FIFRA SCIENTIFIC ADVISORY PANEL
OPEN MEETING

REVIEW OF WORKER EXPOSURE
ASSESSMENT METHODS

U.S. ENVIRONMENTAL PROTECTION AGENCY
CONFERENCE CENTER- LOBBY LEVEL
ONE POTOMAC YARD (SOUTH BLDG.)
2777 SOUTH CRYSTAL DRIVE
ARLINGTON, VIRGINIA 22202

JANUARY 12, 2007

FIFRA SCIENTIFIC ADVISORY PANEL (SAP)
Review of Worker Exposure Assessment Methods
January 12, 2007

DR. HEERINGA: Everybody's in a hurry to get out of there. That quieted it down and, okay. I think we're qualified to begin, here, so, let's get underway. Good morning, everyone, and welcome to the fourth and final day of the meeting of the FIFRA Science Advisory Panel, and the topic of review of worker exposure assessments.

As most of you well realize by now, we've had three days, I think, of very successful and informative presentations and discussions. And we are wrapping up this morning with an update, potentially some new information, added information, initially in the session, and then we are going to turn to the sixth and final charge question that was presented to the panel by the EPA. And that's the one dealing with sample size.

I think a very critical, very critical issue. This point in time, I'd like to turn to, first, Jeff Evans and Jeff Dawson, to see if there are any follow up comments or information that you'd like to provide us from the previous dates, and Jeff Dawson.

MR. DAWSON: Jeff Dawson. No, I think

we're fine, and we appreciate the level of the discussions. Thank you.

DR. HEERINGA: I, we have one additional, Tim Leighton from the antimicrobials division of the EPA, spoke to me at the end of the day yesterday, wants to also bring in a few points about their design, because it's relevant to the sample size discussion. Again, we probably shouldn't confuse it with the discussion on the age, ETF, the ag handlers exposure task force plan, but I want to make sure that we also address it, because they're part of this meeting as well.

So, Tim, if you would just like to give your introduction, we'll make sure that, before we wrap up today, that we get a specific discussion and focus on your issue, to the extent that it may differ from the AGTF plan.

MR. LEIGHTON: Thank you. Again, I'm Tim Leighton from the antimicrobials division. Day one, back on Tuesday, Dr. Cassie Walls made a presentation and discussed the antimicrobial side. She went over the similarities and differences. It's been a good meeting for the antimicrobials. We didn't have as much date to go back on to bring up these examples, so we were very happy to piggyback on what HED has been

presenting over the last three or four days. And we're going to have a lot of information to take back, make revisions to the protocols, and move forward.

What I want to do now for this one is just to highlight some of the differences that Cassie had mentioned in day one, so when we discuss clustering and the number of samples that are needed, that we also look at what we have for the antimicrobial database. So, in the next slide, and what this, actually, specifically, pertains to is within the first day as presented by the antimicrobial task force, Dr. Ryan Williams.

On slide number 8, we don't need to pull it out, but it just contained a table of 19 exposure studies that are going to be collected, scenarios, actually; and each scenario has one study and 15 replicates. Next slide, and the basis for why it is, why it is what it is, is that exposure scenarios of interest have been identified. Initially, there, I think there was 23 exposure scenarios.

They have been discussed by EPA, PMRA, and also California Department of Pesticide Regulations. And what we decided on is that we can get what's, what we want as representative antimicrobial type uses with about 17 handler scenarios. The data have been

required or in the process of, through data call ins, through our re-registration eligibility decision documents, also known as reds. Each one of them has call-, confirmatory data call in for each of these, basically, 17 exposure scenarios. So, the task force, instead of having each company repeat the same study 17 times, you know, it's best to do this by task force, and as we've discussed, there's a lot of history to that.

And where the 15 monitoring units came for each study, each scenario, they were, initially, based on EPA guidelines, the guideline recommendations. So, from here, what I want to point out and highlight, there's differences. In the initial plan, I'm calling it the initial plan because I know the task force, EPA, Health Canada, and California, we're going to make modifications from what we've learned.

But, I would just want to go over an example. So, for an aerosol spray study, that's in the sample plan. There's going to be one study, 15 replicates, or 20, they may have said 20. We will be looking at an applicator applying aerosol spray with no gloves. We'll be using a disinfectant. That's the initial plan. And we get into, what we're calling simulating this study in a laboratory. First, there'll be a pilot

study. And we're going to look at 4 scenarios. We'll be looking at hard surfaces, soft surfaces, such as somebody applying this to bedding, to couches, et cetera. Look at aerosols for air sanitizers in a room. Looking at exposures, also, to foaming products they put on a toilet.

And, basically, within the laboratory, they'll be setting up two or three bathrooms. And a bathroom will, each bathroom will consist of a sink, shower, and toilet. So, a configuration there, you can, you know you have single sinks, double sinks, showers, stall, a bathtub. These are things that we're going to consider. And one monitoring event or monitoring unit, however we want to call that, is going to be a separate individual spraying this aerosol product over the sink, shower, toilet. And that will give us, basically, in this context for discussion, one cluster, 15 monitoring events.

Now, other options we're looking at, and I think, this is the reason why I wanted to bring this up for clarification today, if we discuss is there a need for clusters, and using clusters, clusters different sites, if that effects the number of samples versus having one simulated site, how many samples we would need. You know, we had been discussing for the AG

side, looking at orchards. So, I want to bring us out of the orchards, now, and into the bathroom, so, if you actually did a field site for a bathroom, you know, what could you do. We, I, you know, there's different, you know, there's apartment buildings, townhomes, single family homes, would we want to find five apartments, monitor five different people, and so forth.

Next slide, and this is the final slide. So, for consideration by the SAP here today, it would be, it, be good to discuss the clustering options of having a one simulated site. How many monitoring units or monitoring events would we need, versus if we go to three field locations. And field locations, you know, either geographic, I'm not sure how we'd want to define that, but it also could be apartment, townhouses, houses.

And then, final note here, 'cause I know this question came up yesterday, is, well, to figure out how many sample sizes you want, what part of the distribution are you going to regulate on. And for us, now, for the antimicrobials, and I think, you know, for a lot of them, we have not selected that end point, what we're going to regulate on yet. Most of our assessments are deterministic. We have done one

assessment, and that was the, the sheds wood that we worked with, ORD. In fact, it came to the SAP. That would be interesting to find out where we regulate on that one, except for the fact that, that the product already has been canceled, so we're not regulating on it.

We're going out and just show them what we have. If somebody wants to see the 99 percentile, they can see it, but you know, we still haven't made a selection. You know, it would've been nice to, you know, for us to make a decision, but we haven't. And, then, again, and finally, what are we looking at for most of these antimicrobial ones. Lot, you know, most of them are short term, short to intermediate term, actually; but we do have some chronic scenarios that are like a metal work in fluid, where somebody's everyday using a material preservative within that fluid, the machinist, and they're exposed every day. So, I mean, at this point, with the data we have, we want to make an incremental leap to have more data, it's still going to be deterministic.

We'll certainly be interest in the central tendencies. And from there, if we have to go to a higher tier, such as a probable-, probabilistic assessment, we will have the option for that particular

chemical to go back and require more data. So, even if we have, let's say, 15 replicates for painting and PHED, the task force can monitoring, at this point, another 15, now we have 30.

If we need 45 or such, or 50 or 60, for a specific probabilistic assessment, we can make those registrants go back, collect additional data at that time, and continue forward. And that's just what I wanted to say for, to make sure that this type of consideration gets input today for clusters and sample sizes. Thank you.

DR. HEERINGA: Thank you very much, Dr. Leighton. Members of the panel, any questions for Dr. Leighton on the specific elements of the data needs within the antimicrobials division as they may be distinct from that in general agricultural pesticides? Dr. Johnson.

DR. JOHNSON: On the, I'm thinking about the bathroom study and the simulation thing. I guess, I'm wondering whether wiping down a dirty bathroom is different than wiping down a clean bathroom, in terms of the amount of exposure someone might get. And so, how do you, how do you get dirty bathrooms, I guess, and, in a simulation study or in a, in a study that involves clusters.

DR. LEIGHTON: The, although these will be simulating in a laboratory, the toilets won't be connected to the plumbing, so.

[All laugh.]

DR. LEIGHTON: But, I mean, that's a fair question. And it's also a fair question for the mopping study. There's a mopping study that, rent a wedding reception hall and go in and clean the floor. Now, if we do that cleaning the floor 15 times, they're not going to get 15 weddings to go through there.

So, you know, this is a concern, but I think, overall, for mopping, I'd, and also for spraying, you know, if it's dirty or not dirty, as long as somebody is doing the job, at this point, that what's we were going to go with. Might have to think about spraying dirt around. I'm not sure.

DR. HEERINGA: Dr. Portier.

DR. PORTIER: The answer is college students. It's always the answer, right.

[All laugh.]

DR. HEERINGA: I see we're going to have to put a bridle on him this morning. Okay. Dr. MacDonald, please.

DR. MACDONALD: So, would it be fair to say that the sources of variation in these studies are

considerably less than what we've been dealing with up to now, with the farm workers using different spray application, that this is really much more like a controlled experiment. In fact, in practice, it would seem to me that there's going to be less variability in the use of these chemicals?

DR. LEIGHTON: Yes, for some of them, that's certainly the case. When we get to mixing load in studies, there will be some variability, because, again, there are thousands, tens of thousands of products.

They're packaged from small containers up to five gallon buckets for open pouring. So, for that one, we're going to see some more variation. But if you were spraying a sanitizer in a room, and if, and we would control for more of the worse case, where the air exchange rate would be potentially zero. And spray for ten, twenty seconds, whatever's going to be representative, I would think that, yes, we would see less variability.

DR. HEERINGA: Thank you, Dr. Leighton. We'll definitely consider this also, not only in the discussion, but in our report response, too, I think we'll make sure that we single out or distinguish any unique features of your particular situation that are

distinct from those for the general agricultural worker exposure.

DR. LEIGHTON: And if it be considered fairly generically, 'cause, again, we've got 17 of these scenarios, and there is going to be changes. I know there'll be changes from that.

DR. HEERINGA: You have almost as much variability as we've seen in the ag exposure scenario. Well, not quite, but some at task, clearly, appear to be much more systematic, and this is a controlled, as yours suggest, but, okay. At this point, do we have any additional comments from other groups that have presented from either of the exposure task forces, any updates, any corrections or clarifying information? Dr. Landenberger.

DR. LANDENBERGER: I just wanted to add a little bit to what Tim Leighton has presented as well for the antimicrobials. When we initially started with those, we were looking at simulation, because we were trying to cover some very broad waterfront. The initial matrix that we looked at was a, approximately, 19 by 13, 19 application methods, 13 general categories of usage. And the request was that we fill the cells completely, which was beyond the capacity of our task force to do. So, we started looking at breaking down

some of the tasks into segments that we could then combine over particular use patterns. That's why we had a pour liquid study that we were looking at doing, in terms of a simulation mode.

As Tim pointed out, there are tens of thousands of applications that use either pour liquid or pour solid or some of the other studies that we have. And this has caused us to stop and think, how do we get at that type of information. The other issue that comes along with this, that we are looking at simulation in some of these cases as well, is the extremely low quantities that we're using, often milligrams inside of a bucket of active ingredient, and that's it; which has caused us to try and look at simulation, what we can control, some of the confounding factors.

One of the products we've at for some of these studies is didec, which if you're familiar with a spray, trigger spray, a lot of them use it. So, we have to have a controlled environment so we can figure out what's coming from the usage in that particular case, and avoid having interferences from what might just be in the household already presented. So, these are some issues that have come into play in our, in terms of our thinking on how we have to try and get at

this information. And how we can try and get it across this entire grid of 13 by 19 that we initially started out with. Tim's correct. This has been useful for us. I appreciate the discussion that's gone on with the SAP.

We certainly are trying to re-think of some ways we can try and get at this. We have some issues that are going to be hard for us to deal with, just because of our detection limits being low. The quantities used being low. And trying to cover as much as we can, in terms of usage, which is quite, quite broad. Just to give you a little thought starter, if you have something on the shelf that can degrade biologically, it probably has a biocide, and that's in-can preservative and all those need to be covered by a pour liquid study, or they have to be covered by some of the other usages that might go along with that, brush roller if it's a paint.

