REPORT TITLE

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

REPORT DATE

December 7, 2006

SUBMITTED BY

Agricultural Handler Exposure Task Force
c/o Dr. David R. Johnson
Johnson Management & Consulting, LLC
1720 Prospect Drive
Macon, MO 63552
TE (660) 395-9590

Table of Contents

Benchmark Objectives of the AHETF Monitoring Program 3

Assumptions and Sampling Model for MU Exposures 4

Estimates of GSD and ICC from Existing Data 5

Determining Sample Sizes Necessary to Meet the Primary Benchmark Objective 6

Example: Determination of Sample Sizes when K=3 7

Investigating the Secondary Benchmark Objective 9

Example: Power to Distinguish Proportionality from Independence 11

Summary 12

Table 1. Scenarios and Clusters with Available Monitoring Unit (MU) Exposure Results 13

Table 2. Dermal Exposure Variability Estimates for Each Scenario 14

Table 3. Inhalation Exposure Variability Estimates for Each Scenario 15

Table 4. Relative Accuracies of Lognormal Distribution Parameter Estimates When GSD = 4 and ICC = 0.3 16

Table 5. Relative Accuracies of Lognormal Distribution Parameter Estimates Obtained by Varying ICC when GSD = 4, N_c = 5 and N_m = 5. 16

Table 6. Relative Accuracies of Lognormal Distribution Parameter Estimates Obtained by Varying GSD when ICC=0.3, N_c = 5 and N_m = 5. 17

Table 7. Relative Accuracies of Lognormal Distribution Parameter Estimates Obtained by Varying Both GSD and ICC when N_c = 5 and N_m = 5 17

Table 8. Relative Accuracies of Lognormal Distribution Parameter Estimates When GSD = 5 and ICC = 0.5 18

Table 9. Estimated Power 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. 18

Table 10. Estimated Power for Discriminating a Proportional from and Independence Relationship between Exposure and Amount of AI Handled For Various Combinations of N_c and N_m when R_H=10. 19

Benchmark Objectives of the AHETF Monitoring Program

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. Rather, its purpose is to collect sufficient data to meet specific minimum or 'benchmark' adequacy requirements. These data, possibly augmented by additional exposure data from other sources, will then be used for a variety of regulatory purposes by numerous organizations. The benchmark data objective(s) for 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 dermal exposure distribution (e.g., means, percentiles) are accurate to within K-fold when exposures are normalized by (i.e., divided by) the amount of active ingredient (ai) handled.***

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.

Using K-fold accuracy for exposure normalized by amount of ai handled as a benchmark does not necessarily imply that other types of exposure distribution will be estimated less accurately. Because many potential normalizing factors (e.g., time, concentration) 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 amount of ai handled as a data adequacy benchmark.

A secondary benchmark objective is sometimes considered for scenarios for which the practical range of amount of ai handled 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 distinguish between complete proportionality and complete independence of dermal exposure and amount of ai handled.***

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 handled. 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.

It must be emphasized that the use of amount of ai handled in these benchmark objectives is both reasonable and based on considerable historical precedent. This 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 free to consider any (or no) normalization.

Assumptions and Sampling Model for MU Exposures

The AHETF monitoring program uses purposive diversity sampling (a type of non-random sampling) to select a sample of individuals and conditions (i.e., monitoring units). This reflects very complex logistical limitations. For the purposes of determining data adequacy, however, a surrogate cluster-sampling model is used. This surrogate sampling model has the following characteristics:

- Observed exposures can be viewed as arising (at least approximately) from a random sample of clusters and then from a random sample of monitoring units within each cluster. These clusters are merely artifacts of the sampling process and are usually associated with separate studies in geographically separated locations (e.g., different states or regions). Clusters could also be viewed as studies separated by sizeable time periods (e.g., different years).

- The sampling distribution of normalized exposures within and between clusters is, at least approximately, lognormal.

