Background

As part of the meeting on October 17-18, the two industry-sponsored pesticide handler exposure monitoring data development task forces (AHETF and AEATF-II) presented the scopes of their respective projects to populate a scenario-based database to support regulatory exposure assessments. Monitoring data that will be used to populate the database will be collected only once for each scenario, and is expected to remain useful for the foreseeable future, before an update would be considered. Both Task Forces have proposed a purposive diversity sampling approach. The EPA Human Studies Review Board (HSRB) has commented that the purposive diversity sampling approach is inadequate, and recommended that elements of randomization should be incorporated into it to make the resulting database more defensible from a statistical perspective and more useful to EPA.

Follow-up Questions

EPA would appreciate your written thoughts on the following questions. In responding please explain your rationale:

1. What would be the scientific limitations of the inferences about the population of regulatory concern – pesticide handlers in the United States – that could be drawn from databases containing exposure monitoring values collected through the purposive diversity sampling design proposed by the AHETF and AEATF-II? To what extent would these limitations be mitigated by the inclusion of elements of a randomized sampling approach?

Response: The first and foremost limitation would be making any statistical inference. Since there is no randomization, any statistical statements on the basis of the PDS will be hard to validate scientifically. The secondary limitation is the richness of the database. This data may not be useful for any secondary analysis since the quality (statistical) of the data is unknown. Furthermore, it will be hard to build any statistical model on this data base with supplement information from any related database if necessary. For instance, it will be hard for one who likes

to `borrow information' from other resources due to its small sample size, (insufficient information) or like to combine with future data collection. Among many other limitations, there is no statistical way to know the information about the non-sampled units of the universe, or comparing the exposure level from two different sites.

By including a few randomization would positively mitigate the above limitations to some extent, and would depend how and at which level the randomization is made. For example, if a group of states is judged as fairly uniform, and the sampling design allows only one state to be selected due to cost constraint, the state could be chosen purposively due to operation constraint. However, from the *designated* (as opposed to selected) state, a random sample of subjects (farm, day, time, scenario) can be selected. The more details can be worked out depending on the complexity, and the subjectivity at the different level of the universe. This way, the study can be validated at least for the designated state, if not for the group of states. Most importantly, the collected data will be statistically valid and thus can be used for any future and secondary study, at least for the designated state.

During my visit October 17-18, I came to know that the unavailability of proper sampling frame. This limitation often arises in any survey and the decisions are usually made case-by-case basis either by constructing an approximate frame or even from gathering data from multiple incomplete frames.

2. The Task Forces contend that the use of on expert opinion and professional judgment to devise a purposive diversity sampling strategy would be a more efficient method of obtaining useful information about typical single-day exposure values for various exposure scenarios than a simple random sample or stratified random sample. In reviewing the literature on sampling strategies, the EPA found the following:

The concern about the bias of survey estimators from a judgment sample – or any non-probability sample – increases with sample size. Consider the comparison of a sample estimator from a judgment sample and that from a probability sample of the same size. If the sample size is very small, the variance of the probability sample estimator will be very large, so that in

relative terms, the bias of the judgment sample estimator may be unimportant. However, as the sample size increases, the variance of the probability sample estimator decreases while the bias of the judgment sample estimator may change little. This reasoning provides a justification for non-probability samples when the sample size is small with a change to probability sampling for larger sample sizes. Thus, for instance, if a researcher can conduct a study in only one or two cities, it is probably better to select the cities by expert choice rather than to rely on the vagaries of random chance which could easily result in an odd sample. If, however, the sample size is increased to 50 cities, then a carefully stratified probability sample would almost certainty be preferable (Kalton, 1983).

Please comment on this argument in the context of the Task Forces' plans to employ a sampling design for scenarios that involves 5 clusters of 5 subjects (AHETF) or 3 clusters of 6 subjects (AEATF).

Response: Note that in the last part of Kalton's comment, one city is selected from two cities. Clearly one has very through idea about both the cities – meaning having clear idea for the non-sampled one given the sampled one. Now, in case of Task Forces' design, do you have information about the non-sampled clusters or subjects? I believe not. Hence the PDS design may not be suitable here. On the other hand, as is mentioned above, choosing the state purposively, knowing completely about all the states in the group, selecting the subjects randomly would be better design (although, this may not be state-of-the art solution). Along with Kalton's comment, let me provide you the following from Jessen (1978): Judgement selection has its strongest case where (i) sample is small, (ii) the universe is fairly small and visible or known to the selector, (iii) the elements in the universe vary considerably in the character under investigation, and (iv) the selector has great and proven skill in this art. It's EPA's responsibility to judge the quality of PDS in respect of above criterions, but as far as I can tell, the number of subjects is not too small or the selector has complete knowledge for all subjects (sampled and non-sampled). Thus, I would re-iterate that at least selecting the subjects randomly (having complete knowledge about the scenarios) would greatly improve the data quality.

