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 11, 2007

FIFRA SCIENTIFIC ADVISORY PANEL (SAP)
REVIEW OF WORKER EXPOSURE ASSESSMENT METHODS
January 11, 2007
Morning Session

DR. HEERINGA: Good morning everyone. And welcome to the third day of our four day meeting of the FIFRA Science Advisory Panel on the topic of a Review of Worker Exposure Assessment Methods.

I am Steve Heeringa from the University of Michigan. I'm a statistician and Chair of the Science Advisory Panel, the FIFRA Science Advisory Penl for the EPA.

I'll be the Chair for today's session. And I want to thank my colleague, Ken Portier, for stepping in for me yesterday afternoon when I had a teaching commitment. Thank you very much, Ken.

At this point again as we do each morning when we open these meetings I'd like to have the members of our panel introduce themselves and state their affiliation and a little background for the audience. Ken.

DR. PORTIER: I'm Ken Portier, Director of Statistics with the American Cancer Society in Atlanta, Georgia. And my background and interest is in probabilistic issues in risk assessment.

DR. HANDWERGER: Good morning. I'm Stuart Handwerger from the Departments of Pediatrics and Cell and Cancer Biology at the College of Medicine of the University of Cincinnati. My clinical area of expertise is in pediatric endocrinology. My research is in the developmental and molecular endocrinology.

DR. CHAMBERS: I'm Jan Chambers, I direct the Center for Environmental Health Sciences in the College of Veterinary Medicine at Mississippi State University. I'm a pesticide toxicologist specializing in metabolism, neurotoxicology and exposure. And I am a member of the permanent SAP as well as a member of the Human Studies Review Board.

DR. BUCHER: I'm John Bucher with the National Toxicology Program at the National Institute of Environmental Health Sciences. I direct the Carcinogenesis Bioassay Program and the Toxicology Evaluation Programs.

DR. HINES: My name is Cynthia Hines, I'm a research industrial hygienist with the National Institute for Occupational Safety and Health. I conduct occupational exposure studies, including a number in pesticides.

DR. JOHNSON: My name is Dallas Johnson. I'm a retired professor of statistics from Kansas State

University. I've spent more than 30 years consulting on agricultural type problems.

DR. APPLETON: I'm Hank Appleton with the Forest Service of the U.S. I'm a pesticide toxicologist and risk assessor and I've accumulated about 20-odd years of experience in conducting exposure assessments for pesticides.

DR. KIM: I'm David Kim, I'm from the Department of Environmental Health at Harvard School of Public Health and my research area is in human exposure assessment and pharmacokinetics.

DR. BARR: I'm Dana Barr, I'm from the Centers for Disease Control and Prevention in Atlanta, George. I'm the Chief of the Pesticide Laboratory. My primary research focus is in assessing human exposure to environmental toxicants with a specialty in pesticides using biomonitoring.

DR. LU: Good morning. I'm Alex Lu from the Rollins School of Public Health at Emory University. My research interest is using biomarkers to assess pesticide exposures and using the pharmacokinetic model to interpret those biommarker data.

DR. HUGHES: My name is Brian Hughes, I'm a toxicologist. I'm with the Pesticide Section of the

Michigan Department of Agriculture. My interest is in risk assessment and also in doing studies for, field studies for occupational risks in conjunction with Michigan State University.

DR. LANDERS: My name is Andrew Landers. I'm an agricultural engineer at Cornell University. My role there is to head the Application Technology Group where we look at the engineering ways of reducing operator contamination and the environmental pollution. My research is looking at ways of improving deposition and reducing drift.

DR. MACDONALD: And my name is Peter MacDonald. I am Professor of Mathematics and Statistics at McMaster University in Canada. My general expertise is in applied statistics and this is my seventh year on FIFRA panels.

DR. HAMEY: Good morning, I'm Paul Hamey. I'm from the U.K. government's Pesticide Safety Directorate where I'm responsible for Human Exposure Assessments.

DR. ROBSON: Good morning, I'm Mark Robson. I'm the Director of the New Jersey Agricultural Experiment Station and Professor of Entomology at Rutgers University and Professor of Environmental Health at our School of Public Health.

My primary responsibility is administrative, but when I did honest work I did work looking at farmer and farm worker exposure to pesticides.

DR. POPENDORF: And I'm Will Popendorf, Professor of Industrial Hygiene at Utah State University. I focus on exposure assessment and control of chemical hazards and I've got 30 years experience in agricultural exposures to chemicals.

DR. CURWIN: Good morning, I'm Brian Curwin with the National Institute for Occupational Safety and Health. My primary research interest is occupational exposure assessment with a particular interest in pesticide exposure.

DR. HEERINGA: Thank you very much members of the panel. I'd also like to introduce the Designated Federal Official for today's meeting, Myrta Christian. Myrta tells me she has no information to give us this morning so just with the introduction I think we're ready to move on.

As we do each morning we're going to open with a session which gives presenters an opportunity to provide short follow ups on discussions from the previous days to clarify points or provide some additional information that may have been requested or a need arose during the discussion.