But, as you can see, there's just thousands of products that would have biocides in them, and this has made it a little bit more difficult for us to try and figure out how do we get our arms around this. How do we get our arms around just a normal usage pattern for an in-can preservative. These are some issues that are probably a little bit different than what the ag

guys are dealing with. And so, we're looking for some solutions. We'd appreciate input that the SAP could give us on that.

DR. HEERINGA: Thank you very much, Bryce. Okay, at this point in time, I think, again, if we are, this is a critical discussion question this morning. And if we reach any points in the discussion where I feel we need clarification of information, I may invite individuals to return to the mike, but otherwise, we'll confine it to panel discussion.

So, either Mr. Miller or Mr. Dawson, Mr. Evans, could you read the charge question number six into the record for us, please. Before you begin, Dr. Portier just reminded me, too, that we really didn't answer in detail the second part of question number five.

And what I would propose, is that we'll read charge question six into the record, and that panel members, when you respond to charge question number five, can fold in whatever sample size related comments that you had. Excuse me, to question six, fold in whatever sample size related comments you have from question number five. So, sorry for the interruption, please.

MR. DAWSON: Charge question six. The

agency's goal is to ensure that monitoring studies rely on sample sizes that adequately represent the range of exposure of people who engage in a particular handler scenario and activity.

It is also recognized that occupational monitoring studies are costly, and have many logistically obstacles. The agency is also concerned about limiting the numbers of participants in these studies, in accordance with the ethical requirements described in sub-part K, 40 CFR 26, and the recent criteria outlined by the human's, agency's human studies review board.

The agency's current guidelines recommend 15 monitoring units for each scenario. In addition, the AHETF has provided a rationale for the number of samples and their study design. Please comment on the uncertainties associated with the agency's and AHETF's recommended number of monitoring units.

In your comments, please include any recommendations you may have regarding specific statistical analyses that may assist the agency in developing better understanding of these uncertainties and characterizing them in a complete and transparent manner in agency assessments based on these data.

DR. HEERINGA: Our lead discussant is Dr.

Johnson.

DR. JOHNSON: Thank you. That's a tough charge. The, particularly, the last sentence there, complete and transparent manner. I'm not sure I've ever been able to do that. Sample size questions are always tough questions.

I think back a number, over the years, a number of scientists that have come into my office and says, how many samples should I take. And I, usually, respond with, well, how many were you planning to take. And if they answer a thousand, I say, well, you probably don't need that many. If they answer ten, I might say, can you take a few more. So, it's, it's a tough, it's a tough thing to say when you have enough and when you don't have enough.

Particularly, well, you're in better, in a better situation here, because you have some old data to work with to, sort of, help give you some guidance as to what types of samples to have, or how many samples to take. Yesterday afternoon, as we were getting ready to leave, it seemed that maybe the emphasis in this charge might have shifted a little bit from, exactly, how many monitoring units, and how they would be selected, to the process that the task force is planning to use to determine the sample size. And,

that's a lot easier charge. I think that the task force has done an excellent job in looking at past data, doing some simulations, talking about K fold accuracy.

And I'm pretty happy with the process that they've used to, sort of, get some idea as to an appropriate number of samples. And, for the antimicrobial group, I guess, they might need, since they don't have all that much data to start with, they might need to do some pilot studies to, and maybe try to do, go through a similar kind of process that the agricultural handlers exposure task force has done. I think Dr. Portier is going to come at things from a lot different perspective, and probably has a lot more good suggestions to add than what I might have.

So, the rest of my comments are just sort of generic in some sense, things that we've already said, I think, and, but, and may not even need to be re-said, but I'm going to re-say them anyway. First, if you're going to fix the total number of monitoring units, then I agree, that it's generally better to have more clusters, and fewer numbers of monitoring units per cluster, than it is to have fewer clusters and more monitoring units per clusters. I think we've discussed a little bit already, and everybody seems to be in

agreement with that particular statement. In the EPA report, botanical report, there's a table 5.2 that suggests the number of monitoring units that will be selected for each scenario, but that table did not give any information as to the number of clusters and the number of monitoring units within each cluster.

And so, that table might need to be updated when that kind of a decision is finally made. Also, when reading the report, the report seemed to have a fair amount of information about what data will be collected, but there was not much information about how the date will be collected.

There wasn't anything that I could find in the report when I was reading it, as to how the clusters will be selected, how the monitoring units within a cluster will be selected. Will there be control over the amount of ingredient handled. I think the, the task force has indicated that there will. How the participants are selected, and when the data will be collected. Fortunately, the presentations did a lot better.

The presentations that were given seemed to address the answers to a lot of these kinds of questions, and I'm much happier with that now, than I might have been earlier. Now, the problem with this

is, those answers to those questions might, actually, have been in the report, but there was 114 pages of report, and I, there's lots and lots of materials, and I had trouble finding everything.

Sometimes with respect to sample size, last night, I just loaded the report, and then did a search for the word size or sample or and equals to see what I could find, and there, really, wasn't much in any of the reports about that. The latter part, then, and this, sort of, addresses the second question, I guess, include recommendations you may have regarding specific statistical analyses that may assist the agency in developing a better understanding of the inherent uncertainties, and characterizing them in a complete and transparent manner.

As I said before, if we're going to fix the total number of monitoring units, then I agree that it's better to have more clusters and fewer numbers of monitoring units per clusters, than the other way around. You've heard me say this before. I don't know why you need to worry about whether the slope coefficient in the regression equation is equal to one or not.

Just use whatever slope it is to, to predict the exposure level. And it seems to me that would

provide all the information that you need. And, of course, if you're going to estimate that slope coefficient, as the task force has indicated, you need to have a nice spread in the amount of ingredients handled.

And preferably, we want a nice spread within each cluster. And if one is sure that the relationship is linear then you just need small and large values of HIJ of the amount being handled. Maybe to test linearity, you want something pointing in the middle, 'cause if you have a low amount and a high amount and something in the middle, then it gives you some handle on whether you have linearity or not. Also, in estimating the slope parameter, it only used up one degree of freedom in your, in your data analysis. So, you're not, you're not spending much of your data used to estimate that slope parameter.

The next little bit I have here has to do with the primary benchmark as stated in the, in the task force document stated that the number in configuration of sampling, sampled monitoring units should be adequate so that selected measures of the dermal exposure distribution means the percentiles and so on are accurate to within K fold, when the exposures are normalized, that is divided by the amount of active

ingredient handled. And so, the question then is, what value of K is appropriate and reasonable.

And I'm not sure as a statistician that I can answer that question. I think the scientists have to answer that question. But, some of the possibilities that, that might exist that, using some numbers that I took out of Jeff Dawson's presentation as to some representative values, perhaps.

If the geometric mean is equal to 12, then at 3-fold range, would go from 4 to 36. And in that case, if the 95th percentile, and I'm just guessing where it might be, but say it was at 21, then a 3-fold range of that would go from 7 to 63. On the other hand, if the geometric mean is smaller, the exposure rate is smaller at, say, 6, then a 3-fold range would go from 2 to 18, 95th percentile might be 15, and 3-fold range on that would be 5 to 45.

On the other hand, if there's a lot of exposure so that the geometric mean is 900, then a 3-fold range is 300 to 2700, and a 95th percentile might be 1500 and a 3-fold range on that would be 500 to 4500. So, I guess, the agency needs to decide if that kind of accuracy is okay. If it's, if you, if, as I say, as a statistician, I don't really have any way to judge whether that's a reasonable range or not. And so

the agency has to make that decision, and, obviously, the sample size is going to be determined by whatever that value of K is.

The secondary benchmark was that the number and configuration of monitoring units should be adequate so that it is possible to distinguish between complete proportionality and complete independence of dermal exposure and the amount of AI handled. I like this sentence 'cause now I know that there's a difference between proportionality and complete proportionality.

Complete proportionality, I assume, means the slope is one, and proportionality just means it's something but doesn't have to be one. And, as I've mentioned before, I guess I don't think this is worth worrying about too much, but just use a model where the slope is estimated, but it does say something about how you need to collect the data so that you can get an accurate estimate of that slope parameter.

So, I guess the bottom line is, I think that the task force has done an outstanding job, I guess, of looking at past data and trying to use that to get some ideas about what the sample size should be. The simulations that have been done and the estimates have inter-class correlation and so on that seem reasonable.

I think the process that is, has been suggested is, is a good process and I think it can be recommended.

So, I think the issue comes down as to what kind of accuracy do you need. And so, the emphasis, then, is on this value of K. How much accuracy do you need? What value of K is appropriate and reasonable? And as a, unfortunately, as a statistician, I don't know how to tell you that. I'll pass on, then, to the rest of the - -

DR. HEERINGA: Thank you very much, Dr. Johnson. I'll go in order. Cynthia Hines?

DR. HINES: I don't know when H came so close up in the alphabet, but it seems to in all of these. As Dr. Johnson alluded to, I think Ken is going to be having some additional ideas, so I'll just comment for the moment on what has been proposed by the task force. I just have a few additional comments in addition to Dr. Johnson's.

Maybe the first comment would be, and you alluded to this, the selection of the clusters. As I thought about this some more, this is really a critical decision, how you select these clusters, and whether or not they're really building in the stratification that you desire. And it may be somewhere in the large document, but it's not real clear to me what the

criteria are for selecting these clusters in a real field practical sense. Is it going to be a combination of state and crop? Is it, could it go down to the county level? Could you be within a state and would a different cluster be in, say, northern Arkansas versus southern Arkansas?

So, I think there needs to be some tightening up, if this cluster approach is used of what the definition of these clusters are, so that it's not something you're kind of doing on the fly. And, you know, maybe there's no intention of doing that, but it's just not real clear at this point.

Because, and the reason I think this is critical, 'cause in inspecting the data analysis that you've presented, it's very clear, of course, that increasing clusters as opposed to increasing numbers of monitoring units per cluster, gives you the best efficiency in sample size.

So, there may be a tendency to want to maximize those clusters, and so, will you end up doing that, perhaps, in a way that really isn't improving your condition stratification, just because it becomes more practical and more expedient. So, I would give that some serious thought if you continue with the cluster approach. My next comment, there was some

discussion yesterday about whether or not the agency would be using the 95th percentile versus say the geometric mean. And, again, as I inspected the data last night, this is clearly a critical decision, because it has major implications on the sample side. And it, also, means that the selection of the ICC and the GFD is very critical.

And if those numbers are off from what you are projecting at this point, and it is helpful that you have some data already to go on, that could have some consequences for the utility of the data down the line. So, one thought might be, since this is a multi-year study with a lot of scenarios, that, as you progress through the study, that there might be some points where you stop and actually evaluate, what are we seeing for an ICC.

What are we seeing for our GFD's. Are we on target, or do we need to make some adjustments. My next comment refers to this evaluation of proportionality. Again, looking at the data that's been presented, you clearly will help yourself, in terms of power, if you can get to, say, 100-fold difference on your range of active ingredient, as compared to a 10-fold difference. Now, I know there's going to be some practical limitations on that. If you

have, say, a herbicide applied by ground boom to a row crop, and if your threshold applied is 5 pounds, it may be very feasible in those situations to find farmers or applicators who are applying 500 pounds, 600 pounds. And, in reading through the document, it sounds like you are going to evaluate on each chemical what it's range is, and so you have some sense whether you can go to that upper range, and I would encourage you to think about that to the extent that you have people normally exposed to those levels, because it will help with that sample size.

But, on the other hand, there are going to be chemicals, either applied at low application rates, or the application method itself, you would never, with a 5 pound threshold, be able to get to 500 pounds. You know, you may even be struggling to get to 50 pounds on some of those. And so, I don't know whether that means you're going to have to rethink a little bit the 5 pound threshold to really get this 10-fold range that you're interested in.

And, I think, I think, that's my comments for now. I may have additional ones after.

DR. HEERINGA: Thank you very much. Dr. Lu.

DR. LU: I think I'm going to give the

heavy duty statistics to other panel who have expertise on it. I would just like to comment on two things. I think, I agree with task force approach when it come to the sample size determination. And sometimes you just have to take logistic matter into consideration. Sometime you will probably outweigh the science-, um, statistical consideration.

