Agricultural Handlers Exposure Task Force (AHETF)

VOLUME IV

AHETF Revised
Governing Document for a Multi-Year
Pesticide Handler Worker Exposure Monitoring Program

Version Number: 1

April 7, 2008

Table of Contents

Executive Summary 5
1 Introduction and Background 8
1.1 AHETF and AHED® 8
1.2 Purpose of this Document 10
1.3 General Purpose and Description of the AHETF Monitoring Program 11
2 Plan for Submitting Human Research Protocols 13
3 Justification for Human Exposure Data 14
3.1 Regulatory Need for Generic Exposure Data 14
3.2 Alternatives to Additional Human Monitoring 17
4 Study Benefits 18
4.1 Description of Potential Benefits 19
4.1.1 Benefits to Subjects 19
4.1.2 Benefits to Society 19
4.1.3 Benefits to AHETF Members 20
4.1.4 Benefits to Growers or Landowners 20
4.2 Likelihood of Realization of Benefits 21
5 Risks to Subjects in AHETF Studies 21
5.1 Risk of Heat-Related Illness 22
5.1.1 Nature of the Risk of Heat-Related Illness 22
5.1.2 Minimizing the Risk of Heat-Related Illness 22
5.1.3 Nature and Likelihood of Residual Risk of Heat-Related Illness 23
5.2 Risk of Exposure to Surrogate Chemicals 24
5.2.1 Nature of the Risk of Exposure to Surrogate Chemicals 24
5.2.1.1 Acute Effects of Proposed Surrogate Chemicals 25
5.2.1.2 Regulatory Risk Assessments of Surrogate Pesticides 26
5.2.2 Minimizing the Risk of Exposure to Surrogate Chemicals 28
5.2.3 Nature and Likelihood of the Residual Risk of Exposure to Surrogate Chemicals 29
5.3 Risks Associated with “Scripting” of the Field Activities 29
5.3.1 Nature of the Risk Associated with Scripting Field Activities 29
5.3.2 Minimizing the Risk Associated with Scripting Field Activities 29
5.3.3 Nature and Likelihood of Residual Risk Associated with Scripting Field Activities 30
5.4 Psychological risks 30
5.4.1 Nature of Psychological Risks 30
5.4.1.1 Risk Associated with Taking a Pregnancy Test 30
5.4.1.2 Risk of Embarrassment during Changing of Clothes 30
5.4.2 Minimizing Psychological Risks 30
5.4.3 Nature and Likelihood of Residual Psychological Risks 31
5.5 Risk of Exposure to Surfactant during Face/Neck and Hand Sampling 31
5.5.1 Nature of the Risk of Exposure to Surfactant 32
5.5.2 Minimization of the Risk of Exposure to Surfactant 32
5.5.3 Nature and Likelihood of Residual Risk of Exposure to Surfactant 32
5.6 Background Risk of Injury Associated with Agricultural Work 32
5.6.1 Nature of the Risk of Injury Associated with Agricultural Work 32

5.6.2 Minimizing the Risk of Injury Associated with Agricultural Work 33
5.6.3 Nature and Likelihood of Residual Risk of Injury Associated with Agricultural Work .. 34
5.7 Likelihood of Serious or Irreversible Effects..... 34
6 Risk versus Benefit Comparison..... 35
7 Description of and Rationale for Scenarios Considered by AHETF 35
7.1 Handler Scenarios Included in the Monitoring Program 35
7.2 Limitations of PHED 38
7.3 Scenario-Specific Data Needs for AHED® 41
8 Evaluation of Existing Data and Incorporation into AHED® 41
9 Design Concepts Common to All Scenarios..... 43
9.1 Predicting Generic Exposures with Experimental Monitoring Units 43
9.2 MU Selection Procedures Common to All Scenarios..... 45
9.2.1 Two-Stage Selection of MUs..... 45
9.2.2 Diversity Selection..... 46
9.2.2.1 Diversity Selection of Monitoring Sites 47
9.2.2.2 Diversity Selection of Monitoring Units..... 48
9.3 Scenario Sample Sizes 49
9.3.1 Reference Random Sampling Model 50
9.3.2 Benchmark Objectives 50
9.3.3 Sample Sizes that Satisfy the Benchmark Objectives..... 51
10 Implementation of MU Selection and Construction for each Scenario 52
10.1 Defining the Scenario 54
10.2 Defining the Normalizing Factor 54
10.3 Identifying Characteristics that could Impact Normalized Exposure 55
10.4 Likely Ranges of Exposure-Impacting Characteristics in Generic Handler-Days 56
10.5 Sample Sizes that Satisfy Benchmark Objectives 57
10.6 Identify Monitoring Sites..... 58
10.7 Identifying a Potential Worker Pool within each Monitoring Site 59
10.8 Constructing Monitoring Units 61
10.9 Sources of Expert Information..... 63
11 Description and Role of Field Studies under the AHETF Monitoring Program 64
11.1 Relation of Field Studies to Scenarios 64
11.1.1 Descriptions of Surrogate, Locations, and Number of Measurements 65
11.1.2 Documentation procedures 66
11.1.3 Quality Assurance Procedures 67
11.1.4 Quality Control Procedures..... 67
11.1.5 Reporting Process 69
12 Scenario Summary (Monographs) 70
12.1 Assessment of Benchmark Adequacy Objectives..... 70
12.1.1 Relative Accuracy of the Normalized Exposure Distribution 71
12.1.2 Adequacy of the Data for Distinguishing a Proportional from an Independent Relationship between Exposure and the Normalizing Factor..... 72
12.1.3 Interim Analysis of Benchmark Adequacy..... 73
12.2 The Impact of Ignoring Clusters 73
12.3 Other Information to be Included in the Scenario Monographs 74

13 Exposure Monitoring Techniques 74
13.1 Validation of Passive Dosimetry 75
13.2 Description of Techniques 76
13.3 Nature of Testing Guidelines 79
14 Ethical Considerations 80
14.1 Recruitment 80
14.1.1 Recruiting Growers, Landowners, or Commercial Applicators 80
14.1.2 Recruiting Subjects 81
14.1.3 Eligibility Criteria 82
14.1.4 Use of Vulnerable Groups 84
14.2 Informed Consent Process 88
14.3 Subject Remuneration 90
14.4 IRB Review Process 92
14.5 Additional Efforts to Protect Human Subjects 92
15 References 95
16 Glossary of Terms 98
Appendix A. AHETF Acceptance Criteria for Existing Studies 106
Appendix B. Using Monitoring Units to Characterize Generic Exposure 110
Appendix C. Procedures for Determining the Required Number of Clusters and Monitoring Units per Cluster to Achieve Benchmark Adequacy 127

Executive Summary

Section 16 contains a glossary of terms that are used throughout this document.

The Agricultural Handlers Exposure Task Force (AHETF) was formed to develop a generic exposure database for use by the EPA and other regulatory agencies for their assessment of occupational exposures encountered by workers who mix, load and/or apply agricultural chemicals. The new database, AHED® (Agricultural Handlers Exposure Database) will serve the necessary role of supplanting data in PHED (Pesticide Handlers Exposure Database) which has been in use since 1992.

Generating new exposure data involves monitoring potential dermal and inhalation exposure for occupational pesticide handlers performing a variety of mixing/loading and application tasks. The basic experimental construct for exposure monitoring is the monitoring unit, or MU. An MU is an experimental realization of a single worker handling a particular pesticide under a particular set of circumstances that represent a single workday. Thus, every MU will provide an estimate of a single handler-day of exposure to that pesticide. Validated passive exposure monitoring dosimetry techniques are utilized in the AHETF field study program. These techniques are preferred as they are non-invasive and provide exposure estimates for individual body parts that can be useful in determining exposure mitigation strategies.

The foundation and justification for the AHETF exposure monitoring program are presented in this "Governing Document". Field study protocols and other detailed documents are prepared and submitted separately to support specific human research protocol submissions.

The degree to which individual workers are potentially exposed while handling pesticides depends primarily on specific activities and conditions. These include:

- Task performed (e.g., mixing/loading or applying, or both);
- Equipment being used (e.g., open or closed loading systems, aerial or ground application equipment, and open or enclosed cab tractors);
- Amount of contact with the pesticide;
- Use and type of Personal Protection Equipment (PPE) worn; and
- Pesticide product formulation type (e.g., liquid vs. solid)

Exposure potential is not affected by the particular active ingredient (ai) in the agricultural chemical product. Thus, under the same handling conditions, an MU-based exposure from one ai is a generic exposure prediction for any other active ingredient. Volatile compounds such as fumigants however are an exception to this 'generic principle' and will not be addressed by the AHETF Monitoring Program.

Exposure is generally understood to be proportional to a worker's potential for contact with the active ingredient. Consequently, when measured worker exposure is normalized

by an appropriate measure of potential ai contact, the results for that MU provide a generic prediction of exposure at any other level of active ingredient contact. Thus, generic databases provide a powerful tool and are widely accepted by regulatory authorities throughout the world. These generic exposure databases group MUs into distinct exposure situations, or "scenarios". A scenario is a combination of similar work tasks, pesticide formulations, equipment, engineering controls, and work practices (i.e., specific procedures used for a particular task). AHETF has designed a program to monitor worker exposure for a wide variety of occupational handler scenarios using registered pesticide products and typical worker activities and equipment.

AHED® defines an appropriate exposure normalizing factor (or NF) for each scenario. This NF is a measurable quantity available for every MU and is approximately proportional to the amount of worker contact with active ingredient. For most scenarios a reasonable normalizing factor is the amount of active ingredient handled (or AaiH). This is simply the amount of active ingredient that gets mixed into a tank or piece of application equipment and/or applied from a piece of application equipment. Normalized worker exposure is obtained by simply dividing the exposure by the NF for each worker monitored. The normalized exposures from a set of MUs can be thought of as an analogous set of generic exposures for an arbitrary chemical when the value of the NF is equal to one.

A recent EPA Scientific Advisory Panel has affirmed that the currently available occupational pesticide handler data, primarily in the Pesticide Handlers Exposure Database (PHED), are not adequate to meet contemporary regulatory requirements to properly evaluate agricultural chemical handler exposure potential. AHETF has reviewed thousands of handler exposure measurements in existing studies and identified those studies and MUs that satisfy modern acceptability criteria. However, many exposure scenarios still lack the number or configuration of MUs needed for regulatory purposes. The AHETF database will contain data to address 33 scenarios (see Section 7.1). Each scenario will be supported by appropriate documentation that describes and justifies the need for additional monitoring data. The AHETF scenarios are fundamentally similar to those in PHED that have proven to be practical for regulatory use by EPA and other regulatory agencies to conduct product-specific exposure assessments using a generic database.

For most scenarios, regulatory interest is usually focused on characterizing the typical and the larger levels possible in a hypothetical distribution of normalized worker exposures. If the MUs have been designed to reflect the diversity of handler-day conditions that impact exposure, then the MU-based exposures should also reflect the diversity of exposure in the normalized exposure distribution. In this case, the middle and upper MU exposure values should tend to characterize the true (but unknown) middle and upper values. Therefore, it is reasonable to use sample statistics such as means and upper percentiles to approximately characterize these same quantities in the hypothetical normalized exposure distribution. This will be a conservative estimate for the larger values because 'extra-diverse samples' tend to overpredict the true upper percentiles.

For each scenario the AHETF monitoring program will 'select' a set of MUs by choosing, restricting, or scripting handling conditions in a manner that tends to capture exposure diversity. The MUs are always selected in two successive stages: At the first stage, several monitoring sites are selected. At the second stage, a cluster of MUs is constructed within each monitoring site. Methods designed to increase diversity are superimposed on both stages of selection, typically using a combination of random and purposive elements. Random selection reduces intentional selection bias, but purposive selection is usually more practical, cost effective, and can frequently induce greater diversity.

The number of monitoring sites and number of MUs per site targeted for each scenario are based on benchmark accuracy objectives established by AHETF and the EPA for the mean and 95th percentile of normalized dermal exposure. Although the value could vary by scenario, 3-fold relative accuracy in these sample statistics (assuming a lognormal two-stage random sampling reference distribution) is the default benchmark accuracy requirement. A minor secondary benchmark objective is used in the scenario design whenever the NF can vary by at least an order of magnitude. In such cases, it is also desirable that the MU data be adequate to distinguish a proportional from an independent relationship between exposure and the normalizing factor. This minor requirement is usually satisfied whenever the primary objective is met and the NF levels are widely varied within each cluster of MUs.

When a scenario has been completed, the data from its MUs will be added to AHED®. This database will be used by EPA and other regulatory agencies to regulate pesticide usage. Key characteristics of AHETF's multi-year pesticide handler exposure monitoring program and the associated AHED® database are outlined below:

- Exposure potential for 33 scenarios will be addressed
- Most scenarios will contain about 25 MUs in about five clusters
- AHED® will contain approximately 825 MUs
- About 100 existing MUs have been identified as useful and put into AHED®
- To date, 173 MUs have been generated by AHETF
- The remaining MUs will be collected in clusters of about 5 MUs
- Individual AHETF monitoring studies will generally collect 5 MUs at one location for 1 scenario
- Approximately 50 field studies will need to be conducted over the next several years to complete AHED®

Monitoring exposure to professional agricultural handlers who follow their normal practices presents a reasonably low additional risk to participants. The risk from dermal exposure to the pesticide product is actually reduced because subjects must wear an extra

layer of clothing (long underwear that serves as an "inner dosimeter") that will trap chemical before it reaches the skin. The potentially increased risk of heat illness from wearing the inner dosimeter is mitigated by a medical management program which emphasizes measures to prevent heat-related illness and guidelines for stopping participation. Before monitoring workers for any scenario, the benefit to agricultural workers as a whole and to society in general, in the form of more accurate measurements of potential exposure to pesticides, will be compared to assure the benefits outweigh the risks to study participants.

1 Introduction and Background

1.1 AHETF and AHED®

The Agricultural Handler Exposure Task Force (AHETF) was established in December 2001 to generate exposure data for agricultural pesticide handlers and meet EPA registration requirements. Several AHETF member companies had ongoing data requirements resulting from product-specific data call-in notices, reregistration obligations, or prospective registration obligations. These companies agreed to jointly develop generic data in support of their respective registration obligations since existing data are not adequate.

The primary AHETF goal is to collect pesticide handler exposure monitoring data and incorporate it into a new generic database that will be used to characterize future worker exposures from arbitrary pesticides. The database will be called AHED®, Agricultural Handlers Exposure Database. AHED® will be submitted to EPA and other regulatory agencies and used by those regulators to conduct detailed quantitative exposure assessments and make safety determinations for occupational pesticide uses. The AHETF will exercise the rights associated with submission of data under the Federal Insecticide, Fungicide, and Rodenticide Act (FIFRA, U.S. EPA, 1989) in connection with AHED®. AHETF is focused on occupational mixer/loaders and applicators of pesticides on farms, in nurseries, in greenhouses, and in seed treatment facilities. It is not concerned with residential uses, fumigants, or incidental exposures from others using pesticides.

Generic databases were developed over the last twenty years in response to a regulatory need to assess the occupational risks associated with a wide range of pesticide handling situations. The concept was discussed and its development encouraged by a FIFRA Scientific Advisory Panel (SAP) in 1986. In 1992, the Pesticide Handlers Exposure Database (PHED) was first released following a joint effort by pesticide manufacturers, the EPA, and Canadian regulators (Honeycutt, 1986; Lunchick, 1994; Reinert, 1986). Since then, PHED has been used extensively in a generic manner and has successfully supported many occupational risk assessments. Much of the data in PHED are derived from exposure studies that are outdated or scientifically inadequate by current standards (Stasikowski, 2001). In addition, many handler scenarios of interest to EPA are absent or

under-represented in PHED. Other regulatory agencies, including the Canadian Pest Management Regulatory Agency (PMRA) and California Department of Pesticide Regulation (CDPR), have expressed similar dissatisfaction with the limitations of PHED data. In 2007, EPA convened another SAP to discuss the need for new data to replace PHED. The panel agreed with EPA that "additional data could significantly improve the Agency's ability to assess worker exposure" (SAP, 2007). A major purpose of developing the Agricultural Handlers Exposure Database, or AHED®, is to address PHED deficiencies.

Like PHED, AHED® will be populated with exposure data derived from workers who handle pesticides as part of their normal job, so their participation as subjects in the field studies underlying AHED® will not add appreciably to their typical exposure from handling pesticides. All AHETF studies are designed and conducted in accordance with the latest U.S. EPA guidelines for occupational exposure studies.

The development of AHED® is funded and directed by the AHETF. An AHETF Joint Regulatory Committee (JRC) has been established to promote active participation by interested regulatory agencies. The JRC is comprised of AHETF representatives and representatives of the U.S. EPA, PMRA, CDPR, and the United States Department of Agriculture (USDA). This committee meets on a regular basis to review program progress and provide technical and regulatory input to the AHETF.

Between AHETF inception (December of 2001) and April of 2006 (effective date of Final Human Testing Rule), AHETF:

- Defined the scope of AHETF
- Developed standard operating procedures and a standardized exposure monitoring protocol
- Reviewed 216 existing studies against acceptance criteria (about 3,000 worker exposure measurements, or monitoring units) for possible incorporation into AHED®
- Acquired the right to cite 105 monitoring units (MUs) from existing data (from 6 scenarios)
- Collected 173 MUs from 11 scenarios
- Developed and began populating AHED®

In 2006, AHETF submitted five study protocols to EPA and the Human Studies Review Board (HSRB) to continue its monitoring program, but the HSRB concluded it did not have sufficient information about the overall monitoring program to evaluate the scientific soundness and ethical acceptability of the protocols (Fisher, 2006). After that review AHETF devoted its resources to formally documenting all of its study design, conduct, and data analysis procedures. This included an internal examination of the existing experimental worker exposure events (called monitoring units, or MUs) that AHETF acquired and those it had collected in its first few years of existence.

In 2007, AHETF submitted to EPA and the HSRB the final draft of this Governing Document (dated May 22, 2007) and a variety of other documents to support the scientific and ethical validity of the entire AHETF exposure monitoring program. This submission did not include a specific study protocol. The HSRB (Brimijoin, 2007) fundamentally supported the ethical aspects of the monitoring program (e.g., need for data, risk-benefit analysis, and specific research procedures), but recommended adoption of a formal probability sampling approach to obtain existing worker-days instead of the purposive diversity sampling approach proposed by the AHETF. Although simplistic and sometimes misleading, probability sampling is often referred to colloquially as 'random' sampling. After that review, EPA further considered the merits of probability sampling versus purposive diversity sampling and has determined the latter is more appropriate for the worker exposure data it needs to support pesticide product safety assessments. However, the EPA also concluded that random elements should be incorporated into study designs whenever feasible to reduce the potential for intentional selection bias.

This first complete version of the Governing Document (Version 1) addresses the comments of EPA and HSRB and details the procedures AHETF intends to follow to collect additional pesticide handler exposure data. This Governing Document will support new pesticide handler exposure monitoring study protocols for review by EPA and for conduct by AHETF in 2008 and beyond. A detailed MU Selection Plan will be developed for each scenario and will include a justification for new data and a rationale for how much data will be generated. Each submitted study protocol will reference the Governing Document and the appropriate scenario plan(s), and will provide study-specific information to fully address the scientific and ethical validity of the protocol and proposed research.

1.2 Purpose of this Document

This Governing Document describes the overall AHETF exposure monitoring program and plans for developing a generic database, AHED®. It discusses the need for additional human exposure data, the rationale for determining the use scenarios that will be included in the program, how AHETF will generate new data, how the data will be analyzed statistically, and how the data will be used to support regulatory decisions. However, it is important to point out that each distinct handling scenario that AHETF will address (33 scenarios, see Section 7.1) will involve an examination of existing data and detailed plans for collecting sufficient new data to meet regulatory needs (in an MU Selection Plan). This information will be submitted as separate documentation as AHETF studies address the different scenarios. In this sense, each scenario can be considered a distinct research project and this Governing Document alone will not provide complete information to justify any of those projects.

Further, each field study designed to collect MUs from scenario plans (and there will typically be several studies per scenario) will have a protocol that describes study-specific information including which tasks are to be conducted (i.e., scenario included),

how many MUs are to be collected, and study-specific details for test substances and exposure monitoring procedures.

This Governing Document also discusses important ethical considerations including how AHETF will generally recruit growers (or landowners), recruit workers, obtain informed consent, and minimize risks to study participants. Each field study protocol will address specific procedures that will be followed since different handling scenarios will involve special issues for subjects including study-specific risks and grower/participant recruitment procedures.

This document will support each specific study protocol for new AHETF worker exposure studies (in conjunction with the scenario-specific plans and standard operating procedures [SOPs]). Throughout this document, an attempt is made to identify information as describing program-specific approaches (i.e., general information such as the basic sampling method), scenario-specific procedures (such as evaluating the need for additional human data), or study-specific detail (such as the risks associated with the particular task, surrogate chemical, etc.).

This document describes how the AHETF monitoring program will comply with 40 CFR Part 26 regarding intentional human dosing studies. It also addresses concerns raised in the reports of the HSRB meetings of June 27-30, 2006 (Fisher, 2006) and June 27-29, 2007 (Brimijoin, 2007).

1.3 General Purpose and Description of the AHETF Monitoring Program

The goal of the AHETF monitoring program is to develop data that will be incorporated into a generic exposure database, AHED®. These data will consist of dermal and inhalation exposure measurements obtained from basic experimental constructs called monitoring units (or MUs). Each MU is an experimental realization of single worker handling a particular pesticide under a particular set of circumstances that represent single workday. Thus, every MU will provide an estimate of single handler-day of exposure to that pesticide. These exposure data, as well as the database itself, are referred to as generic because the ultimate objective is that the MUs be used to predict future exposure to arbitrary pesticides given some arbitrary, but measurable, amount of active ingredient contact.

When completed, the data in AHED® will be used to support North American regulatory decisions—that is, to estimate exposure for future agricultural uses of existing and new pesticide products in the United States and Canada for a wide variety of pesticide handling situations called scenarios. In general, a scenario is a combination of similar work tasks, pesticide formulation, equipment, engineering controls, and work practices (i.e., specific procedures used for a particular task). For example, two scenarios of interest are:

- mixing/loading dry flowable pesticides using open pouring techniques

- applying liquid sprays using airblast equipment with open cabs

It is anticipated that AHED® will contain sufficient data to support exposure assessments for many distinct handling scenarios.

AHED® defines an appropriate exposure normalizing factor (or NF) for each scenario. This NF is a measurable quantity available for each MU and is believed to be approximately proportional to the amount of worker contact with active ingredient. It is generally accepted that, within limits, worker exposure is proportional to the expected amount of contact with active ingredient. If the NF is approximately proportional to this contact, then exposure should also be proportional to the NF. For most scenarios a reasonable normalizing factor is the amount of active ingredient handled (or AaiH). This is simply the amount of active ingredient that gets mixed into a tank or piece of application equipment and/or applied from a piece of application equipment. Normalized exposure is obtained by simply dividing the exposure by the NF for each worker monitored. The normalized exposures from a set of MUs can be thought of an analogous set of generic exposures for an arbitrary chemical when the value of the NF is equal to one. Multiplying these MU-derived generic exposures by any desired NF level yields a set of predicted exposures for that level.

For each pesticide handling scenario, regulatory interest is usually focused on characterizing certain aspects of the (future) distribution of generic exposure. In particular, it is desirable to have reasonable estimates of both the typical (i.e., 'middle') and the larger levels of worker exposure possible for the scenario. The set of MU-derived normalized exposures are considered to be possible values from this distribution, but are not a true random sample. However, if the MUs have been designed to reflect the diversity of handler-day conditions that impact exposure, then the MU-based exposures should also reflect the diversity of exposure in the generic exposure distribution. In this case, the middle and upper MU exposure values should tend to characterize the middle and upper values in the generic exposure distribution Therefore, it is reasonable to use sample statistics such as means and upper percentiles to approximately characterize the same quantities in the generic (normalized) exposure distribution.

For each scenario the AHETF monitoring program will 'select' a set of MUs by choosing, restricting, or scripting handling conditions in a manner that tends to capture exposure diversity. The MUs are always selected in two successive stages: At the first stage, N_c large staging areas, termed monitoring sites, are selected. At the second stage, a cluster of N_m MUs are constructed within each monitoring site. Methods designed to increase diversity are superimposed on both stages of selection. The procedures for selecting both monitoring sites and MUs within sites will typically use a combination of random and purposive elements. Random selection reduces intentional selection bias, but purposive selection is usually more practical, cost effective, and can frequently induce greater diversity.

The number of monitoring sites and number of MUs per site targeted for each scenario are based on benchmark accuracy objectives for the mean and 95th percentile established

by AHETF and the EPA. Although the value could vary by scenario, 3-fold relative accuracy in these sample statistics (assuming a lognormal two-stage random sampling reference distribution) is the default benchmark accuracy requirement. A minor secondary benchmark objective is used in the scenario design whenever the NF can vary by at least an order of magnitude. In such cases, it is also desirable that the MU data be adequate to distinguish a proportional from an independent relationship between exposure and the normalizing factor. This minor requirement is usually satisfied whenever the primary objective is met and the NF levels are widely varied within each cluster of MUs.

The monitoring activities for each monitoring site are conducted as a GLP field study. As each field study is completed, it will provide an additional cluster of MUs for the scenario. As scenarios are populated with clusters of MUs consistent with the scenario-specific MU Sampling Plans (see Sec. 10), AHED® will gradually be completed. When a particular scenario is completed, it will be documented in a Scenario Monograph as described in Section 12 of this document. This Scenario Monograph and all scenario-related study reports will be formally submitted to EPA and other regulatory agencies to support their use of the AHED® data for that scenario. Given a complete set of generic exposure data for all scenarios, regulatory users of the AHED® database will then be able to estimate individual worker exposures for a single workday given only:

- A mixer/loader and/or applicator pesticide handling scenario of interest and
- A particular value for the scenario-specific measure of expected active ingredient contact, such as the amount of active ingredient to be handled by the worker

Such information is expected to be of great value to the EPA and other regulatory agencies for assessing risks to workers who handle pesticides.

2 Plan for Submitting Human Research Protocols

While this Governing Document presents general guidelines for AHETF to plan and conduct human exposure monitoring studies, this section will describe the plan for submitting to EPA all the necessary information for review and approval of new human studies. Several distinct studies will generally be conducted to fulfill the data needs for a particular scenario. With each new study protocol, AHETF plans to submit a package of documents that presents all the necessary information to evaluate the scientific and ethical validity of the study and the scenario involved in the study. This submission will include the documentation specified by EPA and the HSRB, which presently includes this Governing Document, the MU Sampling Plan for the applicable scenario, the study protocols and supporting documents, the applicable SOPs, specific reference material, and all of the transmittals to and from the Institutional Review Board.

AHETF intends to submit new human study protocols regularly over the next several years as it works toward building the next-generation generic database of pesticide

handler exposures. New study protocols will generally be submitted to the JRC for a scientific evaluation before being submitted for approval by the HSRB.

3 Justification for Human Exposure Data

A necessary condition for both scientific and ethical acceptability of the AHETF program is that the use of humans to generate these data be justified. This requires that human data are necessary for the regulatory risk assessment process and that adequate alternatives to conducting additional human exposure monitoring are not available. This section discusses the program-level need for human exposure data and the value of generic databases to EPA and other regulators. Section 7 discusses the procedures used to examine the need for data on a scenario-specific basis. Each scenario plan will be accompanied by documentation that confirms the need for additional human exposure data for that scenario.

3.1 Regulatory Need for Generic Exposure Data

FIFRA requires the U.S. Environmental Protection Agency to assure that any pesticide registered in the United States does not have unreasonable adverse effects on workers handling that pesticide. The Pest Control Products Act requires a similar determination be made by Health Canada. This safety determination is generally made by means of quantitative risk assessment and risk management procedures. Risk assessments require a detailed evaluation of the toxicity of the pesticide and an estimation or measurement of the exposure potential for workers (and/or amount of pesticide absorbed by the workers as a consequence of its use). Exposure or absorbed dose estimates are used in conjunction with no-effect exposure levels and/or cancer potency factors, Q1*, for hazards identified in standardized toxicology studies. During the risk evaluation, the likelihood of the expression of any toxicological effect on the workers and a comparison of the risks and benefits are considered. This basic paradigm (hazard identification, dose-response assessment, exposure assessment, and risk characterization) was summarized by the National Academy of Sciences and has become the standard for risk assessment by regulatory agencies (NAS, 1983; NAS, 2006). More recently, the pesticide handler risk assessment process was fully described in a summary document prepared for a Scientific Advisory Panel review of exposure methodologies (U.S. EPA, 2007). This summary also describes the tiered approach to handler exposure assessments that includes baseline assessments based on generic data (e.g., PHED or AHED®) with minimal PPE and no engineering controls and, when needed, followed by assessments using additional PPE and/or engineering controls, followed by product-specific information (including perhaps biomonitoring data).

AHED® is intended to provide the North American regulatory agencies with the potential exposure data necessary for them to perform the handler exposure assessment portion of these safety determinations. Toxicology data and benefit information are product-specific and must be provided by individual pesticide product registrants.

When estimating exposure to workers who handle pesticides, a major challenge to overcome is that several parameters contribute to the likelihood and level of exposure. These include factors such as handling liquids versus solids, product packaging type, using open versus closed systems, applying with various equipment types, amount of product handled, use of personal protective equipment (PPE), and whether the worker mixes/loads or applies or does both. The number of combinations of these parameters makes it impractical to generate human exposure data for all situations, so a number of simplifying approaches have been adopted. These include:

- Establishing various 'scenarios' that cover common combinations of these parameters and generating data for those scenarios
- Generating data with workers wearing minimum PPE
- Using data for one chemical/product as a surrogate for another (similar) product
- Assembling generic databases (e.g., PHED) that contain surrogate data applicable to many products

Since the early 1980's it has been the consensus of the scientific community that the amount of residue that contacts a worker's clothing and skin, and the amount of residue that is available for inhalation, are primarily a function of physical rather than chemical factors. That is, the chemical nature of the active ingredient in a pesticide product has little influence on the extent of exposure compared to physical parameters associated with the use of the product. The physical parameters include formulation type (e.g., liquid or granule product), method of application, and the way in which a person handles the pesticide during mixing, loading and application. Because of this, exposure potential is considered "generic" since it is independent of the specific active ingredient (Hackathorn, 1985; Honeycutt, 1985 and 1986; Reinert, 1985). Generic exposure data may therefore be used in lieu of product-specific data for most safety assessments. One major exception is that exposure to highly volatile compounds, such as fumigants, is not considered generic, and so will not be addressed by AHETF.

The use of generic data enhances the efficiency of regulatory agencies in conducting exposure assessments. Rather than relying on individual studies to evaluate case-by-case uses of each pesticide product, a single, comprehensive database of high quality data applicable to most products can be used. The broad applicability of generic data and the resulting efficiency of their use in regulatory safety assessments led to the widespread acceptance of PHED. PHED components were created by assembling exposure data from studies that had already been conducted and submitted to EPA.

Most of the pesticide exposure data available at that time had been submitted by individual pesticide manufacturers who designed their studies to support the registration of a specific product or a group of similar products. It was very common for these companies to generate a set of exposure data that represented the worst case for exposure potential incorporating design features such as the maximum use rate and minimum engineering controls. If a risk assessment was acceptable for such a situation, then it was argued that an assessment involving lower use rates, additional PPE, and additional engineering controls would also be acceptable. However, this meant it was common for a

study to involve 15 or more measurements of essentially the same situation where each person handled the same product, in the same packaging, in the same amount, using the same equipment, and for the same amount of time. While these studies are useful for product-specific cases, they are less useful for making generic estimates of exposure. Nevertheless, many of these types of studies were assembled to form PHED and, collectively, the database did improve the risk assessment process as regulators could often rely on larger data sets to estimate potential exposure.

As discussed in detail in Section 7, PHED has several technical limitations since the studies included in PHED were not designed to meet the needs of a generic database. In addition, it is outdated and many agricultural practices have changed. The written summary of a recent SAP meeting (SAP, 2007) concluded:

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

Exposure monitoring methods have also changed since most PHED studies were conducted. Basic passive dosimetry methodology has long been accepted as a standard, reproducible procedure that provides accurate and reliable data and does not underestimate exposure. Even though basic passive dosimetry methodology yields a very sound measure of exposure, there have been some improvements. In particular, much of the data in PHED are based on patch dosimetry and exposures were often not measured on all body areas. The recent SAP (2007) also concluded:

> The inclusion within PHED of studies where either not all parts of the body were monitored or a substantial number of exposures were undetectable do not allow the results to yield accurate exposure statistics of interest for regulatory assessments.

Nonetheless, PHED provided reasonable estimates of exposure based on the technology of the 1980's. Today, whole-body garment dosimetry is used instead of patches to improve the ability to estimate whole body exposures as well as the distribution of total body exposure.

There is consensus among regulatory agencies that the most efficient means of generating handler exposure data is to pool technical resources and assemble a generic database. This consensus, EPA's recognition of the limitations of PHED, and their intention to use additional data to augment PHED, led to the formation of the AHETF in December, 2001. The task force database, AHED®, was designed to reflect a logical set of use scenarios with adequate data in each scenario to provide good estimates of exposure potential and its distribution. Individual measurements will involve separate workers and

more diversity in equipment and conditions than in PHED, especially for the amount of product handled. The recent SAP (2007) concluded:

> The AHED study design will also include more reliable exposure assessment methods (especially of the hands; see also Charge #2) and newer ("modern") pesticide application equipment and techniques (see also Charge #4).

3.2 Alternatives to Additional Human Monitoring

Regulatory agencies are charged with assuring that registered uses of a pesticide will not cause unreasonable adverse effects to pesticide handlers. As part of such determinations, regulators and risk assessors must be able to estimate with confidence the levels of occupational exposure. Information now available to support these estimates comes primarily from the data in PHED, but also from pesticide-specific exposure studies and published literature. Modeling or animal data are of limited use in estimating occupational exposure of workers. The best estimates of worker exposure are based on monitoring the activities of people who handle pesticides as part of their regular job. This is what the AHETF program involves.

The only alternatives to the conduct of new human monitoring studies appear to be:

- Continued reliance on existing information sources
- Acquisition of additional handler exposure data from other existing product-specific studies that meet established acceptance criteria and that have generic applicability

The recent SAP (SAP, 2007) endorsed the need for new worker exposure data:

> This Panel is clearly of the opinion that additional worker exposure data collected on human volunteers under field conditions and label requirements on chemicals that have been approved by the Agency are necessary.

The limitations of PHED are discussed more thoroughly in Section 7.

In the first stage of the AHETF program, and prior to the conduct of any field studies with human volunteers, the AHETF reviewed existing handler exposure data from various sources (primarily from AHETF members, CDPR, and the open literature) and acquired data that met established acceptance criteria. These activities are described in Section 8 below. Although some useful worker exposure studies were acquired by AHETF, most of the existing data were not sufficient to meet the generic data needs identified in advance by the AHETF and the Joint Regulatory Committee.

A recent SAP (2007) evaluated the AHETF acceptance criteria and concluded:

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

Given the limitations of PHED and limited useful existing data, no viable alternatives to performing additional human monitoring studies exist for generating an updated exposure database.

It should also be pointed out that pre-requisite studies, necessary for the conduct of AHETF exposure monitoring, do not involve human participants. These pre-requisite studies include analytical method validations, field recovery validations, and toxicity studies that support the registrations of the test materials used. Therefore, the exposure measurements (MUs) proposed by this document reflect the entirety of human participation proposed by the AHETF.

4 Study Benefits

A critical principle of ethical human studies research is that the benefits to the subjects and to society must outweigh the risks to the subjects. To approve proposed research with human subjects, an Institutional Review Board must determine that "risks to subjects are reasonable in relation to anticipated benefits, if any, to subjects, and the importance of the knowledge that may reasonably be expected to result" (40 CFR §26.1111(a)(2)).

AHETF believes the incremental risks to professional pesticide handlers participating in this monitoring program are outweighed by the societal benefits expected to be gained from increased knowledge of typical exposure levels in representative agricultural use scenarios. This knowledge will improve the EPA's ability to assess occupational pesticide exposure and better protect workers.

It is more appropriate to discuss these societal benefits at the scenario level for it is at this level that the validity of the scientific design must be judged; taking into account existing data, the appropriateness of normalization factors (such as AaiH), the scenario-specific MU selection design, etc. Furthermore, it is only when the scenario-specific MU selection design has been fulfilled—typically only after several discrete field studies are conducted—that the benefits of the research can be realized. The discussion below outlines in general terms how new scenario-specific exposure data will provide a societal benefit, however each field study protocol will detail the benefits applicable to that study and the scenario it supports.

4.1 Description of Potential Benefits

4.1.1 Benefits to Subjects

None of the studies in the AHETF monitoring program will provide direct benefits to the study participants. This means risks to participants must be justified by the societal benefits that are anticipated to come from a successful study (NAS, 2004). Information from this monitoring program will be used to estimate the exposure risk to agricultural workers who mix, load, and apply pesticides. This may lead to safer pesticide handling practices that indirectly benefit the participants and other agricultural pesticide handlers.

4.1.2 Benefits to Society

The AHETF exposure monitoring program will significantly improve the ability of EPA and other regulatory agencies to estimate the risks to professional pesticide handlers from handling agricultural pesticides. This is a benefit to society and these benefits accrue on a scenario-specific basis.

Benefits of human dosing studies have been examined at length by the National Research Council of the National Academy of Sciences (NAS, 2004). These discussions are also applicable to the occupational exposure monitoring that AHETF plans to conduct, including the following:

> Any human dosing study, regardless of its risk category, must have a useful purpose and convey some benefit to the participants and/or society. As discussed earlier, the committee concludes that under the risk-benefit balancing required by the principle of beneficence and the Common Rule, personal benefits to participants are insufficient by themselves to justify human dosing studies conducted for EPA regulatory purposes. This means that risks to participants imposed by human dosing studies must be justified by the societal benefits that are anticipated to come from a successful study, if they are to be justified at all.

The NAS concluded that improving the accuracy of the science employed in regulatory decisions "constitutes a societal benefit", but also indicated several ways that particular studies can generate societal benefits beyond the minimal benefit of increased knowledge, including studies that:

- result in more stringent regulatory standards;
- enable EPA to adopt a public health measure it otherwise could not adopt;
- support approval of a product that protects public health; and
- improve the scientific accuracy of risk assessment for a class of chemicals and/or EPA decisions

Knowledge gained from the AHETF monitoring program will be applicable to a variety of pesticides, and will be used to assess risks of new pesticides and new uses of registered pesticides. Knowledge gained from the monitoring program could also be used by EPA to impose stricter safety standards on currently used pesticides, when appropriate (Resnick, 2005). Consequently, agricultural pesticide handlers could be better protected. Some information from AHED® will also support pesticide products that control disease vectors and protect the public health. These are all examples of situations which provide societal benefit that go beyond the basic benefit of increased knowledge.

The data developed in the AHETF monitoring program will also improve the scientific basis for EPA's occupational risk assessment because worker exposures will be measured under modern, common, and actual conditions. The data collection will reflect current agricultural practices, equipment, and techniques. Monitoring techniques are also of high quality and have been standardized for use across the AHETF monitoring program. AHED® will become the best available data to support assessments of agricultural pesticide handler exposure.

AHED® will not repeat the limitations of PHED. In particular, the AHED® database will include only data for individuals with dermal exposure data for all sampled body parts (unlike PHED where many records reflect exposure data for only some body areas). Improved estimates of whole-worker exposure, with a sense of the potential distribution between workers, will now be possible. In addition, to the extent the generic database approach proves successful; it will reduce the need for product-specific worker exposure studies conducted by individual registrants for new products and uses.

