF's consultations with agricultural experts whose advice informed the choices of crops and growing regions to monitor for the closed-cab airblast application scenario.

Battelle should focus primarily on pages 14 – 18 in Part A. These pages are specific to the two protocols (AHE55 & AHE56) under consideration for the closed-cab airblast application scenario.