The task force group justification of how they come to the conclusion of using 5 pound to a unit per cluster for 5 cluster per scenario seems adequate and feasible. The group has assured the panel that if the proposed 5 monitor unit / 5 cluster are deemed grossly inadequate, they will seek for additional monitoring unit or cluster. I think unclear to how inadequate is grossly inadequate. I think that's a little bit too conservative.

And, also, I don't know how this inadequacy will be assessed, whether it's at a monitoring unit level, or the cluster level or in combined. The other concern is, I mentioned yesterday, is the selective target populations. As the task group people represented a way that it seems like it's not finalized yet, but they're going to follow this criteria as a guideline. The concern, actually, is related to the sample size and the overall data distribution as well.

I think I, my concern is that the task group may end up selecting a group of pesticide handler and pesticide applicator that is deviated from the true population. Meaning, I mean, I really cannot tell whether the distribution will be higher or lower.

The language speaking ability is a concern. English, Spanish and vices versas. So, again, since it's not finalized yet, the task group need to present a much clearer guidance. And as I can tell, that many people have told me they have Hispanic people working in the orchard, and they get a much better job than anybody else.

So, if your delivery is through this group of people, the, my prediction will be, the data will look much worse than it should be. So, those concern need to be take into account.

And I totally agree with the agency's charge question is that, especially for these two, the two, my two concerns is that, it has to be complete and transparent, that the agency and the public should know how you're going to assess the inadequacy, and also, how you select those people, and basis on what criteria. That's it.

DR. HEERINGA: Thank you very much, Dr. Lu. Dr. MacDonald is the next associate discussing.

DR. MACDONALD: You know, to design the monitoring program that will then be used for a variety of regulatory purposes by various organizations, challenges the developers to anticipate all possible future applications, while keeping costs in mind. The cluster sampling design proposed by the AHETF makes good sense, as there are cost savings in sampling a number of pesticide handlers in a single field operation.

The usual practice and survey design when there is inter-class correlation within clusters is to consider the costs of getting to a cluster relative to the cost of sampling individuals within the cluster. The optimal cluster size and number of clusters can then be chosen to minimize the variance of the estimate, subject to a constraint on the total cost. For this study, the task force has determined from experience that the inter-class correlation is modest, and that it is usually practical to monitor five pesticide handlers at a time, so no further argument is needed for a cluster size of five. This means that only the number of clusters needs to be chosen.

The first benchmark objective is to estimate the parameters of the distribution of dermal exposure to an adequate level of precision. The criterion

chosen that the upper 95 percent confidence bound for the parameter being no more than K times the parameter, and the 95 percent lower confidence bound be no less than the parameter divided by K, makes sense under the log normal assumption, and we were told that regulatory personnel have not had difficulty in specifying what, for them, would be an acceptable value of K. A closely related criterion giving similar results is to require that the upper 95 percent confidence limit be no more than K squared times the lower 95 percent confidence limit.

I think that this might be easier to communicate, and has the advantage of not requiring the true parameter value explicitly in the formula. I have no problem with the values of geometric standard deviation and interclass correlation used for these examples. However, I would expect that as more monitoring data are collected in this program, it will become evident that some scenarios may have very different interclass correlations from others.

I think the variation in interclass correlations observed to date comes from sparse data on variability and monitoring methods, and can't be attributed to specific scenarios. The examples shown to the panel assumed that the most extreme upper

percentile of exposure that anyone would want to estimate was the 95th, in which case, K equals 3-fold relative accuracy can be achieved with 5 clusters, which means 25 monitoring units per scenario.

However, this sample size will be inadequate if, at a future time, it is necessary to estimate the 99.9th percentile. This example was not included in the tables, but using the sass coat provided to the panel, it appears that ten, eleven, or twelve clusters would be needed to achieve K equals 3. So the total number of monitoring units would be more than doubled to 55 or 60.

The second benchmark objective, testing a proportionality of amount of AI handled, does not seem to me to be so interesting. It is clear that, at best, the amount of AI handled is a weak surrogate for potential exposure. There is error in every variable measured, and it seems to me, unreasonable to expect perfect proportionality in the regression line with unit slope.

I don't think it is worth testing. If you chose a large enough sample, you could end up projecting the hypothesis that slope equals zero and the hypothesis that slope equals one, so what would you do then? The only analyses of past data that showed a

clear unit slope were combinations of several studies spanning an extreme range of amount of AI handled. As far as I could see, no single study gave any indication of proportionality.

If the data will be used to compare scenarios, for example, to compare different application methods with the same pesticide, then the design needs to be considered more as a stratified sample, and there have to be enough observations in each stratum to make the test powerful enough to be worthwhile. I suspect if the sample size meets the first benchmark objective, it will also be good enough for this, but it would be worth checking this out.

The panel has talked a lot about measuring within worker variants and determining all three variance components that is between clusters, between workers within clusters, and within workers. It might be worthwhile to carry out some limited studies, but I can't see that it would be worthwhile to say triple the size of the entire study by monitoring every individual three times.

Note, by the way, that within worker variation, is still confounded with error in the monitoring technique and the chemical analysis. So, I've given three reasons why it may be advisable to

have at least 50 monitoring units per scenario; estimation of upper percentiles of exposure, effective comparisons of scenarios, and the possibility of measuring within worker variations.

I think the database will be of greater value into the future if costs are controlled by a thoughtful choice of scenarios, rather than by using small samples. If, at a future date, it is found that a sample size is inadequate for regulatory purposes, it will be impossible to return and get more observations that are consistent with the original sample. It will be much easier to do a complete study of new scenarios as they are needed.

In contrast, the AEATF study plan is dealing with a very different situation, and is much more amenable to experimental control. In particular, it should be feasible to increase the sample size for any scenario at a future date if more observations are needed. The proposal to take 15 monitoring units initially is adequate to give an overview.

For probabilistic assessments and determinations of extreme upper percentiles of exposure, 15 units will not be enough. We were asked whether the AEATF study should be one simulated site or three field locations. The simplest way to answer this

is to try both a few times in a pilot study and compare.

Perhaps, the three field sites should be treated as blocs or strata rather than clusters. But, I think, in summary, the biggest difference between the two plans is the possibility of increasing the sample size in future. It's much easier when you're working with an experimental bathroom than when you're working with real life crop scenarios.

DR. HEERINGA: Thank you very much, Dr. MacDonald. A question to you, which I think the statements you made regarding the 95th percentile and sample size, that's all conditioned on the log normal distribution model.

So, once we estimate its mean and geometric standard deviation, we have assumed a log normal distribution, and assumed the appropriate properties for the 95th percentile. If, empirically, the world differs in the tails from a true log normal, we have different conclusions on that.

DR. MACDONALD: Yes, that's correct. But, remember, too, if you have a small sample size, then any extrapolation into the tails is going to be heavily dependent on the distribution you've assumed. And your larger samples, certainly, if you're talking about the

95th percentile, you're talking, sort of, wha-, 19 out of 20, so you don't want to be using less than 20 observations. You talked about 99.9th. You don't really want to be talking about less than 1,000. So, if you want to get up into the extremes, you need sample sizes that are large enough that you're not just purely extrapolating into an assumed model.

DR. HEERINGA: Right. I just wanted that clear here. We, many sample size calculations, including the information Dr. Holden presented yesterday, and our statements made here, are obviously, conditioned on the log normal probability distribution model holding.

And I think, we can ask the panel, but I think that's an assumption we have to live with, but we want to be explicit that that's the model that is driving, in fact, not only sample size estimation, but the estimators of the point values and their confidence bounds. Turn now to Dr. Portier, please.

DR. PORTIER: I hate going last. You know, it's kind of like sitting in the first pew in church. Everybody's looking at you.

DR. HEERINGA: You only sit there if you get there late, though.

[All laugh.]

DR. PORTIER: Not everybody. I'm going to preface my remarks by saying my concentration's going to be on the AHETF task force studies, but I don't, it, I think it holds very much for the antimicrobial exposure studies as well.

My feeling, I think, mirrors what was said, that we feel that the antimicrobial stuff, we have a little more control, and maybe have a little better handle on the scenarios and the situations, so we're probably not as worried about them as we are with the agricultural task force because of the, just the sheer number of factors and the complexity of what's going on.

The discussion in the AHETF background document on numbers of clusters and monitoring units, was relatively straightforward, clear, proper, and what I think is representative of good statistical thinking. I compliment Dr. Holden on creating a clear, conceptual model for the sampling process and following it through to the particulars of the sampling design.

My issues are not with the sample size determination methodology, but with the assumptions underlying the sample size analysis. And we just talked about one, which is a log normal distribution, but there are a couple others that really need to be

addressed. In particular, I want to address the statement that users, and this is a quote, must also assume that the purposive sampling, sampler of the MU's approximates some type of probability sample from the target population.

My understanding of risk assessment is that the exposure value input into the risk equation is expected to be representative of the average exposure that would be experienced by the population potentially exposed to the chemical. For probabilistic risk assessments, individual exposure values are drawn from a distribution of exposures that are expected to describe the distribution of long-term, average exposures for individuals in the population potentially exposed.

So, I look at the proposed sampling design through this lens of representativeness. So, I'm thinking more, as much statistically as, you know, is this design going to produce a representative distribution of exposures that really reflect the population that are going to be exposed. The first assumption made is that a surrogate of cluster sampling model, which assumes underlying random selection, can be used to estimate sample sizes, even though the proposed sampling methodology does not advocate random

sampling for clusters. The second assumption relates to the normality of variance components and the nested effects linear model on log normalized exposure. I don't take issue with the second assumption, but I have some real concerns with the acceptability of using the surrogate random sampling model.

The discussions in section 5152 of the AHETF technical summary background document is excellent in that it provides a good framework for thinking about sampling for exposure assessment. I'm going to use a slight modification of their conceptual model to illustrate my concerns with the sampling protocol proposed. And you might want to get the slides up at this point.

The goal of the AHET-, AHEP data set, is the estimation of the true exposure E, for a specific handler task. To collect these data, AHETF proposes a cluster sampling or hierarchal sampling design in which clusters or studies, I might refer to them as studies, rather than clusters, but so, if I say studies, think clusters, are essentially examinations of handlers performing the handler tasks of interest at specific locations in time.

All right, so this is the illustration that was from the presentation yesterday. So, as mentioned

in the background documentation, there exists a very large number of potential studies, right.

So, in this space, you know, we can move each of these studies a little bit, and we're in a different time, a different location, all right. And there's, since it's a continuous space, there's an infinite number of studies that are out there. Conceptually, each of the cease of eye, each of the clusters or studies, is characterized by specific settings for a large number of factors. For example, climatic conditions, environment combinations, task times, et cetera, and we've talked a lot about that in the last few days.

The no-, each se-, each unique set of factor conditions, we'll call each set of factor conditions as a scenario, and I'll refer to that as S of I. So, in theory, if we know all the conditions that effect exposure, we could compute a true average exposure concentration for each scenario. Next slide. So, let's change the space from a location time, now, to a condition space. So, this is, you know, if we can conceptually think of what the population is not being places and times, but now they're situations, they're scenarios. And this is following the discussion in the background document, so, I'm just, kind of, slightly

changing it. So, those studies that looked nice and round in the previous series, now have some meaning on a particular measure. So, the top one is climate, for example.

Study C-3 was done at a particular location and time, so it had a fairly tight climate range. It doesn't have a lot of variability on climate, but the other conditions could have been quite variable. All right, the equipment they dealt with; the individuals, the people, the workers, the handlers, themselves, have situations that cause variability. So, C-3 is, has more variability in one dimension than in the other. Each cluster study is, essentially, a replicate of some scenario.

Since many of the factors that impact exposure are continuous, theoretically, there are an infinite number of scenarios, and, hence, there's an infinite number of potential studies. So, we've just kind of shifted the space to talk about conditions and scenarios, rather than locations and times. But we haven't reduced the complexity of problem any at this point.

Click one time. So, if you think about the PEHD data set, when it was created, it was created, supposedly, to address the conditions at the time that

people were sampling, and they were doing the sampling to handle, to prove lack of risk under the situations that the handlers were going to be doing with, this pesticide.