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 amount of ai handled by worker for MU j in cluster i
Q_{ij} = the exposure for MU j in cluster i normalized by amount of ai handled
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

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 population standard deviation, SD. The quantity GSD_Q = antilog (SD) is the population geometric standard deviation of normalized exposure. The intra-cluster correlation (i.e., the intraclass correlation due to clusters) is defined as

(2) $$ICC = V_c / V$$

Under this sampling model, the only quantities needed to determine relative accuracy of population parameter estimates are reasonable values for GSD_Q and ICC.

Estimates of GSD and ICC from Existing Data

Estimates of the 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 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 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 only 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.

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.

Determining Sample Sizes Necessary to Meet the Primary Benchmark Objective

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

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

Thus, if θ denotes the distribution parameter of interest and T is the estimate of that parameter obtained from the monitoring data then we would like there to be at least a 95% chance that T/θ is between 1/K and 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 this benchmark objective just says that

(3) $$RA_\theta = \text{Max}\ (\ T_{97.5} / \theta,\ \ \theta / T_{2.5}\ ) \leq\ 3$$

The quantity RA_θ in (3) is defined as the relative accuracy of the estimate for the distributional parameter θ.

A straightforward simulation approach can be used to determine relative accuracy of T relative to the parameter θ:

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, respectively.
5. Calculate the relative accuracy RA_θ from definition (3) above.

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

Relative accuracy can be computed once the samples sizes and the method for computing the various T from the data are specified. This raises the question of which types of estimates of the geometric mean, arithmetic mean, and 95th percentile to use. For example, one must choose between:

- Empirical or lognormal estimates, and
- Simple random sampling versus cluster random sampling estimates.

The relative accuracies could differ depending on the particular estimates used. At first glance it might seem best to use the estimation method that is most closely aligned with the sampling model used. In this case one should probably use 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 get the arithmetic mean (AM_Q) and 95th percentile (Q_{95}) one could then use the lognormal relationships:

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

(4)

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

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

Alternatively, 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 would be used (e.g., the standard formulas used in spreadsheet programs such as Excel). The 95th percentile could also be calculated empirically as if the data were a simple random sample from an arbitrary distribution. But data users might be less inclined to do this, especially with smaller sample sizes. The lognormal percentile estimate in (4) above would then still be used but 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, these will be labeled the 'simple random sampling estimates'.

Example: Determination of Sample Sizes when K=3

As an example, consider the case where an approximate 3-fold relative accuracy is desired. It is assumed that the true GSD_Q and ICC are 4 and 0.3, respectively. The simulation method above was implemented in SAS with M=10,000 and relative accuracies were calculated for different combinations of N_c and N_m. These results are listed in Table 4. Both simple random sampling and cluster random sampling estimates

were computed. However, since both types of estimates gave very similar relative accuracies, only the simple random sampling results are presented.

From these results it appears that, on balance, a sample of 5 clusters with 5 MUs per cluster (a total of N=25 monitoring units) will achieve around 3-fold relative accuracy under sampling model (1). The geometric mean will 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) will have greater accuracy than the 95th percentile and higher percentiles will be less accurate.

Increasing either N_c or N_m will improve the accuracy, 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. A configuration of 6 clusters with 4 MUs/cluster is slightly better than (N_c, N_m) = (5, 5).

It is important to consider how sensitive the relative accuracy with (N_c, N_m) = (5, 5) is to the assumed values of GSD_Q and ICC. Table 5 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 reduces the relative accuracy slightly and decreasing the within-cluster correlation improves it. These changes are modest, however.

Table 6 shows the effect of changes in GSD_Q on relative accuracy. As was the case with ICC, increasing GSD_Q when ICC is fixed at 0.3 makes relative accuracy worse and decreasing GSD_Q improves accuracy. Overall, however, the effects of changes in the GSD of ±1 do not appear substantial.

Table 7 shows more extreme situations in which both GSD_Q and ICC are perturbed together. As would be expected the worse case is seen when both GSD_Q and ICC increase. 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 accuracy will occur. Of course, if both parameters decrease, great improvements in accuracy can occur.