I'd like to turn to Jeff Dawson and Jeff and Evans to see if there are some initial follow ups from the EPA team at this point.

MR. DAWSON: I just had one followup thought from yesterday about, and there was some discussion about this issue yesterday about why a kind of broadly generalizable generic database for exposures is very important for us as a program.

I just put together a couple slides to kind of illustrate, you know, where we are as a program and the level of effort we have around conducting occupational exposure assessments.

So just to give everybody some perspective, over the last 10 years basically under the re-registration program in EPA we essentially conducted occupational exposure assessments related to the reevaluation of about 10,000 food tolerances, so quite a lot of work as you can probably understand. And that involved looking at basically hundreds of active ingredients, I don't have an exact count, I apologize for that. But, you know, it's 300 to 400 or so active ingredients over that time frame.

And also in the registration program with new chemistries coming online we've looked at over 100 cases or so also over that 10 year period. And the one

thing to point out related to the new chemistry cases is that, and I think somebody brought this up yesterday, is that, you know, we don't have monitoring data for those specifically so it becomes very important because we're very limited as to how we can evaluate those prior to them, you know, being released into the marketplace.

And then the antimicrobial folks also have, you know, hundreds of active ingredients that they're going through and doing reevaluations on at this point. So that, to put it in context.

And I had one more slide as well, so John Worgan on Monday had talked about, or Tuesday, sorry, had talked about the tiered approach. So basically on our tiered approach, tier 1 is the lowest amount of resources and data needs and complexity and chemical specificity related to the exposure data so we use that.

This is a kind of tool for that. And also for tier 2 we begin to increase the amounts of data. We may have a specific monitoring study or something that goes with that chemical or more information about use patterns and such that we try to incorporate. But we don't move up the tiered process unless we need to. If something passes our tier 1 approach with flying

colors we're not going to invest any more resources in it. And then as we move up to tier 3 that tends to be just out of whatever context where we tend to see the investment on the registrant's part and others and doing more complex kind of analyses where you get more chemical specific data.

It may be more passive dosimetry monitoring studies, it could involved biomonitoring and even we've done, you know, a few cases over the last couple years where we've done kind of highly complex multi compartment pharmacokinetics modeling to look at dose levels in the tissue, those kind of things. So our target organs. But they become highly complex cases and there's a lot of resource investment and we tend to only get to that point if we need to.

As far as validation tools, I mean we certainly want to bring all the tools available to use related to, you know, validation and evaluation of the methodologies as well. So I'm not saying that we wouldn't use, on a routine basis we would use, try to use the generic approach as much as possible but as far as the analysis and validation steps we want to bring everything to the table. Thanks.

DR. HEERINGA: Jeff.

MR. EVANS: Yes, thank you. I'm Jeff

Evans with the Health Effects Division. And I just want to clarify a point that I believe is on page 61 of the background document where we talk about how to resolve the issues of hand rinse performance and other breakthrough issues with the passive dosimetry.

And what we were thinking about in addition to perhaps correcting based on physical chemical properties but also perhaps incorporating into the study design a biological monitoring component, not necessarily to compare with the passive dosimetry measurements were, but in fact to confirm if there was breakthrough.

So if there was anything that was not recovered from the face wipes or the hand rinses it perhaps would show up in the urine and that would be a way just to kind of confirm those issues of breakthrough and inefficiencies in hand rinse performance. So I just wanted to make that clear. Thank you.

DR. HEERINGA: Thank you very much. At this point too in terms of additional followup there is I've asked Doctor Ross, representing the AG. Handlers Task Force to come forward. They have a few additional followup pieces of information that they as a task force would like to provide to the panel too. So

Doctor Ross.

DR. ROSS: Thank you, Mr. Chairman, members of the panel. As you recall there were questions about the data that was excluded from the comparison of passive dosimetry and biomonitoring yesterday in my presentation and I referenced Table 7 which is shown on this slide, at least half of it.

The reasons for excluding these various studies basically fall into three categories. One is the lack of primate dermal absorption and/or pharmacokinetics. Another is the biomonitoring being less than or equal to the limit of quantification at many of the intervals at which samples were collected. And finally one of the significant issues was studies in which significant body parts that we would expect to have high levels of exposure were either not represented or sporadically represented in the data.

For example if you look at the right hand column of the first row on Cowel 1987, there was no hand monitoring done in the studies of mixer/loader/applicators. Now we know from other studies of this nature that hands can contribute up to 50% of the dosage. So if the hand was not monitored and in addition the legs, the lower legs were not monitored and there was no inhalation monitoring in the

study so it has significant shortcomings, defaults.

In the case of the next study, also by Cowel, there was extremely low primate dermal absorption, and by extremely low I mean .08 percentage of an applied dosage was absorbed. And as Doctor Popendorf indicated yesterday, the error associated with extrapolation from that kind of low dermal absorption is going to be phenomenal. So we didn't feel that it would be appropriate to include that data set within the data that we did the comparison of the passive dosimetry and biomonitoring.

Next study, Fenske, 1988, there was incomplete data reporting and by incomplete I mean that there were data provided for body regions in terms of micrograms per centimeter squared, but the body regions were not representative of an entire body.