Click it one time. What happens over time is that the core conditions, my blue circle, which, might have represented conditions in 1990, now in 2006, for some of these factors, the conditions have changed. The equipment's changed. The scenarios, the scenarios under which the workers are working has changed, and so, the database is now out of sync with the conditions that the workers are looking at. And so these studies, they haven't changed, right. They're the same data we've had since '85, but they no longer are representative of the core conditions that we're worried about, right.

So, what the AHETF is trying to do is put studies back in the center of the blue. Click it one more time. So, they're trying to come up with studies that, actually, fit, you know, fit in the middle here, that are more representative of current conditions, current scenarios, right. One way to think of the true exposure, the parameter we're trying to get at, would be to av-, would be the average exposures for the handler task across all possible scenarios. So, my

little equation up at the top here, sums one to infinity, actually, I'm going to show you an integral next. I'm sorry for that, but really, since the space is infinite, you can't sum over an infinite space, but you can integrate over it.

But, if we have a, if we could generate an exposure for every one of these possible scenarios, really, what we're trying to do is average across all those things, and get that estimate. So, theoretically, that's what we're trying to do. And you can think of the exposure as a function of the scenario. Again, this comes out of the background document.

That G of S just says that if we knew this relationship between how exposures effe-, how exposures are a function of scenarios, we could calculate that, right. But this equation assumes that each possible scenario has an equal probability of frequency of occurrence. And we know this is not true. Best application task of certain climatic conditions that define when they must and can be performed, sometimes their equipment are more common than others.

If we knew the relative frequency of each of the scenarios, then the true exposure would be estimated as an average weight, a weighted average, I'm

sorry, and I think, I hope that's on the next slide. Next slide.

Oh, just to point out that, again, my concern here is, that while AA-, the study conceptually, we're thinking the study is being designed to have the studies in the blue circle, and conceptually in the current conditions. One of my concerns is that, well, let me just get back to that, so, this is finding an average. Let's go back to the sampling and talk about the surrogate random sampling model means in terms of clusters and scenarios.

With random selection of clusters, we're, essentially, randomly selecting scenarios for inclusion in the study in proportion to their relative frequency in the population of interest. If the scenario is a high relative frequency, in a random sampling design, there will be a number of studies included that are replicates of that frequency.

So, if it were a common scenario, by random sampling, we'd have replicates of that common scenario in proportion to its, if you like importance of relative frequency in the population. This means that averaging the study's specific average exposures would produce an unbiased estimate of the true handler task exposure. In a sense, random sampling self-weights all

the scenarios, and when we just, kind of, add things up like that equation, we get the right estimate of the overall average exposure.

Now, consider the diversity sampling approach as proposed by AHETF. The approach does not include randomness. Though thoughtful consideration of location and time, it is possible that a large number of scenarios will be examined. In fact, the background document to the panel seemed to indicate that the locations and time would be selected to ensure the different scenarios would be considered. The problem with the approach is that the relative frequency of scenarios will not, necessarily, be considered in the selection of the scenarios.

If it's possible that only one of a really common high-relative frequency scenario will be included in the sample set at the same time as one of a really rare, say, low-frequency scenario is included. When the sample average of estimated study exposures is computed, estimated exposure for the rare scenario is weighted equally with the estimated exposure for the common scenario, and as a result, the sample average will be a biased estimate of the exposure that we're really trying to get at.

So, the concern is that, that the diversity

sampling approach won't, kind of, properly represent the population.

The issue is even worse if what we want is not just the mean exposure, but an estimate of the true dist-, exposure distribution, or some other distributional parameters, like the standard deviation and the upper or lower quartiles, or some upper percentile. The non-probability sample will not produce a fateful estimate of the population distribution. Worse, we cannot predict the directions of the bias.

The study design could produce over-estimates. For example, if the rare scenarios produced high exposures, our design could produce under-estimates, for example, if only the common scenarios are included and the common scenarios have low risk. So, the problem with a non-probability based sample is that we know it's going to produce biases and we don't know which way it's going to produce biases.

This, then, is the basis for the statement made by the AHETF statistician that, in quotes, non-random sampling means that statistical methods alone are insufficient for generalizing to the target population. Most statisticians and many risk assessors are aware of this problem. And the problem's not new.

Almost every environmental data set has this problem, right. The question is whether we want to support the creation of another environmental data set with this problem, right. So, my concern is that we're going to spend $18,000 dollars a person to generate another data set, which will produce biased estimates. And I, and so my issue here is to think, can we, can we move forward.

AHETF acknowledges the above problems in the background document, and points out that rarely are the relative frequencies of the scenarios known. At the same time, it's not possible to create a simple random sample of studies that are guaranteed to approximately represent the scenarios.

So, the goal of the diversity sampling approach propose for populating AHED, is to achieve a diversity of major factors that are likely to influence exposure, and again, that's a quote. And to attempt, in quote, to capture the major aspects of the actual distributions of exposure.

In essence, AHETF will attempt to identify specific CI's to sample in a representative of the whole set of possible conditions and such that the distribution of exposures from the diversity sample is approximately equal to the distribution exposures

appropriately weighted for all scenarios. Statisticians have heard this kind of proposal many times before, and have never really seen true success. It's actually impossible to purposely define a sample that produces a distribution of exposures that duplicates the true population distribution, when one has no knowledge of the true population distribution to start with.

So, it's like a, you know, it's a catch 22. I want to produce this distribution. I don't know what it looks like, but I think I can create a set of samples that are going to pro-, it's actually a common exercise we do in sampling class with our students to see if they can purposely sample to produce a true distribution. And they're surprised every time how far off they are.

Rare events are seldom given proper consideration, and common events are often under-represented. Selecting to get true representation does not work. There needs to be random selection used somewhere in the process. So, is this really a hopeless situation. I don't think so. We have, at this point, the opportunity to rethink these issues and possibly come up with some new approaches that might get us closer to our stated goal. So, what might I

suggest. Consider the following approach. Create a list of all the factors that are known to impact exposure levels. The list may be long, but it's not infinite.

So, you just sit down and start writing all these things, and I think, I get a feeling that the antimicrobial task force has done this. I wasn't as sure that the agricultural handlers task force has, and it probably has, we just haven't seen it. But it's going to be a long list, right. There's a lot of factors involved here.

Rank order the factors by their expected magnitude of impact on exposure variation. I might suggest using something like a Delphi approach with a panel of expert risk assessors to accomplish this ranking. So, you've got this list. Let's get some people who know what's going on, and discuss this, and come up with a rank ordering.

What do you think's most important, down to, what do we not care about. You know, we've had a lot of this discussion over the last few days, but we really, I still don't know what's the most important condition, what's the most important factor. Is it climate. Is it equipment. I think that's another four day task for a panel of people to come up with that

list. Select the top two to four factors and identify, for each factor, two to three categories or levels. Next slide.

Essentially, what we're going to do is create a set of possible co-, all possible combinations of these important factor levels. Consider these combinations as strata of the population of interest. In a sense, these become the scenarios or scenario categories of interest. And I'm going to call them SI* star, now, because they're not points in this space. They're not chunks, right, they're areas, right. So, we've taken this condition space.

We've reduced its dimensionality by selecting the most important factors, and now we've stratified that, that space. Next assign weights to each scenario that approximate their relative frequency in the population. Sampling theory tells us that the weights don't have to be exact for us to gain a large improvements in the estimated precision.

Here a panel of agricultural experts could help, right. So, now we've reduced the dimensionality, but we still have to figure out which of these are important scenarios, my blue circle; and which of them are less important. Now, Cynthia Hines tells me that this is not as easy to do as I might think. Her last,

well, you just say your last use of agricultural experts, they made recommendations, and then when you actually went into the field, you found out they were wrong.

So, we may make mistakes, there, yeah, there are experts and there's experts, right. So, I'm not saying this is easy, but at least it produces a, it would produce some weights that would tell us what's important and what's not important. And at this point, you have, kind of, two options. In option one, you could go in and select at random studies and or MU's for each scenario, since the relative number of MU's or exposure estimates obtained for that scenario equals its weight.

The population exposure estimate is then the, just the simple average of the estimated exposures for the MU's. So, here's an example where C1 may not have, I mean, that first strata in the top left hand corner, might not be a very heavily weighted strata, so we only do one small study with one person in it, just to kind of see what's going on.

We don't really care where in that box that C1 is. It could float around, right. But at least we've got some representation in that part of the space. C3, C5, C6 are representing the core scenarios,

the more common, highly weighted ones. So, we're putting more of our sampling effort into that area. And then we're allocating stuff around it. And there's no more than 24 MU's in this study here, right. So, that's one option. Next slide.

Option two is to select, at random, a fixed number of studies, or MU's, for each scenario and assign each scena-, and then assign to each estimate estimated exposure value the weight of that scenario, in which case, now the exposure estimate is going to be a weighted average.

A kind of, I don't like this option as much as the previous one, because I think the users of the AHED data set are not going to be thinking in terms of weighted averages for estimation. They're going to want to do some kind of simple averaging. So a self-weighted, stratified design, which was the previous slide, works just fine then. You don't have to be worrying about weights every time you're calculating some exposure.

This in, these approaches incorporate both representation and randomness into the creation of the database, because at any stage, you could define a couple of locations and times that match the scenarios in a particular strata, and you could randomly select

from those, right. It's a little bit of work to develop that, but at least you could be randomly selecting them. And, at this point, even if you didn't randomly select within those studies, I'm happy, because I, you've done a lot more toward creating a more representative sample, than the approach that, I'm afraid, was going on before.

The above approach might be quite similar to what AHETF is actually doing. I mean, it's a little hard to tell. They may actually be doing this, in which case, I'm happy, and I'll erase everything from the report here. But, I have a feeling they haven't quite gone to this level of design thinking to make sure of what's going on.

The major differences that, in the approach outline here, an attempt has been made to first map out the possible condition space, although in a rough categorized way, to assign relative importance to each category, and finally to sample according to that relative importance.

The above approach is almost, certainly, not the best design that could be created. There are a large number of statisticians out there much cleverer than I am, who could produce some sampling protocols that would be much more efficient, and produce less

biased estimates, if someone would only ask. All I'm asking is that EPE and AHETF give some thought to this kind of approach. And, I picked up a word from, I think it was what Dr. Lu said, complete and transparent.

One of the things I like about this approach is that, for any user of the database, it's very transparent what conditions you're covering. And as the conditions shift, it also becomes very transparent that your database is losing its representation and where we have to go in the future.

One of the things Dr. MacDonald talked about is, how does this design help us in the future. Well, one of the things this design does is helps us identify where we are, what the database is expected to cover, and as change, things change, we know where to fill in. Maybe we have to add another set of strata and do samples as the, as the core shifts, but at least we have a, we have, kind of, a transparent picture.

I feel that this is much more dependable in a sense of the science of the database, and the science of the utili-, the statistics of the utilization of the database, than something a lot looser, which is what I've felt is occurring in the diversity sample. And I think that ends my comments.

DR. HEERINGA: Thank you very much, Dr. Portier. Dr. MacDonald.

DR. MACDONALD: It's, using the weights when you're estimating an overall mean level, of course, is a very well-known problem in survey design. Can you say anything about how you would approach estimating upper percentiles, using the weights?

DR. PORTIER: I can think of at least one way to do it, right, well, and one way would be, again, working within the log normal distribution. You could use the weights to calculate the parameters of the distribution, and then use, through that fitted distribution, estimate the upper percentiles.

That's, I mean, it's doable, so at least we have a proof that it can be done once, right. Now, and I was sitting here thinking. I looked at this one time trying to figure out, can you get directly to estimating upper percentiles with weighted observations without having to specify the distribution. And I don't think there's any theory out there that supports that.

DR. MACDONALD: Yeah, I guess I was thinking in terms of fitting a model, and then getting percentiles on the residuals, and then working back from that. And that's a little more distribution free.

DR. PORTIER: So, your, yeah, I was trying to avoid, I mean, we don't want to talk about research, right. And I, what Dr. MacDonald was saying is, well, you already know what these factor levels are.