4.1.3 Benefits to AHETF Members

As described in previous sections, pesticide handler exposure data are required by Federal laws. Pesticide registrants (e.g., manufacturers, producers, and marketers) have the obligation to submit such data. Therefore, generating new human exposure data from AHETF studies is a benefit to the members of AHETF as they are all pesticide registrants that need data to meet regulatory requirements.

4.1.4 Benefits to Growers or Landowners

As described in Section 14, AHETF needs to locate growers or landowners who are willing to accept a pesticide product being applied to their crop or land. In the majority of studies, this will be a crop, grown either indoors or outdoors, and a grower will need to give AHETF permission to have their crop treated. However, pesticide treatments may also be made to rights-of-way, pasture land, forests, and other non-crop areas which may be owned by corporations or governmental agencies instead of growers. In many cases, AHETF is also interested in using workers that handle pesticides for that grower or landowner as part of their job.

Growers have considerable expense in cooperating with an AHETF study: time with the study team; lost productivity of their handlers while discussing the study with the study team; lost productivity on the day(s) of monitoring; wear on tractors and sprayers, fuel costs, etc. Therefore, it has become common practice for AHETF to provide the cost of test substance to the grower or landowner. Paying for cost of product for the acreage treated is viewed as reasonable compensation for the inconvenience associated with the study (see also Sec. 14.1.414.1.4). To the extent the compensation exceeds the inconvenience, paying for the pesticide product may be a benefit to the grower or the landowner.

4.2 Likelihood of Realization of Benefits

The generation of worker exposure data that can address the data needs of the regulatory community and membership of the AHETF is considered extremely likely. It is also very likely that regulators and risk assessors will use these data extensively. This has been the case for previous FIFRA joint data development task forces of many types, including those developing data for generic exposure assessment (e.g., for reentry and residential worker exposures). Regulatory agencies are strongly committed to using generic exposure databases as an important component of risk assessments. The use of worker exposure data in a generic manner has been generally accepted since 1986 when the concept was discussed and supported by a FIFRA Scientific Advisory Panel. In addition, the successful development and release of PHED in 1992 and its subsequent use by regulators to support many occupational risk assessments strongly suggests that the AHED® database will find even greater use.

5 Risks to Subjects in AHETF Studies

For the AHETF monitoring program, risks to subjects occur at the field study level and must be assessed and minimized at that level. This section of the Governing Document indicates in general terms the qualitative nature of the risks that are expected to be encountered and how they will be reduced in the design and conduct of specific field studies. Risks, and how they will be minimized, will be fully addressed in each field study protocol.

In summary, risk to subjects is classified as “greater than minimal”, primarily since agricultural work is considered a high risk occupation where the likelihood of harm or discomfort is greater than what is encountered in ordinary daily life. People who handle pesticides in their occupational activities are at greater risk than the ordinary public for accidents and physical injuries associated with equipment use or shop activities, and for adverse health effects due to the use of chemicals, which might include pesticides. In addition, AHETF believes the risk of heat-related illness (since workers must wear an extra layer of clothing to trap chemical) will be increased due to study participation (as compared to the risk associated with the job itself).

5.1 Risk of Heat-Related Illness

The risk of heat-related illness will be increased due to participation in some AHETF studies from the combination of hot and humid climatic conditions, extra clothing (in the form of inner dosimeters worn under normal clothing), and scripting of handling activities. Some of these aspects can be controlled by location and ventilation, yet other aspects are a direct result of the study design and cannot be easily altered without compromising the quality or value of the data collected. In particular, all participants will be asked to wear the inner dosimeters that they normally would not wear, but without compliance with this criterion a full set of dermal exposure measurements cannot be obtained. In addition, some studies will require the use of open cabs or cockpits, so workers will not have the choice of using an enclosed cab or cockpit during the study. As discussed later in this document, scripting in AHETF studies will be minimized and will primarily involve design features that ensure monitoring intervals that represent a typical day's duration (i.e., not excessively short or long) and coverage of the practical range for amount of product handled within each handling scenario. In some cases, this will increase the length of the work day which might add to the risk of heat-related illness. All of these factors can contribute to the risk of workers overheating and suffering from heat-related illness. Therefore, AHETF has developed an extensive program designed to minimize this risk.

5.1.1 Nature of the Risk of Heat-Related Illness

Heat stress is the build-up in the body of heat generated by the muscles during work and of heat coming from the environment. Heat illness (e.g., heat exhaustion and heat stroke) can result when the body is subjected to more heat than it can accommodate. Weather, workload, clothing/PPE, and worker conditioning can influence the risk of a worker experiencing heat-induced illnesses. In addition to causing serious physiological conditions, early symptoms of heat illness such as dizziness and confusion can lead to an increased risk of occupational accidents beyond that which is already present. Most early and mild heat-related illness conditions are reversible, but it is possible to have irreversible effects, especially if it contributes to an accident or injury. That is why the AHETF program for minimizing heat stress focuses on early identification and intervention.

5.1.2 Minimizing the Risk of Heat-Related Illness

Study participants are asked to wear an extra layer of clothing (whole body inner dosimeter) under their normal work attire which could increase the risk of heat-related illness. Efforts are made to schedule studies during cooler times of the year as much as practical to help minimize this risk. As stated in the informed consent form, heat-related illness is a potential health hazard that may be associated with participating in the study, so AHETF takes steps to prevent such illness. First, the on-site medical professional and study observers are provided with guidance to recognize symptoms of heat stress. Second, researchers always have plenty of water and sports drinks available and workers

are encouraged to drink some before and throughout the monitoring period. Most importantly, environmental conditions (temperature and humidity) are regularly monitored and operating procedures are in place to define when a study can start and when monitoring must be stopped. SOP AHETF-11.G addresses identification and control of heat stress in detail and a summary of that procedure is presented below.

In summary, the heat stress management plan includes the following procedures for researchers to prevent illness in study participants:

- Ensure plenty of water and sports drinks are available for the workers.
- During worker orientation immediately before participation in the study, remind the workers of the risk of heat stress, suggest they drink some water before they start work, and let them know how/where they can get water during the monitoring period.
- Urge workers to drink water during the monitoring period and remind them that thirst does not give a good indication of how much water a person needs to drink. There is no need to take hand washes or stop inhalation monitoring during a water break.
- Observe workers during the monitoring period and be aware of the signs and symptoms listed below.
- Require workers to take rest breaks when early signs or symptoms of heat illness are present.
- Monitor the heat index (based on air temperature and relative humidity and derived from a National Weather Service heat index table) at least hourly whenever ambient temperature is at or above 70 ºF.
- Stop the monitoring when the heat index (adjusted for direct sunlight, if applicable) reaches 120 ^{o}F.
- Have a medical professional on site to observe for signs of heat-related illness
- Know the location of the nearest medical facility

During the study, the Study Director (or a designee) is responsible for monitoring ambient temperature (ºF) and relative humidity (%) using portable devices (at least every hour when temperature is at or above 70 ºF). Based on these measurements, a heat index is determined from the National Weather Service Heat Index chart (reproduced in the SOP). In addition, the heat index is increased by 10 degrees if any study participant is working in the direct sun. If the adjusted heat index is at 120 ºF or above, monitoring cannot be started, and any ongoing monitoring will be stopped.

5.1.3 Nature and Likelihood of Residual Risk of Heat-Related Illness

AHETF tries to avoid scheduling studies in locations and at times where very hot and/or humid conditions are likely. However, it is probable that AHETF will be conducting some studies during summer months when climatic conditions will require increased vigilance by AHETF researchers to prevent heat-related illness. It is therefore possible that some monitoring events will need to be cancelled or stopped due to excessive temperature and humidity conditions. Adherence to the procedures developed to identify and control heat stress (SOP AHETF-11.G) will minimize the risks, but early signs of

minor heat-related illness could occur on very hot and humid days. This would lead to mandatory rest breaks and other preventive measures. Serious heat-related illness is not likely since researchers and the on-site medical professional will always be nearby to prevent the advancement of heat-related illness.

During 173 MUs monitored by AHETF, there have been no instances of workers experiencing heat stress or other physical injury that required them to stop to rest or to stop their monitoring altogether.

5.2 Risk of Exposure to Surrogate Chemicals

The surrogate chemical products selected by AHETF have all been through complete regulatory exposure and risk assessment processes and are approved for use in AHETF studies since the products are always used in accordance with the labels. This means that handler exposure levels are not expected to reach a "level of concern" to the EPA. Margins of exposure (MOE), usually based on no observable adverse effect levels (NOAELs), for the labeled uses and rates are considered adequate by the EPA. In addition, whole body dosimeters further mitigate dermal exposure, thus increasing the confidence that there is no reasonable foreseeable risk to workers handling the products utilized by AHETF. However, the product labels do provide precautionary statements indicating the most likely acute toxicity effects which are usually eye and/or skin irritation, and sometimes effects due to cholinesterase inhibition. These risks will be identified in the consent form and will be discussed with study participants. These are study-specific (and product-specific) requirements.

Since participants are generally allowed to participate on just one day of the study, the impact on their risk of long term toxicity is negligible.

5.2.1 Nature of the Risk of Exposure to Surrogate Chemicals

AHETF monitors exposure to workers who handle commercially available pesticide products. In general, useful surrogate chemicals have multiple uses (e.g., several crops or sites), multiple formulation types, minimal PPE requirements, and reliable and validated analytical methods. To date, AHETF has approved, or is in the process of developing, the following active ingredients as surrogate chemicals for use in its monitoring program:

- Acephate: an organophosphate insecticide
- Carbaryl: a carbamate insecticide
- Chlorothalonil: a substituted benzene fungicide
- Chlorpyrifos: an organophosphate insecticide
- Diazinon: an organophosphate insecticide
- Glyphosate: a glycine analogue herbicide
- Malathion: an organophosphate insecticide
- Mefenoxam (metalaxyl): an anilide fungicide
- Simazine: a triazine herbicide
- 2,4-D: a phenoxy herbicide

A discussion of the likely acute toxicity effects and the status of regulatory exposure assessments for each of these chemicals are presented below, however it should be emphasized that acute effects are formulation-specific, so each field study protocol and consent form will necessarily discuss the particular effects for the test substance used. In addition, scenario-specific MU Selection Plans will identify the pesticide active ingredients that might be used and will include MOE calculations (dermal, inhalation, and both routes combined) for the highest AaiH that is planned for each chemical and the task involved in the handling scenario.

5.2.1.1 Acute Effects of Proposed Surrogate Chemicals

The AHETF monitoring program is designed to generate exposure data for workers who handle a pesticide in a variety of ways for a period of time representative of a single (scenario-specific) work day. AHETF does not plan to use workers for repeated measurements, so the acute toxicity effects are of primary importance in selection of surrogates for exposure monitoring studies. In addition, AHETF uses only currently registered pesticide products, requires workers to follow all label and Worker Protection Standard (WPS) requirements, and will generally include study participants who would be using that product in the normal course of their job regardless of their participation in the AHETF monitoring study. Therefore, the one-day exposure will have a negligible effect on each participant's chronic risk.

The use of federally registered products means that the EPA has determined that the use of the product will not cause "unreasonable adverse effects on the environment" (which includes humans). This determination is made for both shorter term toxicity endpoints (including reproductive, neurological, and systemic toxicity) and repeated exposure endpoints. This includes cholinesterase inhibition which is a common endpoint of concern for several AHETF surrogate chemicals.

The table below summarizes the signal word and label precautionary statements (worst case) for formulations of the surrogate chemicals which AHETF intends to use in its monitoring program. These are generally based on the acute toxicity profile of the end-use formulation containing the active ingredient (as prescribed in 40 CFR 156) and provide guidance to AHETF about the relative risks to handlers.

Surrogate Chemical	Signal Word	Label Precautionary Statements
Acephate	CAUTION	Harmful if swallowed. Causes eye irritation.
Carbaryl	WARNING	May be fatal if swallowed. Harmful if absorbed through skin, inhaled, or in eyes.
Chlorothalonil	WARNING	Causes substantial but temporary eye injury. May be fatal if inhaled. Harmful if absorbed through skin. May be a potential skin sensitizer.

Surrogate Chemical	Signal Word	Label Precautionary Statements
Chlorpyrifos	WARNING	May be fatal if swallowed. Harmful if absorbed through skin or inhaled. Causes eye irritation.
Diazinon	CAUTION	Harmful if swallowed, absorbed through skin, or inhaled. Causes moderate eye injury.
Glyphosate	WARNING	Causes substantial but temporary eye injury. Harmful if swallowed, inhaled, or absorbed through skin.
Malathion	CAUTION	Harmful if swallowed, inhaled, or absorbed through skin.
Mefenoxam	WARNING	Causes substantial but temporary eye irritation. Harmful if swallowed or absorbed through skin.
Simazine	CAUTION	Harmful if swallowed, inhaled, or absorbed through skin. Causes moderate eye irritation.
2,4-D	DANGER	Corrosive. Causes substantial eye injury. May be fatal if absorbed through skin. Harmful if swallowed or inhaled.

It should be noted that signal words and precautionary statements do not provide complete information about the relative risks to handlers since PPE and/or engineering controls may be mandated based on other toxicology concerns, such as developmental toxicity. Different formulations of an active ingredient can have different signal words or precautionary statements on their labels. The specific chemical risk of each surrogate chemical formulation used in a study will be addressed in the field study protocol and consent form.

In addition, this is not necessarily a complete list of surrogate chemicals that AHETF will utilize during its monitoring program. Additional surrogates may be utilized and their registration status and risks will be fully discussed in scenario plans and study protocols for which they will be used.

5.2.1.2 Regulatory Risk Assessments of Surrogate Pesticides

All of the surrogate chemicals listed above were originally registered by EPA before November 1984 and were therefore subject to reregistration review, including the considerations dictated by the Food Quality Protection Act (FQPA) of 1996. During the reregistration process, EPA performed complete risk assessments and determined whether any risk mitigation procedures were necessary to ensure safety for that chemical (and perhaps others that operate by a similar mode of toxicity). These assessments included an evaluation of the entire toxicity database including potential chronic effects; developmental and reproductive effects; neurological effects; and other systemic effects. EPA findings were published in Reregistration Eligibility Decisions (REDs) or Interim Reregistration Eligibility Decisions (IREDs) which were sent to all registrants of the particular active ingredient. For handler situations, mitigation could involve reducing

application rates, eliminating uses, requiring PPE be worn by workers, or requiring engineering controls for mixing/loading and/or applications. Although registrants generally have some time to evaluate reregistration decisions, respond to EPA, and modify their product labels, AHETF intends to conduct all of its monitoring studies in accordance with published REDs or IREDs. The table below summarizes the pesticide reregistration status for the AHETF surrogates. For example, the diazinon IRED required that wettable powder (WP) products be packaged in water-soluble packets to reduce mixer/loader exposure potential, so AHETF will not use diazinon WP products that are not in water-soluble packets in a monitoring study even though those products may still be in the channels of trade and legal to use.

Surrogate Chemical	**Status**	**Changes which may affect AHETF**
Acephate	IRED signed 09/2001	Soluble Powders must be in water soluble packaging (WSP), enclosed cockpits for all aerial applications, eliminate low pressure hand wand applications to trees/shrubs/outdoor flora, no belly grinder application of granules
Carbaryl	IRED signed 06/2003 Revised IRED 10/2004	Added PPE for wettable powders (WP), some aerial applications eliminated
Chlorothalonil	RED signed 09/1998	Current labels meet all requirements
Chlorpyrifos	IRED signed 09/2001	Added PPE for most uses, WP must be in WSP, enclosed cockpits for all aerial applications
Diazinon	IRED signed 05/2004	Eliminate all aerial uses, eliminate most foliar applications to vegetables, WSP or lock-and-load for all products, enclosed cabs only for ground applications
Glyphosate	RED signed 09/1993	Current labels meet all requirements

Surrogate Chemical	Status	Changes which may affect AHETF
Malathion	RED signed 07/2006	WP must be in WSP, CR headgear required for all airblast applications, enclosed cockpits required for all aerial applications
Mefenoxam	RED signed 09/1994	Current labels meet all requirements
Simazine	RED signed 04/2006	Eliminate aerial applications
2,4-D	RED signed 06/2005	WP must be in WSP

All of the occupational risk mitigation steps presented in these REDs and IREDs will be followed during AHETF studies.

5.2.2 Minimizing the Risk of Exposure to Surrogate Chemicals

The primary method for preventing chemical toxicity when handling formulations of these surrogate pesticides is to ensure the workers follow the label requirements for clothing and personal protective equipment (PPE).

The Study Director or designated member of the study team must approve all clothing worn by study participants prior to the start of monitoring to ensure compliance with WPS (SOP AHETF-8.G.). Clothing with large tears, holes, rips, several missing buttons, or other defects that present a significant exposure to the worker's skin or inner dosimeter will not be accepted for use during the study.

During study conduct, researchers assigned to observe participants in AHETF studies will ensure the workers wear all the required PPE while handling product. Non-compliance on the part of the worker will result in discontinuing the monitoring for that worker. For example, the labels for formulated products containing several AHETF surrogate active ingredients include the requirement for all handlers to wear protective eyewear. AHETF will ensure that all study protocols involving any of these active ingredients require workers to wear protective eyewear that meets the requirements of the WPS. Researchers are also reminded when particular products may cause allergic skin reactions and that exposure monitoring will be discontinued for any worker that shows signs of such a reaction (including formation of a skin rash and/or itchy skin). Finally, researchers will remind workers just prior to participation about general ways to minimize exposure to chemicals such as washing their hands before eating and removing clothing/PPE that get contaminated by spills.

5.2.3 Nature and Likelihood of the Residual Risk of Exposure to Surrogate Chemicals

Since study participants will be handling approved pesticides for just one day, and AHETF researchers will ensure they wear label-required PPE, it is very unlikely that any serious or irreversible toxic effects will be encountered by study participants. However, acute toxicity effects, primarily reversible skin or eye irritation, may occasionally occur. During 173 MUs that AHETF has already monitored, there was only one report of eye irritation (not confirmed as associated with pesticide handling) that was reported the day after participation in a study. This suggests the probability of even minor effects is very small.

5.3 Risks Associated with "Scripting" of the Field Activities

During the conduct of some studies, AHETF will ask some participants to perform their job in certain ways that might not reflect their usual practice; that is, some procedures will be scripted. Scripting is primarily utilized to achieve diversity in certain factors that might have an impact on exposure potential for a particular scenario. In particular, scripting may be needed to ensure that at least three loads are handled or to ensure that certain amounts of active ingredient are handled.

Scripting may lead to increased AaiH and/or slightly longer work periods which may increase the risks of acute toxicity to the surrogate chemical or heat-related illness. These increases will be discussed in each field study protocol.

5.3.1 Nature of the Risk Associated with Scripting Field Activities

Scripting to handle at least three loads per work period often involves making modifications to the mixing/loading and/or application parameters such as using a smaller tank or increasing the application volume per acre (for liquids). Such changes will always be consistent with label directions and common regional practices. These changes will lengthen the work day somewhat, but do not increase the AaiH. Therefore, the added risk will come in the form of increased risk to heat-related illness on hot days.

Scripting to achieve diversity in AaiH may involve increasing or decreasing the amount of product a participant might handle if he weren't participating in the study. In the case of an increase in AaiH, the risks of acute toxicity to the surrogate and the risk of heat-related illness may be increased.

5.3.2 Minimizing the Risk Associated with Scripting Field Activities

The same general procedures discussed above that are designed to minimize the risks of surrogate chemical risks and the risk of heat-related illness will be followed in all studies, especially for participants who have their work day lengthened or AaiH increased due to scripting of activities.

5.3.3 Nature and Likelihood of Residual Risk Associated with Scripting Field Activities

Scripting will generally increase the risks associated with study participation only slightly. Each field study protocol will describe those situations where scripting is anticipated and the increase in risks associated with that scripting.

5.4 Psychological risks

Participating in AHETF exposure monitoring studies involves a couple of activities that are unusual and might lead to psychological concern for subjects. These include:

- Requiring prospective female participants to perform an over-the-counter urine pregnancy test prior to participation
- Allowing a researcher to assist participants with removing their long underwear (i.e., inner dosimeter)

Every field study protocol and consent form will identify the risks associated with taking a pregnancy test and changing clothes.

5.4.1 Nature of Psychological Risks

5.4.1.1 Risk Associated with Taking a Pregnancy Test

Female workers may be uncomfortable performing a supervised over-the-counter urine pregnancy test. A female who discovers she is pregnant may be disturbed by that information.

5.4.1.2 Risk of Embarrassment during Changing of Clothes

Subjects may be embarrassed to undress (they wear undergarments of their choice under the long underwear) in the presence of a researcher who will carefully collect the whole body dosimeter for analysis.

5.4.2 Minimizing Psychological Risks

During the consent process, female volunteers are informed they cannot participate if they are pregnant and they will have to take a urine pregnancy test within 24 hours of participation. The pregnancy test is generally performed after a consent form is signed since consent is usually taken more than a day ahead of planned monitoring. If a woman is uncomfortable with the thought of taking such a test, the person obtaining consent will urge her not to sign the consent form. However, women who have consented to participate may discover (often on the day of monitoring) that they are pregnant. This could cause significant psychological harm to some women.

Pregnancy tests are self-administered by the subject, but must be supervised by a female researcher. This researcher will explain how to take the test, escort the subject to a bathroom, and wait outside while the subject takes the test. The researcher will also explain that the subject does not have to disclose the results of the test, but that she will be asked to indicate after the test whether she wants to continue with the study. Only if the subject indicates an interest in continuing will the researcher verify the results (i.e., to confirm a negative result, see SOP AHETF-11.D).

The primary mechanism for minimizing the psychological harm of a positive pregnancy test result is to ensure no one but the subject herself is aware of the positive test. However, this risk cannot be eliminated and that point will be emphasized during the consent process.

When subjects complete their assigned handling tasks, they are required to allow researchers to collect OVS tubes, take a final face/neck wipe, take a final hand wash, and remove the whole body dosimeters for analysis. Undressing takes place in a private location such as the inside of a research truck or trailer enclosed by a curtain. Once inside this privacy area, a researcher of the same sex will assist the subject with removing the dosimeter. At this point, the subject will get dressed in the clothes he/she arrived in.

The primary mechanism for minimizing the risk of embarrassment during undressing is to provide a private area for undressing and to ensure the researcher is of the same sex as the subject.

5.4.3 Nature and Likelihood of Residual Psychological Risks

AHETF experience is that handlers rarely exhibit any concerns with the undressing procedure.

None of the women subjects in prior studies have expressed concern about self-administering the pregnancy test. Only a few women have been monitored and none have had a positive test result. The likelihood of a positive test causing psychological harm could be significant. Thus, the risks will be described to potential test subjects for all studies.

5.5 Risk of Exposure to Surfactant during Face/Neck and Hand Sampling

For all AHETF studies, a very dilute surfactant solution in water is used for face/neck wipes and hand washes. The only variation between MUs is in the duration of exposure to the surfactant solution since longer work periods or frequent eating breaks can lead to multiple hand washes and/or face/neck wipes. Every field study protocol and consent form will identify this risk of skin or eye irritation.

5.5.1 Nature of the Risk of Exposure to Surfactant

During face/neck wipes and hand washes, AHETF uses a very dilute solution of a non-ionic surfactant called sodium dioctyl sulfosuccinate (CAS No. 577-11-7). In its undiluted form, this surfactant causes mild to moderate skin and eye irritation in animals, depending on the concentration and duration of exposure. When used by AHETF to remove pesticide residues from workers' faces, necks, and hands a 0.01% v/v dilution in water is used and the duration of skin exposure is about 2 minutes per hand wash and 1 minute per face/neck wipe. Longer work periods or frequent eating breaks can lead to multiple hand washes and/or face/neck wipes and an increase in duration of exposure, but total dermal exposure to surfactant is generally less than 10 minutes for an entire monitoring period.

5.5.2 Minimization of the Risk of Exposure to Surfactant

The diluted surfactant is always prepared for use very shortly before being used in the field, generally within 24 hours of the first monitoring event. Distilled or deionized water is required by SOP and the dilute solution is typically prepared in clean 1 gallon jugs. Eye exposure would only be accidental and researchers are trained to be sure the amount of solution in face/neck wipe gauze will not cause dripping into the workers' eyes. The amount of surfactant is limited by SOP to approximately 4 mL (of the 0.01% v/v surfactant in water solution). AHETF will have a portable eye rinse system on hand at all studies in case such an accident does occur. Finally, when subjects complete their participation and leave the site in their street clothes, researchers remind them it is always advisable to shower or bathe as soon as possible.

5.5.3 Nature and Likelihood of Residual Risk of Exposure to Surfactant

This surfactant use represents a very dilute solution and a very short exposure period. A long history of using this surfactant in pesticide exposure monitoring studies indicates the likelihood of skin or eye irritation is negligible.

5.6 Background Risk of Injury Associated with Agricultural Work

Agriculture (i.e., farm occupations, see Bureau of Labor Statistics) remains one of the country's most dangerous occupations. It perennially ranks in the top ten occupations measured by fatality rate (on-the-job deaths divided by total number of workers) or injury/illness rate. These risks will be present for all AHETF field studies, but the nature of the risks will be scenario- and study-specific. This section describes in general the types of risk that are anticipated during the AHETF monitoring program.

5.6.1 Nature of the Risk of Injury Associated with Agricultural Work

The most common risks for serious injury to farmers are vehicular accidents (especially tractor rollovers, but also accidents while driving machinery on roads) and entanglement

with moving parts of farm machinery. Farm workers are also commonly exposed to a variety of chemical products that present increased risks compared to the general public. These include pesticides, fertilizers, solvents, lubricants, fuels, etc.

More than half of the scenarios of interest to AHETF involve some type of application equipment including handheld devices, ground rigs, or aircraft. The risk of injury will probably be greatest for studies involved with these scenarios since they involve intimate contact with large pieces of mechanical equipment. Mixer/loader activities probably involve a lower risk; however these tasks often involve getting close to the application equipment as well. Mixer/loaders are also more prone to lifting injuries since they may be handling containers with several pounds of product inside and sometimes have to move around hoses, pumps, or other equipment as they conduct their work.

Study participants might have an agronomic need to mix/load or apply other pesticides, fertilizers, or adjuvants in the same equipment as the surrogate pesticide. Thus, there might be an added risk of toxicity from these products.

5.6.2 Minimizing the Risk of Injury Associated with Agricultural Work

In general, background risks associated with agricultural work are out of the control of AHETF. However, study plans take into account that the risk of injury may be increased by scripting field activities such as handling more loads or more product than usual. The primary consideration when planning and conducting AHETF field studies is to have subjects use clothing, PPE, equipment, and facilities they are familiar with.

One very important situation that increases these background risks is the potential for early heat illness to cause dizziness or confusion that could lead to an accident or injury. On hot and humid days, researchers will be extra vigilant to prevent heat-related illness and follow established procedures to minimize the risk of heat-related illness.

In addition, care is always taken to keep air sampling pumps and tubes out of the way of workers so they don't interfere with their tasks and increase the likelihood of accidents.

The choice to add **tank mix** products is determined by the worker (or his supervisor) and generally can not be controlled by AHETF. Researchers always ask in advance whether the grower thinks tank mix additives will be needed, however it is common for these decisions to be made on the day of application or shortly before. In particular, weather conditions and pest pressures will often determine what non-surrogate chemicals are needed. AHETF will allow the use of such tank mix "partners" so long as they are legal uses, don't interfere with chemical analysis of the AHETF surrogate pesticide being applied, and do not require the worker to wear any additional PPE. Prior to allowing the use of tank mix partners AHETF researchers will ensure that none of these situations exist. Since AHETF does not require addition of tank mix products, any risks associated with exposure to such products would not result from a worker's participation in this research, and would simply be among the background risks normally experienced as part

of the job. Nevertheless, a researcher will review the label precautions for all tank mix products with the worker prior to their handling the products. This discussion will be documented by the researcher and ensures the workers are informed of the risks associated with these tank mix products. Tank mix products will not be allowed if the PPE required are contrary to the study objectives, in particular coveralls or chemical-resistant clothing requirements are usually unacceptable for AHETF study designs. If the grower insists such a tank mix product is necessary, and the Study Director determines extra PPE would be contrary to the study design, the study will not be conducted using that tank mix combination.

Individual study protocols and consent forms will address the risk of other tank mix partners in a general sense only, since it is impractical for AHETF to identify all potentially useful products and to define the risks associated with those products in advance.

5.6.3 Nature and Likelihood of Residual Risk of Injury Associated with Agricultural Work

Since AHETF studies involve common agricultural equipment and practices, physical injuries should be considered a possibility. In addition, it is common for growers to tank mix various chemicals to improve the effectiveness of the surrogate chemical or to provide other benefits (e.g., a different pesticide activity or nutrients) and these risks will be discussed with the subject. Field study protocols will also indicate that hot conditions and heat stress can increase the likelihood of physical injuries.

5.7 Likelihood of Serious or Irreversible Effects

As discussed above, participation in AHETF exposure monitoring studies might have an impact on the likelihood for acute toxic effects. The most likely routes of exposure are dermal and inhalation, not oral. For some of the surrogates listed above, their labels list possible serious or irreversible effects from eye exposure, inhalation exposure, cholinesterase inhibition, or skin allergies. These types of effects would generally be a result of misuse or accidental spills, not from prescribed label use. Since all pesticides will be handled in accordance with label instructions, misuse should not occur; however, spills are still possible. Accidental exposures of sufficient magnitude for these serious or irreversible effects are unlikely.

Farm accidents are relatively common and can result in serious or irreversible effects, including death. Accidents are generally out of the control of AHETF, so these effects are considered somewhat likely. In addition, heat stress might cause dizziness or confusion and increase the risk of mechanical accidents that could be serious or irreversible. However, the heat stress management procedures are designed to minimize this risk through prevention and early detection practices, so serious heat-related accidents are unlikely.

6 Risk versus Benefit Comparison

In general, the risks to participants in all AHETF studies are outweighed by the benefit to society in the form of high quality exposure data for use in evaluating pesticide safety (as described below). If there is no need for new exposure data for a particular scenario, no studies will be proposed by AHETF relating to that scenario.

Each field study protocol will discuss the particular benefits, risks, and risk/benefit comparison to ensure the benefits outweigh the risks. This comparison must balance the societal benefit of new scenario-specific exposure data with the study-specific risks to subjects.

There are no direct benefits to subjects, but some future indirect benefits to handlers as a whole are anticipated as new data are used to regulate pesticides. There are also benefits to growers, the EPA, AHETF members, and society in general.

AHETF's monitoring program presents a greater than minimal risk to participants. The primary risk comes from their employment as an agricultural worker where accidents and chemicals contribute to injury and illness. The increased risk to heat-related illness caused by the extra layer of clothing is mitigated by a medical management program which emphasizes prevention measures and guidelines for stopping participation when warranted based on environmental conditions.

The benefit to agricultural workers as a whole and to society in general, in the form of more accurate measurements of potential exposure to pesticides, must outweigh the risk to study participants.

7 Description of and Rationale for Scenarios Considered by AHETF

The handling scenarios selected for inclusion in the AHETF program reflect logical classifications of tasks, equipment, and formulations. Many of these scenarios are similar to those in PHED which have proven to be practical for regulatory use (to support product-specific exposure assessments using a generic database). A few others cover use situations that are not included in PHED.

In addition, AHETF evaluates deficiencies in PHED data for each scenario before any monitoring is done for that scenario to be sure adequate data for exposure assessments do not already exist.

7.1 Handler Scenarios Included in the Monitoring Program

This section discusses how scenarios were chosen for inclusion in the AHETF monitoring program.

As discussed above, EPA (and other regulatory agencies) generally utilizes a scenario approach to exposure and risk assessments which is logical and has proven to be practical. The various scenarios reflect logical categories since task, equipment type, formulation type, and engineering controls can greatly impact the potential for handler exposure. These categories are also practical since product labels are formulation-specific and often address only certain types of application equipment. In addition, regulators find scenarios convenient for product-specific exposure assessments and must ensure worker safety for all scenarios in which a product could be used. In practice, regulators can often mitigate exposure by requiring engineering controls for certain mixing/loading or application techniques (i.e., changing the scenarios that are allowed), limiting use rates, eliminating some uses, or a combination of these methods. All of these mitigation decisions can be supported by scenario-specific exposure data such as those in PHED and AHED®. This scenario approach is outlined in the PHED Surrogate Exposure Guide (Keigwin, 1998) and is consistent with Agency-wide guidelines for exposure assessment.

Collecting occupational pesticide handler exposure data by scenario therefore reflects the following parameters that conventional wisdom and experience indicate have a large impact on the potential for handler exposure:

- Task (e.g., mixing/loading or application)
- Application System (e.g., aircraft, ground, or hand-held equipment)
- Product Formulation (e.g., liquid, powder, or granule)
- Engineering Controls (e.g., open or closed loading and open or enclosed cabs)

AHETF member companies have examined their products and uses so the task force could collectively define the scope of their project in terms of which handler scenarios will be represented by AHED®. Since monitoring MUs is very costly, AHETF does not include scenarios in its scope that EPA rarely considers and that don't represent major use patterns in agriculture. The current scope includes the following 33 scenarios:

7 Mixer / Loader Scenarios:

M/L System	Product Formulation	PHED Scenario Number
Open Pour	Liquid	3
	Dry Flowable	1
	Wettable Powder	4
	Water-Soluble Packets	5
	Granule	2
Closed	Liquid	6
	Granule	No Scenario

17 Applicator Scenarios:

Application System	System Specifics	Formulation (As Applied)	PHED Scenario Number
Aerial	Open Cockpit (rotary-wing only)	Liquid	No Scenario
		Granule	No Scenario
	Enclosed Cockpit (fixed-wing)	Liquid	7
		Granule	8
	Enclosed Cockpit (rotary-wing)	Liquid	9
		Granule	
Airblast	Open Cab	Liquid	11
	Enclosed Cab	Liquid	12
Groundboom	Open Cab	Liquid, no SI[a]	13
		Liquid, SI	
		Granule, no SI	15
		Granule, SI	
	Enclosed Cab	Liquid	14
		Granule	16
	Rights-of-Way	Liquid	24
Hand-Held	Low Pressure	Liquid	18
	High Pressure	Liquid	19

[a] Soil Incorporation

5 Mixer / Loader / Applicator scenarios:

Application System	Product Formulation	Mix / Load System	PHED Scenario Number
Belly Grinder	Granule	Open Pour	30
Backpack	Liquid	Open Pour	34
	Granule	Open Pour	No Scenario
Mist Blower	Liquid	Open Pour	No Scenario
Chemigation	Liquid	Open Pour	No Scenario

4 Seed Treatment scenarios:

Seed Treatment Location	Product Formulation	PHED Scenario Number
Commercial	All	No Scenario
On-Farm, Includes Planting Seed	Solid	No Scenario
	Liquid	No Scenario
	Treated Seed	No Scenario

Each of these scenarios can be viewed as a distinct research project for which AHETF will: a) develop a comprehensive MU Selection Plan designed to address the primary and secondary objectives, b) conduct one or more studies to collect MUs from different locations, and c) analyze the exposure data from the collected MUs to evaluate whether the objectives were met.

Technical guidelines for how scenario MU Selection Plans will generally be developed and analyzed are provided in Sections 9 and 10, Appendix B, and Appendix C of this document. For any scenario that requires modifications to the standard statistical design or analysis process, the scenario MU Selection Plan will detail the reasons why and provide complete justification for the modifications. Each scenario MU Selection Plan will be referenced by each field study protocol that supports that scenario.

7.2 Limitations of PHED

Since 1992, the EPA has conducted agricultural mixer/loader and applicator exposure and risk assessments relying primarily on the data in PHED. PHED version 1.01 was released in February 1992. It was followed by PHED version 1.1 in February 1995. PHED version 1.1 was described by the Agency as an incremental improvement over the 1.01 version (Pesticides Handlers Exposure Database, User's Guide Version 1.1, Health Canada, U.S. Environmental Protection Agency, American Crop Protection Association, February 1995). The forward to Version 1.1 User's Guide cautions the user that the database still has some limitations and should not be considered a panacea in estimating pesticide handler exposure. Noting the limitations, the guide states that a goal was to release a PHED version 2.0 in 1997. However, no subsequent version of PHED has been released.

By 2000, the U.S. Environmental Protection Agency began evaluating alternatives to PHED. On 16 March 2001, the Agency outlined its intentions regarding PHED (Letter from Margaret Stasikowski, Director, Health Effects Division to Daniel Fay, Valent USA Corporation, 16 March 2001). The letter stated EPA's intention to drastically overhaul PHED version 1.1 because many of the existing exposure studies in the database were outdated or scientifically inadequate by "today's standards". In addition, many exposure

scenarios that are being assessed by the Agency are under-represented in PHED version 1.1.

In summary, PHED suffers from a number of limitations regarding its use as a generic exposure database, including:

- Inadequate number of measurements for one or more body areas (that reduces the confidence in exposure estimates for those areas);
- Inadequate quality assurance or quality control data (that sheds doubt on the reliability of all measurements);
- Use of patch dosimeters instead of whole-body dosimeters (that don't accurately reflect entire body exposures and requires extrapolations from a small patch area to the entire body area);
- Lack of entire body dermal estimates for workers (i.e., not all body parts monitored for dermal exposure) (that reduces the confidence in exposure estimates for those areas);
- Many non-quantifiable residues on dosimeters (that can lead to overestimates of actual exposure by assuming one-half the limit of quantification is present on all dosimeters with non-quantifiable residues when it may be even lower – this overestimate can also be magnified when patch dosimeters are used and the assumed residue is extrapolated to an entire body area);
- Lack of diversity for study conditions (e.g., same workers used repeatedly or all workers handling the same amount of product) (that reduces the confidence that measurements are reflective of a variety of common practices); and
- Lack of representativeness of study conditions (e.g., products or procedures that are no longer in common use) (that sheds doubt on exposures for modern agricultural equipment and practices)

The U.S. EPA recently convened a FIFRA Scientific Advisory Panel (SAP) to review the current worker exposure assessment methods (U.S. EPA, 2007), including a summary of the technical plan for AHETF. Regarding adequacy of handler exposure data in PHED, the SAP (2007) concluded:

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

In particular, the SAP (2007) also concluded:

> The inclusion within PHED of studies where either not all parts of the body were monitored or a substantial number of exposures were

undetectable do not allow the results to yield accurate exposure statistics of interest for regulatory assessments.

Issues regarding the adequacy of the data in PHED can also be illustrated by reviews of Registration Eligibility Decision (RED) documents issued by EPA as part of the recently completed FQPA reregistration process. These documents have characterized the existing PHED data as low confidence for the following important use patterns. Confidence ratings are based on "number of replicates" (quantity) and "QA/QC Grades" (quality). In general, low confidence scenarios have fewer than 15 replicates and/or barely acceptable laboratory fortification recovery data (or worse).