3. Can you recommend a feasible random sampling strategy that would result in a database of equivalent size, obtained at comparable cost to what the Task Forces' proposed to do using a purposive diversity sampling strategy?

Response: This has been indicated in above two points and clearly partial randomization, for example, selecting the subjects randomly is one good possibility. The detail sampling design would require more in depth thought on cost structure and group efforts. For example, the states and the scenarios can be *identified* by the exposure experts and the farms, individuals can be selected randomly. This procedure can also be driven by cost and time constraint to some extent to keep everything into a manageable level.

4. Please comment on the HSRB suggestion that adding some elements of randomization to the purposive diversity approach could improve the utility and reliability of the information generated by the Task Forces? What specific advantages would a partially randomized purposive diversity sampling design have over the purposive diversity sampling design proposed by the Task Forces? What possible elements of randomization do you think could be incorporated into the proposed purposive diversity approach to add to the statistical utility of the data, within the budget constraints discussed at our meeting on October 17-18? Can the impact of including such elements be quantified? If you think these goals cannot be achieved within the current budget, how much more investment do you think would be required and how long would it take to develop such a design? What advantages would a fully stratified random sampling design have over a purposive diversity sampling design incorporating such elements?

Response: I think my response for this point lies in point 1 -3. The fully randomized sample with effective sample size would definitely resolve all the issues, and could be the best option. However, that would definitely increase the total cost and time significantly (I am unable to give exact figures at this time). As I have indicated, some level of randomness (selecting the subjects randomly) may be the best option and explain the top level selections (state, scenarios) using four criterions mentioned in point 2. I am sure, for this, cost cannot be sky-rising. I can work on this with you to come up with a partial random sampling design, cost and time.

5. The Task Forces reported at our meeting about their experience with very low response rates (≤ 5%) to random contacts. How would you recommend dealing with this problem in the context of the AHETF and AEATF-II monitoring program sampling designs? In the absence of any follow-up on non-respondents, to what extent would such a low response rate detract from any advantages that a partially randomized purposive diversity sampling design would have over the purposive diversity sampling design proposed by the Task Forces?

Response: Such a low response rate would even go against the PDS design, because, the participants are evidently from the response group, and there is no information who are potentially non-respondent and clearly only a small segment of the population is represented, not the entire population. For the sake of arguments, if EPA is happy with only this small segment of population, then even taking a random sample from this segment would provide better validity with a clause that the study valid for `specific' segment of universe. The Task Force's effort on collecting information from the non respond group is not convincing.

6. At our meeting we discussed in some detail the process that would be required to identify groups of appropriate pesticide handlers from which test participants could be recruited. There is limited scientific data to define which factors other than the amount of active ingredient handled potentially affect handler exposure, but expert judgment can guide the selection of factors likely to be influential in each defined scenario. Once those factors are identified, wouldn't the purposive diversity sampling approach and a stratified random sampling approach both require similar follow-up steps to identify appropriate small groups from which to recruit participants or conditions for monitoring? If similar expert judgment is required to define potential pools of individuals to be monitored, what does this mean for the relative reliability of each sampling approach? Are the two approaches effectively identical through the point of identification of these potential pools of candidates from whom to select test participants? If not, how would they differ?

Response: Actually, the expert judgment would be very useful in identifying the appropriate pesticide handler groups, but this would be only meaningful when EPA is certain that this is the population you want to survey, as the rest is not important. Once this *effective* population is defined, the random sampling design should be adopted to have scientifically (statistically) valid data set. Thus the two approaches are not identical rather expert judgment must be used for a better (random) sampling design.

7. We understand that it is not possible to quantify the uncertainty or bias resulting from decisions (based on expert judgment) to define sampling pools under a purposive diversity scheme or to define strata in a random sampling scheme. Please discuss how to judge whether the potential for greater statistical reliability achievable from a random sampling design—or from a purposive design incorporating some random elements—would justify the likely increase in the cost of the research. For example, how could EPA judge the relative utility of a database containing fewer monitoring units of greater statistical reliability obtained at higher unit cost compared to one containing more monitoring units of lesser statistical reliability obtained at lower unit cost?

Response: There is no clear answer, since PDS and random sample are not comparable unless one knows the truth. For a given scenario, the PDS can be right strategy in light of four criterions mentioned in point 2 and may not be useful for a deviation from the scenario, but the random sample is more on all-purpose for constituting a rich database for future research, and general applicability and validity.

References

Jessen, R.J. (1978). *Statistical Survey Techniques*, Wiley

Kalton, G. (1983). *Introduction to Survey Sampling* (Sage University Paper Series on Quantitative Application in the Social Sciences, No. 07-035). Beverly Hills, CA: Sage.