Why don't you go ahead and fit a model that removes the factor level effects from the exposure estimates, and what's left is the residual. That residual is likely to be normally distributed or have nice properties, and then you could, kind of, figure out the upper percentile value from that. And then you back, go backwards through the model to a percentile estimate. Something like that, right.

I don't know. I want to make the point that this kind of looks like research, I know, right. I'm not really, I'm not really in a model-fitting sense here, though. I'm not using it from the context of trying to understand and predict exposures from factor levels. I'm using the factors really to stratify the sampling space so that we get appropriate coverage, right.

DR. HEERINGA: Steve Heeringa, here. I guess I'll weigh in with a few comments, sort of, prompted by Dr. Portier's recommendation. And I, in general, agree with the principle of what he's driving at, in that is, I think it would be beneficial, in

terms of the ultimate utility of the AHED database to think about, you're trying to span ranges of variability, to span the distribution of exposure under appropriate and realistic end relevance sets of operating and other conditions. One area, I think, we ought to be very clear about it.

Introducing randomization into the sampling. If your, if you cannot observe more than 15 to 20 clusters, it's not a beneficial thing to do. And, Ken's right. You may not, you may have biased estimates, but with fewer than 20 clusters, you cannot, your variances are going to be enormous, and they're going to swamp the bias.

You're worried about total error, variance and biased squared. And, until you can get, and there's no better source than Ed Demming for this comment, and that is, that if he were forced to choose more, fewer than 10 observation units, or I would extend that to 15, he would rely on judgment sampling. And what Ken's point is, is that, we're really in that place at this point, unless we start to think about going to 20, 30, 40, 50 or even hundreds of clusters of observation, which is where the survey world lives, and probability sampling inference lives.

We have forced, with some sort of judgment

or, I forget Dr. Holden's term kind of dressed that up, but I still like judgment sampling. But it's expert judgment, potentially, as Ken said. And we may not be in a position to determine all of those factors that appropriately define the distribution.

DR. HEERINGA: I'm talking sites, these clusters. The things that are generating these ICC's of .3, and by the way, .3 in a probability sampling framework is an enormous inter-class correlation. Voting behavior only has inter-class correlations of something like .05 or .06, so the types of inter-study correlations that we are seeing in these data sets, as they've been estimated, and I think Dr. Kim even showed that, I guess, that was worker inter ICC's in there. But my sense is that, until we get into this range where you could, within each of these scenarios, work with 20, 30, 40, or more clusters, that to actually think about a detail probability sampling approach, while theoretically, potentially, desirable, in practice, doesn't work; because variability and instability of variance estimates just swamp the potential bias that you might even get from even the worst of judgment sample.

So, that's my comment on that. The average cluster size of five, I think, I've actually brought it

with me. I've got class notes from last term here. I don't commit these optimum sample sizes for two-stage samples to memory, unfortunately. My memory's failing. I can't use that many cells to do that, but I'll look that up, but an optimal cluster size of five, I agree with Dr. MacDonald.

I think it's, obviously, you've sort of tested it empirically in your minds. I suspect it's fairly close to the optimum for a cost structure in which the analytic costs are roughly 50 percent of the total cost of an observation.

The, therefore, increasing the precision and the effectiveness, both in terms of estimating the geometric mean and standard deviation, but also in terms of Dr. Portier's push for ensuring representation suggests that I wouldn't increase cluster sizes, but if you extend the sampling, you would move to adding additional clusters.

One thing that I would also recommend, based on Dr. Holden's presentation yesterday is, we know these clusters are not going to come in a nice, neat units of five. He did a nice simulation, effectively simulating the distribution of estimates under a fixed distribution, samplings with five clusters and five units. I might suggest it's possibly exploring

something in which we have clusters that average five clusters, but you allow a little bit of variability on the actual number of observations, 'cause in reality, you won't get exactly five observations.

Some places you may get two, some you may get four, some you may get seven. It's a, I don't know what that'll show, but it's worth doing, just to make sure that you have a sense that there is some robustness to these results that have been developed for equal size clusters.

I agree with, with Dr. Portier that if you are able to develop, within each scenario, a fairly gross level prioritization with some measure, not only of, essentially, a stratification of type, with some measure of the frequency of that type of application, that might well guide your choice of the cluster units. Again, with only five clusters per scenario, you only cut the pie five ways, and it's very difficult to, sort of, proportionately allocate five units to even as few as two or three strata.

So, again, that's a challenge. I agree in principle with what he has suggested as a way to go, but I think you're going to find it very difficult with only five clusters per scenario to do much of this type of work. But I think the thought process that he's

outlining clearly is beneficial on this. My sense is, and I know money's a restriction here, that you would be greatly served for each additional cluster you add to this, I think that we could certainly go to much higher levels, but my sense is that, in some of the critical scenarios, why assign five clusters to every scenario.

If you as a task force, and the EPA, Health Canada, California DPI, could think through these different scenarios and, you know, which are those most critical scenarios, in terms of total population exposed, allocate resources into these different scenarios. Maybe you want ten or fifteen clusters for the most common.

If they, you know, if twenty or fifteen percent of applications are in one scenario, or fifteen percent of population exposed are in one scenario, I would certainly not constrain that to five clusters, and then have another, say, .5 or one percentile scenario, also with five clusters in terms of resources.

Again, some of this will depend on how data sets come to the task force. The ta-, how the task force can purchase data sets, but I think if you're planning about new work, I would look at putting the

effort, essentially, where the population of workers, exposed workers, is best served. And that, you know, relates not only to the potential severity of exposures, but also the extent of the population exposed. So, those are comments that I had, and turn it over to other members of the panel at this point. Dr. Bucher.

DR. BUCHER: So, I weigh into this as a non-statistician, however, for many years, I reported to a statistician, and for many, many, many years, I've been married to a statistician, so I understand the territory that I'm going to try to tread on. So, we're dealing with a question of sample size.

But I'd like to go back and look at the benchmark objectives for data adequacy that were presented yesterday by Dr. Holden. And I wonder, these benchmarks, basically, stated that the primary objective is to select measures of a distribution means percentiles that should be accurate to within a certain degree of accuracy.

And the secondary objective is to use, the users of the data should be able to distinguish between complete proportionality and complete independence of exposure and amount of the active material handled. So, we've been asked to comment on sample sizes for,

what I would consider to be, a fairly modest objective as spelled out here.

I would, I would hesitate to endorse Dr. Johnson's suggestion that the slope of the line be calculated and used for all different proportions material, active ingredients over the actual amount of material used, because I don't think that the overall objectives for data adequacy really are precise enough to believe that the data that has been, will be generated under this program are going to be sufficient to be able use, to be able to be used in that manner. So, what I would suggest, and this has troubled me throughout the entire meeting, is that it seems to me that, that EPA needs to look at these benchmark objectives for data adequacy very closely, and decide whether, in fact, you believe that the data that are generated, using this as a target, are really going to be adequate to be used to set up a generic database that's going to be used for thousands, potentially, thousands of materials in the future.

Because, if, in fact, you're accepting something that isn't as good as it should be, it's going to be, and I understand that you're dealing with a database right now that's very inadequate and very limited, but there has to be some compromise, and I

think I may be saying the same thing that some of the statisticians have been saying, but in a, in a, certainly, a more ignorant manner.

But, I really would hesitate to, to move forward with a program like this, unless you a priori set some guidelines on how you're going to use the data. How you're going to interpret the adequacy of the data for a prospective use in a database that's going to be used in the future.

DR. HEERINGA: Cynthia Hines.

DR. HINES: Just a clarifying question of Ken on his proposal, where you have factor A and factor B, I assume that you want them to be fairly uncorrelated, because if you do a list of factors, you're going to have quite a number of them correlating highly with each other.

DR. PORTIER: Clearly, using factors that would be correlated with each other would not benefit. I mean, I, if we got into the mathematical thing, I'm thinking, principle component A and B, you know, kind of orthogonal dimensions that use it, but I'm not expecting them to do that.

I'm just, most of the people who look at these factors know which ones co-vary, right. So, I'm just saying, pick one of them, and divide the space on

that one. Again, I bring it, I bring it down to this complete and transparent manner.

It, part of what we've been talking about is being able to make transparent to future users what you've really sampled. And I think the location time space doesn't make it very transparent what exactly you're sampling. It tells you where you sampled and what you sampled, but it doesn't necessarily say how you covered this condition, or this space.

DR. HEERINGA: Dr. Popendorf.

DR. POPENDORF: Might as well add my two cents here to this, as another non-statistician visit. I think a lot of us are, really, also, saying the same things from our individual perspectives. And I, you know, the value of that virtual study that we were shown the other day seems, seems very, very useful, whether or not we did it.

I think it was a good exercise to give you that sense of in. When you look at the numbers like Cindy mentioned, the, this range issue, you, five pounds and I think, somewhere, one of the documents says range of something like trying to get fi-, between 5 and 2,000, and that's a good range. It's a little over two orders of magnitude, but I kind of wonder about the real feasibility of doing that, within the

context of a given application method in trying to assure that you get a half-day's worth or more. And what artifacts are you adding by putting that into, which kind of gets to Ken's point of study design. It's not representative if, if, you know, how many would apply 2,000.

How often does that happen. So, you end up with this bias, and then you end up with the upper percentile issue. You know, a lot of this whole idea of clustering was driven by question three, having to do with linearity, and the idea of active ingredient handled being the driver, which is an agency driver. And I can see some rationale for that, in terms of using other compounds or f-, I guess I'm looking down the line in terms of how it's being used in terms of registration and label restrictions.

'Cause I, you need to assume some maximum like agricultural acreage or something along those lines. The concentration, then, will drive the amount used. So, you're going to run a proportion. I don't know, I haven't really thought through, from the regulatory agency perspective, other options that you might be able to put on the label that might be more driven by the kind of physical models that were suggested. I think you suggested, Ken suggested

experts. The idea, you know, the experts, I've been there and done that, too. The experts, you know, if we really had experts to do what we were, what happens out there, we wouldn't need to take measurements.

And you're looking at observations that you think you know what's going on, but you don't really know what the, where the chemical is. A lot of things happen that you can't see, so the experts, really, don't know until they take measurements. So, it's really tough to derive those categories. And, again, I, sort of, put some of the burden back on the agency to think about, or perhaps, explain.

I was talking about the idea of putting the, walking through a physical mechanisms modeled for each of those scenarios that would, sort of, justify why active ingredient handled is a variable, but, perhaps, think about other ways that you could use this type of data that might effect, you know, do the clustering another way, perhaps, along the lines that were suggested, that would be more, well, wouldn't bias the data as much.

It needs to be used in a practical sense, in terms of, you know, labels or restrictions. So, I think that through. What are other options. I'm not sure.

DR. HEERINGA: Dr. Johnson.

DR. JOHNSON: Yes, I'm thinking about this amount of active ingredient handled, if you're going to vary that within each scenario, each scenario, probably, has some, some range that is reasonable, may differ from scenario to scenario, but it maybe is reasonable for that scenario. And I don't think it matters whether that's 100-fold range, a 10-fold range, a 5-fold range, whatever range it happens to be. There's a high value.

There's a low value. There's places in the middle, and those are probably the three main places that ought to be, ought to be considered. I wonder, Steve, if the, if it might be helpful for the panel to, to give opinions or thoughts about the value of this K, in terms of getting K-fold accuracy. I don't know whether that's part of the, par-, something that the EPA needs to decide upon, or whether members of the panel want to weigh in on what values of K might be reasonable to use.

DR. HEERINGA: I think that if there, if the panels' members are willing to, sort of, stick a foot out there on that issue, I think it's fair game to heard at this point. Do you personally have - -

DR. JOHNSON: I don't have any

suggestions.

DR. HEERINGA: I suspect there aren't, nobody's going to step in the breach here, then, so. Dr. Popendorf, do you, or Dr. Portier. I think that, you know, that issue of K really gets down to the data and all of the other sources, and, Mr. Dawson.

MR. DAWSON: Jeff Dawson. I think we would concur that K is going to be, ultimately, it's going to be a policy call for us. So, maybe the more, the most utility for this discussion would be, maybe, outline some factors that you think we should consider in our discussions around how we select the K. That would be very useful for us.

DR. HEERINGA: Yes, Mr. Villanueva.