Low Confidence Scenarios in PHED include:

- Mixing/loading of wettable powder in water soluble packaging (Scenario 5);
- Aerial application of a granular formulation (Scenario 8);
- Application by rotary aircraft (Scenario 9);
- Enclosed cab airblast application without gloves (when inside the cab as allowed by the WPS) (Scenario 12);
- Application of granular formulation by broadcast spreader (Scenario 15);
- Low pressure hand spray applications for greenhouses (Scenario 18);
- High pressure hand spray applications for greenhouses (Scenario 19);
- Application by backpack sprayer (Scenario 20); and
- Application to rights-of-way (Scenario 24)

For reference, PHED confidence ratings can be summarized as:

Confidence Rating	**Number of Measurements**		**QA/QC Grading**
High	>= 15 per body part	And	Good laboratory plus good field fortification data (or better) (Grade AB)
Medium	>= 15 per body part	And	Moderate laboratory fortification data plus either poor field fortification or moderate storage stability data (Grade ABC)
Low	< 15 per body part	Or	Barely acceptable (or unacceptable) laboratory fortification data (Grades D or E = All Grades)

In addition, it should be noted that PHED provides dermal exposure estimates, and confidence ratings, for several distinct clothing situations:

- no clothes (i.e., based on outer dosimeters or clothing);
- single layer of clothing, no gloves;
- single layer of clothing, with gloves; and
- coveralls over single layer of clothing, with gloves (some scenarios)

Therefore, PHED can have low confidence for one clothing/PPE situation and high confidence for another within an exposure scenario. While protection or penetration factors can be used to estimate protected exposure from non-protected exposure results, or vice versa, this creates additional uncertainty for exposure estimates and may not be appropriate for all risk assessments.

7.3 Scenario-Specific Data Needs for AHED®

Before AHETF collects additional exposure monitoring MUs for any particular scenario, it will provide a justification that additional data are needed for that scenario. This will primarily involve a detailed analysis of the quality and quantity of exposure data in PHED as well as any data that have already been purchased by (see Section 8) or conducted by AHETF. AHETF will then propose a plan for generating additional MUs sufficient to define the expected distribution of exposure for that scenario (see Sections 9 and 10 and Appendices B and C for more details). These scenario-specific data evaluations and plans for new MUs will be submitted as separate documentation with each new field study that AHETF proposes to conduct.

These scenario MU Selection Plans will discuss the status of available knowledge; the goals of further research; the most appropriate normalization factor in terms of the primary and secondary objectives (i.e., AaiH, other, or none); and the justification for the number of clusters, number of MUs per cluster, and the factors that will be used to select both clusters and MUs under the diversity selection approach. As each field study is completed, the scenario plan will be reviewed to reflect what has been learned; for example, to ensure appropriate diversity of conditions within the scenario, especially for the appropriate normalization factor such as AaiH. However, the statistical design will not be modified as MUs are collected since it is only after all planned MUs for a scenario are collected that an evaluation can be made to determine if the data meet the benchmark objectives.

8 Evaluation of Existing Data and Incorporation into AHED®

Following the determination that PHED version 1.1, as a whole, did not meet the current needs for the conduct of handler exposure assessments, the AHETF evaluated existing handler exposure data available to the task force. Each of the studies considered had the potential to provide exposure data and supporting information from MUs for a proposed AHED® scenario. The evaluation process involved the following steps:

- ***Development of data acceptability criteria***: The existing data acceptability criteria addressed general study design and exposure monitoring techniques, including the analytical and quality control aspects of the studies. These acceptability criteria are detailed in Appendix A.
- ***Primary review***: A process that involved the screening of handler exposure data from PHED version 1.1, publicly available data, and compensable data owned by AHETF members. Approximately 145 studies (about 1,800 MUs) were rejected during this process for not meeting basic design criteria.
- ***Secondary review***: A detailed evaluation of data that passed the screening process for acceptability under the acceptance criteria. Decisions and rationale for those decisions were documented in a report for each study reviewed. Approximately 71 studies (about 1,200 MUs) were reviewed during this more detailed process.
- ***Final review***: A process that involved concurrence by the Joint Regulatory Committee (JRC) on acceptance of the data for use within AHED®.

A total of 216 existing studies (about 3,000 MUs) were evaluated from which a total of 105 MUs were deemed to be suitable for the AHETF generic database. It should be pointed out that 19 of these studies reflect data from PHED and only one of these studies met the acceptance criteria. AHETF eventually acquired the rights to use these 105 MUs and they were added to AHED® under the appropriate scenarios. In most cases, these existing data are not sufficient to satisfy the full MU selection requirements for the respective scenarios. These scenarios will be supplemented by obtaining new human exposure monitoring data (see Appendix C for more details on sampling designs when there are some existing data).

During the secondary and final review stages, much of the existing data were deemed unsuitable for a generic database (and were not acquired) due to poor QA/QC (generally low or insufficient field fortification results), a preponderance of non-quantifiable residues, or the use of study conditions that do not represent current agricultural practices in North America. However, the technical issue that eliminated the most existing data was the decision to exclude exposure data for workers who wore more than a single layer of clothing. This decision was discussed with the JRC who agreed that a modern generic database would be most useful if it contained exposure data for minimal clothing and PPE situations. Regulators are generally more comfortable estimating exposures to protected areas (e.g., dermal exposure under coveralls plus normal clothing) using exposure measurements from unprotected areas (e.g., dermal exposure under just one layer of normal clothing) than vice versa. Therefore, AHED® has been designed so that clothing/PPE protection factors can be specified by a user in order to estimate protected exposures, but not vice versa.

Each scenario-specific MU Selection Plan will determine whether data already exist that are suitable for a generic pesticide handler database and will summarize those evaluations. Then a suitable plan will be made to generate sufficient data to meet the benchmark objectives as described in Appendix C.

9 Design Concepts Common to All Scenarios

Each AHETF agricultural handler scenario will be addressed by several studies. The purpose of each study is to obtain experimental monitoring units (or MUs) for incorporation into a generic exposure database, AHED®. The specific details for each scenario design will necessarily differ from scenario to scenario. In addition, individual studies within each scenario will usually employ variations within the parameters of the scenario design. The scenario design documents and study protocols will provide the rationale and description of the specific MU selection procedures.

This section summarizes the scenario design concepts that are common to every AHETF scenario. Appendix B of this document provides a more detailed rationale for using experimental monitoring data to characterize future worker exposures when performing scenario-related tasks. Appendix C discusses details of the statistical basis and general methodology used by the AHETF to determine the sample size and configuration for each agricultural handler scenario.

9.1 Predicting Generic Exposures with Experimental Monitoring Units

For the purposes of the AHETF Monitoring Program, the basic element of every scenario is considered to be the handler-day (or HD). Each handler-day corresponds to a particular worker and the scenario-related activities that he or she performs during a single work day. Regulatory interest for each agricultural handler scenario is centered on predicting occupational exposure under a specific set of generic future handler-day conditions. In particular, it is desired to characterize exposures resulting from the future use of an arbitrary (and perhaps currently non-existent) pesticide active ingredient at some arbitrary, but quantifiable, amount of expected active ingredient contact.

The MU is the basic experimental tool used by the AHETF Monitoring Program to predict future exposures. Each MU is a set of scenario-specific handler-day conditions that have been selected (i.e., chosen, simulated, scripted, or otherwise constructed) to represent expected future HD conditions. Every MU is also monitored to obtain an experimental measurement of the actual exposure resulting from the simulated or chosen HD conditions.

An MU can predict future exposure only when the handling conditions of the MU are similar to those of future HDs. Because each MU is expensive, it will be possible in practice to construct only a small number (N) of MUs. Obviously, any set of only N MUs could not hope to predict every conceivable future HD condition. But it is feasible to characterize the diversity of future scenario-dependent conditions in a way that will have practical regulatory value. The successful use of a small set of MUs to represent the diversity expected among future handler-days is aided by the following (see Appendix B):

1. Dermal and inhalation exposure to the chemicals are considered generic (i.e., independent of the particular active ingredient used). This generic principle permits use of a small set of surrogate active ingredients to predict exposure from other active ingredients.

2. Exposure is considered to be proportional to the true amount of active ingredient contacted by the agricultural handler. The exposures obtained from MUs can be expressed relative to a measurable normalizing factor (NF) that is expected to be at least proportional to the true amount of active ingredient contact. Then exposures for any future NF level of interest (such as the maximum amount of active ingredient that might be handled for a new product) can be predicted by multiplying the normalized exposure by this future level.

3. The AHETF has, or has access to, subject matter experts who are knowledgeable about the range of future handler conditions expected for each scenario and which of these are likely to impact exposure. This permits construction of a set of MUs that are diverse with respect to HD conditions likely to impact exposure. This, in turn, results in MUs more likely to span the range of possible future exposures.

A standard device for thinking about all possible future handler-day exposures for a scenario is the exposure distribution. The future HD distribution describes the likely exposure that would result if one were to randomly pick a future HD among those workers using ai X when the level of the normalizing factor is H_X. This distribution does not describe a currently existing population in the sense used by sample surveys. It only describes handler-days that have the potential to exist in the future. Because of the generic principle (see 1 above) each future distribution can apply to any active ingredient, including those that are not currently in use.

There are actually a series of future exposure distributions of regulatory interest, one for each possible value of the normalizing factor. However, since exposures are considered proportional to the NF, (see 2 above) it is only necessary, and simpler, to focus on the distribution of normalized exposures. The distribution of normalized exposure (i.e., the exposure divided by the normalizing factor for each worker) is, in a sense, a generic exposure distribution. It does not depend on the particular active ingredient and can be used to predict the exposure for any specified amount of active ingredient contact (i.e. a particular NF level).

In practice, the complete normalized exposure distribution is rarely needed. Regulatory interest is most often focused on two general aspects of this distribution:

- The **middle values** such as the mean or median. These exposure values tend to characterize average or 'typical' exposure levels.

- The **larger values** of exposure possible, such as the 95th percentile of the distribution. This aspect better characterizes the extreme, one-time, worker exposures.

The normalized exposure measurements obtained from a set of N experimental MUs are also considered generic since they can be used to predict a corresponding set of N exposures for any active ingredient at any level of the normalizing factor. Thus, when populated by data derived from MUs, AHED® can be properly termed a generic exposure database. The MU-derived normalized exposures for a particular scenario in AHED® are considered an aggregate characterization of future normalized exposures. However, they are not a true random sample from the exposure distribution. This is because, the HD conditions used to construct MUs can never be 'sampled', randomly or otherwise, from HD conditions that do not yet exist. Therefore, technically, the MUs comprise a set of synthetic handler-day conditions that are used as a starting point from which to generically predict a set of future HDs for arbitrary pesticides.

However, some type of random sampling interpretation might still be a convenient and reasonable approximation for how a set of MU-derived predicted exposures relates to the generic exposure distribution for a scenario. This permits the use of conventional statistics, such as means and percentiles, calculated from the observed MU exposures to approximate the 'middle' and 'larger' values expected in future HDs. For the AHETF Monitoring Program, confidence in this approximation is improved by using a nested reference random sampling model rather than assuming simple random sampling (see Sec. 9.3.1 and Appendix C). In addition, diversity selection (using both purposive and random components) is used whenever possible (see Sec. 9.2.2 and Appendix B). This increases the likelihood that the range of conditions expected to impact exposure in the future HD population is reflected in the 'pseudo-sample' of MUs as well.

9.2 MU Selection Procedures Common to All Scenarios

The process of obtaining scenario-specific handling-day conditions (including workers) needed to create monitoring units is a complicated process. This process is termed MU selection and refers to any method that establishes the handling-day conditions for a particular monitoring unit. Selection may involve choosing, restricting, or 'scripting' HD conditions to insure that the complete set of MUs better reflects the diversity expected of future (and generic) handler-days. The specific MU selection methods used will vary from scenario to scenario and can involve both random and nonrandom selection elements. However, there are two general aspects of the MU selection common to all scenarios. These are (1) a natural two-stage selection structure, and (2) the superimposition of diversity selection methods onto this basic structure.

9.2.1 Two-Stage Selection of MUs

The final set of MUs for all AHETF scenarios can be envisioned as the result of two successive stages of selection. The first stage of selection consists of identifying geographic areas with a corresponding range of potential monitoring dates within that area. Each such local area is called a monitoring site and serves as a base from which to stage monitoring operations. For AHETF scenarios, it is common for monitoring sites to

extend over multiple counties and have potential dates of monitoring that cover about a week.

The second stage consists of selecting agricultural workers and other handling conditions within each monitoring site that are then used to construct the MUs. In general, N_c sites are selected at the first stage and N_m monitoring units will be obtained within a site at the second stage. N_m will not necessarily be the same at all sites. When N_m is greater than one, the set of MUs at the same site is referred to as an MU cluster. In general, the handling conditions for MUs in the same cluster are expected to be more similar than those in different clusters. Thus, there is expected to be some degree of within-cluster correlation of exposure. Theoretically, such within-cluster correlation favors the use of only a single MU per monitoring site. However, there are often substantial overhead costs per site that make multi-MU sites more efficient (see Appendix C).

9.2.2 Diversity Selection

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-sample will tend to be more variable with respect to exposure than would a same-sized representative sample. As a result, a diversity selection 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. For regulatory purposes the important aspects of the distribution of exposures are central tendencies and upper percentiles. In addition, overestimation of these characteristics is less of a regulatory problem than underestimation since it usually results in overprotective restrictions on pesticide use. Thus, the fact that diversity selected MUs are biased towards greater exposure variation is of little practical concern to the potential users of AHED®.

All scenarios will employ some form of diversity selection independently at each of the two stages of selection. Thus, at the first stage, monitoring sites will be selected in ways that are expected to result in average differences in exposure between clusters of MUs. At the second stage, the handler-day conditions obtained for MUs at each monitoring site are selected in ways to promote within-cluster diversity in exposure. At either stage, diversity selection can employ random and/or purposive methods. Random diversity selection avoids the appearance of intentional bias that can result when researchers choose some HD conditions and exclude others. On the other hand, purposive diversity selection can be more efficient and cost effective whenever the possible choices of HD conditions are non-equivalent in terms of cost or convenience.

In a small sample it would be impossible to obtain diversity with respect to every characteristic that can potentially impact exposure. The number of possible ways to obtain MUs that differ grows exponentially with the number of factors and the number of possible levels of these factors. Only a few of these possibilities could be selected with a small set of MUs. When MUs differ with respect to many characteristics, it is difficult, if not impossible, to determine which configurations are more diverse than others. In addition, some combinations of factors might not be expected to occur in a future handler-day population. Therefore, not every characteristic that may impact exposure can be, or even should be, considered in diversity selection.

Whenever possible, therefore, only a few characteristics, preferably meta-factors, can be used effectively in diversity selection. Meta-factors are characteristics that directly or indirectly influence a number of other factors. For example, a worker is a meta-factor because substituting one worker for another alters a number of factors (e.g., behavior, physical appearance, stamina, and equipment available) that might affect exposure. Other common meta-factors are geographic location and season. When MUs are different with respect to a meta-factor they will likely be different with respect to most of the factors determined by the meta-factor. More importantly, varying a meta-factor is unlikely to result in combinations of MU conditions that are impossible or rare in future handler-days.

There are a number of ways to achieve diversity among MU handler-day conditions. The most straightforward approach is to purposively select MUs that appear to be sufficiently different with respect to the characteristic(s) of interest. It is often practical to implement some form of stratified diversity selection. In this approach available selection units (e.g., monitoring sites or monitoring units) are partitioned into strata based on characteristics likely to impact exposure. Each potential selection unit must belong to one and only one stratum. The number of strata must be at least as large as the number of units that will be selected. Diversity could be achieved by selecting (purposively or randomly) no more than one unit from each stratum. If there are more strata than units to be selected, then a subset of the strata should be selected first. This could be done either purposively (to increase diversity or reduce costs) or randomly (to reduce intentional selection bias).

9.2.2.1 Diversity Selection of Monitoring Sites

Independent monitoring sites for the same scenario are expected to differ as much as separate studies of worker pesticide exposure. The AHETF and others have found that such studies typically demonstrate substantial differences in mean normalized exposure. If all MUs for a scenario were obtained from the same monitoring site, the normalized exposure could be quite non-representative of the entire scenario and the exposure variation would be too small. By distributing MUs over multiple monitoring sites, the total HD diversity in the set of MUs is increased. Maximum diversity could be obtained if no more than a single MU were obtained at each monitoring site. However, as shown

in Appendix C, the considerable fixed costs per monitoring site make multiple MUs per site more efficient.

Diversity selection of monitoring sites means obtaining those that are different from each other, on the average, with respect to some handler-day characteristics expected to impact exposure. If there are a number of potential monitoring sites available and a set is to be selected (randomly or purposively), then a stratified diversity selection of sites, based on geographic regions or other important characteristics, is a feasible approach to obtaining the final set of N_c monitoring sites. Typically, merely changing geographic locations varies exposure potential that comes from numerous regional differences in workers, equipment, or environmental conditions. Because geographic location is a strong meta-factor, selecting monitoring sites from different geographic (and possibly time period) strata should provide substantial diversity at this stage of selection.

9.2.2.2 Diversity Selection of Monitoring Units

For all scenarios, the handler-day conditions for MUs will be diversified independently within each selected site. In most cases, within-site diversity selection of MUs will focus on only two primary meta-factors: worker and normalizing factor.

Every worker contributes a unique set of physical and behavioral characteristics to a monitoring unit. For AHETF scenarios in particular, workers also are strongly associated with subsets of possible scenario conditions. For example, a worker might only have available a limited set of equipment or be able to handle a limited range of active ingredient. This means that worker is a very strong meta-factor since a change in worker will invariably result in changing many factors (known and unknown) that will impact exposure. All MUs constructed with the same individual worker would be expected to show similarities in handler-day conditions and in exposure. Therefore, HD diversity is increased by simply requiring that each MU be constructed with a different individual worker.

In principle, additional diversity might be obtained by also selecting workers from different strata based on characteristics such as stature, ethnicity, etc. However, because the MUs in each cluster are also diversified by the normalizing factor (see below), additional diversification by such 'worker type' categories increases complexity with little additional benefit. A possible exception might be the use of simple restrictions to reduce obvious sub-clustering of HD conditions at the same monitoring site. For example, the number of participants recruited from the same grower or monitored on the same day could be limited. This reduces the likelihood that any two workers will be similar in characteristics such as environmental conditions, worker training, equipment types, or standard pesticide handling practices that could impact exposure.

The AHETF Monitoring Program will also attempt to diversify, as much as possible within each cluster, the single normalizing factor (NF) deemed most appropriate for the scenario. The NF is a strong meta-factor since different values of the NF will be

naturally associated with specific sets of handler-day conditions in the future population of HDs. For example, the equipment and handling activities could be quite different for handling small amounts of active ingredient than for handling larger amounts (e.g., in size or type of application equipment into which the pesticide is mixed/loaded). By insuring that the levels of NF are varied, the set of MUs indirectly captures diversity in those components naturally associated with the normalizing factor. In addition, for many scenarios, a secondary benchmark objective is to ensure there are adequate data to allow AHED® users to determine if, on the average, exposure appears proportional to the particular normalizing factor used (see 9.3.2 and Appendix C). This will also require obtaining a sufficient number and type of MUs over the practical range of the normalizing factor so that if exposure is truly unrelated to NF, a test of consistency with a proportional relationship will be rejected.

In most cases the normalization factor used is the amount of active ingredient handled (AaiH). It is also conceivable, albeit unlikely, that no normalizing factor would be appropriate for a particular handling scenario. In such an event, the AHETF recommends that the meta-factor AaiH be diversified by default. To accomplish diversification of the normalizing factor, the default approach for all scenarios is to partition the practical range of the NF at a monitoring site into strata and use only a single NF level from each stratum within each cluster. Selection of the actual NF level from each stratum can be random or purposive. However, since AHETF MUs incorporate actual commercial work activities, fine control over the NF level is not always possible.

For most AHETF scenarios, worker availability strongly depends on the particular NF level chosen. Some individuals might only work with higher NF levels and some with only lower NF, for example. When there are multiple workers available for a particular NF level, they, and their associated HD conditions, could be randomly selected. However, when such strong associations between subject, NF level, and other HD conditions exist, purposive allocation of workers and NF levels to MUs might result in a more cost effective and practical configuration.

In addition to the three primary meta-factors (i.e., monitoring site, worker, and NF), AHETF experts do attempt to identify and list other factors expected to vary in the future handler-day population for a scenario and to have an impact on worker exposure. These might include factors such as equipment used, specific worker techniques, and number of product containers used. For the most part, such factors should vary naturally with the three primary meta-factors and additional perturbation should be unnecessary. In the event two putative MUs have excessive similarity with respect to such factors, consideration will be given to reducing this similarity.

9.3 Scenario Sample Sizes

In the strictest sense, statistical sample size determination methods require random, representative sampling from a population. Only then is statistical theory able to predict how increasing sample sizes yield better data-based estimates of that population's

characteristics. For all AHETF scenarios, such rigorously-controlled probability sampling cannot be used. Each set of MUs is not sampled from an existing population but actually constructed to predict a set of generic exposures for a non-existent future population. The MUs include some HD components that are scripted and some chosen from existing conditions by a mixture of random and purposive selection. In such cases, sample size can only be determined using a reasonable 'random' situation as a reference model. The random reference model is defined so that it reflects the actual selection process as closely as is practical. The sample size that is appropriate for the reference model is then used as a 'benchmark' for the actual study design. In a real sense, then, this reference random sampling model is used to establish benchmark sample sizes that can satisfy benchmark objectives. Although rarely stated explicitly, the use of reference sampling models and benchmark objectives is quite common.

9.3.1 Reference Random Sampling Model

Because all AHETF scenario designs will have a two-stage selection structure, they all assume the same form of reference sampling model. For each scenario, two-stage random nested (or cluster) sampling is the reference model used for the combination of purposive and random two-stage diversity selection that can actually occur. This reference model assumes that:

1. Exposure, normalized by an active ingredient contact factor, is lognormally distributed with geometric standard deviation GSD.

2. There are N_c clusters (i.e. monitoring sites) randomly sampled from all possible sites.

3. There are N_m MUs randomly sampled per cluster. Unequal numbers of MUs per cluster are possible.

4. There is a within-cluster (i.e., within-site) correlation of log normalized exposure.

This reference sampling model incorporates the two-stage selection structure and the potential for within-cluster correlation but ignores any effects of diversity selection. A more extensive description of this reference model is given in Appendix C. As described in Appendix B, the normalizing factor (NF) for exposure can be any measured quantity that is expected to be proportional to the potential amount of contact with active ingredient. This normalizing factor is scenario-specific although in most cases the amount of active ingredient handled (i.e. 'processed') by the MU is used.

9.3.2 Benchmark Objectives

The purpose of the AHETF Monitoring Program is to collect sufficient data for each handler scenario to meet specific benchmark adequacy objectives. The primary benchmark objective for all scenarios will be that a sample from the hypothetical reference sampling distribution above be of adequate size to describe selected measures

of the (normalized) exposure distribution with a pre-determined level of accuracy. The design benchmarks are not intended to address all possible ways the exposure data could be used. Rather, they are established to ensure that the data will at least be adequate to meet common regulatory needs.

The primary benchmark objective for each scenario in the AHETF program will be that the geometric mean, arithmetic mean, and 95th percentile of the reference distribution of normalized dermal exposure be accurate to within K-fold, 95% of the time. The desired relative accuracy, K, can be scenario dependent. Currently, however, there is a consensus that, for regulatory purposes, 3-fold relative accuracy (i.e., K=3) is a reasonable default for all scenarios. A more detailed discussion of this primary objective and the relative accuracy concept on which it is based is provided in Appendix C.

A secondary (i.e., less important) benchmark objective is considered for scenarios where the practical range of the normalizing factor is at least one order of magnitude. In such cases it should also be possible to statistically distinguish between complete proportionality and complete independence of dermal exposure and the normalizing factor (e.g., amount of ai handled). A more detailed statistical description of the secondary benchmark objective is given in Appendix C. This secondary objective merely ensures a level of adequacy to illuminate the relationship between NF and dermal exposure. It is not the objective of the AHETF monitoring program to guarantee that generic data for each scenario discern more complicated relationships between exposure and amount of ai contact. Nor is it the intent to guarantee that future analyses of the data will be able to choose between several potential normalizing factors or combinations of factors.

9.3.3 Sample Sizes that Satisfy the Benchmark Objectives

Methods for determining the number of monitoring sites (i.e., clusters of MUs) and MUs per site to meet the primary and secondary benchmark objectives are described extensively in Appendix C. Currently available monitoring data from multiple studies were analyzed to establish default estimates of relative variability and intracluster correlation for the reference sampling model. When the default K=3-fold accuracy requirement is reasonable, simulations in Appendix C demonstrate that the primary objective can be satisfied cost-effectively using a standard design having 5 clusters and about 5 MUs per cluster. It is also shown that the same benchmark accuracy can be obtained when cluster sizes are unequal as long as the total number of MUs is 25 and no cluster has more than 5 MUs (in this case, there must be more than 5 clusters of MUs).

The investigations described in Appendix C also examined the effect of sample size and NF configuration on the secondary benchmark objective. An analysis using default values for relative variation and correlation indicates that the standard design that satisfies the primary benchmark objectives should also satisfy the secondary objective as well. That is, it should be possible to distinguish a proportional from an independent

relationship between dermal exposure and NF with 5 clusters and 5 MUs/cluster provided that:

1. The range in the normalizing factor is nearly two orders of magnitude, or

2. The range in the normalizing factor is one order of magnitude but there is strong overlap between the ranges of NF levels in different clusters.

Condition (2) above is often partially satisfied because NF is a meta-factor that is diversified within each cluster. Therefore, the ranges of NF levels in one cluster are expected to have considerable overlap with those of different clusters.

It must be emphasized, however, that a sample size of 5 clusters with 5 MUs/cluster is considered the 'default' or 'standard' configuration. It strictly applies only to scenarios without existing data and when the default variability structure and 3-fold benchmark accuracy are considered reasonable. In other cases, the AHETF will use the simulation techniques described in Appendix C to develop optimal selection plans for each scenario it addresses. When some MUs already exist for a particular scenario, they also will be considered in the determination of number of additional clusters and MUs per cluster.

10 Implementation of MU Selection and Construction for each Scenario

The previous section discusses the common conceptual basis for constructing monitoring units that characterize normalized exposures resulting from the use of an arbitrary pesticide. The procedures used to obtain an appropriate set of MUs for each scenario will conform to these general principles but the specific details will vary from scenario to scenario and possibly also from monitoring site to monitoring site (i.e., study to study). However, the implementation of MU selection for any scenario will always consist of the following basic steps:

1. Definition of the handling scenario

2. Definition of a practical measure of active ingredient contact for the scenario that would be appropriate to use as an exposure normalizing factor (NF)

3. Identification of major handling-day (HD) characteristics and meta-factors that have the potential to impact normalized exposure

4. Determination of likely ranges for these characteristics whenever an arbitrary active ingredient is handled in a manner consistent with the scenario definition. This includes determining the practical range of the normalizing factor.

5. Specification of benchmark objectives for the scenario and determination of the number of monitoring sites and target number of MUs/site that satisfy these objectives

6. Selection of a diverse set of monitoring sites appropriate for the scenario

7. Identification, at each monitoring site, of a working pool of growers (and/or application contractors) where appropriate HD conditions are likely to occur naturally (or with minor scripting) and who are willing to participate in the program.

8. Construction of actual MUs from the potential MUs associated with the working pool. This includes recruiting workers and selecting (choosing, restricting, or scripting) handler-day conditions to represent the diversity expected in scenario-specific HDs with an arbitrary active ingredient.

Each of these steps is described in greater detail below. Some of the above steps apply to the scenario as a whole, while others (especially steps 7 and 8) are specific to a particular monitoring site.

The AHETF will document the design of the common scenario-level activities in an MU Selection Plan. Each scenario-specific MU Selection Plan will contain the following components:

- Summary of existing MUs and cluster structure (e.g., from PHED or AHETF)
- Rationale for the proposed number of additional clusters and MUs/cluster
- Identification of the most appropriate normalization factor (e.g. amount of ai handled, AaiH)
- Identification of characteristics (other than the normalization factor) that may influence the distribution of normalized exposure and a summary of available information about the expected ranges of these characteristics
- Identification of monitoring sites (or of target areas from which monitoring sites will be selected)
- Proposed methods for obtaining diversity in the normalizing factor within each monitoring site
- Guidelines for implementing second-stage diversity selection of monitoring units within monitoring sites

Each MU Selection Plan will be developed by a Field Studies Subcommittee that is made up of agriculturalists and exposure assessment professionals from the member companies. These individuals have considerable experience conducting crop residue, environmental fate, and exposure studies and have a good working knowledge of North American agricultural practices. External agricultural experts will also be consulted as needed.

In general, the selection of MUs and subsequent monitoring activities within each monitoring site occurs in a separate GLP field study (see Sec. 11 below). Consequently,

plans for these activities will be documented in GLP field study protocols. If necessary, each study protocol will also describe the process used to select the monitoring site from a larger target area.

10.1 Defining the Scenario

The AHETF will attempt to define, *a priori*, what handling conditions will comprise each scenario. Such a description will include details regarding scenario-specific tasks as well as typical equipment types, product formulation, engineering controls, etc. When relevant, scenario-related restrictions on crop, climate, and geography will also be listed. A scenario definition will certainly include some handling conditions that are more common than others. In general, likelihood-of-occurrence information, if known, is irrelevant to the definition of the scenario, but in some cases a scenario definition might exclude conditions that are less likely to occur and expected to have no negative impact on exposure.

In most situations, a scenario definition will be based on information that is publicly available (e.g., USDA statistics, government reports, or literature references) and by consultation with appropriate experts (see 10.9 below). However, it may also be based, in part, on discussions with EPA and other regulators so that the scenario definition, and hence the data collected according to the definition, will be useful for regulatory purposes.

The set of MUs cannot include every possible handling condition in a scenario. Therefore, the scenario definition must not be envisioned as consisting of only those conditions represented in the MUs. However, a scenario does consist of all those conditions that would be considered valid for selection before the actual MUs are constructed.

10.2 Defining the Normalizing Factor

An appropriate normalizing factor is defined for every scenario. As discussed in Sec. 9.1 above, the normalizing factor (NF) is an experimentally measured MU characteristic that is expected to be (at least) proportional to the potential daily amount of active ingredient that a worker contacts during scenario-related tasks. For most scenarios, the NF used is the amount of active ingredient handled (or AaiH). This simply means the amount of active ingredient that gets mixed into a tank or piece of application equipment and/or applied from a piece of application equipment.

However, for some scenarios (e.g., closed-system mixing and loading) a worker might actually process (i.e., 'handle') a large amount of active ingredient, but may have the opportunity to contact only a small fraction of this amount. In such cases, there may be other measures of potential contact that are more appropriate than AaiH. Such measures usually involve the concentration of active ingredient in the processed product multiplied by some factor related to the expected amount of contact the worker might have with the

product. For example, (concentration)×(number of loads processed) might be a reasonable ai contact measure for closed-system loading systems.

Although very unlikely, it is also possible that no quantifiable measure is available to gauge the potential worker contact with active ingredient. If such were believed to be the case then absolute, rather than normalized, exposure might be appropriate for predicting generic exposure.

10.3 Identifying Characteristics that could Impact Normalized Exposure

One group of characteristics that clearly affect exposure is those involved with defining the scenario, including:

- Task (e.g., mixing/loading or application)
- Application System (e.g., aircraft versus ground equipment)
- Product Formulation (e.g., liquid or granule)
- Engineering Controls (e.g., open or closed loading and open or enclosed cabs)

A second group includes the three primary meta-factors that will be formally diversified within all scenarios (see Sec. 9.2.2 and Appendix B):

- Monitoring site (i.e., geographic location and dates of monitoring)
- Workers
- Level of the normalizing factor

These are called meta-factors because diversity in these parameters is expected to result in diversity of other conditions that directly or indirectly impact exposure potential. Varying location could affect crop and equipment choices. Varying workers could result in different worker practices and/or behaviors. Varying the NF (e.g., AaiH) could affect parameters such as tank size, number of loads applied, etc. All of these component factors could impact normalized exposure potential. Altering a meta-factor can certainly change many factors simultaneously, but it will also reduce the chance of aberrant combinations that can result when the component factors are diversified independently.

Each specific scenario will be evaluated to determine whether other characteristics exist that impact (or might be hypothesized to impact) the potential for exposure. This evaluation will be made primarily by exposure assessment experts from AHETF member companies in consultation with exposure assessors from North American regulatory agencies. In some cases, these characteristics can be used to augment the MU selection process, especially at the first-stage selection of monitoring sites. For example, exposure potential during some application scenarios, like airblast spraying, might be impacted by the particular crop or a characteristic of the crop such as height or foliage density. In that case, potential monitoring locations could be stratified by crop type and one monitoring site could be selected from each stratum. This evaluation and how it impacts the selection of study conditions will be described in scenario-specific MU Selection Plans.

10.4 Likely Ranges of Exposure-Impacting Characteristics in Generic Handler-Days

For characteristics identified as possibly having a major effect on exposure (Sec. 10.3 above), AHETF will predict the likely range of values such characteristics might have in a future handler-day (HD) population. This future (or 'generic') population for a scenario consists of possible HD conditions that might occur when an arbitrary pesticide is used at any level of potential ai contact (e.g., at any AaiH level). Some indication of how these characteristics are expected to vary in the future HD population provides a basis for diversity selection of MUs. (See Sec. 9.2.2 and Appendix B.)

The AHETF will base their predictions on reviews of public information, discussions with agricultural experts in the particular handling scenario, and discussions with local growers or crop consultants. Depending on the scenario involved, agricultural experts might include university contacts, crop and applicator associations, crop consultants, equipment manufacturers, independent research organizations, and field experts employed by AHETF members.

As discussed in 9.2.2.2 and in Appendices B and C, AEHTF will take special care to vary the normalizing factor (NF) over a wide range. This diversifies an important meta-factor and helps to insure that AHED® users will be able to examine the relationship between NF and exposure. To do this, AHETF must first establish the practical range of the NF for each scenario. For the default normalizing factor, amount of ai handled (AaiH), the low end of the range is generally set at 5 pounds of active since handling amounts below this level may result in residue levels that are not quantifiable. Non-quantifiable exposures are not very useful for a generic database since the true exposure is somewhere below the analytical limit and some assumption about the actual level of exposure for the MU is needed in order to make a reasonable estimate of predicted exposure. The upper end of the AaiH range is generally based on high end product use rates and maximum acres treated per day for various crops and application techniques. However, AHETF sometimes sets the practical range at a somewhat lower value than the calculated maximum in order to prevent workers from handling excessive amounts of product that might take significantly longer than 8 hours. Pesticide use rates are generally found by surveying member products and there are a number of publicly available resources for estimating acres treated, including Policy 9.1 (U.S. EPA, 2001) that EPA uses. Similar methods are used when the normalizing factor is something other than AaiH.

Example: Dry flowable, open pour, mixer/loader

When designing studies for this scenario AHETF used AaiH as the normalizing factor and established a practical range of ai handled as 5 to 2,000 pounds of active ingredient. The upper limit was based on a high end use rate of 2 lb ai/acre, the assumption that up to 1,200 acres of crop could be treated by air but reduced somewhat since open pouring small containers would probably not be the technique of choice for the very highest use level (i.e., bulk packaging of liquids would be more common).

Although expected to impact (unnormalized) exposure, 'duration of work period' is a special characteristic that will intentionally not be diversified among MUs. AHETF designs MUs to represent a normal workday for the particular task being monitored. Generally, this will involve monitoring periods of between four and eight hours, since most activities can be performed all day long. For tasks that typically involve shorter time periods, such as cleaning out seed treatment equipment that takes just 1 to 2 hours, workers will generally be monitored for the entire time it takes to perform the task. In some situations, such as handling large amounts of product, workers may need to work and be monitored for more than 8 hours. However, long work-days are not uncommon in commercial agriculture.

Minimum work periods will be specified for each task in each field study protocol. This will be based on the tasks to be monitored in each study and the amounts of active ingredient to be handled. The minimum will usually be 4 hours. This is designed to overcome the criticism of early exposure studies where many of the sampling regimes monitored workers for only a few minutes. Avoiding very short monitoring intervals will ensure that daily exposure estimates are not biased by unusual conditions during that short interval. For tasks where a typical work period is markedly different from about 8 hours, a specific study protocol may indicate a more appropriate minimum monitoring time (i.e., other than 4 hours).

10.5 Sample Sizes that Satisfy Benchmark Objectives

Benchmark objectives (Sec. 9.3.2) will be defined for each scenario. For scenarios with no existing MU data, the approach used to determine the number of monitoring sites and the target number of MUs per site is described in Appendix C. In most cases, the standard design of 5 clusters with 5 MUs/cluster should be sufficient to meet the primary benchmark objective of 3-fold accuracy. When the practical range of the normalizing factor is at least an order of magnitude, the secondary objective applies. In this case the standard design should also be acceptable.

When MUs for a scenario are available from previous studies, the sample size methods described in Appendix C require modification. In most cases, it is reasonable to assume that MUs in the same study are analogous to an AHETF cluster since they were conducted in a geographic location and time frame that is similar in extent to planned AHETF monitoring sites. Unfortunately, a number of older studies did not construct MUs with diverse handler-day conditions. For example, many studies were conducted using a narrow range of the normalizing factor. More importantly, however, many studies used the same worker repeatedly and, in several cases, these repeated-worker MUs were conducted on the same day and in close proximity. Such extensive sub-clustering increases the potential for within-cluster correlation and, in effect, reduces the effective number of MUs in each cluster. The existing data for a single study are usually

too sparse to estimate within-worker, within-day, and other components of variation. Thus, the actual magnitude of the within-cluster correlation is, in general, unattainable.

However, using similarities among existing MUs with respect to worker, day, and locale, it is possible to derive a reasonable value for the increase in the average within-cluster correlation. Given this assumed average correlation increase, an effective number of AHETF-like MUs in a previous study, N_e, can be calculated. Assuming that the existing study provides a single cluster of N_e MUs, the methods described in Appendix C can then be employed to determine the additional sample size needed to satisfy the primary benchmark objective.

In many cases, existing studies are limited to only a narrow range of the normalizing factor. Such studies often made no attempt to diversify NF within clusters as is done in the AHETF Monitoring Program. As a result, when the secondary benchmark objective is used, these pseudo-clusters will not generally be as effective in distinguishing a proportional from an independent relationship between exposure and NF. When the NF configurations in the existing pseudo-clusters are included, the simulation methods described in Appendix C can also be used to determine the additional number of clusters, the number of MUs per cluster, and the within-cluster NF configurations needed to satisfy the secondary benchmark.

10.6 Identify Monitoring Sites

As discussed in Sec. 9.2.1 and Appendix B, the monitoring site is the first-stage unit of MU selection. Each monitoring site is associated with a particular broadly defined location and time period(s) when monitoring is possible. Because site is a meta-factor associated with numerous factors influencing exposure, diversifying monitoring sites by geographic location can make a big difference to total diversity in the set of MUs.

Of course, monitoring sites will be restricted only to locations (and often by other broad conditions) where the particular handling situation of interest will commonly occur. Some application scenarios may be focused only on specific crops. In such cases monitoring sites would only be selected from within areas of the country where these crops are predominantly grown. As discussed in Appendix B, selection procedures could have both random and purposive components. The MU Selection Plan for the scenario will describe how the selections will be made.

Example: Airblast applications.

Airblast applications are commonly made to orchards and to trellis crops, but not to field crops. Therefore, monitoring sites will only be located in states where one of these two crop categories is common. This also suggests that a diverse set of monitoring sites should include both categories of crops. The USDA National Agricultural Statistics Services (NASS) database can be used to determine that the largest trellis crop acreage is in California (primarily grapes). This is a strong indication that

at least one monitoring site for airblast application MUs on trellis crops should be selected from within California.

For some scenarios such as mixing/loading of water soluble packets, it is less important to restrict monitoring sites to particular geographic regions since this task is performed in virtually all areas of the country. When such flexibility in location is possible, AHETF can sometimes increase total efficiency by synchronizing the collection of mixing/loading scenario MUs with applicator scenario MUs.