Table 8 demonstrates what configurations of (N_c, N_m) would be necessary to achieve 3-fold accuracy for the worst-case situation shown in Table 7. It appears that with 5 MUs/cluster, approximately 9-10 clusters (45-50 MUs total) would be necessary. With 3 MUs/cluster it would take about 11 clusters (N=33) to achieve the same degree of accuracy.

In this example, 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. In this case, greater gains can be achieved by focusing on more clusters with fewer MUs/cluster.

Investigating the Secondary Benchmark Objective

A more precise characterization of the secondary goal first requires the sampling model (1) be recast in a more general form assuming that log exposure is linearly related to log amount of ai handled:

(5) $$\text{Log } E_{ij} = \alpha + \beta \text{ Log } H_{ij} + C_i + W_{ij}$$

When exposure is proportional to amount of ai handled, then $\beta=1$, $\alpha = \text{Log } GM_Q$ and equation (5) reduces to (1). That is,

(6) $$\text{Log } E_{ij} - \text{Log } H_{ij} = \text{Log } Q_{ij} = \text{Log } GM_Q + C_i + W_{ij}$$

If $\beta=0$ in (5), then exposure is unrelated to amount of ai handled. In this case (5) simplifies to:

(7) $$\text{Log } E_{ij} = \text{Log}_e GM_E + C_i + W_{ij}$$

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 sampling model (5) 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 $\beta=0$.***

In addition to the number of clusters (N_c) and the number of MUs/cluster (N_m), the power to discriminate proportionally from independence also depends on the specific values of ai handled, H_{ij}. This ai-configuration has several different 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 (5), 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:

(8) $$H_{min},\ H_{min}\times\Delta,\ H_{min}\times\Delta^2,\ H_{min}\times\Delta^3,\ \ldots,\ H_{min}\times\Delta^{N-1} = H_{max}$$

where

(9) $$\Delta = (H_{max}/H_{min})^{1/(N-1)}$$

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})

In practice, exact control over the amount of ai handled is difficult. Consequently, the spacing of amount of ai handled will only be approximately logarithmic and configurations will be intermediate between A and B.

A simulation method for determining the power for rejecting the null hypothesis H_0: $\beta=0$ when $\beta=1$ 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 (5) above with $\beta=1$.
2. For each set of simulated data, perform a mixed-model regression analysis using model (5) above. Determine if the slope is significantly greater than zero at the 5% level. Do this for both configurations A and B
3. Repeat steps 1 and 2 above for M times and tally the proportion of time that a significant result is obtained. This proportion is the estimated power to reject $\beta=0$ when $\beta=1$ is true.

Example: Power to Distinguish Proportionality from Independence

As an example, 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. Table 9 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 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 power of 0.82 when R_H is only 8. But the range in ai handled must be nearly 50-fold before acceptable power is obtained with configuration A. The reason for this disparity is straightforward: 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 10 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 can be overcome by increasing the number of clusters. As N_c increases from 5 to 12 the 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% power for both configurations.

The simulation results in Table 10 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% power for configuration A. 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 few MUs per cluster. Table 10 shows that when R_H=10 adequate power can be obtained for configuration A with N=36 monitoring units if there are 18 clusters with only 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.

This example illustrates that acceptable power to discriminate between proportionality and independence can be attained with N_c=5 and N_m=5 if R_H is sufficiently wide or if the ai-configuration is closer to B than to A. Increasing the number of clusters will improve power, but little advantage is realized by increasing the number of MUs/cluster.

Summary

Nested lognormal variance component assumptions can be used in a surrogate-sampling model to determine the sample sizes necessary to achieve any desired level of accuracy of distribution parameter estimates. The only requirements are reasonable values for the geometric standard deviation of normalized exposure (GSD_Q) and the within-cluster correlation (ICC). Analyses of both dermal and inhalation exposures from several scenarios suggests that GSD_Q=4 and ICC=0.3 are reasonable for both dermal and inhalation exposure. An example analysis using a 3-fold accuracy requirement suggests that 5 clusters with 5 monitoring units per cluster achieve this benchmark goal.