MR. VILLANUEVA: Yeah, just a point of clarification on Larry's simulations with the K-fold and everything. I guess, one of the caveats that the panel members mentioned was that the model was based on the log normal distribution. I think that's only true for the data generation process. If I read it correctly, the confidence intervals that Larry presented empirical confidence intervals?

DR. HEERINGA: Right, and that's a good clarification. The data generation for the 1,000; 10,00 simulations that produce the empirical

distribution, that underlying distribution was log normal, and assumed the inter-class correlation, and cluster sizes of 5 clusters and 5 elements per cluster. That's my understanding. So, the final results, the coverage properties and the confidence bounds were based on the empirical simulation distributions, right.

DR. VILLANUEVA: Right, which would imply that the estimate of the 95th percentile would have been an empirical estimate, so I think, as far as the sampling weights go, Dr. MacDonald mentioned, would just be estimating empirical percentile, based on the sample, so that's pretty easy to incorporate weights as opposed to, I guess, a parametric estimate of the percentile incorporating no weights.

DR. HEERINGA: I think the difference, if I can answer. Ken can correct me there. When you say, if you refer to empirical, there's a world out there, that if we drew 10,000 samples from the real world, that would be the empirical distribution. 10,000 samples from a log normal distribution, with a fixed mean and variance, gives us a simulation under a world that perfectly follows that log normal distri-, they're two different things. So, a true empirical distribution, we would interpret as 10,000 samples from the real world, and then I would be happy with the

confidence bound set. So, I think that's the distinction. It's a good point to make. The people agree with me on that, with regard, it's a good point to raise, but a, the simulations that are being done here, really, put all of the world in the form of a log normal distribution. And all we're trying to do is to estimate the two parameters of that log normal distribution.

DR. PORTIER: An interesting exercise would be to replace that log normal with something like a gamma. And then see if that, and see if those, those bounds are relatively robust. So, if you did a log normal, you did a gamma, and things didn't change all that much, then we'd be much more likely to believe that, because, because those kind of represented extremes of possible bi-, yeah, what, skewed distributions that we might encounter, right. So, if those bounds were not that driven by the shape of the distribution, then, then I might be happier, right.

DR. HEERINGA: I think the suggestion there would be, actually, two. If you are going to, pretend that you are going to analyze it as though it's log normal, but then, in your simulations, generate from slightly deviating distributions, maybe with longer tails, sticker tails, and then see under the

long normal distribution, just how your confidence bound coverages for the true values actually conform. Peter, do you have comments on that?

DR. MACDONALD: Yeah, I did try some simulations where I was simulating from the log normal, and then using the empirical distribution to get upper percentiles. And the, from modest samples of about 25 or more, they came out very close to the distributions you got extrapolating with the log normal. So, that was interesting, but I was still generating the data from the log normal, but it does show that the analysis doesn't have to be log normal dependent. I just don't know where that information comes from to get the extreme tail, though, when you've only got 20 observations.

DR. HEERINGA: Any other questions on that. Yes, Dr. Appleton.

DR. APPLETON: Well, if Will Popendorf's not a statistician, I guess I know where I stand on the food chambers statistics, but.

[All laugh.]

DR. APPLETON: This is really a belated question that I'll try to couch as a comment, but I may be the only person on the panel that represents the regulated community, such as it is. The government

agency that actually uses pesticides, and occasionally sponsors worker exposure studies. And, I'll address this, primarily, to EPA representatives, past and present. I wasn't with the pesticide agency in the, say, in the period of 1984 to '86, when the original selection of 15 replicates or monitoring units per scenario was chosen.

I would presume there was something more rigorous than, well, 10's not enough, and 20 is what we want; 15 sounds good and the industry will buy it. But, presuming that the task force recommendations of 10 monitoring units per scenario is sufficient and EPA agrees.

The question I really have is, will the EPA need to revise subdivision U guidelines again to address only the agricultural uses. It looks like the antimicrobial group is going in its own direction, in terms of staying with 15, at the moment. And, how will non-agricultural uses, whether those are home uses, you know, wasp sprays, foggers, crack and crevice, or some of our oddball forestry uses, be addressed.

Will they still fall under subdivision U requirements of 10, if that's what it turns out to be for a database? Again, it's, maybe we're too far down the line to look at that yet, but.

MR. EVANS: This is Jeff Evans. I guess I'll step into that a little bit. We certainly want to update out guideline requirements, and I wish I could say more clearly why the 15 for handler and 9 for aerial applicators was chosen, but I think it was some sort of mixture of logistics, feasibility, costs, and you know, I guess some robustness. And, also, the fact that we do mostly pouring estimates, we do central Tennessee values, and kind of beef up our estimates with higher estimates of amount AI handled and acres treated, things like that.

Obviously, we're getting much more into a, maybe, more representative sample. And, I mean, just thinking about what Dr. Johnson pointed out, the comparison between what we have now, and what we plan on getting, and how we think about that, with respect to clusters. And quite rightly so, you put the onus on us to make us think about how many we really need, and how many we desire. And where we do want to regulate in the future. So, the short answer is yes. We are going to update our guidelines, but just what that's going to end up being, will require a fair amount of thought on our part.

DR. HEERINGA: I wonder at this point, I could ask Dr. Johnson to try to summarize the sort of

views of the panel.
[All laugh.]
DR. JOHNSON: You could.
DR. HEERINGA: I think, specifically, with regard, the one question that came forth yesterday is, is the task force and the EPA and Cal DPA and Health Canada working with them on the right track in terms of thinking through these issues, regardless of what K winds up being, et cetera. Are we approaching it in the correct.
DR. JOHNSON: Well, I guess, I, I'm know, think I know where the panel is kind of heading. I think that Dr. Portier's suggestions and Dr. MacDonald are both know a lot more about sampling than I do. Most of my career has been spent in designed experiments. And without looking at observational type studies, and observational type studies like this are always a lot harder to design and a lot harder to sample. And so, I defer to them, with respect to the kinds of recommendations that they made.
They sound reasonable to me, and are recommendations that should be considered. I think the panel is, generally, in agreement that, that the, given a choice between increasing the number of units per cluster, and increasing the number of clusters, they

would go to increasing the number of clusters. In terms of the K-fold accuracy, I guess, that, the panel's not expected to, sort of, answer that question, which is a good, I'm happy with that. I don't know how I would want to estimate that value of K anyway.

I think that, that the goal is to consider different amounts of ingredient handled, and you can get that, get that 100-fold range, I guess, by decreasing the amount handled, as well as, increasing it. And, perhaps, that as the EPA comes up with their plan, or the pesticide handlers exposure task force comes up with their plan, they might address it from the point of view of picking some maximum amount that's legal to use, and then decreasing that, rather than, and use the words decrease, rather than increase, and they might have a better chance of getting by the human factors board, human studies board.

I guess that, I know, I'm kind of stumbling around a little bit, because I'm, I've, I think I, kind of, get the sense of where the panel is, but I'd really need some time to get the responses that, in writing, and then put it all together in our final document.

DR. HEERINGA: Thanks. I didn't mean to put you on the spot, but I, I guess, you're up to it, I think. And, Dr. Portier.

DR. PORTIER: Something Dr. Johnson just said kind of clicked off a thought. You know, we've got this primary and this secondary objectives, and if it were my druthers, I would forget the secondary objective's regards to the database. I would do a separate study to figure out whether proportionality seems to hold.

I mean, that, the agency has assumed proportionality for ye-, for decades now. That question, obviously, wasn't that important twenty years ago, and is it that important for us to spend a lot of money on that today. I mean, let's do a study, and if it still looks reasonable, that's fine. And you have your justification.

If I'm going to spend a unit on a increased amount of material, versus a replicate of a scenario, I'd rather do a replicate of the scenario. I'd rather not spend a lot of, my personal preference would not be to spend a lot of effort trying to do the secondary question. And, I don't know, John may argue with me on that, but, I, you know, it's what's the purpose of the database, and where is it going to, and whether we're served by diverting our attention onto that secondary issue.

DR. HEERINGA: Dr. Bucher.

DR. BUCHER: Well, I'm not sure I can answer the question of which, where I would put my money, but if one is going to generate a database, and I, I presume that most of the new data and the new methodologies are going to, in terms of the proportionality of active ingredient handled versus the total amount of material, as the, I would imagine that as the, with the new technologies, as the amount of material goes up, the proportion of active ingredient is, actually, going to go down for the exposure.

So, that's going to be a very important aspect of the overall risk assessment. And what I'm afraid of is that if we generate a database that just has a very few data points, and it changes the shape of that slope of that curve in a quite a substantial way, you're going to really affect the overall risk assessment.

And, unless the agency has criteria for the acceptance of data that are going to actually move them away from this very conservative position that they're taking now, then, it's a troubling, troubling situation in my mind.

DR. PORTIER: Ken Portier, here. I don't disagree, but I, and I'm not thinking about the secondary study having a few data points. I think you

need a lot of data points.

The problem is you-, you know, it's like Dr. Johnson said. Some scenarios are going to be able to only accommodate a small range, and other scenarios accommodate a wide range. And if you're allocating a lot of effort to making sure that's covered across the board, we're going to lose a lot of representativeness. So, I'd rather do a side study on one set of scenarios, with a lot of, a lot of power for actually looking at that secondary question, and answer it.

And, you're right. If the question is, there is no proportionality, that changes your whole, underlying risk equation, right. And we may need to know that answer first, before you go through much further in a representation.

DR. BUCHER: So, it, could I?

DR. HEERINGA: Tom Bucher, yes.

DR. BUCHER: It could be then, that money would be well spent in picking out those scenarios, where you think that that proportionality is, is absolutely wrong.

The linear and, you know, the slope of one is absolutely wrong, and verifying that, in a very limited sense. And then, but, but for the vast majority of all of these unstudied or poorly studied scenarios, one

might want to continue to, to take the conservative approach of linearity with a slope of one.

DR. HEERINGA: Yes, Paul Hamey.

DR. HAMEY: It struck me when I was listening to Dr. Portier's stratification approach, that in some instances the amount handled might be one of those factors it would include, and in others, that it wouldn't. So, I just put that comment on the table.

DR. PORTIER: Ken Portier here. You mean for a certain handler task, that may be a major factor. Another handler task, it has no impact on it, right.

DR. HAMEY: Exactly, yes.

DR. HEERINGA: I think there's a general sense in the conversation this morning that, probably, at, relative to the costs of these data, the cost of your time, which are relatively inexpensive for us, more expensive for you, to invest some time in thinking through the 30 scenarios that have been outlined, and I would say the same applies for the antimicrobials, that to look at those in the context of these things. Because, I think, while I understand fully that expert judgment on actual exposure levels and those models, and those models may not even be that good in the end, because of all the other variables involved, but to think through that, and to think through

prioritization, in terms of your ultimate objectives, with the task force, to say, you know, this is where people are being exposed. This is where, really, the greatest impact for our 30 times 25 observations times $18,000 dollars is going to do.

And so, I think that's worth doing, and maybe it's already been done very much internally within the task force. And I think that affects a number of things. Not only where the priority for the sample cluster placement, you're really restricted in the numbers of clusters, and we understand that. But where you place those, but also for this issue of really studying these secondary objectives. And, as I've thought about it, too, I think Dr. Holden had it right, theoretically, if we could completely manipulate the world out there, we would put the range of X in each of the separate clusters. That's clearly the most efficient.

He demonstrated that. But if you think about that, if you have to manipulate the essential application conditions, in other words, if somebody wants to apply at a certain rate, plus and minus 10 percent, that's not going to give you the 100-fold range that we saw was really needed to achieve power in a number of situations. And if you had to manipulate

that by having somebody apply twice as long as the next person, that would affect this. That changes, I think, another very fundamental condition, which the task force said, we want these people, you know, working a normal workday or a normal half workday in this application.

So, I think you don't want to get into a situation in these scenarios where you change the application conditions or the measurement conditions so much, simply to get a variability on the applied active ingredient. I think that's fairly intuitive. But I think this, so I think some thought through each of these scenarios like that, based on the discussions of the last four days, and all of you have a lot more insight in some of the particulars here, to think through where best to invest the effort.