Another goal is to choose monitoring sites and areas where a sufficient number of growers and workers can be recruited to conduct a study in a reasonable amount of time. This also involves limiting the monitoring site to a reasonably-sized geographic area so that MU identification and selection operations can be conducted efficiently. This is necessary primarily to keep the costs of study conduct reasonable so the AHETF monitoring program can address the desired number of scenarios and obtain an adequate number of MUs per scenario.

A final practical consideration for selecting monitoring sites is an AHETF preference for areas where local researchers are available who can help with some of the GLP requirements for field studies. Not coincidentally, these researchers, which AHETF calls Local Site Coordinators (LSCs), are typically located in important growing regions in North America and are familiar with common growing practices for local crops. LSCs can provide storage facilities that are monitored for temperature and have access restrictions (ambient for test substances, refrigerated for field fortification solutions, and frozen for exposure dosimeter samples). They also tend to have many contacts within the local grower community and, therefore, can assist with the listing of suitable growers within the monitoring site.

10.7 Identifying a Potential Worker Pool within each Monitoring Site

Once the monitoring sites have been selected (by a process documented in the scenario-specific MU Selection Plan), all subsequent second-stage MU selection activities at each monitoring site are conducted as part of a separate GLP field study (see Sec. 11). The conduct of each field study is governed by a separate study protocol.

The first activity at this stage is to identify those workers within the monitoring site that could potentially participate in monitoring events. Ideally, this pool of potential workers would consist of all eligible individuals within the monitoring site who commonly perform the types of pesticide handling tasks for the particular scenario being planned. For practical reasons, however, the potential worker pool must be defined in terms of a set of eligible employers of these potential workers. Employers typically will be either growers or commercial application firms. For this purpose, self-employed ‘workers’ (e.g., farm owners or operators) are also considered to be employers. Once identified,

this pool will provide the potential set of workers and handler-day conditions from which the final set of MUs is constructed (see Sec. 10.8).

The steps involved in obtaining the potential worker pool at each monitoring site are:

1. Using publicly available information and a variety of local sources, construct a list of employers in the monitoring area that are likely to be involved in scenario-specific activities.

2. Reduce the list of employers to those who are eligible and (initially) willing to participate in the monitoring program.

It is important to reduce, as much as is practical, the possibility of intentional biases in the selection of employers. Such biases are likely, for example, when only growers known to the Local Site Coordinator or other study personnel are contacted. Consequently, the preferred approach is to obtain as complete a list as possible of relevant employers within the monitoring site. For some highly specialized scenarios, such as open cockpit rotary-wing aerial application, there may be only a few employers available within the monitoring site. In such cases this listing step is quite simple.

Employers will be screened for characteristics required for program participation. These include:

- Initial willingness to participate in the program;
- Availability of, or access to, suitable acreage of a specific crop or crop type (or other pesticide application site such as a greenhouse);
- Availability of suitable equipment and engineering controls that are consistent with the scenario definition;
- Availability of workers who are experienced with those particular equipment and practices; and
- An anticipated need for the application of a pesticide product that meets AHETF needs as a surrogate chemical, but not tank mix products that would require additional PPE.

As indicated above, it can sometimes be more effective to define the potential worker pool in terms of commercial applicators rather the growers. For some scenarios, such as aerial applicators, it may be just as important to identify suitable commercial applicators (e.g., pilots) since growers rarely apply pesticides by air themselves. In these cases, the grower provides suitable acreage and the need for the pesticide, while the application and/or labor contractor provides the pesticide handlers themselves and equipment they typically use. Even when the potential worker pool is identified only by growers, the actual application or mixing/loading of the pesticide could be contracted. When growers and commercial applicators are so linked, both must agree to participate in the program at some point.

When the list of employers is small (e.g. aerial applicator firms), then AHETF can easily determine if they meet the screening criteria. In such situations it is entirely possible that the potential worker pool would consist of all employers within the local area of the monitoring site. It is more typical, however, for there to be a large number of growers that perform scenario-related pesticide handling tasks (e.g., closed cab airblast applications to a particular crop type). In this event, it may be more practical to use only a randomly or purposively selected subset of the employers for the potential worker pool. For example, the list of growers could be first randomized and the growers screened (in random order) until an adequate pool of growers is obtained. This approach insures that every eligible grower in the initial list has an equal chance of being in the pool.

When desired, limitations on the degree of similarity allowed between employers successively added to the random subset could also be used to increase diversity. For example, further addition of larger growers to the pool could be discouraged once a certain number have been selected. Each scenario MU Selection Plan will discuss the relative merits of methods to identify and select growers or commercial contractors and each study protocol will describe specific procedures that will be followed. In particular, some guidance will be provided for how large the working pool of growers must be, generally involving more growers and potential workers than will be needed for each stratum of the NF.

The choice of surrogate chemical is limited by the set of active ingredients that have been deemed suitable by AHETF (see Sec. 5.2.1) and involves discussions between the grower and researchers to find a chemical that meets both of their needs. If this is not possible, the grower or contractor will be excluded from the working pool. The choice of the particular product to be used can be influenced by the scenario being addressed (e.g., mixing/loading scenarios necessitate a particular formulation type) or the need for diversity of conditions within a scenario (e.g., container sizes). However, growers/applicators may choose between equivalent or similar products (e.g., manufacturer or strength), if available.

10.8 Constructing Monitoring Units

After the pool of growers and/or contractors has been obtained as described above, researchers (e.g., Local Site Coordinator, Study Director, and/or principal field investigator) will examine its potential to supply workers and handler-day conditions for MUs. They will then identify a configuration of MUs (e.g., growers, chemicals, workers, AaiH, timing) that will result in an efficient monitoring study. An efficient configuration will generally consist of a second-level pool of growers/contractors that:

- Are in the same geographical area,
- Can provide separate workers for all the strata of AaiH,
- Involve some diversity in equipment, and

- Are expected to make applications within a narrow time frame.

Choosing a cost-effective configuration of MUs is necessary since costs escalate rapidly when a research team makes several visits to a location in order to monitor the desired five MUs. Cost-effectiveness is obviously maximized when all MUs are collected during the same visit so researcher salary, travel, food, lodging, and field fortification expenses are minimized. The table below provides relative total costs for the items listed above associated with collecting 5 MUs in the field for various degrees of efficiency. These efficiency ratios highlight the importance of selecting growers in a defined area that plan to handle product at approximately the same time.

Most Efficient ……..…… Least Efficient				
1 Trip	2 Trips	3 Trips	4 Trips	5 Trips
1.0	1.3	1.7	2.1	2.5

The chosen configuration should include excess workers because participation is always difficult to guarantee. For example:

- Growers or commercial application firms might change their mind about cooperating,
- Workers might not volunteer to participate,
- Expected applications might not be made due to lack of pest pressure,
- Various growers have different application timing, etc.

The growers and/or commercial applicators in the chosen configuration will provide a second-level pool of workers from which study participants will be recruited. When constructing MUs, additional restrictions will generally be enforced to reduce sub-clustering within the cluster of MUs. These restrictions will be discussed in the MU Selection Plan and specified in study-specific protocols, but might include things such as:

- A worker may be used for only one MU in the cluster.
- A major piece of equipment (e.g., a particular airblast sprayer) may be utilized only once.
- No more than two MUs may be obtained from any grower or contactor.

From the second-level pool based on a cost-effective configuration, workers will be recruited as described generally in Section 14.1 with specific details provided in study protocols. In general, this might begin with sending a recruitment flyer to growers in the

eligible pool, followed by site visits where the Local Site Coordinator, Field Investigator, and/or Study Director meets the growers and confirms the suitability of their crops, equipment, and willingness to cooperate (including discussions about non-coercion of workers). Then the workers associated with the chosen growers and/or commercial applicators may be contacted directly (e.g., by Study Director) to begin recruitment. At this point, recruitment might involve purposive or random methods for selecting workers. Each scenario MU Selection Plan will discuss the relative merits of methods to select workers and each study protocol will describe specific procedures that will be followed. It is anticipated that some form of random selection will often be appropriate for selecting workers (e.g., selection by lottery when excess workers are available from a selected grower).

10.9 Sources of Expert Information

As indicated above, AHETF will seek expert advice on a variety of topics to support scenario plans. Two basic types of information are needed:

- Information on agricultural practices that will support the definition of each scenario and identify likely conditions associated with the scenario tasks.
- Information on characteristics that might be important to exposure potential and will support the diversity selection procedures.

The exact sources of expert information will generally be scenario-specific, but may also be study-specific. The information is used to guide scenario plans (e.g., diversity of conditions) and/or study-specific procedures (e.g., selecting growers or study conditions). Following are some examples of the types of experts that might be consulted during the planning process:

- For most scenarios, general agricultural information that will characterize the handling scenario (e.g., common sites, equipment, workers, and practices) will be obtained from AHETF member company experts, USDA County agricultural extension agents, professional crop consultants, and academic and governmental research organizations

- For many scenarios or studies, planning will be based on a particular crop or a crop group. For example, airblast applications are unique to orchards and trellis crops. In these cases, grower associations, governmental agencies, and applicator associations might be contacted.

- For other scenarios, the equipment or worker practices rather than crops may determine what expert advice is needed. Examples are: chemigation application, hand-held sprayer application, and enclosed mixing/loading systems. In these cases, farm equipment manufacturers, pesticide dealers, and commercial applicators might be contacted.

- For worker issues (e.g., handling procedures, diversity, recruitment, etc.), pesticide safety training organizations and farm workers themselves may be contacted.

- When identifying factors that might impact exposure, exposure assessment professionals from AHETF member companies and from North American regulatory agencies (e.g., EPA, PMRA, and CDPR) will be consulted. Together, these experts will determine which factors must be diversified within a scenario and will discuss appropriate methods to achieve appropriate diversity.

The information collected will be used to guide diversity selection procedures and will be summarized in the MU Selection Plan prepared by AHETF for each scenario. This summary will include identification of the experts (when permitted by the expert), their expertise, years of experience, and what specific information was provided.

11 Description and Role of Field Studies under the AHETF Monitoring Program

A field study is the component of the AHETF exposure monitoring program that actually involves selecting MUs and collecting exposure measurements. Within the diversity selection design, each field study typically represents a cluster of MUs from one location for one scenario.

Each field study conducted as part of the AHETF monitoring program will be conducted in compliance with Good Laboratory Practice (GLP) standards issued by EPA (40 CFR 160). As required by GLPs, Standard Operating Procedures (SOPs) have been developed that address many routine elements of AHETF field study operations and these SOPs will be submitted to EPA along with new study protocols and other supporting information. AHETF field studies meet the definition of "study" in the GLPs at 40 CFR §160.3, which reads: "Study means any experiment at one or more test sites, in which a test substance is studied in a test system under laboratory conditions or in the environment to determine or help predict its effects . . . or other characteristics in humans . . . or media."

Each field study will develop data for MUs in one scenario. However, when feasible, MUs for multiple studies will be collected at the same time and locations. For example, a study designed to monitor helicopter pilot exposure might be conducted at the same time and location as a mixer/loader study since a worker will necessarily be preparing the spray mixture for the aerial applicator. Each study protocol will address how many and what type of MUs will be selected, what AHETF SOPs will be followed, and study-specific issues such as recruitment, consent, benefits, and risks.

11.1 Relation of Field Studies to Scenarios

AHETF study timing and location is usually dictated by the seasonality of the tasks to be performed, pest pressures, and crop growth stages. This often allows a relatively small

window of time each year when a study can be conducted. Finding sites and arranging for studies can be challenging, especially when special efforts are made to monitor workers under actual conditions with minimal scripting. For example, a pilot may spray from hundreds, to sometimes over a thousand, acres during a typical workday. The AHETF must identify sufficient crop acreage to allow a full day of application for each worker (i.e., each exposure measurement or MU) while trying to capture the range of potential acres treated (the typical way to vary AaiH for a particular surrogate product). Since data requirements for most use scenarios cannot practically be accomplished in a single study, most individual AHETF study protocols are part of a multi-study and multi-year plan designed to generate a wide range of data for activities associated with that use scenario. An individual field study typically will not represent a complete stand-alone scenario; but will provide a limited number of MUs, monitored at a specific site (typically a cluster within one scenario-specific plan), often with a single pesticide active ingredient. Data from multiple field studies conducted at multiple sites/clusters will typically be combined to complete the scenario data set for most use scenarios.

Each worker exposure field study is performed in accordance with EPA guidelines for conducting worker exposure studies (Durham, 1962; Wolf, 1967; WHO, 1975 and 1982, OECD, 1981; NACA, 1985; Chester, 1993; Worgan, 1995) as described in Series 875: Occupational and Residential Exposure Test Guidelines (U.S. EPA, 1986 and 1996). These guidelines are consistent with guidelines used in other countries such as Canada, Australia, and members of the European Union.

Many aspects of individual field study protocols, especially sample collection and analysis methods, will be standardized to ensure consistency and uniformity in the resulting data. Exposure monitoring protocols differ mainly in the specific product used, the amount of active ingredient handled, equipment used, timing of the study, and location and activity performed (including the level of scripting). In addition, there may be differences in other aspects of protocols, e.g., recruitment methods and study specific heat stress and medical management plans.

11.1.1 Descriptions of Surrogate, Locations, and Number of Measurements

Since a major component of quality assurance is to perform field fortifications to measure stability of the surrogate chemical (see below), AHETF field study protocols will generally involve a single pesticide active ingredient. The protocol will generally name specific products that are acceptable for use (recognizing that formulation type, container design, and/or product concentration may be important study design factors) and identify the active ingredient. On some occasions, a protocol may identify two or more potential active ingredients which could be used. For example, AHETF has a set of validated analytical methods which can quantify two organophosphate insecticides, malathion and diazinon, which have many approved uses in common. Therefore, some protocols may specify the use of malathion or diazinon. The raw data collected during the study will always identify exactly what product was used and results of the chemical analysis of active ingredient content.

In some cases, the grower may need to add other chemicals to a tank mixture, including other pesticide products, fertilizers, and adjuvants. These decisions are outside of AHETF control and are often not known in advance, so all protocols and consent forms will address this possibility in general terms. AHETF will ensure that only products approved for use by EPA and that don't require added PPE will be utilized. Researchers will also discuss the label precautions for any such tank mix product with the subject before the subject handles that product.

As described in Appendix B, varying the location of monitoring is an important study design parameter, The location of field studies within a scenario is generally varied to obtain variability in exposure potential that may come from differences in workers, crops, equipment, or environmental conditions as outlined in the scenario selection plan. For example, studies involving airblast application to orchards might be conducted both in the west and in the east, so that differences in exposure caused by weather, orchard type, orchard canopy management practices, spray equipment size, or other factors would be reflected in the exposure data for those scenarios.

Each field study protocol lists a target number of MUs that will be generated. The actual number of MUs completed may sometimes be less than the target number due to factors such as availability of workers and equipment, crop acreage available for treatment, worker decisions not to participate or to withdraw, or adverse weather. In these cases, sufficient MUs to fulfill the scenario-specific MU selection design must be obtained during a future field study.

11.1.2 Documentation procedures

Exposure monitoring studies conducted by the AHETF are designed to measure exposure to workers as they perform their normal work functions in a particular handling scenario. All aspects of study conduct will be fully documented in compliance with 40 CFR §160 GLP requirements. Most of the information collected in the field during each study is entered by hand by researchers on standard data forms provided by AHETF. Much of this information is also entered into the generic database, AHED®, for use in data analysis and for examination by database users in conjunction with data from other AHETF studies.

Raw data are collected in a study notebook which will be retained indefinitely in AHETF archives (see Chapter 6 of SOPs for more information about archiving procedures). In addition, a certified copy of the data set is made during report writing and report review in case the original is lost during transit to archives.

These procedures are detailed in Chapter 9 of the AHETF SOP Manual.

11.1.3 Quality Assurance Procedures

A very important requirement of GLPs is rigorous quality assurance to ensure the quality and integrity of the data that will be relied upon for pesticide handler exposure assessments. All aspects of the studies are monitored while they are in progress by appropriate quality assurance units (QAUs) to ensure compliance with GLP regulations and adherence to the protocol and relevant AHETF Standard Operating Procedures. This will typically involve three different QAUs: one from the exposure monitoring contractor that conducts the study in the field, one from the analytical laboratory that determines the level of pesticide residues in field samples, and one contracted directly by the sponsor. For each field study, the following specific activities are conducted by these QAUs:

- Sponsor-contracted QAU inspects all contract research organizations and laboratories prior to use in a study to ensure that those researchers operate in compliance with GLPs
- Sponsor-contracted QAU reviews protocols prior to finalization
- Sponsor-contracted QAU and/or field contractor QAU observes study conduct in the field
- Field contractor QAU audits the raw data file from the field and audits the Field Report
- Analytical laboratory QAU audits the raw analytical data and audits the Analytical Report
- Sponsor-contracted QAU reviews the Final Report which includes the Field Report and Analytical Report as appendices

Each QAU submits an inspection report(s) to the Study Director and AHETF Sponsor Representative and any exceptions to full GLP compliance are listed in the Final Report associated with each protocol.

Chapter 5 of the AHETF SOP Manual provides details about the responsibilities of and procedures to be followed by the sponsor-contracted QAU. Field contractor and laboratory QAU's follow similar SOPs from their own facility.

11.1.4 Quality Control Procedures

In addition to the formal quality assurance efforts discussed above, there are a number of important analytical and field sampling quality control procedures which are followed in order to assure that exposure measurements are accurate and precise and to define what those exposure measurements represent. These include complete validation of all analytical methods; extensive documentation of exactly what the participant does while handling the pesticide product; field fortification and control samples designed to estimate stability of chemical residues during sampling, transit, and storage; laboratory fortification and control samples designed to establish efficiency of the analytical

methods on a day-to-day basis; and detailed guidelines on the use of calibration curves to determine chemical residues found on all sample matrices.

In the field during each study, a researcher prepares exposure matrix positive control samples that are fortified with a known amount of active ingredient (SOP AHETF-8.E). These matrices include whole body dosimeters (WBD, cotton long underwear), hand wash surfactant solution, face/neck wipes moistened with surfactant solution, and inhalation tubes (referred to as OVS tubes which stands for OSHA Versatile Samplers). Some studies may also involve sock dosimeters or head patches. OVS tubes are fortified in the laboratory by injecting diluted analytical grade active ingredient onto the sorbent in the tube while all other matrices are typically fortified in the field with a solution or suspension of diluted test substance or diluted active ingredient (usually from individual vials prepared in the laboratory). Each matrix type is generally fortified at three levels of active ingredient designed to span the range of residues anticipated to be collected from workers. At each level, triplicate samples are fortified. In addition, negative control samples (i.e., blanks) are prepared for each matrix to determine whether background levels of active ingredient may be present in the study environment. Field control and fortification samples are always collected on at least one day during each study and whenever significantly different weather conditions are expected. It should also be noted that similar samples were usually generated as part of the method validation process under conditions that were anticipated in the field to establish, in a preliminary sense, the stability of residues on field sampling media during and after an exposure period.

Fortified WBD and OVS tubes are "weathered" in the field since these sample types involve collection of residues during the monitoring period. For WBD, this involves laying a fortified section of long underwear onto a table in a sunny location (unless it is a cloudy day) and covering that sample with a single layer of outer shirt material. For OVS tubes, this involves drawing air through the tube in the same manner as done for workers. Fortified hand wash and face/neck wipe samples are not weathered since these samples are collected at specific time points during the monitoring period and immediately placed into frozen storage.

Analysis of field fortification samples provides a "recovery" value which will quantify stability of the active ingredient during sample collection (for weathered samples), storage in the field, shipment to the laboratory, and storage in the laboratory freezer. Therefore, field fortification samples serve as a type of positive control. Field fortification samples are analyzed along with worker exposure samples and it is assumed that the worker samples experience similar stability as the field fortification samples. Therefore, residues found in worker samples are adjusted by appropriate average field fortification results to estimate the residues actually collected in the field. These practices are standard in pesticide exposure monitoring studies and are discussed in detail in internationally accepted testing guidelines.

Similar quality control procedures are followed in the laboratory, including control and fortification samples which are designed to detect background residues, monitor the

performance of the method, and detect matrix or reagent interferences which may be present. These samples serve as types of positive and negative controls. In addition to the detailed analytical methods for each surrogate and each matrix, all analyses must follow detailed AHETF analytical guidelines which specify procedures related to standard curves (e.g., chromatographic response to solutions of known chemical content), extract handling, documentation, etc.

11.1.5 Reporting Process

A detailed report is generated for each field study, a "Final Report" in GLP terminology. AHETF calls these "Summary Reports" which include a text and tabular summary; and detailed appendices including a Field Report and an Analytical Report. Summary reports are formally submitted to EPA, California DPR, and PMRA as they are completed. Summary reports submitted to EPA will also be reviewed by the Human Studies Review Board under 40 CFR 26.1602. In general, these documents will report exactly what was done in the field, the results of analyses of residues, and what information will be entered into AHED®. However, since individual field studies do not provide all the data necessary for a complete scenario, these Summary Reports of individual field studies will not include an analysis or interpretation of the exposure data that were generated. Scenario summarization activities are described in Section 12 of this document.

Field reports will document the conduct of exposure monitoring, including:

- Identification of the location of the study, and the environmental conditions during the exposure monitoring period(s)
- Descriptions of the participants in the study
- Description of the test substance and packaging
- A record of the mixing, loading, and/or application, including a description of the workers, equipment, and worker activities
- A summary of worker observations identifying any specific occurrences that may contribute to unusual worker exposure
- Descriptions of the work clothing and personal protective equipment worn by each worker
- A detailed summary of the amount of test substance handled or applied for each worker
- A detailed summary of the length of time each worker was monitored
- A complete description of the field recovery evaluation with a summary of specific handling and weathering of all field samples
- A complete description of collection, handling, storage, and shipping of field samples
- A complete description of the ethical conduct of the field study, including all elements specified in 40 CFR 26.1303

Analytical reports of individual field studies will document the handling and analysis of residues in all samples collected in the field, including:

- Results of analysis (e.g., µg/sample)
- A detailed description of the analytical instrumentation and methods
- A detailed description of sample storage conditions and storage intervals
- Example calculations
- A summary of field and laboratory fortification recovery data
- Representative chromatograms of control, treated, fortified samples and calibration standards
- A typical standard curve

Summary reports summarize the field and analytical aspects and include calculations of adjusted residues found in all collected samples (i.e., adjusted for field fortification recovery); total dermal exposure for each worker; and the air concentration of active ingredient associated with each worker's monitoring period. Study reporting requirements are described in SOPs AHETF-4.A and AHETF-9.I.

12 Scenario Summary (Monographs)

As part of the documentation supporting AHED®, AHETF will generate scenario monographs for the benefit of regulators and other potential database users. Each monograph will include a description of the scenario as well as an assessment of the data adequacy within that scenario. Based on this analysis, the monograph may include additional recommendations concerning the use of the MU-derived data. Scenario monographs will be formally submitted to the regulatory agencies when AHETF determines the data collection for a particular scenario is complete and suitable for use in exposure assessments.

AHETF will not perform any statistical analyses of the scenario data for the purposes of exposure characterization or risk assessment. Such analyses are the responsibility of regulators and other potential users of the generic database. However, as part of the generic database development and documentation activities, AHETF will evaluate how well the collected data for each scenario satisfy the benchmark adequacy objectives. In addition, the AHETF will quantify the impact of ignoring clusters and treating the data as a simple random sample of MUs.

12.1 Assessment of Benchmark Adequacy Objectives

Section 9.3.2 describes AHETF's primary and secondary benchmark objectives for data adequacy. The primary benchmark requires that the relative accuracy of selected parameter estimates of the normalized (or possibly non-normalized) exposure distribution be within specified limits. A secondary benchmark requires that the data be adequate to distinguish a proportional from an independent relationship between exposure and the

normalizing factor, usually amount of ai handled (AaiH). Appendix C discusses both benchmark objectives in detail.

The benchmark objectives are necessarily based on pre-monitoring assumptions about the relative variation, within-cluster correlation, and the ability to obtain the target number of clusters and MUs/cluster. Achievement of the secondary objective also depends on the ability to obtain the targeted within-cluster and between-cluster diversity in the normalizing factor. It would be unlikely for all of these assumptions to be exactly satisfied for every scenario. Although slight deviations will have little or no impact, large deviations from the assumptions might result in data that deviate too far from the benchmark objectives. Consequently, it is also of value to assess the benchmark requirements in light of the data actually obtained.

In Section 9.3.2 and Appendix C, both benchmark objectives are defined in terms of the reference two-stage random sampling model (9.3.1) and a calculated probability that certain characteristics should be observed in the data if it were generated by this reference sampling model (see Appendix C for details). For the primary objective, the characteristics are relative accuracy of several normalized dermal exposure statistics. For the secondary objective the characteristic of interest is the rejection of a hypothesis. Once the data are available, however, such probability statements are less relevant than confidence statements calculated from the actual data. Consequently, evaluation of the benchmark objectives will be based solely on confidence intervals.

12.1.1 Relative Accuracy of the Normalized Exposure Distribution

As defined in Appendix C, the primary benchmark objective specifies that selected MU-based parameter estimates for the normalized dermal exposure reference distribution be accurate to within K-fold, at least 95% of the time. The key benchmark estimates are the geometric mean, arithmetic mean, and the 95th percentile. Of these estimates, the 95th percentile and the arithmetic mean will always have the worst relative accuracies.

To assess this benchmark objective, a 95 percent bound on relative accuracy will be calculated from the confidence interval for each of the three parameters given above. For a particular parameter, θ, let T denote its estimate calculated from the fit of a cluster sampling reference model to the normalized exposure data. Further, let θ_a and θ_b denote the upper and lower limits, respectively, of a 95% confidence interval for θ. In most cases, the confidence interval, (θ_a, θ_b), will be a parametric bootstrap percentile interval obtained by resampling from a lognormal cluster sampling model. (For the geometric mean only, a Satterthwaite t-based confidence interval can be calculated directly from the mixed model.) The 95% upper confidence bound on realized fold relative accuracy (*fRA*) is then calculated as:

$$UCL_{95}(fRA) = \text{Max}\ (\ T / \theta_a,\ \ \theta_b / T\ )$$

The values of $UCL_{95}(fRA)$ will then be compared with the pre-specified relative accuracy benchmark objective, K. In Appendix C this primary benchmark was expressed in terms of the 95th percentile of fold relative accuracy, fRA_{95}. However, fRA_{95} is only relevant when discussing future results. In contrast, $UCL_{95}(fRA)$ relates to the realized relative accuracy and is the post-data analogue of fRA_{95}.

This primary benchmark objective strictly applies to only dermal exposure. However, for consistency, the 95 percent confidence bounds on the three reference distribution parameters will also be computed for inhalation exposure.

12.1.2 Adequacy of the Data for Distinguishing a Proportional from an Independent Relationship between Exposure and the Normalizing Factor

This secondary benchmark objective (Section 9.3.2 and Appendix C) applies only to handler scenarios for which the practical range in the normalizing factor (NF) exceeds an order of magnitude. In such cases it is reasonable to consider the linear regression of log dermal exposure on log NF. Such a regression would use a two-stage cluster sampling reference model in order to incorporate random cluster effects. In the regression model the true slope, β, would be equal to one if dermal exposure were directly proportional to the normalizing factor. If exposure were independent of NF, then β=0.

For applicable scenarios, this benchmark objective requires that the number of clusters and the allocation of NF levels to MUs be adequate to ensure that such a regression analysis has at least 80% power to reject the hypothesis that β=0 when β is actually equal to one. By symmetry, the mixed model linear regression would also have the same power to reject the hypothesis that β=1 when β=0.

As described in Appendix C, the number of clusters, number of MUs/cluster, and the NF configurations that satisfy this power objective will be determined and used to establish sampling targets for each scenario. This pre-data power analysis also requires an assumed true variance structure (i.e., the residual geometric standard deviation and a within-cluster correlation). After the MUs have been sampled, the actual MU and NF configurations are available. The true residual variance structure is still unknown, although estimates of this structure would now be obtained from the exposure data using a mixed model regression analysis. In principle, this 'updated' information could be used to re-calculate the power to distinguish proportionality (β=1) from independence (β=0). However, such post-hoc power analyses can be somewhat non-intuitive, especially if the data suggest that both hypothesized values of β are false. When the data are available, confidence intervals obtained for the parameters of interest (e.g., β) are considered more relevant than an updated power analysis (Hoenig and Heisey, 2001).

Consequently, for each scenario, a mixed model regression of log dermal exposure on log NF will be performed and a confidence interval obtained for β. As shown in Appendix C, the secondary benchmark power requirement is equivalent to stating that the mean width of a 95% confidence interval for β is approximately 1.4. (The expected width is 1.6 in

the case of one-sided hypothesis tests.) Therefore, the width of the 95% confidence interval for slope will be used to gauge the adequacy of the data with respect to the second benchmark. For example, if the width of the confidence interval obtained from regression on the actual data is 1.4 or less, then the data would be judged adequate with respect to the secondary benchmark.

Note that the adequacy of the data depends only on the width of the confidence interval, not on the particular values of the endpoints of the interval or on the observed slope, b. Although the full results of the regression analysis will be summarized in the scenario monograph for completeness, they are not relevant to the question of benchmark adequacy. It is also possible for both $\beta=0$ and $\beta=1$ to be within the 95% confidence interval for the true slope. Such a result is still consistent with benchmark adequacy since a confidence interval whose width is 1.4 could potentially contain both 0 and 1. However, when $\beta=0$ then a value of 1 is expected to be outside such an interval. Similarly, when $\beta=1$ a 95% confidence interval of 1.4 is unlikely to include zero.

As was the case for the primary objective, the secondary objective only applies to dermal exposure. However, for uniformity, the same regression analysis and assessment of the confidence interval will be conducted for inhalation exposure.

12.1.3 Interim Analysis of Benchmark Adequacy

No formal analyses of data adequacy will be conducted until the planned number of clusters and monitoring units have been obtained. In general, the sample sizes for a partial scenario would be too small and variation estimates too imprecise to be of any benefit in improving the design 'mid-stream'. Equally important, any such repeated or sequential uses of the data would be inconsistent with the fixed-sample-size methods used to determine the number of clusters and MUs. If the adequacy analyses described above indicate that the target benchmark objectives for a particular scenario are not met, then the AHETF, in consultation with the JRC, could decide to conduct new studies in order to obtain additional clusters.

12.2 The Impact of Ignoring Clusters

As described in 9.2.1 and Appendix B, the AHETF monitoring design for all scenarios involves selecting MUs in clusters. A cluster is the set of MUs obtained from the same monitoring site at approximately the same time. Clusters are not a characteristic of the future handler-day population, but merely necessary artifacts of the two-stage MU selection process. It is expected, and existing data confirm, that exposures for MUs in the same cluster are correlated to some degree. In principle, estimates of parameters and regression analyses should be based upon a reference random sampling model that accommodates this correlation. If the two-stage sampling reference model is a reasonable approximation of the actual MU selection process and the within-cluster correlations are non-zero then this reference model is also appropriate for analyses of the data. In this case, ignoring clusters and treating the set of MUs as a simple random

sample may bias the parameter estimates and give confidence intervals that are too small. On the other hand, when the within-cluster diversity among MUs is large the within-cluster correlations are reduced and the impact of clusters may be small and of little practical importance. When this is the case, analyses of the data can be simplified considerably. Consequently, to aid regulators and other potential AHED® users, the impact of ignoring clusters will be examined and reported in each scenario monograph.

Estimates and confidence intervals for the three parameters of the normalized exposure distribution listed above will be calculated assuming no cluster effect (i.e. assuming simple random sampling). These estimates will be compared to those obtained under the cluster-sampling model. The differences obtained by ignoring clusters will be summarized. Future AHED® users can then decide, based on their specific needs, if such differences are of practical importance.

Similarly, a simple linear regression analysis (i.e., without a cluster effect) of log exposure on log normalizing factor will be performed. The estimated slope and confidence interval will be compared with those obtained with the mixed model regression described above. Again, the impact of ignoring a possible cluster effect will be summarized for the benefit of future users of these data.

12.3 Other Information to be Included in the Scenario Monographs

In addition to an evaluation of data structure and adequacy described above, the monograph reports will also summarize various aspects of the study designs used for each scenario. These include:

- A graphical summary of how all MUs in a scenario are structured into clusters and how each cluster is further structured into workers, days, and locations
- Representative use information for AHETF member products to define crops, rates, sites, etc. (used for scenario selection plans)
- Information about the diversity of equipment and procedures currently in use in North America for this scenario
- Characterization of the important features of any existing data acquired by AHETF for this scenario
- Summary of the design targets developed and used for this scenario

13 Exposure Monitoring Techniques

The AHETF monitoring program is designed to develop exposure data using passive dosimetry rather than biomonitoring techniques. The passive dosimetry methods provide a reliable measure of individual worker exposure and are amenable to determining exposure to a variety of body areas. In addition, AHETF desires to utilize the same methods for all MUs so exposure levels from one handling scenario can be compared to

those from all others. In particular, AHETF has selected the following standard dosimetry techniques:

- Whole body dosimeters (cotton long underwear)
- Hand washes (surfactant in water)
- Face/neck wipes (cotton gauze pad moistened with surfactant in water)
- Head patches (cotton inner dosimeter pieces, both inside and outside of chemical-resistant headgear; used when headgear is required by the surrogate product label)
- Socks (cotton, as a dosimeter to measure foot exposure in some scenarios)
- OVS tubes with filter and sorbent

Each field study protocol will specify exactly what dosimeter types will be used and how they will be handled in the field and in the laboratory.

13.1 Validation of Passive Dosimetry

Because it is difficult to isolate and validate particular components of dermal dosimetry methods, the best validation is a comparison of the sum of passive dosimetry methods against the biomonitored dose. The data examined in a recent review of both proprietary and published studies demonstrated an excellent correlation between passive dosimetry and biomonitoring (Ross et al., 2007). Passive dosimetry as a measure of dosage appears to be consistent with biomonitoring with no bias, i.e., no tendency to over- or underestimate exposure. This evaluation demonstrated that the total absorbed dose (or daily dosage) estimated using passive dosimetry for important handler and reentry scenarios is generally similar to the measurements for those same scenarios made using human urinary biomonitoring methods. Further, this is strongly supported by statistical analysis of individual worker passive dosimetry to biomonitoring ratio and variance within and between studies. The passive dosimetry techniques currently employed yield a reproducible, standard methodology that accurately and reliably quantifies exposure and does not underestimate daily absorbed dose. The dermal exposure monitoring techniques used in the studies in this comparison were whole body dosimeters or patches; hand washes; and face/neck wipes or head patches.

In 2007, EPA convened a Scientific Advisory Panel to discuss validity of passive dosimetry techniques to collect worker exposure data (SAP, 2007 and U.S. EPA, 2007). The Panel concluded that:

> Although a bias may exist, no bias between dermal exposure monitoring and biological monitoring could be detected in large part because of the statistical uncertainty inherent in the exposure data. (SAP, 2007).

13.2 Description of Techniques

Whole Body Dosimeters (WBD)

WBDs (i.e., cotton long underwear) are preferable to patches since they overcome the need to extrapolate residues on a small patch to entire body part areas. They are also easy to cut into sections to determine residues on various body parts. This provides information that can be used to estimate exposure for workers wearing various PPE items, such as aprons. For most studies, AHETF will section the dosimeters into six parts for separate analysis: upper arms, lower arms, front torso, rear torso, upper legs, and lower legs. For some studies (generally when very low exposures are anticipated), AHETF will section dosimeters into two sections: upper body and lower body. Each GLP study protocol will specify how the dosimeters will be sectioned (in the field) for analysis.

AHETF triple-washes the cotton long underwear in detergent and water before use to prevent analytical interference due to chemicals remaining from the manufacturing process (SOP AHETF-8.J). The use of these WBDs is described in detail in SOP AHETF-8.A.

During data analysis, residues from all sections will be summed. Any dosimeter section with residues below an analytical limit (limit of quantitation or limit of detection) is normally assumed to contain half of that limit when summing residues.

Hand Washes

AHETF has selected a standard hand washing technique using a dilute surfactant solution. This technique is more efficient than wiping of the hands, and is preferable to cotton glove dosimeters, which overestimate hand exposure by absorbing and trapping more chemical than skin would.

Hand washing involves first having the worker rub his hands together while a researcher slowly pours 400 mL of solution over his hands. The wash solution is collected in a bowl placed under the worker's hands. Researchers are trained to take about 30 seconds to pour the liquid. The worker is then asked to immerse and rub his hands in the wash water in the bowl for about 30 seconds. Finally, 100 mL of fresh surfactant solution is poured over the worker's hands to rinse them and the rinse water is collected in the same bowl. The combined wash/rinse water is transferred to a sample container and analyzed to determine total residues.

Since this washing procedure is dependent on cooperation from the subjects, the technique is practiced before monitoring until researchers are comfortable the worker is appropriately rubbing his hands and is not likely to interfere with the collection process. This also creates some consistency in how vigorously workers rub their hands which could impact removal efficiency.

Whenever a worker would normally wash his hands a researcher will collect a hand wash sample instead. This will be done before any eating break and any other time the worker says he would like to wash his hands, so each MU could include several hand wash samples. In addition, a hand wash will be taken if a product label specifically requires that hands be washed prior to eating, drinking, etc.

AHETF believes that repeated hand washes during a work period will overestimate exposure (Ross, et. al. 2007), so AHETF will not require a hand wash except before eating breaks. In particular, workers will be urged to drink water or sports drinks during the day to avoid heat illness (see Sec. 5.2.1) and it would overestimate exposure to stop the monitoring process to collect an interim hand wash every time a subject wanted to drink something. AHETF has also observed that applicators often carry water with them during aerial and ground applications and it would not always be possible to enforce a rule to collect a hand wash for these workers before every drinking occasion. Finally, the risk of heat illness is often an important consideration and AHETF is aware (based on previous experiences in the field) that some workers will decrease their fluid intake or avoid consuming fluids altogether if stopping work and providing hand washes is required before each drinking break. Thus, AHETF chooses not to require hand washes before drinking breaks so workers will not have any incentive to avoid drinking fluids during monitoring. The use of hand washes is described in detail in SOP AHETF-8.B.

During data analysis, residues in all washes will be summed for each MU. Any wash sample with residues below an analytical limit (limit of quantitation or limit of detection) is normally assumed to contain half of that limit when summing residues.

Face and Neck Wipes

AHETF has selected a standard face/neck wiping technique involving two successive wipes with cotton gauze pads wetted with a dilute surfactant solution (same solution as hand washes). The entire face, front of the neck, and back of the neck are wiped two separate times and the two gauze pads are combined for analysis. This technique is preferred to using hat patches since no extrapolation of residues to the entire face/neck area is needed.

Whenever a worker would normally wash his face a researcher will collect a face/neck wipe sample instead. This will be done before any eating break and any other time the worker says he would like to wash his face. Another sample will be collected at the end of the monitoring interval. These multiple face/neck wipe samples will be combined for analysis, so each MU will have just one face/neck residue value. The use of face/neck wipes is described in detail in SOP AHETF-8.C.

During data analysis, any wipe sample with residues below an analytical limit (limit of quantitation or limit of detection) is normally assumed to contain half of that limit.

Head Patches

For a few scenarios, AHETF will employ a combination of inner and outer head patches to estimate exposure to non-face and non-neck head areas. This will usually be limited to studies that involve overhead exposure where EPA regulations often require chemical-resistant headgear. It will also be utilized any time a chosen surrogate requires chemical-resistant headgear for the task being studied. This will be specified in the field study protocol.

An example is the open cab airblast applicator scenario. In that case, AHETF utilized both inner patches (under the hat) and outer patches (on top of the hat) that were arranged so one is not right on top of the other. Residues found on these patches provide a measure of inner and outer potential exposure. This allows estimation of head exposure to workers who do or do not wear chemical-resistant headgear. The latter requires summing extrapolated exposure from the inner patch with that from the outer patch. Any patch sample with residues below an analytical limit (limit of quantitation or limit of detection) is normally assumed to contain half of that limit.