The same model, along with two assumed configurations for amounts of ai handled, can be used to investigate the power for distinguishing between proportionality and independence between exposure and amount of ai handled. An example analysis using GSD_Q=4 and ICC=0.3 indicates that adequate power is possible with 5 clusters with 5 MUs/cluster if (1) the range in amount of ai handled is greater than an order of magnitude or (2) there is strong overlap between the ai levels in different clusters.

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]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

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]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

Table 4. Relative Accuracies of Lognormal Distribution Parameter Estimates When GSD = 4 and ICC = 0.3.

# Clusters	# MUs per Cluster	Total # MUs	Relative Accuracy for		
			Geometric Mean	Arithmetic Mean	95th Percentile
3	5	15	2.8	3.9	4.3
4	5	20	2.4	3.3	3.5
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	2.4	3.3	3.5
4	6	24	2.4	3.1	3.3
4	8	32	2.3	3.0	3.2
4	10	40	2.3	2.9	3.0
4	4	16	2.6	3.5	3.7
5	4	20	2.4	3.1	3.3
6	4	24	2.2	2.8	2.9

Table 5. Relative Accuracies of 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	Relative Accuracy for		
		Geometric Mean	Arithmetic Mean	95th Percentile
4	0.5	2.5	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	2.8
4	0.1	1.9	2.4	2.5

Table 6. Relative Accuracies of Lognormal Distribution Parameter Estimates Obtained by Varying GSD when ICC=0.3, N_c = 5 and N_m = 5.

Geometric Standard Deviation	Intracluster Correlation	Relative Accuracy for		
		Geometric Mean	Arithmetic Mean	95th Percentile
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

Table 7. Relative Accuracies of Lognormal Distribution Parameter Estimates Obtained by Varying Both GSD and ICC when N_c = 5 and N_m = 5.

Geometric Standard Deviation	Intracluster Correlation	Relative Accuracy for		
		Geometric Mean	Arithmetic Mean	95th Percentile
5	0.5	2.9	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	2.1

Table 8. Relative Accuracies of Lognormal Distribution Parameter Estimates When GSD = 5 and ICC = 0.5.

# Clusters	# MUs per Cluster	Total # MUs	Relative Accuracy for		
			Geometric Mean	Arithmetic Mean	95th Percentile
5	5	25	2.9	4.7	4.7
7	5	35	2.5	3.8	3.8
8	5	40	2.4	3.5	3.4
9	5	45	2.2	3.3	3.2
10	5	50	2.1	3.1	3.0
10	4	40	2.2	3.1	3.1
10	3	30	2.3	3.3	3.2
11	3	33	2.2	3.1	3.1

Table 9. Estimated Power 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.

# Clusters, N_c	# MUs per Cluster, N_m	Total # MUs, N	Relative Range in Amount of AI Handled, R_H	Power when the Configuration of AI Levels is	
				A: Minimum within-cluster differences	B: Maximum within-cluster differences
5	5	25	5	0.26	0.63
5	5	25	8	0.37	0.82
5	5	25	10	0.41	0.89
5	5	25	50	0.76	>0.99
5	5	25	100	0.87	>0.99

Table 10. Estimated Power for Discriminating a Proportional from and Independence Relationship between Exposure and Amount of AI Handled For Various Combinations of N_c and N_m when R_H=10.

# Clusters, N_c	# MUs per Cluster, N_m	Total # MUs, N	Relative Range in Amount of AI Handled, R_H	Power when the Configuration of AI Levels is	
				A: Minimum within-cluster differences	B: Maximum within-cluster differences
7	5	35	10	0.55	0.96
10	5	50	10	0.70	0.99
12	5	60	10	0.80	>0.99
5	7	35	10	0.42	0.95
5	10	50	10	0.48	0.99
5	12	60	10	0.50	>0.99
5	50	250	10	0.75	>0.99
5	60	300	10	0.80	>0.99
15	3	45	10	0.79	0.98
18	2	36	10	0.79	0.95