And, I think, particularly on the cluster sampling, to the extent that you have a few of these scenarios that really dominate your registration, dominates the health concerns, I think, clearly it would pay to invest there, maybe, slighting some of those that, while important and define scenarios, really, in terms of overall aggregate risk for the population aren't quite as important.

MR. MILLER: So, in essence, kind of be a

little bit more strategic in our thinking, so.

DR. HEERINGA: Yeah, and where, I don't think we're being critical of you. You've done a lot of thinking to this point, but I think, given the cost and the fact that this is a process that stretches out over time, you can and should afford to be strategic a little bit at this point, too. Dr. Lu.

DR. LU: I think I would agree that EPA's approach to thinking about the disproportionality issues, I mean, as Dr. Johnson point out, if slope equal one, and yesterday the agency justified why they want to stay with this slope one, instead of 0.8 because to work to the, you know, higher amount of active ingredient handled become protocol, and they want to protect the high end exposures.

But if you think about this, and then the proposal they make by the task force that, they're going to only include data that's been more than four hours. If anything below four hours, they will just kind of cut off.

This kind of, it pose problem because, and not to get in a lot of, you know, high amount of active ingredient handled data, and if we compromise the slope equal to one, then what happens. And to me, I think, these four hour application time is rather conservative

or strange, because in the real world scenario, for example, if the manager called for the applying pesticide today, everything is just fine. There's no wind, sunny, and so on and so forth. And in the middle of the two hour application, they have to call off because wind started picking up. And it's not ideal situation.

Then what happens. Base this on the task force criteria, the data will be thrown away, because no, it's only applied two hours, but everything is fine until that moment. So, I think, in this case, the task force has to modify their criteria so, in this case, the data will be in there, and they will be on the lower end of the active ingredient handled, and that will be just fine.

DR. HEERINGA: What I'd like to do at this point, we're at 20 after 10:00. I'd like to have a fifteen minute break. And then we'll come back to wrap up on this question, and get any concluding remarks and comments from the panel. And I would aim to finish by 11:30 or 20 to 12:00. So, let's take a fifteen minute break and reconvene at 25 of 11:00.

(WHEREUPON, a break was taken.)

DR. HEERINGA: Okay, welcome back everybody to the final late morning session of our

four-day meeting of the FIFRA Science Advisory Panel on the topic of worker exposure assessments in pesticide handling. I want to pick up where we left off and wrap up at this point. I think before we continue our conversation, that Jeff Dawson had one point of clarification.

MR. DAWSON: Thank you. I'm Jeff Dawson, HED. We've been talking amongst ourselves over here with regard to the latest conversations around proportionality, and several panel members have commented around the fact that our working assumption at this point of a proportionality of one to one is a conservative approach, so it would be good if that, somehow, made it into the record.

And the other, I guess, issue around proportionality for us is, we're certainly not wed to that over time, and appreciate all the suggestions as to, you know, how we might evolve for certain scenarios, depending up on, you know, the nature of the data and such as we move forward, so thank you.

DR. HEERINGA: Thank you very much. Just one additional point that I'll add. I think Peter MacDonald had mentioned, too, that there is, you know, theoretically, a formula for optimum cluster size, given cost structure and, over the break, I just did

the calculation, and it, 4 or 5 is in the ball park. It is what, it'd be closer to 4 than to 5, but the optimum is generally very flat for this over a narrow range.

So, the point that we made that the optimum cluster size, given your cost structure that Dr. Canez, sort of, hinted at yesterday, that's crude, we know, but it suggests that this cluster size of 5 is probably, fairly, near the optimum, in terms of your resource expenditure on the data collection, under the current cost structure.

Additional comments from the panel on the general issue of sample size determination, and the process by which the task force have gone about thinking about sample size determination, and its relationship to the precision or ultimately accuracy of final decisions made with these generic databases. Dr. Portier.

DR. PORTIER: Ken Portier. I wanted to clarify something just to make sure. I used the term scenario, and I used the term handler task. And I realized that in the discussion up until today, the scenario, really, was a handler task in the terminology used in the task force. And I'll have to come up in my report. I'll change my scenario term to something

else, but I don't want to use condition, because that implied, maybe, one dimensions, and it's really a multi-dimensional problem. I'm going to come up with another word. It might not be an English word, but it will be [laughs] another word, well, maybe Native American word for scenario.

The whole idea, though, is the stratification would have to be looked at by handler task. So, it's not, it's not everything thrown in together. It's kind of looked by task, or goal, task group by task group, rather than everything in one big picture. That doesn't mean that the studies, themselves, might not go across these individual strata.

I mean, one study might address a strata for a mixer operation, and at the same study, you may be doing an individual in another, another task that would be in a different strata for that task. So, it's, it's a little more two-dimensional than I, three-dimensional than I, than I illustrated it, but English is hard to come up with a lot of good words, so I'll come, I'll correct that so it doesn't get confusing.

DR. HEERINGA: There must be a Cajun word for scenario. Dr. Lu.

DR. LU: I think I would like to put this point in record. In terms of a sample size

determination, my concern is that if the task force has come up with a very narrow defined selection criteria. Chances are the data that they were going to go out and collect would probably satisfy their assumptions, and that's why they lead to the 5 monitoring units, 5 cluster, which is okay, but the overall concern is that the pop-, the subject they're going to include may not be representative of the true work force in the field. And it satisfy their need, but not necessarily the agency's desire.

So, I think the selection criteria should be phrased like, any pesticide handler or pesticide applicator, as long as they are licensed to do their job, they should have the equal opportunity to be included in the study, except for the human subject review board concern like pregnancy and so on and so forth.

DR. HEERINGA: Thank you very much. And that's consistent with your earlier comment, too, I think, don't you. Dr. Robson, did you want to make a comment about the active ingredient issue?

DR. ROBSON: Yeah, I - -

DR. HEERINGA: Microphone.

DR. ROBSON: Oh, sorry about that. I was thinking more about Dr. Johnson's comment earlier, and

I have to respectfully disagree with Dr. Portier, and support Dr. Johnson on the AAIH.

I think, I had the opportunity a few years ago, to participate in an evaluation of agricultural health study, which many of the people from the agency and the regulator, regulated community are very involved in as well. And as we struggled with trying to reconstruct pesticide histories that go back decades, and people tried days of applications and realized that some days, a day was an hour, and some days a day was twelve and fourteen hours, that probably the thing that many people argued for was to do acres treated by crop, which really translates into active ingredients.

So, I, as I think about this more, I really am still pretty convinced that active ingredient is one that we want to weigh in a little heavier on, or at least, I would like to weigh in a little heavier on, and support comments that Dr. Johnson made earlier in the day, and throughout the week.

DR. HEERINGA: At this point, I think I'd like to turn to the EPA scientific staff, to see if they have any questions of clarification for the panel. Obviously, we've had a lot of discussion, and it's covered a lot of topics. With regard to the sample

size issue, is there anything that you still have questions about, or you, confusion on some of the responses?

MR. DAWSON: No, no, I think we're good, and I think the topic's been covered in a very thorough manner.

DR. HEERINGA: And I believe I had promised, too, that with the antimicrobials, that we would, in fact, address that, and I think that Dr. MacDonald's comments, initially, covered that. And that, maybe, Dr. Leighton, do you have?

DR. LEIGHTON: Tim Leighton from the antimicrobials. Yes, you did cover this a lot. And one thing I do want to make sure is the follow up written report also includes some of the comments that were made on experimental design versus, you know, going out in the field observational type.

DR. HEERINGA: We'll make a note of that, and I think working with Dr. Portier and Dr. Johnson on these last two responses, that we'll make sure that that does get covered for you.

At this point in time, are there any other comments on charge question number six? I think that we have, at least in our discussion, covered a fairly wide range of views and opinions and information, and

that our written comments will reflect that, and will reflect consensus or lack thereof on the part of panel members.

Hopefully, that has been informative. What I'd like to do now is, I'd like to go back through the panel systematically, to see if there's anything over the course of the past three and a half days where you would like to make some additional, sort of, concluding remarks or bring forth something that maybe, at this point, that you haven't had a opportunity to say. And I'll start over here with Dr. Landers.

DR. LANDERS: Thank you, Mr. Chairman. I've no extra remarks to make, only just to confirm the remarks I've made already. And I think we've seen that as the week's progressed. But there is such variability in application technology, that the idea of some formal matrix which will allow you to categorize tec-, application techniques into high risk, low risk, old or new, or whatever criteria you choose. I think that's the main conclusion I would draw. I agree with other speakers who've mentioned that there is such variability.

The variability in people is one of the greatest concerns. And that opens up a whole can of worms, if you start discussing how well trained

operators are in different parts of the country. And so, how you cope for that is up to others to decide. At any given time, of course, there's always danger with these pesticides and machines from things falling off. And so, what might be a perfectly good study which shows that the limited amount of exposure is correct until it goes wrong. And then, of course, we can't cope with that. And so, I leave that as a thought with you. Thank you.

DR. HEERINGA: Dr. MacDonald.

DR. MACDONALD: Pass.

DR. HEERINGA: Mr. Hamey, Paul Hamey.

MR. HAMEY: Nothing technical to add, but I would just like to thank all the people that submitted documents to us with a high quality of that material. I thought that was very good.

DR. HEERINGA: Dr. Robson.

DR. ROBSON: Mark Robson. One of the things that we talked a little bit about during the break was, as we had a chance to reflect on four days of excellent presentations, as Paul just mentioned, is just some of the terminology. I think it helped us all remember, those of us that have been doing this for a while, that we use different words to describe different things. And I think back to Jeff, one of

your earlier slides, where we gave a range of time for people loading. I think it was one of your earlier slides in the case study.

We talked about someone who ranged from two and a half hours to just a few minutes. And what we were talking about was probably the person wasn't really load-, and it's, we're not, I'm not singling out the presentation you made.

It just reminded us of some of the terminology. That was probably two and a half hours of a monitoring event, versus two and a half hours of that activity. And, of course, since none of us in the room were there to witness the study or participate in it, we have to read it as we review the study as activity and loading or, so I think as the regulated community goes forward, and as the agency requires information, that we just try to be better housekeepers and define things better so that, when other folks read this, as well as yourselves, that we, when it, if it's a monitoring activity, that's really what it is.

The time that one spends to actually do the task, which, I just could not imagine that someone is, actually, consistently for two and a half hours loading granular material. It just doesn't, seems to be, a piece of farm equipment on the other end that could

receive that much material, but, you know, that's the term that was used, and we have to take it as written. So, I think it's a real opportunity for everybody to, to come up with a set of terms that we agree on and descriptors that help us understand exactly how the study was carried out. But I, like Paul, am very grateful for the really thoughtful and well-organized presentations. For those of us that teach everyday, it reminds us of how poor some of our teaching is, and we'll steal some of your Power points.

DR. HEERINGA: Dr. Popendorf.

DR. POPENDORF: Yeah, thank you. Will Popendorf. Just two points come to mind. One is, sort of, for the record is the comments about that K value, and a couple of conversations out during the break and whatnot, but the idea, I think, seems to be pretty consistent, and I, maybe, this might not be news to you, perhaps, to the panel is the idea of just, this is all going to be used back to eventually look at that MOE, and the idea, then, if you have a very large MOE, you can tolerate a very large K.

As your, in your confidence, it doesn't make that much difference. And I think that just, you know, make sure that it was sort of on the record that you guys are thinking along that way, too, in terms of

letting that K value fall into play. The other thing that I've thought about earlier, and have not, really, commented on, and the ideas have been floated, particularly in this last question about representativeness.

The reality is, we do not, there is no data that indicates what is representative. You know, it's that judgmental type of thing. That's really what you're looking at, and I don't know how the agency or, it's not really a requirement of the task force. I don't know what the mechanisms are, but it would certainly be nice if there were data that would, survey type data, that would say, what are the equi-, what is the equipment that are out there.

What is the range of equipment and the frequencies. What are those application rates, the kinds of things that go into that selection process. We really don't know, and I don't know how they really generate that, but it would be nice if it were available.

DR. HEERINGA: Dr. Curwin.

DR. CURWIN: Just to echo what Dr. Popendorf just said. I think we had this in discussion on data needs and we're largely focusing on this database in exposure, but there is certainly a big

portion of the exposure estimate equation is what's called pesticide use information, and that hasn't really been discussed so much during this meeting, but it's certainly something that's critical in the exposure assessment.