AHETF uses whole-body inner dosimeter pieces for head patches and triple-washes them in detergent and water before use to prevent analytical interference due to chemicals remaining from the manufacturing process (SOP AHETF-8.J). The use of hat patches is described in detail in SOP AHETF-8.H.

Socks

For a few scenarios, AHETF will employ sock dosimeters (new cotton socks provided by researchers) to estimate exposure to the feet. Exposure to this body area is generally insignificant for pesticide handlers since shoes and socks (required by the WPS) provide considerable protection. However, scenarios that involve handheld application equipment and workers who walk while they spray (such as backpack applications) might include this dosimetry technique. This determination will be made by AHETF in conjunction with EPA and other regulators and will be specified in the field study protocol. Cotton socks are worn under normal footwear (i.e., shoes/boots and socks) and are collected at the end of the day for analysis, similar to the WBDs. Any sock sample with residues below an analytical limit (limit of quantitation or limit of detection) is normally assumed to contain half of that limit.

AHETF triple-washes socks before use to prevent analytical interference due to chemicals remaining from the manufacturing process (SOP AHETF-8.J). The use of sock dosimeters is described in detail in SOP AHETF-8.I.

Total Dermal Exposure

During data analysis, exposure levels determined from WBDs, hand washes, head (may be based on face/neck wipes and/or hat patches), and socks are summed to provide a

`measure of total potential dermal exposure. This value is generally used by EPA and other regulators in exposure assessments for individual pesticide products. Dermal exposure can be normalized, for example, by the amount of active ingredient handled.

Air Sampling Tubes with Filter and Sorbent

AHETF has selected a standard inhalation dosimetry technique that has been used extensively and that involves a personal air sampling pump and a collection device (OVS tube) attached to the collar of the worker facing downward. Air is typically drawn through the device at 2 L/min. The collection device is a glass tube containing a sampling train consisting of a filter followed by two sections of sorbent. The tube is mounted in a plastic protective sleeve which is clipped to the collar of the worker. The filter and sorbent type are selected for each surrogate chemical to maximize sorbent retention and analytical removal efficiency. Validation studies are performed to ensure that residues trapped early in the monitoring period will not be stripped away by the constant airflow through the device during an entire day of monitoring. In the laboratory, the filter and both sorbent sections are extracted together to yield a total residue value (i.e., for particulates and vapors combined). The use of OVS tubes is described in detail in SOP AHETF-8.D.

During data analysis, total potential inhalation residues collected for the entire monitoring period are reported along with a calculation of the average air concentration (by taking into account the measured airflow rate and the time of monitoring). This probably overestimates exposure, and is therefore conservative, since even large particles that would not be respirable might be trapped on the filter and assumed as potential exposure. Any OVS tube with residues below an analytical limit (limit of quantitation or limit of detection) is normally assumed to contain half of that limit when summing residues. During exposure assessments, potential inhalation exposure to the worker is calculated by taking into account the breathing rate for the level of work activity being performed. The result can be expressed in total mass for the monitoring period (e.g., µg) or normalized by amount of active ingredient handled (e.g., µg/lb ai).

13.3 Nature of Testing Guidelines

Regulatory agencies frequently collaborate to make exposure monitoring guidelines harmonized. A good example is the Series 875 guidelines of US EPA that were designed with multi-national input starting with a meeting in The Hague in 1992 and punctuated with meetings in Ottawa, Toronto, and Washington, DC that culminated with the issuance of OECD and EPA guidelines that are very similar (OECD, 1997; U.S. EPA, 1996). All of the dosimetry techniques utilized by AHETF are consistent with these most recent guidelines.

14 Ethical Considerations

All AHETF field studies will be designed and conducted in compliance with the requirements of 40 CFR Part 26, subparts K and L, and will be documented in compliance with the requirements of 40 CFR Part 26, subpart M.

14.1 Recruitment

Recruitment generally occurs in two phases. In the first phase AHETF identifies growers (or land owners or farm operators) who might use the surrogate chemical under conditions consistent with the scenario-specific selection design for the desired MUs. In some cases, commercial application services are also recruited that can provide equipment and workers (e.g., for aerial applications). In the second phase, workers are recruited that have experience with the desired handling activities and that are associated with the growers, land owners, or commercial applicators that have agreed to cooperate.

14.1.1 Recruiting Growers, Landowners, or Commercial Applicators

When searching for cooperating growers, landowners, or commercial applicators, AHETF needs to locate certain combinations of conditions including appropriate equipment types, sufficient crop acres, and the number of workers required. For most scenarios, that search begins by selecting growers or landowners who might be willing to accept a pesticide application within a particular site (that has already been determined by the scenario plan) as described in Sec. 10.8. AHETF generally begins by contacting local resources for assistance. These may be agricultural researchers, County Agricultural Extension agents, farm equipment dealers, farm supply dealers, and others in the local area where the study will be conducted. In some cases, AHETF may also rely on commercial pesticide application companies that may be necessary for some scenarios (e.g., aerial applicators).

AHETF will often rely on one or more "local site coordinators (LSCs)" or researchers from "contract research organizations" (CROs) to help identify appropriate growers or commercial applicators and to explain to them the need to conduct research with volunteer workers who are willing to be monitored while they perform their regular activities of handling agricultural pesticides in a defined scenario. During initial contacts, growers or commercial application companies are not asked if their employees might be willing to participate, nor are they urged to ask their employees if they might be willing to participate. Cooperating employers are asked only for their permission for study staff to contact their employees at a future time concerning possible participation in the research.

The final step in this phase is generally for the Study Director, LSC, and/or CRO to visit the growers and formally solicit their cooperation in the research study. The AHETF research program and the need for a suitable site are explained. The growers are advised

of the research program benefits to agriculture. The growers must agree to the lost productivity of his workers and the disruption of the daily routine by the field study team. Additionally, they must assent to all of the ethical terms regarding non-coercion, voluntariness, non-retaliation, and full pay for the handler employees. Only when these conditions have been met can handlers be approached to participate in the research study.

14.1.2 Recruiting Subjects

A study-specific recruitment plan will be developed for each field study protocol. This will include the methods used to locate and contact growers and/or commercial applicators as described above. No recruitment of growers, commercial applicators, or subjects will be initiated until the protocol and all supporting materials have been approved by the IRB and reviewed by EPA and the HSRB. If the study is to be conducted in California these documents will require prior approval by the California Department of Pesticide Regulation.

In a typical field study, recruitment of subjects would begin by obtaining permission from all cooperating employers to contact employees who may be interested in participating in the research. These employers must sign a statement indicating they will not encourage or discourage employees from participating in the research, and that workers' decision to participate, not participate, or withdraw will not affect their employment or their pay. The form for this statement is included in SOP AHETF 11.B and will be provided in Spanish when Spanish-speaking subjects are involved. It should be emphasized that some growers themselves may also be asked to participate in the study, for example farm owners who handle pesticides themselves.

The next step typically would be to distribute to the identified handlers (e.g., employees) a flyer summarizing the nature of the study and the eligibility criteria, and inviting interested candidates to contact the Study Director directly to find out more about the study. The flyer is available in English and Spanish. The flyer will be reviewed and approved by the IRB before use, and included in the package of materials supporting the proposed protocol when it is submitted for EPA and HSRB review. This "self –identify" approach reduces the potential for coercion and undue influence exerted on the workers by his peers and employers to participate (or not participate) in the study. It also enhances the opportunity for the potential subjects to make a completely voluntary decision about participation.

Interested candidates will typically be invited to a meeting without supervisors being present where general information about the research will be provided. They will also be provided a copy of their employer's signed "non-coercion statement" discussed above and a copy of the consent form to take home and review.

Candidates expressing further interest in the study will then be invited to a private discussion with the Study Director or other responsible research staff, at which the eligibility criteria will be reviewed, the informed consent document will be presented and

discussed, and all the candidate's questions will be addressed. The candidate's consent to participate in the research will be sought and documented during this meeting, or arrangements will be made to allow the candidate more time for consideration. See Section 14.214.2 for more details on the consent process.

In most cases, more workers than will ultimately be needed are identified during the recruitment process. In addition, MU selection will generally limit participation to a maximum of 2 workers from any grower or commercial applicator. This will often create an opportunity to randomly choose workers to participate in the study, for example if there are more interested employees of a particular grower than are needed. Selecting fewer workers than are available also lessens the pressure on individual workers to volunteer and thus protects their right to decline to participate.

The recruitment process is described in more detail in SOP AHETF-11.B, however study-specific recruitment processes will be detailed in each study protocol.

14.1.3 Eligibility Criteria

The Study Director is responsible for ensuring the following inclusion criteria are met by all participants.

Willingness to Participate

All monitored workers must be freely willing to participate in a study of this type and to sign the approved informed consent form (including appropriate attachments such as a product-specific toxicity risk statement and, for monitoring conducted in California, the Experimental Subject's Bill of Rights).

Occupation

All monitored workers must be professional agricultural pesticide handlers - that is, their regular job must include handling agricultural pesticides. They may be farm owners, farm operators, farm employees, contractors, commercial applicators, etc.

Training

All monitored workers must have received basic pesticide handling training in accordance with the Worker Protection Standard (WPS) or equivalent Canadian regulations, or must be exempt from such regulations. Each participant must confirm that they have received the required training or that they are exempt from the requirement.

Experience

All monitored workers must have recent experience (i.e., within the last year) performing the particular task they would be performing in the research and with the particular

equipment to be utilized. No minimum level or amount of experience is required, but researchers will document how much experience (e.g., in months or years) each participant has for the particular task being monitored.

Age

All monitored workers must be at least 18 years old, and able to verify their age with government-issued photo identification.

Health Status

All monitored workers must consider themselves to be in good general health with no medical conditions that could impact their ability to participate in the study. This precludes selection of workers who are chronically ill or terminally ill. See SOP AHEF-11.C for more details.

Product Label Conformance

All monitored workers must agree to perform pesticide handling tasks in conformance with label and WPS requirements. In particular, monitored workers must agree to wear all PPE required by the label. Researchers will remind participants to use required PPE should they be observed not wearing it. Any workers who fail to follow label requirements during the study will be removed from the study.

In addition to the inclusion criteria just listed, the following exclusion criteria will be enforced:

Normal Use of Extra Clothing or PPE

A worker who normally chooses to wear more protective clothing or PPE than is required by the label may be allowed to wear the additional items, or may be excluded from the study, depending on the goals of the specific field study design. No one included in an AHETF study will be required to wear less protective clothing or PPE than he or she normally wears when performing the activities monitored in the research scenario.

Pregnant or Nursing

Female workers who are pregnant or nursing are ineligible to participate as subjects in any AHETF study. Nursing will be self-identified. The pregnancy status of all potential female subjects will be ascertained using a supervised over-the-counter urine pregnancy test within 24 hours before the initiation of monitoring. See SOP AHETF-11.D.

Do Not Understand English or Spanish

Candidates who do not understand English or Spanish are ineligible to participate as subjects in an AHETF study. See also SOP AHETF-11.B.

Employed by a Member of AHETF

Employees of any member of AHETF, or of any organization or investigator under contract to AHETF, are ineligible to participate as subjects in any AHETF study.

14.1.4 Use of Vulnerable Groups

AHETF excludes as participants people who are ill (self-identified), cognitively impaired (deemed by person obtaining consent), pregnant (based on over-the-counter pregnancy test), nursing (self-identified), minors (based on government-issued photo identification), employees of the Principal Investigator (for AHETF studies, the Study Director), sub-investigators, or AHETF member companies. However, AHETF often identifies the following potentially vulnerable groups which may be necessary to reflect the future handler-day population:

- Employees of a grower or commercial applicator;
- Limited or non-readers;
- Poor people;
- Workers without [health] insurance; and
- Illegal immigrants

Employees of a grower or commercial applicator

Employees of growers or commercial application contractors are allowed to volunteer as study participants. This group may be vulnerable to coercion from the grower who may get the benefit of free pesticide for his crop(s) or the applicator who gets a contract for application services. These study participants will be classified as a “farm employee” or “commercial applicator” in data records, and will provide many of the monitoring units for most scenarios.

The desire to collect exposure data for professional workers who use their own (or their employer’s) equipment necessitates that AHETF work with their employers (for example, the grower who plans to use the pesticide) since sending workers to another grower would generally involve a change in equipment used. For example, AHETF would not want to contract with one aerial application company and have their aircraft operated by employees of a different application company. This is not only for safety reasons, but because the generic database is designed to include workers using equipment they are familiar with. AHETF therefore intends to sample MUs from the employees of cooperating growers and commercial applicators, but will take care to prevent coercion of or exertion of undue influence on these workers by their employers.

AHETF recognizes the need for extra care to avoid the potential for coercion and undue influence from growers or custom applicator operators who might benefit from cooperating with an AHETF study. However, these types of handlers constitute a significant portion of the future handler-day population and exclusion of these individuals would unjustifiably diminish the value of the data collected by the entire program. Therefore, all cooperating employers of potential subjects will be required to sign a statement that they will neither encourage nor discourage their employees to participate, and that the decisions of their employees to participate, not participate, or withdraw will have no impact on their employment or work opportunities. A copy of this statement will be provided to all affected candidates in the informed consent interview. (See also SOP AHETF-11.B).

Regarding the benefit to growers, in the form of cost reimbursement by the sponsor for pesticide product, AHETF believes the magnitude of this benefit is not likely to result in coercion of employees to volunteer to participate in an AHETF study. As described below, this benefit is appropriate compensation for grower inconveniences, but generally accounts for just less than 1 to 3% of production costs for the acreage that is treated.

The AHETF generally reimburses the grower or landowner for the cost of the surrogate pesticide that is used to conduct field monitoring studies. The cost of chemical compensates growers for the inconvenience, potential loss of productivity, and potential risk that the product may not perform as adequately as alternative products.

The cost of chemical is not a major inducement to cooperate because it is a small percentage of the overall cost of producing a crop. The table below lists the cost of five surrogate compounds used in previous AHETF studies along with production costs for four representative crops. The production costs are based on data from universities in the states specified in the table for each crop. The chemical costs were provided by Loveland Products, Inc. The application rates are the maximum labeled rates for each of the compounds and crops.

Crop Type and Surrogate ai	Formulation Type	Production Costs ($/A)	Chem Cost ($/lb ai)	Appl Rate (lb ai/A)	Chem Cost ($/A)	Chem Cost (% of Production Costs)
Corn in IA[1]						
Carbaryl	4 lb/gal liquid	$500	$8.28	2.00	$16.56	3.31
Chlorothalonil	6 lb/gal liquid	$500	$3.97	1.50	$5.96	1.19
Diazinon	50% solid	$500	$10.54	0.75	$7.91	1.58
Malathion	5 lb/gal liquid	$500	$4.92	1.25	$6.15	1.23
Simazine	4 lb/gal liquid	$500	$3.66	2.00	$7.32	1.46
Grapes in CA[2]						
Carbaryl	4 lb/gal liquid	$1,700	$8.28	2.00	$16.56	0.97
Chlorothalonil	6 lb/gal liquid	$1,700	$3.97	NA		

Crop Type and Surrogate ai	Formulation Type	Production Costs ($/A)	Chem Cost ($/lb ai)	Appl Rate (lb ai/A)	Chem Cost ($/A)	Chem Cost (% of Production Costs)
Diazinon	50% solid	$1,700	$10.54	1.00	$10.54	0.62
Malathion	5 lb/gal liquid	$1,700	$4.92	NA		
Simazine	4 lb/gal liquid	$1,700	$3.66	2.00	$7.32	0.43
Tomatoes in FL[3]						
Carbaryl	4 lb/gal liquid	$13,500	$8.28	2.00	$16.56	0.12
Chlorothalonil	6 lb/gal liquid	$13,500	$3.97	1.50	$5.96	0.04
Diazinon	50% solid	$13,500	$10.54	0.75	$7.91	0.06
Malathion	5 lb/gal liquid	$13,500	$4.92	1.56	$7.68	0.06
Simazine	4 lb/gal liquid	$13,500	$3.66	NA		
Apples in WA[4]						
Carbaryl	4 lb/gal liquid	$10,000	$8.28	3.00	$24.84	0.25
Chlorothalonil	6 lb/gal liquid	$10,000	$3.97	NA		
Diazinon	50% solid	$10,000	$10.54	0.50	$5.27	0.05
Malathion	5 lb/gal liquid	$10,000	$4.92	NA		
Simazine	4 lb/gal liquid	$10,000	$3.66	2.00	$7.32	0.07
Average						**0.76%**

Footnotes:

[1] Iowa State University – University Extension. "Estimated Costs of Crop Production in Iowa – 2007", Updated February 2007. Available from: www.extension.iastate.edu/agdm. Accessed May 16, 2007

[2] University of California Cooperative Extension. "Sample Costs to Establish a Vineyard and Produce Wine Grapes", 2001. Available from: http://coststudies.ucdavis.edu/. Accessed May 15, 2007.

[3] University of Florida, "Tomatoes: Estimated production costs in the Dade County area, 2004-2005", Updated February 2005. Available from: www.agbuscenter.ifas.ufl.edu/cost/cop03-04/DadeTomato.doc. Accessed May 16, 2007.

[4] Washington State University. "Tree Fruit Horticulture Page - Apples in Washington State", Updated August 2002. Available from: http://www.ncw.wsu.edu/treefruit/aplcrop. Accessed May 16, 2007.

NA indicates the compound is not registered for use on that crop.

The value of the surrogate chemical is usually less than 1% of the total production cost with an overall average for the examples in the table of 0.76%. The exception is large acreage agronomic row crops (e.g., corn), but the value of the chemical is still low (approximately 1.2 to 3.3%) compared to production costs.

It should also be emphasized that the cost for chemical is only provided for a single application, is only enough to compensate for the actual amount needed for the study, usually does not cover the cost of the actual application itself, and does not cover the total chemical costs to the grower since they usually apply multiple pesticides each year. This is not an unreasonable benefit for growers and is not large enough to cause coercion or undue influence on an employee to participate in a study.

Limited or non-readers

AHETF does not intentionally recruit subjects with limited literacy, but pesticide handlers occasionally fall into this category and may therefore be included in the monitoring program. AHETF has procedures in place to include an impartial witness (i.e., unassociated with the conduct of the research, not employed by the sponsor or any of its contractors, and with a general understanding of agriculture) in the consent process for any candidates who are not comfortable reading the informed consent document. These procedures are discussed in the following section (and in SOP AHETF-11.B).

Poor people and uninsured workers

Another potentially vulnerable group that might be part of the future handler-day population is poor/uninsured workers. AHETF does not intentionally recruit these individuals and will not inquire as to the economic or health insurance status of potential study participants. Therefore, this category will not be identified to the IRB as one that is intended to be recruited. As discussed below, the remuneration being offered (generally for just one day of participation) is believed to be not high enough to induce otherwise reluctant workers to participate, so the economic status of participants in these studies is not a concern. In addition, AHETF will cover all costs of injury or illness that workers experience because of participating in the study (that are not covered by the worker's or employer's insurance).

Illegal workers

Another potentially vulnerable group that might be part of the future handler-day population is illegal workers. For example, illegal workers may feel obligated to participate (e.g., in order to protect their job) or be reluctant to accept medical treatment. Federal laws give employers the responsibility for ensuring their workers are legal and AHETF does not employ workers. AHETF will therefore assume workers are legal and will not ask about their status. In addition, should researchers become aware of an illegal worker they will not share that information. Workers who might be illegal will be protected from coercion primarily via the mechanism described below where the Study Director (or designee) will discuss the voluntary nature of study participation with the worker's supervisor/employer.

In addition, the consent form indicates workers may refuse medical treatment unless they appear to be suffering acute toxicity from the pesticide product or from heat stress, or if

they are unable to make a rational decision. The ability of a subject to make a rational decision will be assessed by the Study Director or designee (in cooperation with the on-site medical professional). For example, a subject who can't answer simple questions like, "What day is it?" might be considered unable to make a rational decision. In addition, any unconscious subject will be considered unable to make a decision.

14.2 Informed Consent Process

When sites have been selected and potential participants have been identified, the Study Director (or a designee who has appropriate ethics training) is responsible for obtaining informed consent from all study participants. Any materials used during recruitment or consent meetings will be approved by an IRB before use.

As discussed above, the researcher obtaining consent first has a discussion with the supervisor of each potential participant to ensure the supervisor understands that workers should not feel any coercion to participate in the study. The supervisor must confirm there will be no adverse impact on a worker who does volunteer, who does not volunteer, or who withdraws from the study for any reason. This extra care to prevent coercion from employers, especially from growers and commercial applicators, will be documented on a form which the supervisor, grower, or commercial applicator must sign. Then, each volunteer is provided with the supervisor's signed form (if not already provided during recruitment as described above), the IRB-approved consent form, and a full explanation of the study, its requirements, and any potential risks. This occurs during a confidential and private discussion with the researcher obtaining consent.

Volunteers are advised of their right to withdraw from the study at any time and for any reason without jeopardizing their job position with their employer. Volunteers are also informed during the confidential consenting process that they will receive an $80 payment for beginning study participation, even if they withdraw for any reason or the sponsor stops the monitoring for any reason. They will also be informed that their supervisor has certified there will be alternate work available if they don't participate, or sufficient work on monitoring days so they will not be denied a full day's pay.

AHETF consent forms follow a standard format, but are unique to individual studies. All consent forms will contain the elements required by 40 CFR 26.1116. This includes discussions of the following aspects of the research: purpose, voluntary nature, benefits, risks, alternatives, withdrawal, compensation, confidentiality, where to direct questions, etc.

During the discussions between potential participants and the Study Director (or designee), ample time is provided for questions and the person obtaining consent will provide any additional information or clarification that is requested. These discussions typically take place at the worker's location, in a private setting. Consent is generally obtained within one to three days of actual study conduct, but sometimes earlier. If the worker agrees to participate he/she is asked to sign and date the informed consent form

and the person obtaining consent provides a copy of the signed form to the worker. Within 24 hours prior to monitoring, any women who agreed to participate will be asked to take a urine pregnancy test (over-the-counter variety) and will be allowed to participate only if the test is negative. This test will be supervised by a female researcher. To protect the privacy of the worker, the test results are not revealed to the employer or co-workers. If the worker chooses to proceed with the study then a female researcher will confirm the test is negative and record this in the study raw data. No positive test results will be documented and the worker will be allowed to withdraw from the study without stating a reason. See SOP AHETF-11.D for details of the pregnancy testing procedures.

For workers whose preferred language is Spanish, AHETF obtains an IRB-approved translation of the consent form and utilizes a bilingual researcher (who is familiar with the study procedures) to obtain informed consent in Spanish. When a bilingual researcher is obtaining consent, the Study Director may be present in the private meeting. If all reasonable efforts to locate a bilingual researcher have been exhausted, it is acceptable to use an interpreter as long as they are accompanied by a researcher with appropriate ethics training that has been designated to obtain consent. Interpreters will translate the (English) discussion from the person obtaining consent into Spanish during the consent process. They will also be utilized during the study should any issues arise which can't be resolved directly with the worker.

In situations where a potential participant is not comfortable reading English or Spanish, but can understand the spoken language, an impartial witness will observe the consent meeting. The witness must not be affiliated with the research or investigators, but must have a general understanding of agriculture. If the volunteer understands English, the person obtaining consent will read the (English) consent form to the volunteer. If the volunteer understands Spanish, the bilingual researcher obtaining consent will read the (Spanish) consent form to the volunteer and the required witness will be bilingual. The role of the witness is to observe the consent interview and evaluate whether the information in the consent form and any other written information was accurately explained to, and apparently understood by, the subject and that the subject freely consented to participate in the research study. The witness must sign the consent form confirming these conclusions; otherwise the volunteer cannot sign the consent form. Interpreters for Spanish speakers (along with an English-speaking researcher) may be utilized only if all reasonable efforts to obtain a bilingual researcher have been exhausted. When both an interpreter and a witness are needed, they may not be the same person. If all questions have been adequately answered and there are no remaining concerns, the witness and the volunteer, as well as the person obtaining consent, will sign the consent form.

In all situations, the Study Director (or designee) will not sign the consent form unless he/she believes the candidate fully understands the information presented and has made a fully voluntary choice to participate. This will be ascertained by providing repeated opportunities to ask questions and by asking a series of standard questions of the potential volunteers that would require a response that indicates understanding of key issues. The

person obtaining consent will document whether or not the subject understands each issue on a standard form and will provide further discussion if needed to ensure understanding (see SOP AHETF-11.B).

During the consent process, volunteers will be asked if they would like their personal results from the study mailed to them. This is entirely optional, but will require they provide their name and address. That personal information will be kept confidential (see Section 14.5).

The process for obtaining informed consent is documented in an AHETF Standard Operating Procedure (AHETF-11.B).

14.3 Subject Remuneration

In almost all cases, AHETF will be monitoring the exposure to pesticides for workers who are performing their usual activities. Monitored workers would be handling pesticides even if they weren't participating in the study. Workers who participate in AHETF exposure monitoring studies are "on the job" and will receive their normal pay and all other compensation they are due, including compensation for any overtime worked according to local laws. This compensation is the responsibility of the worker's employer and not AHETF.

In addition to their normal pay, AHETF will provide payments totaling $100 (U.S. dollars if in the United States or Canadian dollars if in Canada) to each worker who participates in the study. Twenty dollars ($20) will be given when a volunteer attends a private consent discussion with the Study Director (or designee), whether or not the candidate decides to sign the consent form. Eighty dollars ($80) will be given when a volunteer begins participation in the study (i.e., by donning dosimeters and beginning handling activities). These payments are in appreciation for the extra effort and inconvenience associated with the consent process and participating in the study. This includes wearing the inner dosimeter (long underwear, requires undressing in a private area with the assistance of a researcher of the same sex); allowing researchers to wash their hands and wipe their faces; allowing researchers to collect the inner dosimeters at the end of the monitoring period; and wearing a personal air sampling pump and tube throughout the workday.

AHETF has selected standard amounts for all AHETF studies and participants since the inconveniences involved are essentially the same for participants in all studies. In addition, AHETF chooses not to offer an hourly rate since it prefers that workers perform their typical tasks and wants to avoid any incentive for workers to choose a particular task or AaiH since they could "earn more money". The consent process takes about 1 hour. Dressing and undressing workers takes about 20 to 30 minutes total throughout the work day. This includes collecting hand washes, face/neck wipes, and inhalation devices. Collecting interim hand washes and face/neck wipes (e.g., before a lunch break) takes about 5 minutes. Workers may also need to wait about 10 to 15 minutes to be dressed

and/or undressed when there is more than one worker ready to start or stop monitoring at the same time. In total then, up to an hour might be necessary to deal with putting on dosimeters and collecting dermal and inhalation exposure samples. In addition, participants may experience embarrassment or thermal discomfort due to the whole-body dosimeter.

While any standard amount of remuneration could represent a very different proportion of various workers' typical daily pay, fairness suggests that each worker should receive the same amount of remuneration since the amount of inconvenience is essentially the same. AHETF selected the amount of $100 (total payment) at the outset of the task force project and still believes this is an appropriate amount. AHETF believes it is not practical to ascertain individual worker salaries and regional differences in order to determine an amount of remuneration that might be based on a percentage of daily pay. AHETF believes that $100 is not sufficiently high as to create undue influence to participate in the study, especially since workers are generally limited to one day of study participation. However, the IRB reviewing individual study protocols might change this value and if necessary, the standard remuneration amount will be adjusted.

The $20 payment will be provided in cash at the end of the consent meeting. The $80 payment will be provided in cash at the end of the monitoring period (whether ended by design, due to participant withdrawal, or by the researchers).

Generally, a worker will be allowed to participate in a specific field study only one time. This study design principal provides data for separate exposure measurements that reflect different workers in order to capture variability between workers. However, the same worker could participate more than once in a study (or in two studies) as long as the worker performs a different task. For example, one person could be monitored for exposure as a mixer/loader on one day and as an applicator on another day (assuming the worker meets all eligibility criteria for both activities). In this case, that person would receive an $80 payment for participation on each of those two days (in addition to the $20 for attending the consent meeting).

In addition to the remuneration identified above, AHETF will inform all potential volunteers that they will not be deprived of a full day's wages should they decide not to participate, or choose to withdraw from the study, or if the study is stopped short for any reason. First, their employer or supervisor must certify in advance that alternate work will be available for each volunteer should they decide not to participate, or if they withdraw at any time, or if the monitoring period ends before the typical work day is over. Second, if the worker is unable to complete the work day (e.g., due to an injury or heat stress), AHETF will reimburse the employer for the rest of that day's wages.

14.4 IRB Review Process

AHETF will use an Institutional Review Board (IRB) to review and approve each of its study protocols for ethical compliance. Study protocol submissions from AHETF to the IRB typically include the following:

- Field Study Protocol
- Informed Consent Form – English
- Hospitalization Procedures and distance to nearest hospital
- Resumes for researchers
- Copies of signed SOPs cited in protocol
- Copies of product labels and MSDSs for test substances
- Product Risk Statements – English
- Recruitment materials (e.g., flyer)

AHETF is identified as the sponsor and the Study Director as the Principal Investigator. The submissions also identify study facilities (generally local site coordinator research facilities) and provide details about subject recruitment, consent, and payment. The IRB also maintains files containing the curricula vitae and documentation of completion of training in the protection of human subjects for the Study Director and other researchers involved in the study. The IRB is also utilized to translate certain documents (e.g., consent form and product risk statements) into Spanish when Spanish-speaking subjects are anticipated.

Protocols and supporting information for studies to be conducted in California are also submitted to the California Department of Pesticide Regulation (CDPR) for their review and approval. Any changes requested by CDPR are incorporated into the study protocol and/or consent forms that must then be reviewed and approved by the IRB. Only upon receipt of IRB-approved protocol and consent forms will CDPR grant final approval for the study to be conducted in California.

Finally, after IRB approval, protocols will be submitted to EPA (and to the HSRB by EPA) for review. Any resulting changes to the study protocol or consent form must again be reviewed and approved by the IRB (and submitted to CDPR, if conducted in California).

14.5 Additional Efforts to Protect Human Subjects

Additional steps that AHETF takes to ensure the safety and maintain confidentiality of study participants are summarized below.

Minimizing health risks before monitoring

Copies of the material safety data sheet (MSDS) and product label are made available to members of the study team and study participants. During the consent process, the Study

Director (or designee) will discuss the information contained in them. For non-English speakers, a bilingual researcher will have these discussions with the volunteer. Particular attention will be paid to the possible acute toxic effects associated with the pesticide product in the study. Workers will also be reminded of standard practices that should be followed to reduce their exposure to pesticides. For example, provisions of the Worker Protection Standard (WPS) will be cited such as the requirement to wear long pants and long-sleeved shirts and to remove clothing that get drenched with chemical from an accidental spill. See SOP AHETF-11.E for more details.

Study subjects are never asked to wear less clothing or PPE than they would ordinarily wear, even if the items are not required by the product label. In cases where a worker normally wears clothing or PPE not required by the label, the AHETF either allows them to wear the extra clothing (or equipment) or they are excluded from the study, depending on the specific goals of the study. For example, while workers are always allowed to wear a hat of their choice even when headgear are not required by the label, chemical-resistant jackets and hoods would not be allowed unless a particular scenario was designed to include such PPE (as was the case for some MUs in the open cab airblast studies already collected by AHETF). The AHETF may also provide some PPE items required by the product label (e.g., chemical-resistant gloves) to ensure they meet WPS requirements.

Prior to study conduct, the Study Director (or designee) will assess the availability of medical assistance in the locality of the study and identify appropriate emergency medical facilities that may be utilized.

Just prior to participating in an AHETF exposure monitoring study, the worker's normal clothing will be inspected by researchers and the Study Director (or a designee) must approve of that clothing (see SOP AHETF-8.G). To comply with the WPS the outer layer of clothing must be in good condition and be free of large holes or tears or missing buttons which could lead to increased skin exposure. If necessary, a researcher will provide a new shirt or pants for the worker to wear during the study.

Observing subjects during studies for adverse effects

During all AHETF studies, the Study Director and the field investigators share the responsibility for awareness of physical injuries, heat illness, toxic responses, and other adverse effects in study participants. All such researchers are required to complete training on the ethical treatment of human subjects.

As a precaution, a medical professional (emergency medical technician [EMT], paramedic, physician's assistant [PA], licensed practical nurse [LPN], or registered nurse [RN]) will be on-site during the conduct of all AHETF studies while workers are being monitored. The medical professional will be provided the product label, its MSDS, and AHETF SOPS related to pesticide safety and heat stress. The medical professional shall become familiar with these documents and conduct periodic observations of workers

during monitoring and will alert the Study Director to possible signs of illness (especially heat-related or chemical) or injury. This person will also provide appropriate medical care when necessary and help determine when additional medical attention is needed.

During each study, every participant will always have a researcher assigned to observe and document his/her handling practices and this "observer" will have the primary responsibility for detecting adverse effects. Typically this observer is close enough to the worker to have a conversation. Observers are trained to recognize heat stress and are informed of the most likely acute effects of overexposure to the pesticide being used in the study. In addition, the on-site medical professional will check on the workers frequently and help identify early signs of adverse effects.

During study conduct, researchers will also ensure compliance with safety requirements on the product label and with the WPS. For example, workers will be reminded to use the label-specified PPE and to follow use directions on the label. Each worker will be observed by a researcher during the entire monitoring period and the Study Director (or designee) will be present on all days of monitoring. Any study participant who will not follow the label requirements for wearing PPE will be removed from the study. All researchers who interact with the workers must have completed at least one internet-based course in the protection of human research subjects—either the Basic Collaborative IRB Training Initiative Course (CITI) or the Human Participant Protections Education for Research Teams Course (NIH). See SOP AHETF-1.B.

Handling adverse events

Study Directors and other researchers will know in advance where to take workers who might be overheated or who have other medical concerns. If any participant is injured or experiences illness from being in a study, medical treatment will be available at a nearby health care facility. If necessary, AHETF will arrange transportation to receive medical attention. AHETF will cover the costs of reasonable and appropriate medical attention that are not covered by the participant's own insurance or insurance provided by the participant's employer. Treatment records will not become part of the research records for any study, however Study Directors will document that each incident is followed to resolution.

Should an adverse effect occur during an AHETF study, emergency procedures will be implemented. These procedures typically include halting work, removing the worker from the offending environment, and calling 911 (or other local emergency phone number) for medical assistance if needed (see SOP AHETF-11.H). In addition, AHETF has adverse effects reporting policies in place to notify the IRB (SOP AHETF-11.F) and EPA of potential new findings (e.g., as required by FIFRA Section 6(a)(2), SOP AHETF-1.F).

Protecting subject confidentiality

AHETF researchers take several steps to protect the confidentiality of study participant identification. The consent form includes the name and signature of the study participant and is held confidential, separate from other study records. After making a copy of the signed consent form for the participant the Study Director (or designee) will assign a unique worker identification code (generally a letter) to each participant and write that code on the consent form. In addition, study participants have the opportunity to request their personal results from the study by filling out a form that includes their name and address. These documents are the only places where the participant's name and code appear together. All other raw data records and all reports will reference only the unique worker identification code. The Study Director will place the consent forms and result request forms in a sealed envelope to protect the confidentiality of the participant names and addresses. The envelope will be marked as confidential and kept separate from the other study raw data. Only the Study Director will have access to this information until placed into permanent archives where it will remain in a sealed envelope and access will be further restricted (see SOP AHETF-6.D).

Photographs and video recordings that include study participants may be used to document how the study was conducted and may be retained in the raw data. The subject's name will never be associated with these images and no images will be included in submitted reports of study results. However, EPA and other regulatory agencies may audit the raw data from this study, so absolute confidentiality of study participant images cannot be guaranteed.

Post-study follow-up

At the end of each day of monitoring, the Study Director (or designee) will remind participants they should bathe or shower as soon as practical and that they have received a copy of the signed consent form that has phone numbers for reporting any health changes the worker thinks may be related to participation in the study. Any post-study inquiries will be forwarded to the Study Director who will deal with the situation as appropriate and notify AHETF management. When appropriate, the Study Director will report such incidents to the IRB, EPA, and CDPR according to prevailing regulations.

15 References

Brimijoin, William S., "June 27-29, 2007 EPA Human Studies Review Board Meeting Report", Proposed Final Draft V 1 (September 27, 2007).

Chester, G., "Evaluation of Agricultural Worker Exposure to, and Absorption of, Pesticides" Ann. Occup. Hyg., Volume 37, No. 5, pp 509-523 (1993).

DHHS (Department of Health and Human Services). Protection of Human Subjects Common Rule. 45 CFR, Part 46. (2001).

Durham, W. F., and Wolfe, H. R. "Measurement of the Exposure of Workers to Pesticides", Bull., WHO, Volume 26, pp 75-91 (1962).

Fisher, Celia, "June 27-30, 2006 EPA Human Studies Review Board Meeting Report", Final Report (October 6, 2006).

Hackathorn , D. R., and Eberhart, D. C., "Data-Base Proposal for Predicting Mixer-Loader-Applicator Exposure", Dermal Exposure Related to Pesticide Use: Discussion of Risk Assessment, ACS Symposium Series 273, pp 341-355, American Chemical Society, Washington, D.C. (1985).

Hoenig, J. M., and Heisey, D. M. 2001. The Abuse of Power: The Pervasive Fallacy of Power Calculations for Data Analysis. American Statistician 55(1):19-24.

Honeycutt, R., "Field Worker Exposure: The Usefulness of Estimates Based on Generic Data", Dermal Exposure Related to Pesticide Use: Discussion of Risk Assessment, ACS Symposium Series 273, pp 369-375, American Chemical Society, Washington, D.C. (1985).

Honeycutt, R., "NACA Overview on Assessment of Mixer-Loader-Applicator Exposure to Pesticides: Dermal exposure; patch test; biological monitoring; generic exposure data base", Toxicology Letters, Volume 33, 175-182 (1986).

Keigwin, Tracy Lynn. PHED Surrogate Exposure Guide. Estimates of Worker Exposure from The Pesticide Handler Exposure Database. Version 1.1 (August, 1998).

Lunchick, C., Hamey, P. and Iyengar, S., "The Use of the North American (PHED) and United Kingdom (POEM) Worker Exposure Models in Pesticide Registration", Brighton Crop Protection Conference (1994).

National Agricultural Chemicals Association (NACA), "Guidelines for Conducting Biological Monitoring-Applicator Exposure Studies", NACA, Washington, D.C. (1985).

Organization of Economic Cooperation (OECD), "Guidelines for Testing of Chemicals", Organization of Economic Co-operation and Development, Paris, France (1981).

NAS (National Academy of Sciences). Risk Assessment in the Federal Government: Managing the Process, Committee on the Institutional Means for Assessment of Risks to Public Health, Commission on Life Sciences, National Research Council. The National Academy Press, Washington, D.C. (1983).

NAS (National Academy of Sciences). "Intentional Human Dosing Studies for EPA Regulatory Purposes: Scientific and Ethical Issues", Committee on the Use of Third Party Toxicity Research with Human Research Participants, Science, Technology, and Law Program. The National Academy Press, Washington, D.C. ISBN: 0-309-09172-1. (2004).

NAS (National Academy of Sciences). Human Biomonitoring for Environmental Chemicals. Committee on Human Biomonitoring for Environmental Toxicants, Board on Environmental Studies and Toxicology, Division of Earth and Life Sciences, National Research Council. The National Academy Press, Washington, D.C. (2006).

OECD Guidance Document for the Conduct of Studies of Occupational Exposure to Pesticides During Agricultural Application. OECD Environmental Health and Safety Publications. Series on Testing and Assessment No. 9. OECD/GD(97)148. 1997.

Reinert, J., Nielsen, A., Lunchick, C., Hernandez, O., and Mazzetta, D., "The United States Environmental Protection Agency's Guidelines for Applicator Exposure Monitoring" Toxicology Letters, Volume 33, pp 183-191 (1986).

Reinert, J., and Severn, D., "Dermal Exposure to Pesticides: The Environmental Protection Agency's Point of View", Dermal Exposure Related to Pesticide Use: Discussion of Risk Assessment, ACS Symposium Series 273, pp 357-368, American Chemical Society, Washington, D.C. (1985).

Resnik, D., and Portier, C., "Pesticide Testing on Human Subjects: Weighing Benefits and Risks", Environmental Health Perspectives, Volume 113, Number 7, pp 813-817, (July, 2005).

Ross, J., Chester, G., Driver, J., Lunchick, C., Holden, L., Rosenheck, L., and Barnekow, D. (2007). Comparative Evaluation of Absorbed Dose Estimates Derived from Passive Dosimetry Measurements with Those Derived From Biological Monitoring: Validation Of Exposure Monitoring Methodologies, Accepted for publication in J Expos Sci Environ Epidemiol.

Scientific Advisory Panel (SAP) Meeting 2007. Memorandum from M. Christian to A. Lindsay, Acting Director, EPA Office of Pesticide Programs, "Transmittal of Meeting Minutes of the FIFRA Scientific Advisory Panel Meeting Held January 9 - 12, 2007 on the Review of Worker Exposure Assessment Methods", April 2, 2007.

Stasikowski, Margaret, Director of the Health Effects Division (EPA), memorandum to Mr. Dan Fay on behalf of agricultural chemical producers (March 16, 2001).

U.S. Environmental Protection Agency, "Federal Insecticide, Fungicide and Rodenticide Act (FIFRA); Good Laboratory Practice Standards", 40 CFR 160, U.S. EPA, Federal Register Volume 54, No. 158 pp 34052 (August 17, 1989).

U.S. Environmental Protection Agency, "Pesticide Assessment Guidelines, Subdivision U, Applicator Exposure Monitoring", U.S. EPA, Washington, D.C (1986).

U.S. Environmental Protection Agency, "Occupational and Residential Exposure Test Guidelines, OPPTS 875, U.S. EPA, Washington, D.C (1996).

U.S. Environmental Protection Agency, Health Canada, and California Environmental Protection Agency, "Review of Worker Exposure Assessment Methods", presented jointly to the FIRFA Scientific Advisory Panel on January 9-12, 2007.

Wolf, H. R., Durham, W. F., Armstrong, J. "Exposure of Workers to Pesticides", Arch. Environ Health, Volume 14 pp 622-633 (1967).

Worgan, J. P., and Rozario, S., "Pesticide Exposure Assessment: Past, Present and Future", Methods of Pesticide Exposure Assessment, Editor, Curry, P.B., Plenum Press New York, (1995).

World Health Organization, "Survey of Exposure to Organophosphate Pesticides in Agriculture, Standard protocol, VBC/75.9, WHO, Geneva (1975).

World Health Organization, "Field Surveys of Exposure to Pesticides", Standard Protocol, VBC/82.1, WHO, Geneva (1982).

16 Glossary of Terms

AHETF = Agricultural Handlers Exposure Task Force, L.L.C.

A consortium currently consisting of 19 companies that formed a FIFRA joint data development task force to design and develop a database of exposure measurements for agricultural workers during mixing, loading, and/or application of pesticides. The exposure data will cover all important types of mixing/loading systems, application equipment, and formulations. The results will satisfy FIFRA data requirements and be used by U.S. EPA (and other regulatory agencies) to assess exposure potential and conduct risk assessments for most pesticide products marketed by AHETF members. AHETF was formed in December, 2001.

AaiH, Amount of Active Ingredient Handled

A commonly used normalizing factor for pesticide handling situations that is generally defined as a weight of pesticide active ingredient that gets mixed into a tank or piece of application equipment and/or applied from a piece of application equipment by one worker during his/her entire monitoring period.

Benchmark Objective

This is a specific experimental accuracy goal that is used to establish the number and configuration of sampling units. In the AHETF program sufficient data for each handler scenario will be collected to meet one or more specified minimum or benchmark adequacy requirements. Sample sizes are based on applying accuracy objectives to the random sampling reference model. Because the reference model is an approximation to the actual (random and non-random) selection process, these objectives and the resulting sample sizes are referred to as **benchmarks**. The design benchmarks are not intended to address all possible ways the exposure data could be used. Rather, they are established to ensure that the data will at least be adequate to meet common regulatory needs.

Biomonitoring

Measurement of a pesticide or its metabolite(s) in the body of a pesticide handler and the conversion to an equivalent absorbed dose based on knowledge of metabolism and pharmacokinetics. This generally includes measurement of chemical in blood or urine, but does not include measurement of biological effects such as cholinesterase levels. The result is an estimate of total exposure from the dermal, inhalation, and oral routes combined.

Cluster

A set of monitoring units (or MUs) constructed at the same monitoring site (i.e. in the same local staging area over a short period of time such as one to two weeks). Exposures between MUs from the same cluster tend to be more similar than those between MUs from different clusters.

Distribution of Exposure

This is a statistical description of the frequency with which different exposure levels are expected to occur. Interest in this program is centered on the distribution of generic future handler-day exposures, typically normalized for contact to active ingredient. These are predicted exposures arising from future uses of an arbitrary pesticide under scenario-specific tasks and assuming a specific level of active ingredient contact.

Diversity Selection

The set of procedures used to 'select' (i.e., choose, restrict, or script) experimental monitoring units in a way that captures the diversity of conditions expected to exist in the future handler-day population. If the selected MUs differ with respect to factors that are likely to influence exposure, then the MUs will have exposure diversity as well. Diversity selection can have purposive and/or random elements. If there are no random elements then it is called **purposive diversity selection (PDS)**. **Purpose diversity sampling** refers to the special case of choosing only existing handler-days (i.e., not simulating or scripting handler-days to mimic generic conditions.) See 'selection'.

Engineering Controls

Equipment or equipment modifications that eliminate or reduce exposure to a chemical, such as enclosed cabs, ventilation, or closed transfer systems.

Exposure Monitoring

This is the use of passive dosimetry techniques to measure dermal and inhalation exposure to professional, occupational pesticide handlers as they perform their typical activities. Researchers will use a variety of pesticide residue collection devices (cloth dosimeters, hand washes, face/neck wipes, and sorbent tubes) and determine the quantity of active ingredient on each device by chemical residue analysis.

Future Handler-Day Population

The conceptual set of handler-days resulting from the future scenario-specific use of an arbitrary (and perhaps currently non-existent) pesticide active ingredient assuming some quantifiable amount of active ingredient contact.

Field Study

A convenient grouping of MUs that will be sampled during one 'study' in accordance with GLPs (including just one protocol and one final report). Generally the field portion of a study will be conducted over a short period of time (1 to 2 weeks) with one surrogate chemical.

Generic Exposure, Future Handler-Day Exposure – See Distribution of Exposure.

GLP = Good Laboratory Practice Standards

Federal regulations (40 CFR 160) which prescribe good laboratory practices for conducting studies that support pesticide registrations. The standards address the scientific integrity of study conduct and data collection, including specific requirements for study management, equipment calibration, facilities maintenance, record keeping, reporting, and quality assurance. All AHETF studies are conducted in accordance with these standards, both in the field and in the analytical laboratory.

Handler-Day (or HD)

This is a basic conceptual element of a pesticide handling scenario. Each handler-day corresponds to a particular worker and the scenario-related activities that he/she performs during a single work day.

Handling

Generally refers to mixing, loading, transferring, or applying pesticides. However, handling also includes the following common tasks: handling opened containers; disposing of pesticides or pesticide containers; and cleaning, adjusting, handling, or repairing the parts of mixing, loading, or application equipment that may contain pesticide residues.

HSRB = Human Studies Review Board

A board established by EPA under the authority of 40 CFR §26.1603 for the purpose of advising the Agency on scientific and ethical aspects of proposed and completed research involving intentional exposure of human subjects. Research involving intentional exposure of human subjects is defined at 40 CFR §26.1102(i).

JRC = Joint Regulatory Committee

A committee comprised of representatives of the U.S. EPA, the Canadian Pest Management Regulatory Agency (PMRA), the California Department of Pesticide Regulation (CDPR), and the U.S. Department of Agriculture (USDA). This committee meets on a regular basis with AHETF to clarify regulatory data requirements, provide technical and regulatory input on study design, and review progress.

Local Site Coordinator (or LSC)

An agricultural research contractor utilized by AHETF to provide local support with tasks such as receiving and storing test substances, field fortification samples, and exposure samples. These items require storage in restricted-access facilities that are monitored for temperature and include ambient, refrigerated, and frozen conditions. Local Site Coordinators might also help identify growers and equipment in a location where AHETF is considering conducting a field study.

Meta-Factor

This is a characteristic that represents many other known and/or unknown characteristics. For example, an individual worker is considered a meta-factor because substituting one worker for another can alter a number of physical and behavioral conditions. Other examples of meta-factors are geographic location and normalizing factor.

Monitoring Program

The monitoring program consists of all the GLP studies that will be conducted by AHETF to monitor exposure to agricultural pesticide handlers and that will be used to develop a generic database to support pesticide registrations. The planned monitoring program will cover many handling situations, or scenarios.

Monitoring Unit (or MU)

The basic tool used to predict future exposures. An MU is a set of scenario-specific handler-day conditions that have been experimentally selected (i.e., chosen, restricted, or scripted) to represent future handler-day conditions. Each MU is a collection of all exposure monitoring activities pertaining to a single worker for a time period that represents a typical workday, including the exposure measurements for the worker involved. (An MU was formerly called a 'replicate'.) A specified number of monitoring units will be selected for each scenario to adequately characterize the expected diversity of exposure. The term **monitoring event** (used by AEATF II, another pesticide exposure task force) is synonymous with AHETF's monitoring unit.

Monitoring Site

This refers to a well-defined location that is used as a staging area for monitoring activities. The definition of monitoring 'site' also includes a general range of dates for which monitoring could possibly occur. For example, a monitoring site might be defined to be three specific counties in Iowa and the first two weeks in June. The temporal extent of a monitoring site merely reflects the general time-frame within which monitoring activities are planned, not specific dates. In principle, monitoring site could include an entire application season. In practice however, monitoring activities within a site must be restricted to much shorter time frames. The non-specific term "location" is sometimes used synonymously with the local staging area defined by a monitoring site, but is more often used to define a larger area, such as a state, from which a monitoring site could be selected.

Normalized Exposure, Normalizing Factor

Normalized (or unit) exposure refers to daily worker exposure expressed relative to (i.e., divided by) a useful measure of potential active ingredient contact. This contact measure is called the normalizing factor (or NF). The most common normalizing factor is the amount of active ingredient handled during the workday. Multiplication of normalized exposure by any value of NF yields a predicted generic exposure from handling an arbitrary active ingredient at that NF level.

Passive Dosimetry

This is a set of techniques for measuring pesticide exposure to humans that do not involve invasive collection techniques such as collecting urine or blood. In particular, AHETF studies involve whole-body garments that serve to collect potential dermal residues, hand washes to collect hand residues, face/neck wipes to collect residues on the face and neck areas, and sorbent tubes which collect air in the breathing zone of a worker. Occasionally, cloth dosimeters will be used to measure exposure to the feet or to the head area with and without headgear.

Population, Target Population – See Selection and Future Handler-Day Population.

Potential Active Ingredient Contact (or PaiC)

The amount of active ingredient that a worker is expected to come into contact with during a workday. PaiC is a scenario-specific conceptual quantity that can rarely be measured directly. However, knowledge of scenario-related tasks often suggests a practical measure of relative ai contact that can be constructed from MU-derived quantities such as active ingredient handled, number of containers emptied, number of loads processed, etc. Such experimental approximations to PaiC are used as **normalizing factors**.

PPE = Personal Protective Equipment

Devices and apparel that are worn to protect the body from contact with pesticides or pesticide residues, including but not limited to coveralls, chemical-resistant suits, chemical-resistant gloves, chemical-resistant footwear, respiratory protection devices, chemical-resistant aprons, chemical-resistant headgear, and protective eyewear (40 CFR 170.240).

Probability Sampling – See Selection.

Purposive Diversity Sampling – See Diversity Selection .

Purposive Diversity Selection (PDS) – See Diversity Selection.

Purposive Selection – See Selection.

Random Selection – See Selection.

Random Sampling – See Selection.

Reference Random Sampling Model

A theoretical statistical construct used to provide an artificial random sampling context for a set of MUs. The reference model captures as much as is practical of the actual (random and non-random) sampling structure. Since all AHETF scenarios select MUs in two successive stages, the random sampling reference model has a two-stage sampling structure. Conclusions drawn from the reference model are applied, by analogy, to the set of MUs. In the determination of sample size and the statistical summary analysis of MU data, uncertainties for distributional parameter estimates will be based on this reference random sampling model.

Sampling – See Selection.

Scenario, Handling Scenario

This is a specific pesticide handling situation that will be represented by data with defined common properties; generally a combination of a work task(s), pesticide formulation, equipment, engineering controls, and work practices. For example, a scenario of interest is 'mixing/loading dry flowable pesticides using open pouring equipment and techniques'.

Scripting, Scripted Study

Scripting is the partial control of the conditions associated with a particular MU. Workers are sometimes asked by AHETF to conduct their work activities under a set of scripted conditions very similar, but maybe not identical, to those they experience in their normal work activities. Scripted studies must be approved for scientific and ethical validity by EPA and HSRB.

Selection

A general term used by AHETF to refer to 'obtaining handler-day conditions by any means'. It applies to any form of choosing, restricting, scripting, or otherwise simulating conditions. Selection can be based on large items (e.g. selection of geographic locations) or on smaller items (e.g. selection of workers or date of application). Selection of items can be **random** or **purposive** or contain elements of both. Selection is referred to as **sampling** when the items can only be chosen from a set of existing items called a **population** (i.e., there are no scripted, restricted, or simulated conditions). If sampling is based on a formal random process in which all items in a target population have a known, positive chance of being selected, then it is called **probability sampling**. (Although ambiguous and sometimes misleading, probability sampling is often colloquially referred to as **'random' sampling**.) When sample sizes are large, statistical theory can be used to assess how well a probability sample represents the population. It should be emphasized, however, that a probability sample of handler-day components (e.g. workers) does not result in a probability sample of handler-days. For example, if a set of MUs is constructed from a random sample of workers, but other MU conditions are scripted, restricted, or even based on other randomly-sampled components, then each MU is still a **synthetic handler-day**. A set of synthetic handler-days, by definition, can never be a sample from an existing population. At best, each MU can only mimic those generic handler-days that are predicted to occur.

Stratified Diversity Selection

A practical form of diversity selection that first partitions all possible selection units into non-overlapping groups called strata. Then one unit is selected, randomly or purposively, from each stratum. For example, the practical range of AaiH is commonly split into 5 equally-spaced strata (that collectively cover the entire range) so one MU from a typical 5-MU cluster can be associated with each stratum. Stratification can also be applied to large geographic units, for

example by sub setting the country into all its states or growing regions, or sub setting a state into all its counties.

Surrogate Chemical

A pesticide active ingredient found in a test material and collected by passive dosimetry matrices during sampling of an MU. AHETF develops validated analytical methods for each surrogate chemical and each exposure matrix so residues collected can be determined. AHETF chooses surrogates which have relatively low volatility and are commercially available in suitable formulations and packaging. Since exposure to handlers is a generic function, exposure measurements from these chemicals are suitable for estimating exposure to other pesticide active ingredients.

Synthetic Handler-Day – See Selection.

Two-Stage Selection of MUs

For all scenarios, selection of monitoring units occurs in two successive stages. In the first stage, monitoring sites are chosen and in the second stage, monitoring units are selected.

WPS = Worker Protection Standard

A Federal regulation (40 CFR Part 170) which contains a standard designed to reduce the risks of illness or injury resulting from workers' and handlers' occupational exposures to pesticides used in the production of agricultural plants on farms, nurseries, greenhouses, and forests, and also from the accidental exposure of workers and other persons to such pesticides. It requires workplace practices for employers and workers that are designed to reduce or eliminate exposure to pesticides, and it establishes procedures for responding to exposure-related emergencies.

Appendix A. AHETF Acceptance Criteria for Existing Studies

Appendix A
AHETF Acceptance Criteria for Existing Studies

General Study Design Criteria

1. All monitored activities and equipment must be well-described and representative of typical agricultural mixing/loading and application practices.
2. It must be clear that the individuals monitored are normally employed in the mixing/loading and/or application of agricultural products.
3. Appropriate supporting information such as the formulation type, mixing and application method, application rate, duration of the work cycle, amount of AI handled/replicate, etc. must be available.
4. The use of personal protective equipment (PPE) is acceptable.
5. The study location and environmental/weather conditions during the monitoring period must be available.
6. All aspects of the study must have been conducted as per GLP regulators or be free of any significant GLP deviations or shortcomings if not conducted under GLP.

Exposure Monitoring Criteria

Field Aspects

1. Field recoveries must have been collected on a site-specific basis for time periods and environmental conditions representative of those during collection of field activity exposure samples.
2. Field fortification data must include at least triplicate samples at 2 rates and duplicate samples of controls.
3. Dermal exposure monitoring techniques must be specified and must include one of the following approaches.
    a. whole-body dosimeters inside of clothing plus hand (gloves cannot substitute for hand exposure) and head/face exposure determinations,
    b. a minimum of 10 patch dosimeters attached inside normal work clothing to the chest, back, both upper arms, both lower arms, both upper legs, both lower legs, plus hand (gloves cannot substitute for hand exposure) and head/face exposure determinations,
    c. combination of patches and clothing that are representative of the whole body, including hand and head/face exposure determinations.

4. Inhalation exposure – Inhalation data are not required. If data were collected, inhalation exposure must have been measured by sampling the person's breathing zone.
5. Exposure monitoring duration – The monitoring period should be at least half of a normal work period duration or half the default acreage.
6. If the exposure monitoring duration does not meet the requirement of item number 5, then the number of non-detects/less than LOQ values should account for less than 20% of the actual dermal exposure.
7. If the exposure monitoring duration and number of non-detects/less than LOQ values do not meet the criteria in items 5 and 6, then the LOQ must be no more than/1.0 ng/cm2 for dermal exposure and no more than 100 ppb for hand exposure.

Analytical Aspects - QA/QC

1. Analytical methods must have been validated for each analyte and substrate by the performing laboratory including establishment of the method's working concentration range to cover values anticipated in the field studies, determination of detector response over a reasonable standard concentration range, and determination of the accuracy and precision of the method within the analytical environment
2. The study must include both field fortification samples and concurrent laboratory spikes.
3. The average recoveries of lab spikes must be between 70-120 percent and the precision value (coefficient of variation) must be less than or equal to 20 percent.
4. Recovery of field fortification samples must be 50-120% with a C.V. ≤25%.
5. Exposure samples must have been analyzed in such a manner that the stability of each analyte in each substrate was assessed for the entire time period from collection to analysis.

Primary Review Process (216 Studies Reviewed)

1. The primary review is conducted by the study submitter (or a designated representative) and provided to AHETF along with a complete copy of the study report at the time the study is submitted to the task force for consideration.
2. The review by the submitter will include all applicable studies, including those that are presently in PHED.
3. The raw data for a study must be available, if requested, in order to be considered for purchase.
4. A list of potential studies must be submitted to the AHETF by March 1 and all primary review forms and reports must be submitted by June 1, 2002.

5. The purpose of the primary review will be to eliminate the submission of studies that clearly do not meet the selection criteria, and to serve as a check on the availability and submission of supporting information.
6. An Excel spreadsheet will be provided to the submitter for use in summarizing the study details and data.
7. A confidentiality agreement will be provided to the submitter to protect the confidential nature of their data.

Secondary Review Process (74 Studies Reviewed)

8. The secondary review will be conducted by a qualified AHETF contractor hired and trained for this purpose.
9. The purpose of the secondary review will be to verify the accuracy of the primary review and, where necessary, provide a more detailed discussion summarizing each specific area of the criteria, including whether each criterion was met and possible deficiencies in the study data.

Final Review Process (6 Studies Acquired)

10. The final review and decision on whether a study is accepted for purchase will be made by the Joint Regulatory/Task Force Technical Committee consisting of representatives from the AHETF, USEPA, PMRA, CDPR, and USDA.
11. The secondary review by the contractor and the regulatory reviews of contributed studies will be made available to the Joint Regulatory/Task Force Technical Committee and will serve as the basis for the final review. The secondary review will be evaluated and a determination made as to whether the study or any of the data could be used in the AHETF database.
12. Studies or portions of studies selected after final review will then be considered for purchase by the AHETF for inclusion in the task force database.
13. Reports for studies that are rejected will be returned to the submitter. Reports of studies that are purchased by AHETF will be placed in the AHETF archives.

Appendix B. Using Monitoring Units to Characterize Generic Exposure

Appendix B
Using Monitoring Units to Characterize Generic Exposure

This appendix describes the concepts underlying the use of experimental monitoring to characterize the future exposure of workers using pesticides under a specific set of handling tasks for a typical workday. Each well-defined group of handling tasks or situations is called a **handling scenario**. The exposure data collected from all scenarios in the AHETF Exposure Monitoring Program will be used to construct a **generic exposure** database. These exposure data (and the database) are termed **generic** because they will not be used to describe past exposures from a particular quantity of chemical. Rather, they will be used to predict the diversity of exposures expected from the future use of an arbitrary (and perhaps currently non-existent) active ingredient. The resulting AHED® generic database will be used by EPA and other regulatory agencies to regulate future uses of pesticides.

B1. Future Exposure and Monitoring Units

A pesticide-handling scenario is a well-defined worker exposure situation, usually characterized by specific agricultural chemical handling tasks and equipment. For the purposes of the AHETF Monitoring Program, the basic element of a scenario is considered to be the **handler-day** (or HD). Each handler-day corresponds to a particular worker and the scenario-related activities that he performs during a single workday.

Implicitly associated with each HD is a complex set of conditions denoted simply by **C**. The practically infinite number of components of **C** includes, but is not limited to, worker behaviors, active ingredient used, amount of chemical contacted, tasks performed, and numerous environmental factors. Some, but certainly not all, of these conditions can actually be observed and measured.

Each particular set of HD conditions results in a particular worker exposure, E=E(**C**). In principle, although always subject to some measurement error, handler-day exposures (e.g., dermal deposition or amount inhaled) can be obtained by actual monitoring.

Regulatory interest for each pesticide handling scenario is focused on predicting occupational exposure resulting from a specific set of **generic future handler-day conditions**. In particular, it is desired to characterize exposures resulting from the future use of an arbitrary (and perhaps currently non-existent) pesticide active ingredient assuming some arbitrary, but quantifiable, amount of active ingredient contact.

A **monitoring unit** (or MU) is the experimental tool used by the AHETF Monitoring Program to predict exposures. An MU is a set of scenario-specific handler-day conditions that have been experimentally '**selected**' (i.e., chosen, scripted, or otherwise constructed) to represent expected future HD conditions. Each MU is experimentally monitored to yield a set of exposure measurements. Thus, each MU provides a measurement of the actual exposure resulting from the selected HD conditions. The MU will also provide a predicted future exposure if the handling conditions used for the MU are similar to future HD conditions.

B2. The Generic Active Ingredient Principle

One obvious handler-day condition is the identity of the **active ingredient** (ai) to be used in the scenario task(s). Every experimental monitoring unit must use at least one active ingredient. It might seem, therefore, that prediction of future exposure to a particular ai would require that MUs used only that active ingredient. If this were the case, then a **generic exposure database** for arbitrary active ingredients would not be feasible.

Fortunately, exposure is not always chemical specific. For compounds with low volatility (which include all those considered by the AHETF), a generally accepted **generic principle** is:

> ***If all other conditions are the same, the magnitude of exposure does not depend on the particular active ingredient used.***

More formally, if A_1 and A_2 are any two active ingredients and if the exposure resulting from any active ingredient X under conditions **C** is denoted by E(X, **C**), then the generic principle is simply:

(1) $$E(A_1, \mathbf{C}) = E(A_2, \mathbf{C}) = E(\mathbf{C})$$

The practical importance of the generic principle is that it permits an MU based on one **surrogate chemical** to be used to predict HD exposure to other active ingredients under the same (or very similar) set of handling conditions.

B3. Normalized Exposure

For prediction purposes, it is useful to express handler-day exposure relative to the value of a **normalizing factor** (or NF). If H is the value of the normalizing factor and **C** represents all other HD conditions then:

(2) $$nE(\mathbf{C}) = E(\mathbf{C})/H$$

is called **normalized exposure**. The quantity nE is also often referred to as **unit exposure** because it is viewed as exposure 'per unit' of the normalizing factor.

In the AHETF Monitoring Program the normalizing factor is always an experimentally measureable quantity that is expected to be proportional to the potential contact the worker has with active ingredient. **Potential ai contact** (or PaiC) is defined as the amount of active ingredient that a worker is expected to encounter during a workday. Because each AHETF normalizing factor is only expected to be proportional to PaiC, it should be considered a **relative measure** of active ingredient contact.

It is generally assumed that, under identical conditions, exposure is proportional to potential ai contact. That is, if P_1 and P_2 are different amounts of PaiC, and **C** represents other HD conditions, then the **proportionality principle** states that:

(3) $$E(P_2, \mathbf{C}) / E(P_1, \mathbf{C}) = P_2 / P_1$$

Exposure may not be directly proportional to contact in extreme situations (e.g., skin saturation). However, such a relationship is expected to hold for levels of active ingredient contact that occur in practice. If the normalizing factor is expected to be proportional to PaiC then the proportionality principle relationship holds for the NF as well since:

(4) $$H_2 / H_1 = (k \cdot P_2)/(k \cdot P_1) = P_2 / P_1$$

Thus, under the same conditions, **C**, exposure for any value, H_2, of the normalizing factor can be obtained from an MU based on a different value of NF, H_1 say, since:

(5) $$E(H_2, \mathbf{C}) = (H_2/H_1) \cdot E(H_1, \mathbf{C}) = H_2 \cdot \{ E(H_1, \mathbf{C})/H_1 \} = H_2 \cdot nE(\mathbf{C})$$

For most pesticide scenarios, a reasonable normalizing factor is just an estimate of the **amount of active ingredient 'handled'** (or AaiH) by a worker during the workday. This simply means the amount of active ingredient that gets mixed into a tank or piece of application equipment and/or applied from a piece of application equipment. However, for some scenarios (e.g., closed-system mixing and loading) a worker might actually process (i.e., 'handle') a large amount of active ingredient, but may have the opportunity to contact only a small fraction of this amount. In such cases, there may be other measures of PaiC that are more appropriate than AaiH.

It is also important to note that the term normalized (or unit) exposure is not always defined as a relative measure of ai contact. A familiar example occurs in studies of exposure to agricultural reentry workers. Here, exposure, E, is often normalized by duration of the reentry activities to give exposure per hour worked. The normalizing factor 'hours worked' is expected to be proportional to amount of contact with pesticide-treated foliage. However, unless the concentration of ai on treated foliage is always constant, 'hours worked' is not expected to be proportional to the amount of active ingredient contact. On the other hand, the quantity:

(6) $$D = (\text{hours worked}) \times (\text{dislodgeable foliar ai residue})$$

is expected to be proportional to the amount of ai contacted by the reentry worker. In agricultural reentry studies, the quantity E/D is often called the **transfer coefficient** (or just TC) and corresponds to normalized exposure in the sense used by the AHETF Monitoring Program.

B4. Using MUs to Predict Future Exposure

The generic and proportionality principles together imply that, under the same conditions, the normalized exposure from any MU can be used to obtain a **predicted future HD exposure**, pE, for any arbitrary chemical, X, at any arbitrary level of the normalizing factor, H_X, simply as:

(7) $$pE(H_X, \mathbf{C}) = H_X \cdot nE(\mathbf{C})$$

Although this relationship permits a single MU to predict exposure for a broader range of future handler-days, it cannot predict all future exposures. Even when H_X is specified, the number of possible 'other conditions' (i.e., different **C**s**)**, is extremely large. (Although some of these **C**s are more likely than others to occur in the future HD population.) Because each MU is expensive, it will only be possible in practice to construct MUs having a limited set of **C**s. This set of N MUs will then be used to obtain a set of N predicted future exposures using (7):

$$\begin{gathered} pE(H_X, \mathbf{C}_1) = H_X \cdot nE(\mathbf{C}_1) \\ pE(H_X, \mathbf{C}_2) = H_X \cdot nE(\mathbf{C}_2) \\ pE(H_X, \mathbf{C}_3) = H_X \cdot nE(\mathbf{C}_3) \\ \vdots \\ pE(H_X, \mathbf{C}_N) = H_X \cdot nE(\mathbf{C}_N) \end{gathered}$$

Obviously, a set of only N predicted exposures cannot cover every possible future HD condition. Nor is it reasonable to expect that a small set of N MUs will provide sufficient experimental material to develop statistical models for exposure as a function of **C**. In fact, only some, but by no means all, of the components of each **C** can be controlled or measured when constructing MUs. These unknown components might have the biggest impact on exposure.

Nevertheless, this set of pEs will need to be sufficient to allow regulatory issues to be addressed in a practical manner. If some components of **C** that can be controlled are chosen appropriately, then a useful set of MUs can still be constructed. In this case the resulting set of pEs will be used to characterize the diversity of future HD exposures.

B5. The Future Exposure Distribution

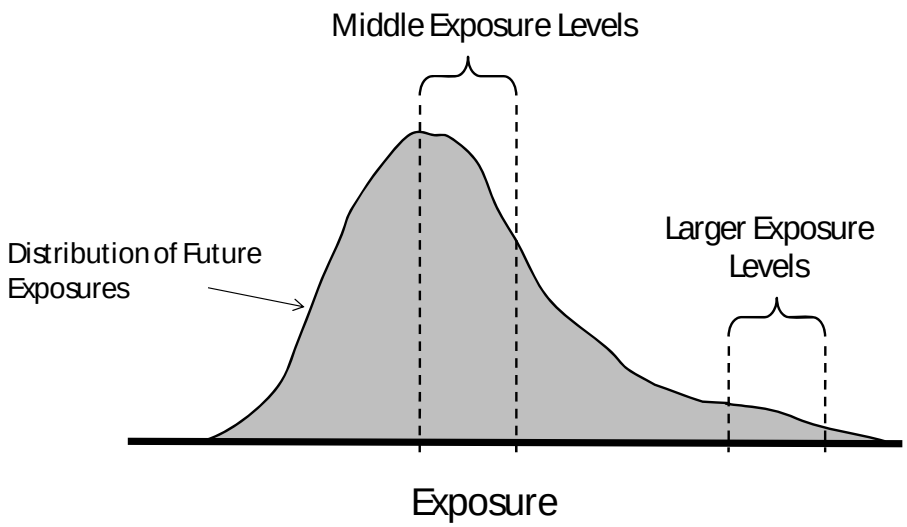
An exposure **distribution** provides a natural aggregate description of future handler-day exposures for a scenario. It is the most common way to think about the set of exposures resulting from all possible future HD conditions. The future HD distribution (Figure B1) describes the likely exposure that would result if one were to randomly pick a future HD from among those using ai X when the level of the normalizing factor is H_X. It is important to stress that these handler-days do not yet exist. They will only exist when and if X is handled at NF level H_X by workers performing scenario-related tasks. Thus, the distribution of future HD exposure is only a potential distribution.

The complete exposure distribution is rarely considered. Regulatory interest is most often focused on two general aspects of this distribution:

- The **middle values** such as the mean or the median. These exposure values tend to characterize average or 'typical' exposure levels.
- The **larger values** of exposure possible, such as the 95th percentile of the distribution. This aspect better characterizes the extreme, one-time, worker exposures.

Obviously the exposures can vary proportionally with the value of the normalizing factor, H. Thus, there is actually a series of potential future exposure distributions, one for each possible value of H_X. Since any predicted exposure can be computed from the normalized exposure using (7), it is simpler to focus only on the future distribution of normalized exposure.

Figure B1: The distribution of future handler-day exposures



B6. The Set of MUs as a Pseudo-Random Sample

The set of predicted exposures obtained from a constructed set of MUs needs to adequately characterize the middle and larger values of the potential exposure distribution resulting from future HD conditions. Apart from measurement errors, each MU-derived predicted exposure, $pE(H_X, \mathbf{C})$, is expected to be identical with a future HD exposure having the same conditions, $E(H_X, \mathbf{C})$. Thus, a set of N predicted exposures is also a set of N future HD exposures. Thus, as an approximation for statistical purposes, one might naively treat the set of pEs as if it were a simple random sample from the future exposure distribution. This cannot be strictly true since exposure is a function of the HD conditions, **C**, and the likelihood of the various **C**'s in a future HD population (for an arbitrary X and H_X) are not known in advance. Therefore the **C**'s for the

MUs can never be randomly chosen with probability proportional to the, as yet unknown, future population frequencies.

However, some random sampling interpretation might still be a convenient and reasonable approximate model for how the set of predicted exposures from the MUs could statistically relate to the distribution of future exposures for an arbitrary ai and H_X. Confidence in this approximation is increased by using a nested reference random sampling model rather than assuming just simple random sampling (see Appendix C). In addition, **diversity selection** (using both purposive and random components) is used whenever possible (see B8). This increases the likelihood that the range of conditions expected to impact exposure in the future HD population is reflected in the 'pseudo-sample' of MUs as well.

B7. The Two-Stage MU Selection Process

Obtaining potential workers and scenario-specific handling-day conditions needed to create monitoring units is a complicated process. This process is referred to as MU (or sometimes HD) **selection** as opposed to MU **sampling**. Selection refers to any method that establishes the handling-day conditions for a particular monitoring unit. Selection may involve choosing an existing handler-day from those already scheduled to occur. However, it may also involve perturbing, limiting, or 'scripting' such HD conditions to insure that the complete set of MUs better reflects the diversity expected of future (and generic) handler-days. Thus, MUs are not technically sampled from an existing population of HDs in the sense of surveys. More correctly, MUs should be viewed as a set of synthetic handler-day conditions that are used as a starting point from which to generically predict future HDs for arbitrary pesticides. Although originating from the survey sampling term 'sampling unit', the term 'monitoring unit' is used by the AHETF to emphasize the difference between conventional sampling and MU selection. (Note: the AEATF II monitoring program uses the term **monitoring event** in the same context. Thus, monitoring event and monitoring unit are synonymous.)

Of necessity, the specific MU selection process used will vary from scenario to scenario and can involve both **purposive** and **random** selection elements. However, the final set of MUs for all AHETF scenarios can be envisioned as the result of **two successive stages** of MU selection. The first stage consists of selecting specific geographic locations (plus a range of dates for monitoring at each location) that serve as areas from which to stage monitoring operations. Each particular combination of local staging area and range of potential monitoring dates is termed a **monitoring site**. Monitoring sites are the **first-stage selection units**.

The second stage consists of selecting agricultural workers and other handling conditions within each site that are then used to construct the MUs. In general, N_c sites are selected at the first stage and N_m monitoring units will be obtained within a site at the second stage. Consequently, MUs are considered the **second-stage selection units**. N_m will not necessarily be the same at all sites. When N_m is greater than one, the set of MUs at the same site is called an **MU cluster**. In general, the handling conditions (i.e., the different **C**) for MUs in the same cluster are expected to be more similar than those in different clusters. Thus, there is expected to be some degree of within-cluster correlation of exposure. Theoretically, such within-cluster correlation favors the

use of only a single MU per site. However, there are often substantial overhead costs per site that make multi-MU sites more efficient (see Appendix C).

B8. Diversity Selection

Diversity selection is an attempt to make a small set of MUs (and resulting pEs) more useful for regulatory purposes when it is treated as an approximate random sample. Often some factors that are likely to influence exposure are known or can at least be reasonably hypothesized. Diversity selection results from any procedure that improves the chance that different MUs differ with respect to such factors. It is an attempt to obtain, as much as is practical for small sample sizes, a diversity of conditions that are expected to influence exposure, either directly or indirectly. With a small set of MUs, it is more practical to construct MUs that differ than to reproduce the (usually unknown) frequencies of future HD conditions.

In the AHETF Monitoring Program, the term **diversity selection** is preferred to the term **diversity sampling**. This is to emphasize the fact that the future HD conditions used for MUs are selected from either existing or from synthesized conditions (or from both) and used to predict future HD conditions generically. This selection of HD conditions can employ **purposive** and/or **random** methods. **Random diversity selection** avoids the appearance of intentional bias that can result when researchers choose some HD conditions and exclude others. When choices of conditions are equivalent and easily listed, this is a natural approach. On the other hand, **purposive diversity selection** can be more efficient and cost effective whenever the possible choices of HD conditions are non-equivalent in terms of cost or convenience. Neither form of selection provides justification for the use of statistical sampling theory. For this to be the case, all stages in the sampling would need to be random, representative, and conform to a rigorous statistical sampling protocol. In addition, all MUs would need to be completely observational. That is, MUs constructed from synthetic (e.g., scripted) components can never be a sampled element from an existing handler-day population.

Diversity selection attempts to create an MU 'sample' that contains as many different expected HD conditions as possible. If these diversifying conditions are associated with exposure, then a diversity sample will tend to be more variable with respect to exposure than would a same-sized representative sample. In effect it will be analogous to representative (but not simple random) sampling from a distribution that is more diverse than the actual future one (Figure B2). As a result, a diversity selection sample should tend to have more extreme exposures (both higher and lower) and fewer exposures 'in the middle'. Thus, a diversity selection sample will tend to predict central tendencies of a future generic exposure distribution better than it will predict either upper or lower percentiles. It will tend to under-predict lower percentiles and over-predict upper percentiles. This general aspect of diversity selection is easily illustrated with the simple normal distribution example shown in Figure B3. A diversity sample of size N=6 can be taken from a normal distribution by first partitioning the range between ±3 standard deviations around the mean into six equal intervals called **strata**. Then, one value is randomly selected (with uniform probability) from each stratum. Compared to the normal distribution, this method of sampling under-represents the values in middle of the distribution and over-represents those at the extremes. In fact, it is effectively a representative sample from the 'more diverse' uniform

distribution. In addition, taking a single value from each stratum gives a sample that is more 'spread out' than would be a same-size random sample from either the normal or the uniform distribution. The impact of this diversity selection on statistics calculated from the six values is illustrated in Figure B4. The increased diversity in the sample causes all statistics calculated from the sample to be less variable over repeated sampling. Sample statistics for central tendencies (e.g., means, medians) will tend to be relatively close to their true values. However, other percentiles will be biased towards more extreme values. That is, upper percentiles are overestimated and lower percentiles are underestimated. Because it is unlikely that such extreme sample diversity can be induced in practice, the actual decreased variation and percentile biases will be less striking than seen in this example. Regardless, for regulatory purposes the important aspects of the distribution of exposures are central tendencies and upper percentiles. In addition, overestimation of these characteristics is less of a regulatory problem than underestimation since it usually results in overprotective restrictions on pesticide use. Thus, the fact that diversity selected MUs are biased towards greater exposure variation is of little practical concern for the potential users of AHED®.

Figure B2: Diversity selection tends to make the distribution of future handler-day exposure appear more diverse than it is.