And as Dr. Popendorf just said, that information needs to be captured and better estimated, in terms of the amount of AI handled, if that's your normalization. I mean, there's things such as the acres applied and the application rates that are being used, so, just to keep that in mind.

Back to the normalization issue, I think it was brought up earlier, it was asked if there was a conceivable time where you might find the exposure, the slope of the proportionality is greater than one, and what comes to mind would be, and this happens, you know, on a regular frequency, I think, is things that aren't anticipated.

What you'll see, you have your exposure going on, and the, your applicator or your handlers doing their basic tasks, but they, they spill that chemical on them, and so you have much more exposure, given a certain amount of handle-, uh, active ingredient handled. And then that also includes things like clean up and repair activities, which hasn't really been

discussed. And I don't know how that information can be captured in this database, but these certain things can actually increase the exposure, without actually having an increase in active ingredient handled.

DR. HEERINGA: I'd like to move over to Dr. Hughes, now, and around the table.

DR. HUGHES: Again, thanks to everybody who made presentations here, and appreciate the fact that what has been brought out with regard to doing bio-monitoring with adding additional monitoring units, that there are just complexities that go on with trying to do this in the field, and not only complexities, but the unpredictableness of weather conditions and other conditions that are outside your control in order to do these studies.

And our charge, I'd also like to appreciate the risk analysis that was brought forward. Whenever you have parameters and a risk analysis, you have to look at those parameters, and look at which ones are the risk drivers, and to which degree they are the risk drivers. Whether they're going to drive your analysis greatly, or whether they're not going to drive your analysis to any great degree.

And then, there's also the risk management decision. If you have something that is a great risk

driver, and you feel that in a tier one, you might have had conservancy in it, where it is a risk management decision to go after a more realistic value.

You may say that, okay, there is a realistic way, I mean, there is enough, how can I say this. It is never a risk driver, and it is highly overestimated that we want to go and spend the resources to do that. Also, to the extent that you might not have a high risk driver, and it might not be very variable away from the assumed risk in a more deterministic approach, and then it might no-, the bark might not be worth the bite. And so, you know, as we go ahead and we have also looked at doing some additional bio-monitoring studies to look at the final risk analysis to see whether or not that impacts the final result significantly.

And so, I think we've made some decisions with regard to our charge regarding, we just don't have enough data, and we could get that data. And I think that question is, do we find that within the certain scenarios, there is enough justification to go ahead and do that. And I think that within our charge, we felt that there was.

And, again, even though we looked at the risk analysis, and our trying to determine from that whether or not, again, the sensitivity is there, and if it

would make a difference in the risk output.

DR. HEERINGA: Thank you very much. Dr. Lu.

DR. LU: I think the agencies and the task force did a great job to outline this piece of the work. And I understand the still ongoing process.

The task force will go back and refine their protocol, and probably going to see for another round of approval from the human subject review board.

I would suggest that by the time the dust will settle, and this final working plan that's agreed upon between task force and agencies, at the end of this data collection period, agency should consider conducting their own studies selectively, in terms of, taking one, select scenarios, base this on what cluster and then monitoring, just to see whether the data that generate by the EPA will fall into the range of the data that generate by, collected by the task force. If, if that's the case, I think the mission has accomplished and it's very successful. Thanks.

DR. HEERINGA: Dr. Barr.

DR. BARR: Thank you. Like everyone else before me, I'd like to congratulate the task force and the appeal for putting together the volume of information, the quality of presentations that really

helped us in evaluating the charges that were put forth to us. In going back and re-evaluating all the data that is a part of our charge, I am convinced that the whole body path of dosimetry is a suitable way to assess overall exposure, and this is a pretty painful thing to admit for somebody who's made their career off bio-monitoring, but it was, it was a great amount of information.

I definitely think the need for additional information is warranted, and I think that's something that you can strongly suggest and strongly argue with the HSIB. And I think that all of the discussions here really helped to provide that input into the new database. I also want to say that, I really hope that somehow, the differences in pesticides can be captured in the generic database, perhaps with the KOW's or some other mechanism, because there are, you know, a large degree of differences between different pesticides. But, again, thank you for the fine presentations and it was a pleasure to be here.

DR. HEERINGA: Thank you, Dr. Barr. Dr. Kim.

DR. KIM: My final comments have to do with how you're going to use the data. We've talked a lot about how to collect the data, and how to measure

and what types of data to use, how to sample, but oftentimes, when we use not the most state of the art scientific equations, or methods to use that data, we're going to make some wrong policy decisions.

So, I think, I'm talking in particular about how to use dermal exposures, inhalation exposure measurements to estimate the internal dose. And I think that there are methods available that are better able to predict the internal dose more accurately, and those should be used.

DR. HEERINGA: Dr. Appleton.

DR. APPLETON: I couldn't have said it better myself, Dr. Kim. I was going to. So, I'll second all that, and bio-monitoring did not die in vain, but I do think that there are techniques available to make a stronger distinction between an externally deposited exposure and an internal dose for the applied quantitative risk assessment. Other than that, my kudos to all the participants.

DR. HEERINGA: Dr. Johnson.

DR. JOHNSON: Yes, I also thank the task forces and the EPA for their presentations, and I strongly encourage them to include some of the material in the presentations into their re-, into the protocols. I think that's, that, there was a lot of

material presented that was very helpful and would have been nice to have seen that in some of the protocols. So, I encourage you to make use of that material, now that you have it generated.

DR. HEERINGA: Dr. Hines.

DR. HINES: I would just like to take the long view and express, despite all the holes that we've been poking at various times in the various proposals, that I have optimism that we are moving toward a better database. There are some issues that need to be worked out, and I think there have been some excellent ideas here at this meeting on, perhaps, how that might be approached. And, looking back, we now have another twenty years of experience, both in the industry and within EPA on how to assess pesticide exposure. We have a much larger literature base to go on, and so, that gives me some optimism that, at the end of this whole process, maybe we won't have completely 30 scenarios rigorously evaluated, but we'll be a lot closer, I'm hoping.

DR. HEERINGA: Dr. Bucher.

DR. BUCHER: I'd just like to add my thanks to the agency and to the industry for the efforts of putting all this together. And thanks for the opportunity to learn a little bit about exposure

assessment.

DR. HEERINGA: Dr. Portier.

DR. PORTIER: Ken Portier. Steve said he wanted to get out of here at 11:30, no, it's.

[All laugh.]

DR. PORTIER: I have a few short things. I'm encouraged that the proposed database will represent a true advance to risk assessment. I mean, I, I'm really convinced of that. Firs-, one thing I'd like to do is request that the HED computer system interface, the, the, when the user's going to interface with this database, there's going to be software, that this software include modern methods for handling non-detects and testing distributional forms in the presence of non-detect data with these modern statistical methods.

The previous PHED system provided pre-packaged reports to users based on selection criteria, and presented such things as estimated means and standard deviations for exposure, and did a Kolmogorov-Smirnoff test for log normality. Today there are much better methods for estimating the mean and standard deviations and replacing the non-detects with half the detection limits, and there are much better statistical tests for looking at normality and log normalities in

the Kolmogorov-Smirnoff test, including test statistics that incorporate or, at least, take a count of the non-detects.

So, I just encourage the developers of the database to not look at the PHED interface, and to talk to some statisticians about incorporating modern estimation techniques, so what you get is, really, the best estimate, not the easiest estimate. And then the last comment is regard to representativeness. I think in the past, representativeness has been defined by the agency risk assessor who is responsible for the analysis.

That person, playing the part of a god, would say, this data's good, let's move forward. I think it, it's, it would nice if there were some kind of external definition or assessment of representativeness with this database, so that all the onus isn't back on the risk assessor to make that decree that the risk assessor can assume, to a large part, that what they're looking at is representative and that doesn't become a point of contention in the risk assessment. You can move on to the risk questions, rather than the data questions, right. That was it.

DR. HEERINGA: Well, thank you very much. And, at this point in time, too, I would like to extend

my appreciation as the Chair of these meetings for the past four days, to all of the presenters, to the task force, to the representatives from the Cal DPA, and Health Canada, and also, obviously, to the Environmental Protection Agency. It's been, sort of, a complex series of presentations. I think it's extremely well organized, given the breadth and amount of detail that we had to go into.

I appreciate all of the contributions of the panel members, too, to, literally, spend a better part of a week at the first of the year, which is a very difficult time to do this, but again, congratulations to the SAP for assembling such an expert panel, and my thanks to all of you for giving your time to, I think, what is a very, very, very important activity in this process.

I appreciate the process. I learn something every time that I participate in one of these. I hope that it's been beneficial. I think, you're, obviously, all in a tough place of trying to maximize a data utility, resource utility in a situation where there's extreme variability and high costs. And that's not an easy world to work in. And I think we all have to, we take that perspective. And I think you heard that here from Cynthia Hines, too, in her comments, that this is

a, we pick apart little pieces during the week, but in the overall picture, I think we represent the importance of studying these efforts, even though they aren't as clean. I'd still like to go back to that one, one regression line, and a, a .7R squared and, sort of, live my life there. But that's not where we live our lives, and - -

[All laugh.]

DR. HEERINGA: - - so, again, I think that, I appreciate the panel's willingness to be open to this. I appreciate everybody's willingness to present their points of view and get them out in the open, and we'll proceed from here. And I wish you all the best as you continue your collaborations on these efforts. I look forward to seeing the results of this. This point in time, I think if there are not any additional comments from the Environmental Protection Agency?

MR. DAWSON: We just wanted to mirror the theme that's been occurring the last couple of minutes. We really appreciate everyone's thoughtfulness related to the charge that we've put to you, and the amount of time and investment that you, clearly, took to address our questions, because this is an extremely important activity for us, and recognizing the time of year and

over the holidays, and such, so we really appreciate it, and we view this activity as highly valuable as we move forward.

DR. HEERINGA: Thank you very much, Mr. Dawson. At this point in time, before we wrap up the meeting, I'd like to turn the mike over to the designated Federal official, Myrta Christian, for just a few last minute administrative notes.

MS. CHRISTIAN: Okay, thank you, Dr. Heeringa. One more time, I want to thank the panel, the presenters, and the public for participating in this meeting. I think it has been a very successful meeting. Also, I would like to remind everyone that the meeting minutes for this SAP meeting will be available in approximately eight weeks. Thank you, again.

DR. HEERINGA: And also, a member, to all of the members of the public in the audience, that materials presented during the course of these meeting will be available on the docket for this meeting, the EPA website, and can be reviewed there. And, again, participants, if you have not submitted materials for that docket, if you would see that they get to Myrta. So, at this point in time, again, I think we've concluded three and a half days, and I want to thank

everyone for their participation and safe travels home today, and best wishes for the start of this new year, so, thank you everyone. Members of the panel, if we could collect in the break room, I, it's not my intent to hold you very long, because I know you have travel plans and things that you've scheduled, but I would like to just get a quick organizational session on the report writing, just to make sure that we don't let anything slip through.

(WHEREUPON, the CONFERENCE was concluded at 11:08 a.m.)

CAPTION

The foregoing matter was taken on the date, and at the time and place set out on the Title page hereof.

It was requested that the matter be taken by the reporter and that the same be reduced to typewritten form.

Further, as relates to depositions, it was agreed by and between counsel and the parties that the reading and signing of the transcript, be and the same is hereby waived.

CERTIFICATE OF REPORTER

COMMONWEALTH OF VIRGINIA

AT LARGE:

I do hereby certify that the witness in the foregoing transcript was taken on the date, and at the time and place set out on the Title page hereof by me after first being duly sworn to testify the truth, the whole truth, and nothing but the truth; and that the said matter was recorded stenographically and mechanically by me and then reduced to typewritten form under my direction, and constitutes a true record of the transcript as taken, all to the best of my skill and ability.

I further certify that the inspection, reading and signing of said deposition were waived by counsel for the respective parties and by the witness.

I certify that I am not a relative or employee of either counsel, and that I am in no way interested financially, directly or indirectly, in this action.

MARK REIF, COURT REPORTER / NOTARY

SUBMITTED ON JANUARY 12, 2007