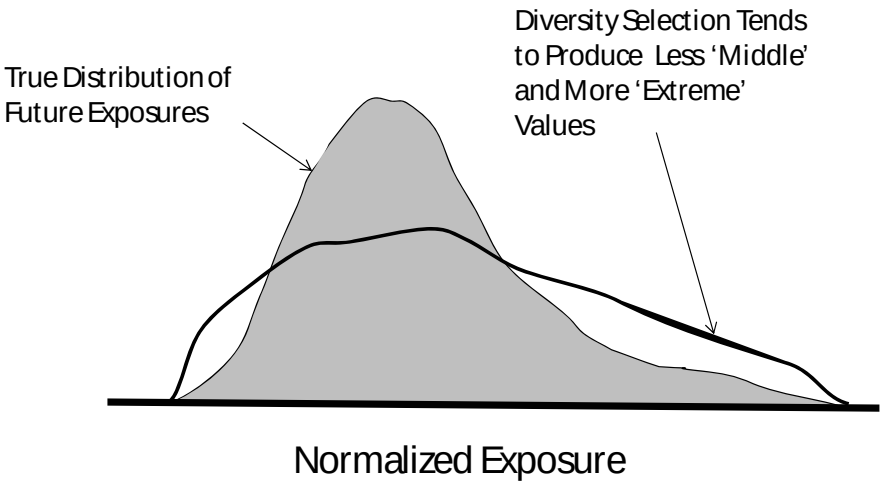


Figure B3: Stratified diversity selection of n=6 values from a normal distribution.

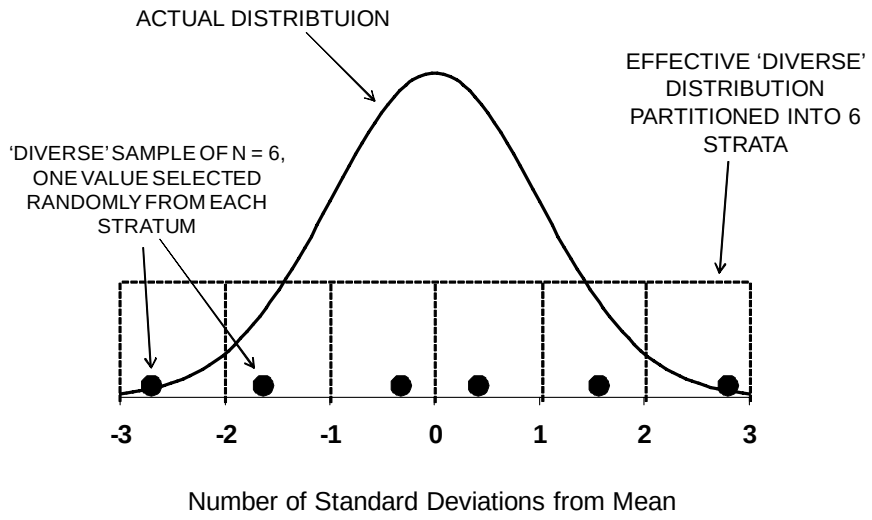
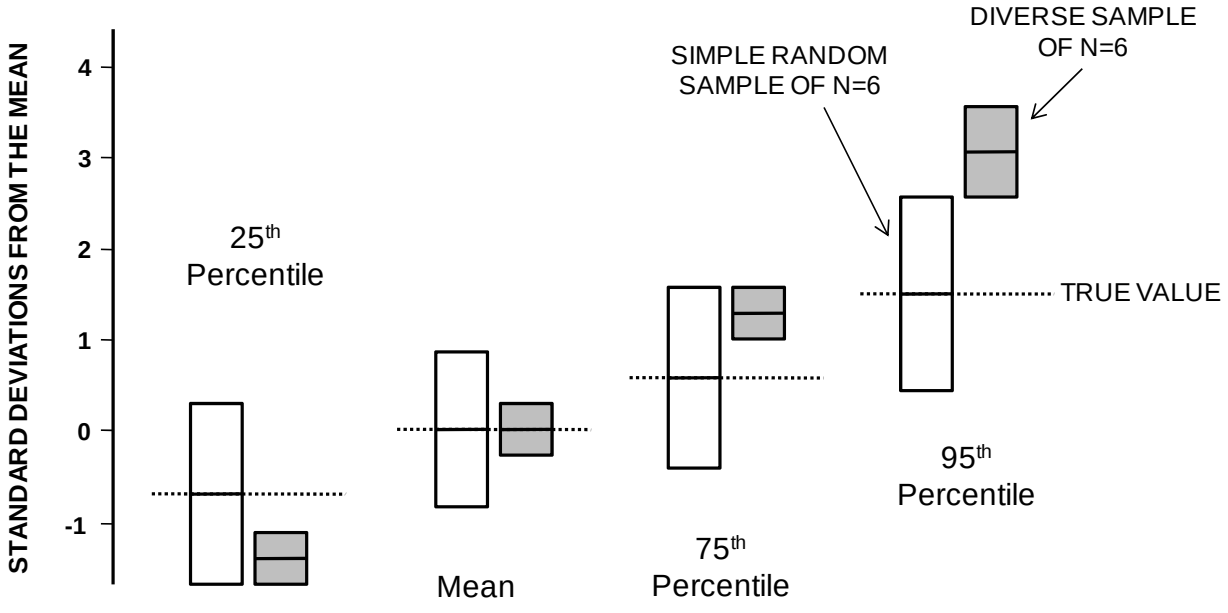


Figure B4: The repeated-sampling bias and variability of the sample mean and selected percentiles from the diverse sample of n=6 shown in Figure 3 compared with a simple normal random sample of the same size. Each probability box contains 95% of the possible values (i.e. the 2.5th to the 97.5th percentile). The middle line in each box denotes the median. The diverse sample gives more consistent estimates than a simple random sample but overestimates upper percentiles (and underestimates lower percentiles).



B8.1. Diversity Selection Methods

Although diversity selection is simple in concept, practical implementation is often complex and usually scenario specific. As described previously, the objective of diversity selection is to obtain a diverse set of handler-day conditions from among those conditions possible when an arbitrary pesticide is used in future scenario-related tasks. These selected HD conditions are then used to construct monitoring units. Diversity selection is done independently at both stages of selection. Thus, a diverse set of monitoring sites is selected followed by a diverse set of MU conditions within each monitoring site.

To obtain diversity in exposure, diversity selection should always be with respect to characteristics (i.e. particular components of **C**) that are expected to impact exposure. However, it would be impossible to obtain diversity with respect to every factor in a small sample. Consider diversifying a set of n=3 selection units (i.e. either monitoring sites or MUs) with respect to a single factor, A. The three values of A could be A_1, A_2, and A_3, say. Similarly, the units could also differ on factor B with values B_1, B_2, and B_3. At first glance it may appear trivial to diversify the three units with respect to both factors A and B jointly. For example, the three units could have values A_1B_1, A_2B_2, and A_3B_3, respectively. Certainly these units would differ with respect to both A and B. But because there are $3\times3=9$ possible combinations of A and B, the three units could also have different factor values. In fact, as shown in Figure B5, there are six 'diverse' **configurations** possible. Only one of these configurations can be selected. Some of these configurations might produce greater exposure diversity than others. In addition, some combinations of factors might not be expected to occur in a future handler-day population. The complexity of this problem increases exponentially with the number of factors considered. Therefore, not every characteristic that may impact exposure can be, or even should be, considered in diversity selection. The number of possible combinations of factors that may impact exposure will always greatly exceed the number of planned MUs.

Figure B5: Six possible 'diverse' configurations of n=3 selection units with respect to three levels of factor A and three levels of factor B.

Configuration	Unit 1	Unit 2	Unit 3
1	A_1B_1	A_2B_2	A_3B_3
2	A_1B_1	A_2B_3	A_3B_2
3	A_1B_2	A_2B_1	A_3B_3
4	A_1B_2	A_2B_3	A_3B_1
5	A_1B_3	A_2B_1	A_3B_2
6	A_1B_3	A_2B_2	A_3B_1

Whenever possible, therefore, only a few characteristics, preferably **meta-factors**, can be used effectively in diversity selection. Meta-factors are characteristics that directly or indirectly influence a number of other characteristics. For example, a worker is a meta-factor because

substituting one worker for another alters a number of factors (e.g., behavior, physical appearance, stamina, and equipment available) that might affect exposure. Other common meta-factors are geographic location and time-of-year. When MUs are different with respect to a meta-factor they will likely be different with respect to most of the factors determined by the meta-factor. More importantly, these sub-factors will tend to be naturally, not artificially, correlated. That is, varying a meta-factor is unlikely to result in combinations of MU conditions that are impossible or rare in future handler-days.

There are a number of ways to achieve diversity among MU handler-day conditions. The most straightforward approach is to purposively select MUs that appear to be sufficiently different with respect to the characteristic(s) of interest. Because purposive selection is, by definition, subjective, documentation for this process should include the characteristics considered and how much these characteristics differ among the MUs. Although flexible, and likely to achieve a set of MUs with a great amount of diversity, this approach may be difficult to reproduce without careful documentation.

More formal approaches to diversity selection are also possible. A general, albeit quite sophisticated, approach is to define **diversity scores** for each possible configuration of selected units based on the characteristics of interest. A total configuration score might be defined as a function of the dissimilarity between possible pairs of selection units. Then, one simply selects (or synthetically constructs) those configurations that give the greatest diversity score. If multiple configurations have the same maximum score, a random selection among these is possible. While achieving diversity in an objective and reproducible manner, this approach, in general, is quite complex and difficult to implement.

A formal approach that is both common and simple to implement is **stratified diversity selection** (Figure B6). In this approach available selection units (e.g., monitoring sites or monitoring units) are partitioned into **strata** based on characteristics likely to impact exposure. Each potential selection unit must belong to one and only one stratum. The number of strata must be at least as large as the number of units that will be selected. For example, if there are five units to be selected, then there should be at least five strata. Diversity could be achieved by selecting (purposively or randomly) no more than one unit from each stratum. 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).

Stratified diversity selection bears some resemblance to the method of stratification used in population sampling. Unlike the case with stratified sampling methods, however, stratified diversity selection increases dissimilarity among the final set of units by choosing only one from each stratum. 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.

Figure B6: Illustration of stratified diversity selection when the number of strata is equal to the number of units (monitoring sites or MUs) to be selected. One unit is chosen per stratum (randomly or purposively).

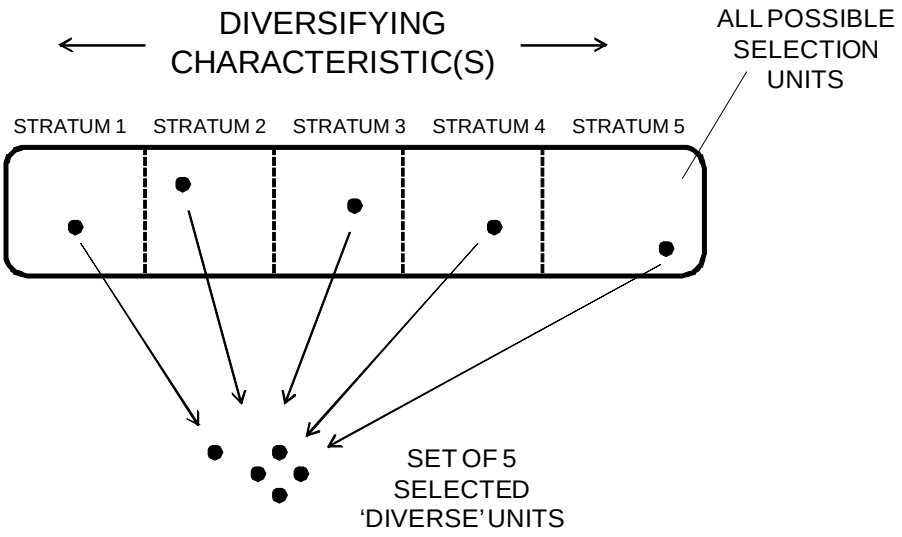
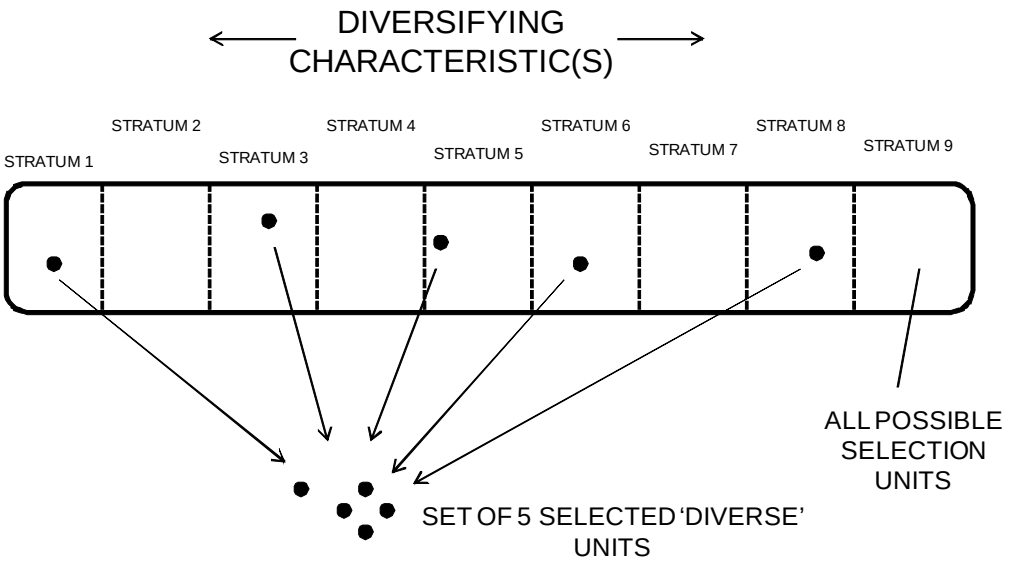


Figure B7: Illustration of stratified diversity selection when the number of strata is greater than the number of units (monitoring sites or MUs) to be selected. A subset of the available strata is first selected (randomly or purposively). Then one unit is chosen per stratum (randomly or purposively).



B8.2. Diversity Selection of First-Stage Units

Monitoring sites are the selection units for the first stage of MU selection. As defined previously, monitoring site is considered a particular combination of experimental staging area and timeframe for monitoring activities within that area. AHETF and others have commonly found significant study-to-study differences in worker pesticide exposure. In most cases each such study was limited to a particular geographic location and date range. Because each monitoring site in the AHETF Monitoring Program will be similar in scope to such a past study, monitoring sites are also expected to differ in average exposure. If all MUs for a scenario were obtained from the same monitoring site, the normalized exposure could be non-representative of the scenario and the exposure variation would be too small. By using more than a single monitoring site for the entire set of MUs, the total diversity in the set of MUs is increased. Greatest diversity is obtainable when each MU is taken from a different monitoring site. However, there is a substantial overhead cost for each monitoring site, and the use of a different monitoring site for every MU is not cost effective (see Appendix C).

Varying the location of monitoring sites sometimes reflects known or suspected regional differences in a particular agricultural parameter. More often, however, changing geographic locations merely varies exposure potential that comes from numerous differences in workers, equipment, or environmental conditions. Thus, monitoring site is merely a surrogate for a cluster of known and unknown factors that affect exposure. Monitoring sites can be temporally as well as spatially distinct. Exposure monitoring conducted at the same general location, but years apart, might still differ in many ways that could impact exposure. Thus, there should be little surprise if exposure, on the average, differs between monitoring sites.

Diversity selection of sites means obtaining sites that are different from each other, on the average, with respect to some characteristic(s) expected to impact exposure. Because geographic location is a strong meta-factor, selecting monitoring sites from different geographic regions (and possibly time periods) should provide substantial diversity at this stage of selection. If there are a number of possible sites available and a set is to be selected (randomly or purposively), then a stratified diversity selection of sites, based on important characteristic(s), is a feasible approach to obtaining the final set of N_c monitoring sites.

B8.3. Diversity Selection of Second-Stage Units

Ultimately, the second stage selection units are the monitoring units. The MUs should be diversified independently within each selected site. In most cases, within-site diversity selection of MUs will focus on only two meta-factors: worker and normalizing factor.

B8.3.1. Workers

Obviously every worker contributes a unique set of physical and behavioral characteristics to a monitoring unit. For AHETF scenarios in particular, workers also are strongly associated with only a limited set the possible scenario conditions. For example, a worker might only have available a limited set of equipment or be able to handle a limited range of active ingredient. Thus, many components of **C** relating to a particular worker will be identical, even on different

days. This means that handling-day exposures for the same individual are expected to be correlated. In contrast, different individuals are expected to show less correlation in exposure. In other words, worker is a very strong meta-factor since a change in worker will invariably result in changing many factors (known and unknown) that will impact exposure. Therefore, diversity can be increased by simply requiring that each MU be constructed using a different individual worker.

The final set of MUs could be further diversified by applying certain additional restrictions when recruiting growers and workers. For example, a scenario plan might restrict the number of participants from any particular grower to just one or two. This reduces the likelihood that any two workers will be similar in characteristics such as worker training, equipment types, or standard pesticide handling practices that could impact exposure. In effect, such restrictions help to reduce possible sub-clustering of MUs within monitoring sites.

In principle, additional diversity might be obtained using stratified diversity selection of workers by specific characteristics such as stature, ethnicity, etc. However, with the small number of MUs chosen per cluster and the diversification of MUs by the normalizing factor (see B8.3.2 below), additional diversification by 'worker type' would increase the complexity of the selection process. The number of possible combinations of worker type and level of normalizing factor would greatly exceed the number of MUs that could be selected in a cluster. It is for this reason that formal diversity selection of MUs will focus on only the two meta-factors worker and normalizing factor.

It is not the intent of the AHETF Monitoring Program to determine how much exposure can vary for the same individual over time. This fact, and the desire to diversify the set of MUs with respect to the meta-factor 'worker', is the reason that the use of the same worker for two or more MUs is strongly discouraged. In an extreme emergency, such as a worker withdrawing from the study at the last minute and there are no 'unused' volunteers available, a worker might be used for a second MU. In this case, however, as many other HD conditions as possible should be varied to reduce the correlation due to the repeated use of the same individual. For example, the MU for the same person might use a different AaiH, different equipment, and occur on a different day. The burden of such restrictions will usually make it more cost-effective to recruit additional workers (or even consider additional monitoring sites) than to 'reuse' the same worker. Regardless, if multiple MUs with the same subject occur at all, they will be extremely rare in AHETF-conducted studies.

B8.3.2. Normalizing Factor

As defined in B3 above, normalized exposure is simply the daily exposure divided by a relative measure of potential active ingredient contact (PaiC) that has regulatory value for a scenario. The set of selected MUs within each cluster should also be diverse with respect to this normalizing factor (NF). It might at first seem pointless to have MUs with differing levels of the normalizing factor when diversity in normalized exposure is the goal. If the NF is approximately proportional to the amount of potential ai contact, then normalized exposure should be

independent of the levels of NF. Consequently, it should make no difference whether all selected MUs are at the same level or at different levels of the NF.

This would be true if the value of the normalizing factor selected had no impact on the other HD conditions selected. However, this is probably not the case and the NF should also be considered a meta-factor. That is, it is not unreasonable to expect that different values of the NF will be naturally associated with specific sets of handler-day conditions, **C**, in the future population of HDs. For example, the equipment and handling activities could be quite different for handling small amounts of active ingredient than for handling larger amounts (e.g., in size or type of application equipment into which the pesticide is mixed/loaded). By insuring that the levels of NF are varied, the set of MUs indirectly captures diversity in those components of **C** that are naturally associated with the normalizing factor.

In addition, for many scenarios, a secondary benchmark objective is to ensure there are adequate data to allow AHED® users to determine if, on the average, exposure appears proportional to the particular normalizing factor used (see Appendix C). This will also require obtaining a sufficient number and type of MUs over the practical range of the normalizing factor so that if exposure is truly unrelated to NF, a test of consistency with a proportional relationship will be rejected. As noted by the SAP (SAP, 2007) and shown in Appendix C, power is greatest when there is a large range in the NF within each cluster.

For both of these reasons, the AHETF Monitoring Program will also attempt to diversify, as much as possible within each cluster, the single normalizing factor deemed most appropriate for the scenario. As noted previously, in most cases the normalization factor used is the amount of active ingredient handled (AaiH). It is also conceivable, albeit unlikely, that no normalizing factor would be appropriate for a particular handling scenario. This might be the case, for example, if potential ai contact was felt to be identical for all future handler-days and, therefore, predicted future exposure could be obtained from MUs using unnormalized exposure. In such an event, the AHETF recommends that the meta-factor AaiH be diversified by default.

To accomplish diversification of the normalizing factor, the recommended approach for all scenarios is to partition the possible levels of NF at a monitoring site into strata and use only a single NF level from each stratum. For practical reasons, the exact mechanism for diversifying this practical NF range will vary between scenarios. Since AHETF MUs incorporate actual commercial work activities, fine control over the NF is not always possible. However the same general guidelines are followed for all scenarios:

1. The practical range of the NF for each scenario (or, if appropriate, for each monitoring site separately) is defined based on information such as current product use rates and assumptions for the amount of crop that can be treated in a day. For example, with AaiH the upper limit of the practical range (ULPR) might include aerial applications of high use rate products. The lower limit of the practical range (LLPR) reflects handling of very low use rate products. In general, however, the LLPR does not extend below 5 lbs ai/day to ensure that quantifiable residues are found for most worker exposure samples and that workers are monitored for a period of time that is representative of a full workday.

2. Several strata are defined such that each stratum represents an equal interval (usually on a log scale) of the normalizing factor and, collectively, the strata cover its entire practical range. The number of strata for each cluster should be the same as the number of MUs in a cluster.

3. Within each cluster select (i.e., choose or construct) one, but not more than one, NF level from each stratum to use for each MU. Selection of NF levels within each stratum can be random or purposive. If purposive, it would also be advantageous to vary the exact levels within each stratum to avoid having several MUs (in the same stratum but in different clusters) with exactly the same level of the normalizing factor.

In rare cases there may be a pool of available workers that can be used for any NF stratum. If all possible configurations of assignment are equivalent, then workers could be randomly allocated to MUs regardless of their NF levels. If some allocations are non-equivalent (e.g. more cost effective or there are scheduling issues) then a purposive assignment of individuals to MUs might be preferable. For most AHETF scenarios, however, worker availability strongly depends on the particular NF level chosen. Some individuals may only work with higher NF levels and some with only lower NF, say. Selection of workers could still be random, although random choice might need to be restricted in some ways (e.g., within each NF level separately). However, when such associations between subject and NF levels exist, purposive allocation of workers and NF levels to MUs might result in a more cost effective and practical configuration.

B8.4. Diversity in Other Factors

As described above, the diversity in handling conditions in the set of MUs for all scenarios should be driven by the formal use of distinct monitoring sites, unique workers, and diversity selection of units (e.g., the normalizing factor stratification imposed on MUs within each cluster). These approaches utilize the three primary meta-factors common to all scenarios: (1) monitoring site, (2) worker, and (3) normalizing factor. These three characteristics might not themselves cause differences in exposure but are certainly associated with factors (both known and unknown) that impact exposure.

In addition to these three primary meta-factors, there could be other factors that have some impact on dermal and/or inhalation exposure for a particular scenario. These might include factors such as equipment used, specific worker techniques, and number of product containers used. There are no formal procedures established to diversify these factors within clusters or scenarios. For the most part, such factors should vary naturally with the three primary meta-factors and additional perturbation should be unnecessary. For each scenario, AHETF experts do attempt to identify and list those factors expected to vary in the future handler-day population and have an impact on worker exposure. (AHETF might also consult with other experts through a variety of sources.) In the event two putative MUs have excessive similarity with respect to such factors, consideration should be given to reducing this similarity.

Appendix C. Procedures for Determining the Required Number of Clusters and Monitoring Units per Cluster to Achieve Benchmark Adequacy

Appendix C
Procedures for Determining the Required Number of Clusters and Monitoring Units per Cluster to Achieve Benchmark Adequacy

This appendix describes the statistical basis and methodology used by the AHETF to determine the sample size and configuration for agricultural handler scenarios. This expanded version of the sample size methodology document has greatly benefited by both formal and informal input from a recent EPA Scientific Advisory Panel review (SAP, 2007) and numerous discussions with the EPA.

C1. The Conceptual Basis for Sample Size Determination

Obtaining potential workers and scenario-specific handling-day conditions needed to create **monitoring units** (or MUs) is a complicated process. Of necessity, the specific MU selection process used will vary from scenario to scenario. However, as described in Appendix B, the MU construction process for all scenarios can be envisioned as the result of two successive stages of selection. The first stage consists of selecting specific geographic locations (plus a range of dates for monitoring at each location) that serve as areas from which to stage monitoring operations. Each combination of local area and range of potential monitoring dates is termed a **monitoring site**. The second stage consists of selecting subjects and other handling conditions within each site that are then used to construct the MUs.

In the AHETF Monitoring Program, N_c monitoring sites are selected at the first stage and N_m monitoring units will be obtained within a site at the second stage. (N_m will not necessarily be the same at all sites.) When N_m is greater than one, the set of MUs at the same site is termed a **cluster**. In general, MUs in the same cluster are expected to be more similar than those in different clusters. Theoretically, such within-cluster correlation favors the use of only a single MU per site. However, there are often substantial overhead costs per site that make multi-MU sites more efficient.

In the strictest sense, sample sizes can only be determined using statistical theory alone when either

1. There is randomization of experimental units to treatments and the goal is only to compare or to contrast treatments in some manner; or

2. There is assumed random, representative sampling from a population and the goal is to estimate some characteristic of that population.

Only in these two situations will statistical theory predict how increasing sample size decreases estimation error. In other experimental situations, sample size must be determined using one of the two 'random' situations above as a **reference model**. The random reference model is defined so that it reflects the actual situation (e.g., a mixture of random and non-random

selection) as closely as is practical. The sample size that is appropriate for the reference model is then used as a **benchmark** for the actual study design. In a real sense, then, the reference random sampling model is used to establish **benchmark sample sizes** that can satisfy **benchmark objectives**. Although rarely stated explicitly, the use of reference sampling models and benchmark objectives is quite common.

The AHETF monitoring program is not an experimental study whose purpose is to test hypotheses about the distribution of exposure or about potential determinants of exposure (i.e., as in situation 1 above). Rather, its purpose is to collect sufficient data for each handler scenario to meet specific minimum or adequacy requirements. This is more analogous to situation (2) above. These data, possibly augmented by additional exposure data from other sources, will then be used for a variety of regulatory purposes by numerous organizations.

C2. The Two-Stage Random Sampling Reference Model for MU Exposures

Because all AHETF scenario designs will have a two-stage selection structure, they all assume the same from of reference sampling model. For each scenario, two-stage random nested (or cluster) sampling is the reference model used for the combination of purposive and random two-stage diversity selection that can actually occur (see Appendix B). This reference model assumes that:

1. Exposure, normalized by an active ingredient contact factor, is lognormally distributed with geometric standard deviation GSD. Equivalently, the logarithm of normalized exposure is normally distributed with standard deviation equal to log GSD.

2. There are N_c clusters (i.e. monitoring sites) randomly sampled from all possible sites.

3. There are N_m MUs randomly sampled per cluster. Unequal numbers of MUs per cluster are possible. When N_m is the same for all clusters, then the total number of MUs in a scenario is $N=N_c \times N_m$.

4. The within-cluster (i.e., within-site) correlation of log normalized exposure is equal to ICC (the 'intra-cluster' correlation).

This reference sampling model incorporates the two-stage selection structure and the potential for within-cluster correlation but ignores any effects of diversity selection. As described in Appendix B, the normalizing factor (NF) for exposure can be any measured quantity that is expected to be proportional to the potential amount of contact with active ingredient. This normalizing factor is scenario-specific although in most cases the amount of active ingredient handled (i.e. 'processed') by the MU is used.

Thus, for determining sample sizes, normalized exposures, Q, are assumed to follow the nested variance component model

(1) $$\text{Log}\,(E_{ij} / H_{ij}) = \text{Log}\, Q_{ij} = \text{Log}\, GM_Q + C_i + W_{ij}$$

where

E_{ij} = the exposure obtained for MU j in cluster i
H_{ij} = the value of NF for MU j in cluster i
Q_{ij} = the normalized exposure for MU j in cluster i
GM_Q = the population geometric mean for normalized exposure
C_i = a random effect of cluster i
W_{ij} = a random effect of MU j within cluster i

As stated above, H_{ij} is usually the amount of ai handled (abbreviated as AaiH). However, it could represent any normalizing factor. In addition, (1) applies to unnormalized exposure when $H_{ij} = 1$. The random effects C_i and W_{ij} are normally distributed with means 0 and variances V_c and V_w, respectively.

The population variance of log Q is then equal to $V = V_c + V_w$ and the square root of V is the true population standard deviation, SD. The quantity GSD_Q = antilog (SD) is the true population geometric standard deviation of normalized exposure. The 'intra-cluster' correlation (i.e., the intraclass correlation due to clusters) is defined as

(2) $$ICC = V_c / V = 1 - V_w / V$$

The ICC is irrelevant to the population distribution of normalized exposure, per se. However, this intra-cluster correlation is a necessary part of the reference sampling model because the MUs are obtained in clusters. Under this sampling model, reasonable values for GSD_Q and ICC will be needed to determine sample size.

C3. Benchmark Objectives of the AHETF Monitoring Program

The purpose of the AHETF Monitoring Program is to collect sufficient data for each handler scenario to meet specific benchmark adequacy objectives. The primary benchmark objective for all scenarios will be that a sample from the hypothetical reference sampling distribution above be of adequate size to describe selected measures of the (normalized) exposure distribution with a pre-determined level of accuracy. The design benchmarks are not intended to address all possible ways the exposure data could be used. Rather, they are established to ensure that the data will at least be adequate to meet common regulatory needs.

The primary benchmark data objective for each scenario in the AHETF program will be of the form:

> ***The number (and configuration) of sampled monitoring units (MUs) should be adequate so that selected measures of the normalized dermal exposure distribution (e.g., means, percentiles) are accurate to within K-fold.***

Throughout this appendix, normalized (or unit) exposure refers to exposure divided by any normalizing factor that represents predicted amount active ingredient contact. By default, this

normalizing factor will be the amount of active ingredient handled by the worker (or AaiH). A benchmark based on AaiH is treated as the default for the primary benchmark because it is currently the most common measure of unit exposure used for regulatory purposes. For some scenarios, however, regulators might prefer to define unit exposure in terms of a different measure of 'active ingredient contact potential'. There could even be scenarios for which users of AHED® prefer to use unnormalized exposure. If, for a particular scenario, the JRC and AHETF jointly decide that a different normalization factor (or none at all) is more valuable for regulatory use then it will be used instead to define unit exposure for the primary benchmark objective.

The desired relative accuracy, K, can be scenario dependent. For example, less accuracy (i.e., a larger value of K) might be tolerated for scenarios that are expected to have lower exposures. Such considerations are often necessary to better allocate limited resources and avoid unnecessary human exposure monitoring. Currently, however, there is a general consensus that, for regulatory purposes, 3-fold relative accuracy (i.e., K=3) is a reasonable default for all scenarios.

Using K-fold accuracy for exposure normalized by amount of ai handled (or a different normalizing factor) as a benchmark does not necessarily imply that other ways of expressing exposure will result in less accuracy. Because many potential normalizing factors (e.g., time worked, loads handled) are usually correlated with the amount of ai handled, similar accuracies are likely in those cases as well. However, for design purposes, the AHETF monitoring program only uses the distribution of exposure normalized by a single measure of contact potential. For simplicity, unless noted otherwise, AaiH will be assumed to be the normalizing factor throughout this appendix.

A secondary (i.e. less important) benchmark objective is considered for scenarios for which normalized exposure is relevant and when the practical range of this factor is sufficiently large. In such cases it is also desired that:

> ***The number (and configuration) of monitoring units (MUs) should be adequate so that it is possible to statistically distinguish between complete proportionality and complete independence of dermal exposure and the normalizing factor (e.g., amount of ai handled).***

If, for a particular scenario, the JRC and AHETF jointly decide that unnormalized exposure has greater regulatory value, then no secondary benchmark will be considered.

It is not the objective of the AHETF monitoring program to guarantee that the data will be able to discern more complicated relationships between exposure and amount of ai contact. Nor is it the intent to guarantee that future analyses of the data will be able to choose between several potential normalizing factors or combinations of factors. The SAP (SAP, 2007) correctly noted that this secondary objective in the AHETF program merely ensures that the data would be adequate to illuminate the relationship between AaiH and exposure. They suggested that a

controlled experimental study, beyond the scope of the AHETF program, might be a better way to address this issue more extensively.

It must be emphasized that the use of amount of ai handled as the default measure of ai contact potential in these benchmark objectives is both reasonable and based on considerable historical precedent. Use of AaiH does not mean, however, that a proportional relationship between dermal exposure and amount of ai handled is assumed to be always the case. In fact, the AHETF monitoring data (assembled in a database called AHED®) will always include measured exposures (generally µg of ai for the entire monitoring period) and the values of many potential normalizing factors, including amount of ai handled. Users of the data are always free to consider any (or no) normalization.

C4. Estimates of GSD and ICC from Existing Data

As discussed in C2 above, use of the reference model to determine sample sizes requires reasonable values for the geometric standard deviation (GSD_Q) and the within-cluster correlation (ICC). Estimates of GSD_Q and ICC were obtained from existing AHETF monitoring data. Although these data are incomplete, they are sufficient to provide reasonable values for normalized exposure variation and the intra-cluster correlation. The scenarios and clusters for which data were available (as of March, 2008) are listed in Table 1.

For each scenario in Table 1, both GSD_Q and ICC were estimated by fitting the variance-component sampling model (1) to the available data. The estimates obtained for total dermal exposure are given in Table 2. For completeness, the estimates obtained for inhalation exposure are shown in Table 3. Two scenarios closed granular ML and hopper box seed treatment MLAP may contain only a single cluster and therefore the ICC cannot be estimated. For these scenarios the GSD_Q is only an estimate of the within-cluster variation. The confidence intervals for GSD_Q and ICC are parametric bootstrap percentile intervals based on N=1000 bootstrap replications. Also shown in both tables are various summary measures of these estimates over all scenarios.

For normalized dermal exposure (Table 2) the GSD_Q estimates range from about 2 to 5 with a typical value slightly less than 4. The ICC estimates range from 0 to 0.66. As the confidence intervals indicate, however, uncertainties in the individual ICC estimates are very large. This is not unusual when the number of clusters is small. The mean ICC is slightly less than 0.3. Table 2 also gives the estimates of GSD_Q and ICC obtained from the fit of a mixed model using all the scenario data together. In this case, the geometric mean was allowed to differ for each scenario but common values of GSD_Q and ICC were required. These common values of GSD_Q and ICC were 3.8 and 0.26, respectively.

Normalized inhalation exposures (Table 3) appear slightly more variable than dermal exposures. The GSD_Q estimates range from about 2 to 6 with a typical value being slightly greater than 4. The ICC estimates range from 0 to 0.71 with a mean around 0.36. For the combined model, the common GSD_Q and ICC estimates are 4.2 and 0.37, respectively.

Table 1: Scenarios and Clusters with Available Monitoring Unit (MU) Exposure Results

Scenario	Clusters	Monitoring Dates	# MUs
Closed Granular ML	AH516-M, 5 towns in NE	Apr-May 1998	15
Airblast Application	AHE07-A, GA	Oct 2003	5
	AHE07-A, ID	Oct 2003	6
	AHE07-A, FL	Dec 2003	4
Granular Backpack MLA	AH207-MLA, Spain	May 1998	16
	AH208-MLA, Martinique	Aug 1998	11
Dry Flowable ML	AHE17 + AHE19, IL	Apr 2005	10
	AHE18, OR/WA	May 2005	5
	AHE20, GA	Jul 2005	5
	AHE21, FL	May-Jun 2005	5
Aerial Application	AH501-A-1, CA, CV	Oct 1991	8
	AHE18-A, WA, CV	May 2005	2
	AHE13-A, TX, ULV	Oct 2004	16
Open-Pour ML Liquids	AH204-M, France	Mar 1997	16
	AH501-M-2, MS	Sep 1991	8
	AHE30, OR	Oct 2005	2
	AHE31, CA	Nov 2005	3
	AHE32, FL/GA	Dec 2005	6
Closed Liquid ML (bulk/minibulk)	AHE13-M, TX, ULV	Oct 2004	15
	AH501-M-1, CA, CV	Oct 1991	7
Hopper-box Seed Trt MLAP	AHE10, AR + TX	Apr-May 2004	16
Open Cab Groundboom App	AHE18, OR/WA	May 2005	2
	AHE20, GA	Jul 2005	1
	AHE21, FL	May 2005	2
	AHE30, OR	Oct 2005	5
	AHE31, CA	Nov 2005	5
	AHE32, FL/GA	Dec 2005	6

Table 2: Dermal Exposure Variability Estimates for Each Scenario

Scenario	GSD	95% CI	ICC	95% CI
Aerial Application	4.2	2.1 – 12.0	0.62	0 – 0.89
Airblast Application				
With Headgear	3.2	2.1 – 4.9	0.00	0 – 0.49
No Headgear	2.9	2.0 – 4.3	0.00	0 – 0.47
Closed Granular ML	2.1[1]	1.6 – 2.8	–	–
Closed Liquid ML	4.2	2.7 – 7.8	0.11	0 – 0.59
Dry Flowable ML	2.5	1.8 – 3.8	0.41	0 – 0.75
Granular Backpack MLA	4.2	2.0 – 16.3	0.66	0 – 0.92
Hopper-box Seed Trt MLAP	3.3[1]	2.2 – 5.0	–	–
Open Cab Groundboom App	3.9	2.4 – 6.3	0.24	0 – 0.65
Open-Pour ML Liquids	5.0	3.5 – 7.8	0.00	0 – 0.30
Mean[2]	3.8		0.29	
Median[2]	4.0		0.24	
Geometric Mean[2]	3.3		–	
Combined Model[3]	3.8		0.26	

[1]Possibly only a single cluster: ICC cannot be calculated and GSD estimates only within-cluster variation
[2]Closed granular ML excluded from calculation of mean, median, and geometric mean GSD; only airblast with headgear values are used
[3]Estimates from a mixed linear model allowing different scenario geometric means but assuming a common GSD and ICC

From this analysis it appears that a GSD_Q of 4 and an ICC of 0.3 are reasonable values of variability and within-cluster correlation to use for planning purposes. Although the benchmark objectives apply only to dermal exposure, these two values should be satisfactory for inhalation exposure as well. The normalized inhalation exposure, at least on the average, appears to be only slightly more variable than dermal exposure.

Table 3: Inhalation Exposure Variability Estimates for Each Scenario

Scenario	GSD	95% CI	ICC	95% CI
Aerial Application	4.4	2.1 – 12.8	0.71	0 – 0.91
Airblast Application	2.6	1.9 – 3.8	0.00	0 – 0.45
Closed Granular ML	4.5[1]	2.6 – 7.6	–	–
Closed Liquid ML	6.0	2.4 – 28.8	0.62	0 – 0.92
Dry Flowable ML	4.6	2.9 – 7.4	0.16	0 – 0.55
Granular Backpack MLA	1.8	1.5 – 2.4	0.23	0 – 0.67
Hopper-box Seed Trt MLAP	3.8[1]	2.4 – 6.2	–	–
Open Cab Groundboom App	5.9	3.1 – 13.3	0.43	0 – 0.78
Open-Pour ML	4.3	2.8 – 7.2	0.38	0 – 0.71
Mean[2]	4.1		0.36	
Median[2]	4.3		0.38	
Geometric Mean[2]	4.2		–	
Combined Model[3]	4.2		0.37	

[1]Possibly only a single cluster: ICC cannot be calculated and GSD estimates only within-cluster variation
[2]Closed granular ML excluded from calculation of mean, median, and geometric mean GSD
[3]Estimates from a mixed linear model allowing different scenario geometric means but assuming a common GSD and ICC

C5. Relative Accuracy and Fold Relative Accuracy

As stated above, the primary objective of the AHETF monitoring program is to achieve adequate relative accuracy of selected parameters of the normalized exposure (reference) distribution. Given the sampling model (1) and the sample sizes (N_c and N_m), this benchmark target can be stated more precisely as:

> ***Estimates of the geometric mean, arithmetic mean, and 95th percentile of an underlying lognormal reference sampling distribution should be within K-fold of the true values at least 95% of the time.***

If θ denotes the distributional parameter of interest and T is the estimate of that parameter obtained from monitoring data, then the relative accuracy of T is defined simply as:

(3) $$RA(T|\theta) = T/\theta$$

Satisfying the benchmark objective above requires that there be at least a 95% chance that T/θ is between 1/K and K. More formally this is stated as:

(4) $$\text{Prob}\ \{1/K \leq RA(T|\theta) \leq K\} \geq 0.95$$

It is more convenient, however, to consider relative accuracy expressed as a 'fold relative difference'. This is because statements such as "T is within K-fold of θ" are more intuitive than the formulation given in (4). The 'fold relative accuracy', *fRA*, is defined as:

(5) $$fRA(T|\theta) = \text{Max}\{ RA(T|\theta), 1/RA(T|\theta) \} = \text{Max} (T/\theta, \theta / T)$$

Then, statement (4) is equivalent to

(6) $$\text{Prob} \{ fRA(T|\theta) \leq K \} \geq 0.95$$

and simply says that the estimate, T, will be within K-fold of the true parameter, θ, at least 95% of the time. The 95th percentile of *fRA*, fRA_{95}, is the specific fold-accuracy value that satisfies (6). Consequently, the benchmark adequacy goal reduces to requiring that:

(7) $$fRA_{95} \leq K$$

If we denote the 2.5th and 97.5th percentiles of the sampling distribution of T by $T_{2.5}$ and $T_{97.5}$, respectively, then the 95th percentile of fold relative accuracy can also be calculated from

(8) $$fRA_{95} = \text{Max} (T_{97.5} / \theta, \theta / T_{2.5})$$

C6. Parameter Estimates

As defined above, relative accuracy applies to the particular quantity T that is used to estimate the distributional parameter θ. Thus, it is important to consider which types of estimates of the geometric mean, arithmetic mean, and 95th percentile are used to evaluate fRA_{95}. The relative accuracies could differ depending on the particular estimates used.

There are often multiple choices for the parameter estimates. The estimators can be broadly grouped into either empirical or parametric. Empirical estimates are the commonly-used statistics available in spreadsheet programs. They do not (explicitly) assume any distribution. However, they can sometimes require simple random sampling for greatest efficiency. Parametric estimates incorporate the fact that the surrogate distribution is lognormal and could also account for cluster sampling being used.

The most straightforward statistic is the geometric mean (GM_Q). In the balanced case, the simple empirical estimate of GM_Q can be calculated by averaging the log-transformed normalized exposures and then taking the antilog of this value. In this case, the empirical and parametric estimates of GM_Q are identical. If the number of MUs per cluster varies, however, one could consider geometric means with different degrees of weighting by cluster size. The arithmetic mean can also be calculated empirically by summing up the normalized exposures and dividing by the total number of MUs. Again, when the cluster sizes differ, other types of weighted empirical arithmetic means exist. In the unbalanced case, neither the weighted nor the unweighted estimates of GM_Q or AM_Q are universally best. Consequently, for this investigation,

the simple (and most common) versions of the empirical geometric and arithmetic means seem preferable. Empirical percentiles could, theoretically be calculated in the conventional manner. However, when there is cluster sampling and the number of MUs are not large, empirical estimates of the extreme upper (or lower) percentiles are not especially efficient. The parametric percentiles (see below) are usually preferred in this case.

Parametric estimates are those closely aligned with the sampling model used. In this case one uses the fit to the variance component model described in (1) above to get estimates for the geometric mean (GM_Q) and the total geometric standard deviation (GSD_Q). To estimate the arithmetic mean (AM_Q) and 95th percentile (Q_{95}) one could then use the lognormal relationships:

$$AM_Q = GM_Q \times \text{Exp}\{ \tfrac{1}{2} (\log_e GSD_Q)^2 \}$$

(9)

$$Q_{95} = GM_Q \times \text{Exp}\{ Z_{95} \log_e GSD_Q \}$$

where Z_{95} is the 95th percentile of the standard normal distribution. For simplicity, these will be labeled the 'parametric cluster sampling estimates'.

It can be argued that few if any users of the AHETF monitoring data will choose to (or be able to) fit variance component models to the data. They will probably ignore the sampling model and use more conventional estimates. In this case empirical estimates of GM_Q and AM_Q defined above would probably be used.

Potential data users might also be less inclined to use empirical percentiles, especially with smaller sample sizes. The lognormal percentile estimate of Q_{95} in (9) above would then still be used but perhaps with the mixed model GSD_Q estimate replaced with the more conventional GSD_Q (i.e., the back-transformed simple standard deviation of log exposures.) For convenience, estimates that assume lognormality but not cluster sampling will be labeled 'simple random sampling parametric percentiles'.

Any or all of the above estimators could be evaluated. However, for the purposes of determining sample sizes, focus will be on the following estimators:

- GM_Q – simple empirical estimate
- AM_Q – simple empirical estimate
- Q_{95} – parametric cluster sampling estimate

C7. Calculation of fRA_{95} given the Number of Clusters (N_c) and a Fixed Number of MUs per Cluster (N_m)

Calculation of the 95th percentile of fold relative accuracy is complex and is usually best accomplished using Monte Carlo simulation methods. However, when the number of MUs per cluster, N_m, is the same for all clusters, the geometric mean, fRA_{95} can be calculated directly from the GSD_Q and ICC as:

(10) $$fRA_{95} = \exp\left\{1.96 \ln GSD_Q \sqrt{\frac{ICC}{N_c} + \frac{1-ICC}{N}}\right\}$$

where N is the total number of MUs (i.e., $N=N_c \times N_m$).

For parameters other than the geometric mean, a straightforward simulation approach can be used to determine fRA_{95}. This procedure is:

1. Simulate a set of normalized exposure data for N_c clusters and N_m monitoring units per cluster using the sampling model defined in (1) above.
2. From each set of simulated data, calculate T, the estimate of θ
3. Repeat steps 1 and 2 above M times to get M values of the estimate T
4. From these M T-values calculate $T_{2.5}$ and $T_{97.5}$, the 2.5th and 97.5th percentiles of T, respectively.
5. Calculate the 95th percentile of fold relative accuracy, fRA_{95}, using formula (8) above.

The number of simulations, M, should be some large number such as 1,000 or 10,000.

C8. Determination of Feasible Values of N_c and N_m

The methods in the previous section can be used to find those combinations of N_c and N_m that will achieve the default 3-fold level of accuracy (or any other desired accuracy goal). For different combinations of N_c and N_m, the simulation method above was implemented in SAS with M=10,000 to obtain fRA_{95} values for the arithmetic mean and 95th percentile estimators. The 95% bound for geometric mean relative accuracy was calculated directly from equation (10). The values used for the true GSD_Q and true ICC were 4 and 0.3, respectively.

Table 4 illustrates how N_c and N_m affect the upper bound of fold relative accuracy. For any configuration, the geometric mean will always be the most accurate since it is in the center of the lognormal distribution. The 95th percentile and, usually, the arithmetic mean estimates tend to be the least accurate for any sample size. Lower percentiles (e.g., 75th, 90th) would have smaller accuracy bounds than would the 95th percentile. The value of fRA_{95} for percentiles exceeding the 95th will be greater.

Increasing either N_c or N_m will improve the accuracy of all estimators, but adding clusters is more effective than increasing the number of MUs per cluster. With just 4 clusters it takes 10 MUs per cluster (N=40) to achieve about the same level of accuracy as with N_c=5 and N_m=5.

The number of clusters needed to achieve an fRA_{95} of approximately 3 was determined for N_m values ranging between 1 to 10. These nearly-equivalent 3-fold accuracy configurations are listed Table 5. Any of these combinations of N_c and N_m would be feasible from a benchmark accuracy standpoint.

Table 4: 95% Relative Accuracy Bounds for Lognormal Distribution Parameter Estimates When GSD = 4 and ICC = 0.3.

# Clusters, N_c	# MUs per Cluster, N_m	Total # MUs, N	95% Relative Accuracy Bound, fRA_{95}		
			Geometric Mean[1]	Arithmetic Mean[2]	95th Percentile[2]
3	5	15	2.8	3.9	4.5
4	5	20	2.5	3.3	3.6
5	5	25	2.2	2.9	3.1
6	5	30	2.1	2.7	2.8
7	5	35	2.0	2.5	2.6
4	5	20	2.5	3.3	3.6
4	6	24	2.4	3.1	3.4
4	8	32	2.3	3.0	3.1
4	10	40	2.3	2.9	3.1
4	4	16	2.6	3.5	3.8
5	4	20	2.3	3.1	3.3
6	4	24	2.1	2.8	2.9

[1]Exact calculation
[2]Based on 10,000 simulations

Table 5: Configurations Yielding Nearly Equivalent 95% Relative Accuracies Bounds of 3-Fold or less when GSD = 4 and ICC = 0.3.

# Clusters, N_c	# MUs per Cluster, N_m	Total # MUs, N	95% Relative Accuracy Bound, fRA_{95}		
			Geometric Mean[1]	Arithmetic Mean[2]	95th Percentile[2]
15	1	15	2.0	2.8	3.0
9	2	18	2.1	2.8	2.9
7	3	21	2.1	2.8	3.0
6	4	24	2.1	2.8	2.9
5	5	25	2.2	2.9	3.1
5	6	30	2.2	2.8	3.0
5	7	35	2.2	2.7	2.9
4	8	32	2.3	3.0	3.1
4	9	36	2.3	3.0	3.1
4	10	40	2.3	2.9	3.1

[1]Exact calculation
[2]Based on 10,000 simulations

C9. Optimal Configuration of N_c and N_m based on Relative Costs

As Table 5 illustrates, there can be many configurations of N_c and N_m that will give acceptable benchmark adequacy. However, some of these feasible configurations are more costly than others. As suggested by the SAP (SAP, 2007), the choice between equivalent configurations is best made on the basis of cost-effectiveness. A commonly-used model approximating the total cost of cluster sampling, C_T is:

(11) $$C_T = C_F \times N_c + C_{MU} \times N$$

C_{MU} is the variable cost per MU and C_F is the fixed cost per cluster. The ratio of C_F to C_{MU} determines which configurations are the most cost-effective. If C_{MU} in (11) above is set equal to one, then $C_F/C_{MU} = C_F$ and C_T can be thought of as the total relative cost as a multiple of the 'per MU' cost. The 10 nearly equivalent configurations in Table 5 are reproduced in Table 6 along with their relative total costs for different values of the fixed-to-variable cost ratio. When the cost-ratio is equal to zero, there are no fixed costs associated with a cluster and total cost is just a multiple of the number of MUs. In this case the N_c=15 and N_m=1 configuration will be optimal since it achieves the desired accuracy with the smallest N. As the ratio C_F/C_{MU} increases, the penalty for each new cluster increases and the total costs are smaller when there is more than one MU per cluster. For C_F/C_{MU}=1, N_m=2 is optimal. When C_F/C_{MU} is in the 5-10 range a

configuration of N_c=5 and N_m=5 is more cost-effective. With a cluster cost 25 times that of the per MU cost, the optimal configuration is N_c=4 and N_m=8.

Table 6: Relative Costs per Scenario for Configurations Yielding Nearly Equivalent 3-Fold Relative Accuracy when GSD = 4 and ICC = 0.3.

N_c	N_m	N	95% Relative Accuracy Bound, fRA_{95} [1]		Total Cost[2] when the Fixed to Variable Cost Ratio is:					
			Arithmetic Mean	95th Percentile	0	1	2	5	10	25
15	1	15	2.8	3.0	15	30	45	90	165	390
9	2	18	2.8	2.9	18	27	36	63	108	243
7	3	21	2.8	3.0	21	28	35	56	91	196
6	4	24	2.8	2.9	24	30	36	54	84	174
5	5	25	2.9	3.1	25	30	35	50	75	150
5	6	30	2.8	3.0	30	35	40	55	80	155
5	7	35	2.7	2.9	35	40	45	60	85	160
4	8	32	3.0	3.1	32	36	40	52	72	132
4	9	36	3.0	3.1	36	40	44	56	76	136
4	10	40	2.9	3.1	40	44	48	60	80	140

[1] Based on 10,000 simulations
[2] Total relative cost assuming the variable cost per MU, C_{MU}, is equal to 1

The AHETF has examined the cost structure of previous studies and concluded that most scenarios should have fixed-to-variable cost ratios between 6.5 and 8. Values of C_F/C_{MU} as low as 4 and as high as 9 are possible, but are expected to be rare. Table 7 lists these same 10 configurations with the total relative costs based on these AHETF cost ratios. Over this entire range, it appears that a configuration of N_c=5 and N_m=5 is the most reasonable choice.

C10. Sensitivity of fRA_{95} to the GSD and ICC

It is important to consider how sensitive (N_c, N_m) = (5, 5) configuration is to the assumed values of GSD_Q and ICC. Table 8 gives the results of simulations varying ICC from 0.1 to 0.5 while GSD_Q is kept at 4. As would be expected, increasing ICC worsens the fRA_{95} slightly and decreasing the within-cluster correlation improves it. These changes are modest, however.

Table 9 shows the effect of changes in GSD_Q on fRA_{95}. As was the case with ICC, increasing GSD_Q when ICC is fixed at 0.3 makes fRA_{95} worse and decreasing GSD_Q improves (i.e., decreases) fRA_{95}. Overall, however, the effects of changes in the GSD of ±1 on the 95th percentile of fold relative accuracy do not appear substantial.

Table 7: Expected Relative Costs per Scenario for Nearly Equivalent '3-Fold Accuracy' Configurations using the AHETF-Estimated Cost Ratio Range

N_c	N_m	N	95% Relative Accuracy Bound, fRA_{95} [1]		Total Cost[2] when the Cluster to MU Cost Ratio is:			
			Arithmetic Mean	95th Percentile	4	6.5	8	9
15	1	15	2.8	3.0	75	112.5	135	150
9	2	18	2.8	2.9	54	76.5	90	99
7	3	21	2.8	3.0	49	66.5	77	84
6	4	24	2.8	2.9	48	63	72	78
5	5	25	2.9	3.1	45	57.5	65	70
5	6	30	2.8	3.0	50	62.5	70	75
5	7	35	2.7	2.9	55	67.5	75	80
4	8	32	3.0	3.1	48	58	64	68
4	9	36	3.0	3.1	52	62	68	72
4	10	40	2.9	3.1	56	66	72	76

[1] Based on 10,000 simulations
[2] Total relative cost assuming the variable cost per MU, C_{MU}, is equal to 1

Table 8: 95% Relative Accuracy Bounds for Lognormal Distribution Parameter Estimates Obtained by Varying ICC when GSD = 4, N_c = 5 and N_m = 5.

Geometric Standard Deviation, GSD	Intracluster Correlation, ICC	95% Relative Accuracy Bound, fRA_{95}		
		Geometric Mean[1]	Arithmetic Mean[2]	95th Percentile[2]
4	0.5	2.6	3.5	3.8
4	0.4	2.4	3.2	3.4
4	0.3	2.2	2.9	3.1
4	0.2	2.1	2.6	2.7
4	0.1	1.9	2.4	2.5

[1]Exact calculation
[2]Based on 10,000 simulations

Table 10 shows more extreme situations in which both GSD_Q and ICC are perturbed together. As would be expected the worst case is seen when both GSD_Q and ICC increase. The bound on

fold relative accuracy for the arithmetic mean and 95th percentile can be nearly 5-fold when GSD_Q=5 and ICC=0.5. On the other hand, when these two variation parameters vary in opposite directions, little change in fRA_{95} will occur. Of course, if both parameters decrease, great improvements in relative accuracy can occur.

Table 9: 95% Relative Accuracy Bounds for Lognormal Distribution Parameter Estimates Obtained by Varying GSD when ICC=0.3, N_c = 5 and N_m = 5.

Geometric Standard Deviation	Intracluster Correlation	95% Relative Accuracy Bound, fRA_{95}		
		Geometric Mean[1]	Arithmetic Mean[2]	95th Percentile[2]
5	0.3	2.5	3.7	3.7
4.5	0.3	2.4	3.3	3.4
4	0.3	2.2	2.9	3.1
3.5	0.3	2.1	2.5	2.8
3	0.3	1.9	2.2	2.4

[1]Exact calculation
[2]Based on 10,000 simulations

Table 10: 95% Relative Accuracy Bounds for Lognormal Distribution Parameter Estimates Obtained by Varying Both GSD and ICC when N_c = 5 and N_m = 5.

Geometric Standard Deviation	Intracluster Correlation	95% Relative Accuracy Bound, fRA_{95}		
		Geometric Mean[1]	Arithmetic Mean[2]	95th Percentile[2]
5	0.5	3.0	4.7	4.7
5	0.1	2.1	3.0	2.9
4	0.3	2.2	2.9	3.1
3	0.5	2.1	2.5	2.9
3	0.1	1.7	1.9	2.0

[1]Exact calculation
[2]Based on 10,000 simulations

Table 11 demonstrates configurations of (N_c, N_m) that would be necessary to achieve 3-fold relative accuracy for the worst-case situation shown in Table 10. It appears that with 5 MUs/cluster, approximately 10 clusters (50 MUs) would be necessary. With 3 MUs/cluster it would take about 12 clusters (36 MUs) to achieve the same degree of accuracy. The relative

costs for the configurations in Table 11 are generally double those for the feasible configurations in Table 7.

Table 11: Expected Range of Relative Costs per Scenario for Nearly Equivalent '3-Fold Accuracy' Configurations when GSD = 5 and ICC = 0.5.

N_c	N_m	N	95% Relative Accuracy Bound, fRA_{95}			Total Cost[3] when the Cluster to MU Cost Ratio is:			
			Geometric Mean[1]	Arithmetic Mean[2]	95th Percentile[2]	4	6.5	8	9
21	1	21	2.0	3.1	3.0	105	157.5	189	210
22	1	22	2.0	3.0	2.8	110	165	198	220
13	2	26	2.1	3.1	3.1	78	110.5	130	143
14	2	28	2.1	3.1	2.9	84	119	140	154
15	2	30	2.0	3.0	2.9	90	127.5	150	165
11	3	33	2.2	3.1	3.0	77	104.5	121	132
12	3	36	2.1	3.1	2.9	84	114	132	144
13	3	39	2.0	2.9	2.8	91	123.5	143	156
10	4	40	2.2	3.1	3.1	80	105	120	130
11	4	44	2.1	3.0	2.9	88	115.5	132	143
10	5	50	2.2	3.1	2.9	90	115	130	140
11	5	55	2.1	2.9	2.8	99	126.5	143	154
10	6	60	2.1	3.0	2.9	100	125	140	150

[1]Exact calculation
[2]Based on 10,000 simulations
[3] Total relative cost assuming the variable cost per MU, C_V, is equal to 1

On balance, a configuration consisting of 5 clusters with 5 MUs/cluster seems to be a reasonable compromise given the existing variation seen in the current exposure data. Obviously, fewer resources would be necessary when it is felt that the GSD and ICC for normalized exposure can be less than the assumed values of 4 and 0.3, respectively. Smaller sample sizes could also be used when K>3 is considered acceptable. For tighter accuracy requirements, additional samples would be needed. The methods described above can be used, as needed, to determine the sample sizes for other combinations of GSD, ICC, and K.

C11. The Impact of Unequal Numbers of MUs per Cluster

In the preceding evaluation of sample sizes, it was assumed that each cluster would have exactly N_m monitoring units. As the SAP (SAP, 2007) correctly noted, equal cluster sizes are not always

possible, or even desirable, in practice. Therefore, the effect of unequal cluster size on relative accuracy should be considered.

When the number of MUs per cluster varies from cluster to cluster, the variation of the distributional estimates is larger than it would be with equal cluster sizes. For example, if m_i denotes the number of MUs in cluster i then the formula for fRA_{95} for the sample geometric mean given in (10) generalizes to:

$$fRA_{95} = \exp\left\{1.96 \ln GSD_Q \sqrt{\frac{ICC \sum m_i^2}{N^2} + \frac{1 - ICC}{N}}\right\} \tag{12}$$

With the same number of clusters, N_c, and total number of MUs, N, formula (12) will always give a larger value of fRA_{95} whenever ICC>0 and m_i are not all equal. When all $m_i=N_m$, then (12) reduces to (10). This is true for other parameter estimates as well. Table 12 illustrates this negative effect on relative accuracy of unequal cluster size with N=25 MUs distributed over N_c=5 clusters. These results were obtained using equation (12) for the geometric mean and an unequal-cluster-size version of the simulation approach described in Section C7 above for the arithmetic mean and 95th percentile. When the variation in cluster size is moderate, Table 12 shows that the loss in accuracy is barely discernable within simulation error. Only when the cluster sizes vary more widely (e.g. from 1 to 9 in this example) are losses evident.

Table 12: Loss in Accuracy with Variation in the Number of MUs per Cluster when GSD = 4 and ICC = 0.3.

# Clusters, N_c	# MUs per Cluster	Total # MUs, N	95% Relative Accuracy Bound, fRA_{95}		
			Geometric Mean[1]	Arithmetic Mean[2]	95th Percentile[2]
5	5 5 5 5 5	25	2.2	2.9	3.1
5	4 5 5 5 6	25	2.2	2.9	3.0
5	3 4 5 6 7	25	2.3	3.0	3.1
5	1 3 5 7 9	25	2.4	3.2	3.6
5	1 1 5 9 9	25	2.6	3.2	3.7

[1]Exact calculation
[2]Based on 10,000 simulations

However, for sample size determination purposes, there is a simple, practical solution to the potential loss of efficiency with unequal cluster size. Suppose that an acceptable configuration of N_c clusters with exactly N_m MUs in each cluster has been found such that fRA_{95} = K. Then, although the actual number of MUs per cluster in an actual design is allowed to vary, we enforce the following restrictions:

- The total number of MUs is no less than $N = N_c \times N_m$
- The number of MUs per cluster is never greater than N_m

Under these conditions, it appears that the resulting fRA_{95} will never exceed K. For the geometric mean, this result follows directly from formula (12). Numerous simulations indicate that this 'accuracy conservation' effect also holds for the arithmetic mean and percentiles as well. The accuracy conservation (or AC) effect is illustrated for several configurations in Table 13. All four configurations have N=25 total MUs and no cluster has more than 5 MUs. All configurations have fRA_{95} values less than or equal to that obtained with 5 clusters of 5 MUs each. The reason for the AC effect is straightforward: Variation in the cluster sizes reduces accuracy. But because of the two restrictions above, reducing the number of MUs below 5 makes it necessary to collect additional clusters. The effect of additional clusters more than compensates for the variation in cluster size. This result holds for all levels of intra-cluster correlation. If ICC=0 then all configurations would yield the same fRA_{95} as with equal cluster sizes. As the ICC increases, the benefit of having more clusters is greater and fRA_{95} is much better with more, but unequal-sized, clusters.

The accuracy conservation effect means that a configuration of equal-sized clusters can be assumed in order to establish the total number of MUs and the maximum number of MUs per cluster (i.e., to develop a scenario-specific MU Sampling Plan as described in Sec. 10 of the AHETF Governing Document). Then, whenever a full set of MUs cannot be obtained for a particular cluster, additional clusters can be used until the total N is achieved. This permits some flexibility in design and/or study conduct at the cluster level.

Table 13: Illustrating the Accuracy Conservation Effect for Unequal Numbers of MUs per Cluster when GSD = 4 and ICC = 0.3.

# Clusters, N_c	# MUs per Cluster	Total # MUs, N	95% Relative Accuracy Bound, fRA_{95}		
			Geometric Mean[1]	Arithmetic Mean[2]	95th Percentile[2]
5	5 5 5 5 5	25	2.2	2.9	3.1
6	5 5 5 5 4 1	25	2.2	2.9	3.0
7	5 5 3 3 3 3 3	25	2.1	2.7	2.8
12	4 3 3 3 2 2 2 2 1 1 1 1	25	1.9	2.5	2.6

[1]Exact calculation
[2]Based on 10,000 simulations

C12. Sampling Model for Investigating the Secondary Benchmark Objective

A more precise characterization of the secondary goal first requires that the sampling model (1) be recast in a more general form assuming that log exposure is linearly related to the log amount of ai handled:

(13) $\text{Log } E_{ij} = \alpha + \beta \text{ Log } H_{ij} + C_i + W_{ij}$

(Note that this model does not state that each individual handler's log exposure is linearly related to log H, only that this relationship holds 'on average'.) The random effects C and W have the same meaning as in (1). If the primary benchmark is based on a normalizing factor other than AaiH then H_{ij} in (13) represents the value of that factor. Of course if non-normalized exposure was used in the primary benchmark then this secondary benchmark would not be relevant.

When exposure is proportional to amount of ai handled, then $\beta=1$, $\alpha = \text{Log } GM_Q$ and equation (13) reduces to (1). That is,

(14) $$\text{Log } E_{ij} - \text{Log } H_{ij} = \text{Log } Q_{ij} = \text{Log } GM_Q + C_i + W_{ij}$$

When $\beta=0$, exposure is unrelated to amount of ai handled and (13) simplifies to:

(15) $$\text{Log } E_{ij} = \text{Log}_e GM_E + C_i + W_{ij}$$

Thus, the difference between a proportional relationship and independence can be reduced simply to whether $\beta=1$ or $\beta=0$, respectively. In this context, then, the secondary goal can be stated more precisely as:

> ***The data should be adequate so that, if the reference sampling model (13) is approximately true, the null hypothesis H_0: $\beta=0$ will be rejected (in favor of H_A: $\beta>0$) at least 80% of the time when $\beta=1$. Because of symmetry, this is equivalent to saying that the null hypothesis H_0: $\beta=1$ will be rejected (in favor of H_A: $\beta<1$) at least 80% of the time when $\beta=0$.***

This secondary benchmark is expressed in terms of a desired power to reject a particular one-sided hypothesis about the regression slope. (A two-sided hypothesis test, probably more familiar to AHED® users, would have lower power.) It is important to note that a complementary relationship exists between this pre-data power statement and the expected width of the confidence interval for β. If the hypothesis test for non-zero slope is one-sided with a 5% significance level, then it can be shown that, to a reasonable approximation, the expected difference between the upper and lower limits of the 95% confidence interval for β is:

(16) $$EWCI = \frac{2Z_{0.975}}{Z_{0.95} + Z_{0.8}} = \frac{2 \cdot (1.96)}{(1.64) + (0.84)} = 1.6$$

The quantity Z_P in (16) denotes the Pth percentile of the standard normal distribution. If the hypothesis test is two-sided, then 80% power corresponds to an expected confidence interval width of, approximately,

(17) $$EWCI = \frac{2Z_{0.975}}{Z_{0.975} + Z_{0.8}} = \frac{2 \cdot (1.96)}{(1.96) + (0.84)} = 1.4$$

Thus, the secondary benchmark can be expressed as either a target power or as an expected width of the confidence interval. From a pre-data standpoint, the power formulation is more common. However, the observed width of the confidence interval provides a simpler and more intuitive mechanism for post-data assessment of data adequacy than would a corresponding 'post-hoc' power analysis.

In addition to the GSD, ICC, N_c and N_m, the power to discriminate proportionality from independence now also depends on the specific values of the normalizing factor, H_{ij}. This ai-configuration has several aspects that need to be considered. Namely:

- the range in the amount of ai handled and
- the degree of confounding of the amount of ai handled with clusters

In general, the wider the range in ai amount the higher the power. Power is also increased when there is a large range in ai amount within clusters. When clusters have non-overlapping ranges of ai handled, then cluster effects become confounded with the effects of ai and power is reduced.

For any given scenario, the relative range in the amount of ai handled, R_H, is defined to be the ratio of the maximum to the minimum amounts. Obviously, there are an infinite variety of H_{ij} levels that can be specified for any given R_H. For the purposes of investigating power under sampling model (13), however, it is sufficient to consider just two standardized configurations of amount of ai handled. For both of these configurations it is assumed each of the $N=N_c \times N_m$ MUs have unique amounts of ai handled and that these levels are equally spaced on a logarithmic scale. That is, if H_{min} and H_{max} are the minimum and maximum amounts of ai handled in the scenario, then $R_H = H_{max} / H_{min}$ and the N different ai levels are:

(18) $$H_{min},\ H_{min} \times \Delta,\ H_{min} \times \Delta^2,\ H_{min} \times \Delta^3,\ \ldots,\ H_{min} \times \Delta^{N-1} = H_{max}$$

where

(19) $$\Delta = (R_H)^{1/(N-1)}$$

The difference between these two configurations is how the N ai amounts are allocated among the N_c clusters. If $H_1, H_2, H_3, \ldots, H_N$ denote the ordered values of amount of ai handled, then the two configurations are defined as follows:

Configuration A (minimum within-cluster variation)

In this configuration the smallest N_m ai amounts are assumed to be in cluster 1, the next smallest N_m ai amounts are in cluster 2, and so forth. For example, if N_c=3 and N_m=4 then configuration A would be:

Cluster 1 = (H_1, H_2, H_3, H_4)
Cluster 2 = (H_5, H_6, H_7, H_8)
Cluster 3 = (H_9, H_{10}, H_{11}, H_{12})

Configuration B (maximum within-cluster variation)

In this configuration, cluster 1 is allocated ai level 1, N_c+1, 2N_c+1, etc. Cluster 2 then gets ai level 2, N_c+2, 2N_c+2, etc. Again, if N_c=3 and N_m=4 then configuration B would be:

Cluster 1 = (H_1, H_4, H_7, H_{10})
Cluster 2 = (H_2, H_5, H_8, H_{11})
Cluster 3 = (H_3, H_6, H_9, H_{12})

C13. Calculating the Power to Distinguish Proportionality and Independence

The simulation method for determining the power for rejecting the null hypothesis H_0: β=0 when β=1 is as follows:

1. For each of the two configurations of amount of ai handled, simulate a set of exposure data for N_c clusters and N_m monitoring units per cluster using the sampling model defined by (13) above with β=1.
2. For each set of simulated data, perform a mixed-model regression analysis using model (13) above. Determine if the slope is significantly greater than zero at the 5% level. Also calculate the confidence interval for the slope and determine its width (WCI). Do this for both configurations A and B.
3. Repeat steps 1 and 2 above M times and tally the proportion of times that a significant result is obtained. This proportion is the estimated power to reject β=0 when β=1 is true. The average of the M WCI values is the expected width of the confidence interval, EWCI.

C14. Ability of Specific Designs to Distinguish Proportionality from Independence

It is again assumed that the residual GSD (i.e., GSD_Q) is equal to 4 and the ICC=0.3. The simulation method above was implemented in SAS using M=1,000. Strictly speaking, the highest power would result when the alternative hypothesis for the test of β = 0 is one-sided (i.e.,

H_a: $\beta > 0$). However, users of the data might tend to use a 2-sided hypothesis (i.e., H_a:$\beta \neq 0$) instead. Consequently, power was calculated for both 1-sided and 2-sided tests.

Table 14 lists the powers obtained for different values of R_H when N_c=5 and N_m=5. When the range of ai handled is only 5-fold there is insufficient power to discriminate between proportionality and independence. The (1-sided test) power for configuration B is considerably better (0.63) but still does not reach 0.8, a conventionally accepted minimum power. As the range of amount of ai handled increases, the power obtained for both configurations increases as well. It is clear, however, that configuration A always has the lower power. Configuration B has a 1-sided test power of 0.82 when R_H is only 8. But the range in ai handled must be nearly 50-fold before acceptable 1-sided-test power is obtained with configuration A. To obtain adequate 2-sided-test power with configuration A, the range in ai handled needs to be at least 100-fold. This disparity in power between situations A and B is quite reasonable: when clusters have non-overlapping ranges of ai handled, the cluster differences will tend to mask the relationship between amount of ai handled and exposure.

Table 14: Estimated Power[1] for Discriminating a Proportional from an Independence Relationship between Exposure and Amount of AI Handled for Different R_H when N_c=5 and N_m=5.

Relative Range in Amount of AI Handled, R_H	Configuration of AI Levels					
	A: Minimum within-cluster differences			B: Maximum within-cluster differences		
	Power (1-sided test)[2]	Power (2-sided test)[3]	Mean width of 95% CI[4]	Power (1-sided test)[2]	Power (2-sided test)[3]	Mean width of 95% CI[4]
5	0.26	0.16	4.4	0.63	0.51	2.0
8	0.37	0.23	3.4	0.82	0.72	1.5
10	0.41	0.27	3.1	0.89	0.80	1.4
50	0.76	0.59	1.8	>0.99	>0.99	0.82
100	0.87	0.72	1.5	>0.99	>0.99	0.70
200	0.94	0.83	1.3	>0.99	>0.99	0.61

[1] Based on 1,000 simulations
[2] 1-sided test of H_0: β=0 vs H_A: β>0 at the 5% significance level when true β = 1.
[3] 2-sided test of H_0: β=0 vs H_A: $\beta \neq 0$ at the 5% significance level when true β = 1.
[4] Average width of 95% confidence interval for β

Table 15 gives the results of additional simulations varying N_c and N_m while assuming an order of magnitude range in the amount of ai handled (i.e., R_H=10). These results show that the masking effect of clusters reflected in configuration A could be overcome by increasing the number of clusters. As N_c increases from 5 to 12 the (1-sided test) power for configuration A

increases from 0.42 to 0.8. The power for configuration B increases with N_c as well, but it was already quite large. Thus, with R_H=10, it would take 60 monitoring units in 12 clusters to achieve 80% 1-sided test power for both configurations. It would require 16-17 clusters (80-85 total MUs) to get 80% power for 2-sided tests.

Table 15: Estimated Power[1] for Discriminating a Proportional from and Independence Relationship between Exposure and Amount of AI Handled For Various Combinations of Number of Clusters (N_c) and Number of MUs per Cluster (N_m) when R_H=10.

			Configuration of AI Levels					
			A: Minimum within-cluster differences			B: Maximum within-cluster differences		
N_c	N_m	N	Power (1-sided test)[2]	Power (2-sided test)[3]	Mean width of 95% CI[4]	Power (1-sided test)[2]	Power (2-sided test)[3]	Mean width of 95% CI[4]
5	5	25	0.41	0.27	3.1	0.89	0.80	1.4
12	5	60	0.80	0.67	1.7	>0.99	0.99	0.90
16	5	80	0.89	0.79	1.4	>0.99	>0.99	0.78
17	5	85	0.91	0.83	1.4	>0.99	>0.99	0.75
5	60	300	0.80	0.68	1.7	>0.99	>0.99	0.40
5	94	470	0.88	0.79	1.4	>0.99	>0.99	0.32
5	95	475	0.89	0.81	1.4	>0.99	>0.99	0.31
16	3	48	0.80	0.69	1.6	0.98	0.97	1.0
19	3	57	0.87	0.78	1.4	0.99	0.98	0.94
20	3	60	0.90	0.82	1.4	0.99	0.99	0.92
18	2	36	0.79	0.70	1.6	0.95	0.90	1.2
19	2	38	0.84	0.74	1.5	0.94	0.90	1.2
22	2	44	0.88	0.79	1.4	0.98	0.95	1.1
23	2	46	0.89	0.81	1.4	0.98	0.95	1.1

[1] Based on 1,000 simulations
[2] 1-sided test of H_0: β=0 vs H_A: β>0 at the 5% significance level when true β = 1.
[3] 2-sided test of H_0: β=0 vs H_A: $\beta \neq 0$ at the 5% significance level when true β = 1.
[4] Average width of 95% confidence interval for β

The simulation results in Table 15 also show that the power can be improved by increasing the number of MUs per cluster (N_m) in lieu of the number of clusters (N_c). However, this approach

to improving power is very inefficient. With 5 clusters, it takes N_m=60 monitoring units per cluster to achieve 80% 1-sided test power for configuration A. (An N_m of at least 94 is needed for 2-sided test power.) Obviously, devoting a total of N=5×60=300 monitoring units to a single scenario is quite impractical. This is especially true when only N=60 MUs in 12 clusters would achieve the same power.

The efficiency of N_c over N_m raises the possibility of reducing the total N by using more clusters with fewer MUs per cluster. Table 15 shows that when R_H=10, adequate (1-sided test) power can be obtained for configuration A with N=36 monitoring units if there are 18 clusters and 2 MUs/cluster. Note, however, that when the size of the cluster (N_m) decreases, the power for configuration B also decreases, albeit only slightly.

These examples illustrate that acceptable power to discriminate between proportionality and independence can be attained with N_c=5 and N_m=5 if R_H is nearly two orders of magnitude. This range could be as small as one order of magnitude if the ai-configuration is closer to B than to A. Increasing the number of clusters would improve power, but little advantage is realized by increasing the number of MUs/cluster.

In practice, exact control over the amount of ai handled (or another normalizing factor) is difficult. Consequently, the spacing of amount of ai handled will only be approximately logarithmic and configurations will be intermediate between A and B. However, as described in Appendix B, scenario and study plans will strive toward configuration B.

C15. Summary

Nested lognormal variance component assumptions were used in a reference sampling model to determine the sample sizes necessary to achieve a 3-fold relative accuracy of distributional parameter estimates. Reasonable values for the geometric standard deviation (GSD_Q) and the within-cluster correlation (ICC) of exposure normalized by the amount of ai handled were obtained from an analysis of existing data. This analysis suggests that values of GSD_Q=4 and ICC=0.3 are reasonable defaults for both dermal and inhalation exposure. Simulation analyses indicate that N_c=5 clusters with N_m=5 monitoring units per cluster will achieve the desired benchmark goal and is more cost-effective that other feasible configurations. As long as a cluster size of 5 is not exceeded, the same total number of MUs (N=25) will also achieve this same level of relative accuracy even if the number of MUs per cluster varies.

An analogous regression model, along with two assumed configurations for amounts of ai handled, was used to investigate the power for distinguishing between proportionality and independence between exposure and amount of ai handled (or other normalizing factor). An analysis using GSD_Q=4 and ICC=0.3 indicates that adequate power is possible with 5 clusters and 5 MUs/cluster if (1) the range in amount of ai handled is nearly two orders of magnitude in general or (2) one order of magnitude if there is strong overlap between the ai levels in different clusters.

It must be emphasized that the recommended sample size of 5 clusters with 5 MUs/cluster is considered a 'default' or 'standard' configuration only. It strictly applies only to scenarios without existing data and when the default variability structure (i.e. GSD=4, ICC=0.3) and benchmark accuracy (K=3) is considered reasonable. In other cases, the AHETF will use the simulation techniques described above to develop optimal sampling plans for each scenario it addresses. This will sometimes involve considering some MUs that already exist (generally that AHETF has acquired or conducted themselves). Other, field-related considerations will contribute to these scenario plans by determining appropriate locations for each cluster, whether or not it is practical for cluster sizes to be the same, and targets for allocating an amount of active handled to each MU in the plan (see Sec. 10 of the AHETF Governing Document).

C16. References

Scientific Advisory Panel (SAP) Meeting 2007. Memorandum from M. Christian to A. Lindsay, Acting Director, EPA Office of Pesticide Programs, "Transmittal of Meeting Minutes of the FIFRA Scientific Advisory Panel Meeting Held January 9 - 12, 2007 on the Review of Worker Exposure Assessment Methods", April 2, 2007.