In other studies like Bernard, 2001, Krager, 1996, the post-exposure biomonitoring data was extremely low. In the case of the Krager study with Borax, we know that the human dermal absorption is 0.2%. We wouldn't anticipate that there would be very much coming out in the urine and in fact, despite high dermal loadings there was not any above background.

For the two Rotunaro studies on venclosylin there is no primate dermal absorption of

pharmacokinetics/metabolism, so we don't have anything with which to reference back to the human study. There might be rat data but as I had indicated before, that can be very deceptive in terms of trying to emulate human kinetics or metabolism.

The case of the Dubelman, 1989 study, again there was a mixer/loader/applicator study with no hand monitoring.

And the last on this page, Levy, there were samples not present, they had been mentioned in the document as being lost. There were no hand residues, no lower arm, lower leg samples of any kind and again, these are significant in terms of the work activity that these people were conducting. So you wouldn't expect to be able to get any kind of meaningful assessment out of this kind of data. You can get a qualitative idea but not a quantitative idea.

And I don't know, want to bore you with going through the rest of these but the story is similar.

At this time I'd like to turn over the comments to Curt Lunchick.

DR. HEERINGA: Yeah, Mr. Lunchick. Thank you, Doctor Ross.

MR. LUNCHICK: Thank you panel. We just wanted to make a clarification in regards to the

comparisons that have been raised between the number of observations in the Pesticide Handler Exposure Database compared to what is the anticipated or possible within the AG. Handler Exposure Database.

Specifically for example, Doctor Fenske in his written comments compared the 1,700 monitoring events in PHED to approximately 600 that would be in the AG. Handler Exposure Database. And I think we need to look at the quality of the data to make a more direct comparison and I think this may be important as we get into discussions on sampling sizes later in this panel meeting.

I actually have a working copy of the database and went into the Pesticide Handler Exposure Database applicator file to use as an example. And there are a total of 715 records in the applicator file which range form aerial and ground boom type of applications all the way down to nonagricultural events such as hose and sprays and aerosol cans. If you look at long sleeved shirt, long pants and no gloves, of those total of 715 records your number of observations ranged from 350 on the thighs up to 564 on the hands. Now that's all data grades.

The Agency prefers grades A and B data and that's typically what the registrants use when they try

to do their assessments too. If you subset this to dermal grade A and B which is just the dermal dosimeter, it's not the hands, you're reduced now to 358 records and a total of 214 to 293 data observations on different parts.

Now, if you look at the hands, and these are totally independent of the dermal areas, grade A and B, there are 365 records, of which 259 provide data without gloves. If you wanted to get down to those records that were grade A and B for both dermal and hands at the same time, you're actually now down to 248 records of which you now have a total of 133 observations to 215 observations of data.

Now as it was mentioned also our selection criteria was to make sure we had some type of dosimetry on all parts of the body. The Pesticide Handler Exposure Database in its statistical function allows you to select for what is called complete data sets and you can define that.

Defining the complete data set as having head exposure, hand exposure and exposure to the 10 parts of the body that are part of the AHETF criteria which is essentially upper arms, lower arms, chest, back, the thighs and lower legs, you're now reduced to a total of 57 records of which the use patterns will range from

agricultural to nonagricultural and to put it in perspective as we talk about proportionality today, out of those 57 only 27 records provided data on the application rate in pounds AI per acre. So there's a very big difference between the total number of data points in the database to those that are comparable to the quality of data that we are looking for in the AHETF program.

DR. HEERINGA: Thank you very much, Mr. Lunchick. At this point I believe that we also have a followup comment from Doctor Doug Baugher.

DR. BAUGHER: Yes, thank you. Yesterday I jokingly voted New York off the island when I learned that they operated at the high end of the label. On reflection I realize that I wrote the model to create the perfect storm, that is the co-occurrence of the high rates of application, the high acreage, the low residue recovery efficiencies and everything else. And therefore the New York scenarios would be included in the upper percentiles where we normally look for safety assessments. So welcome back.

Secondly, yesterday afternoon there was a comment and a concern which was well founded, that with the handler studies where gloves are worn, exposure of the hands may be fairly low but during reentry they may

be fairly high and the residue inefficiencies of hand washes could be important for reentry. I mentioned that I had done a second run with a high exposure reentry scenario and I'll give you a couple pieces of information.

For the air blast applicator wearing gloves, 20% of the exposure was to the hands. For orchard harvesters and thinners, 70% of exposure was to the hands. Counter-intuitively the impact of handwash inefficiency was less in the reentry model than in the handler model. And the reason is pretty straightforward. The reentry model has more parameters with characterized distributions and they just tend to swamp out the uncertainties associated with residue recovery inefficiencies. So the conclusions were the same for reentry as for handler, but even stronger.

DR. HEERINGA: Thank you very much, Doctor Baugher. Based on this additional information, any questions of clarification from the panel members? Okay, well let's move on then to the morning's proceedings and I think that at this point in time we'd like to welcome Matthew Crowley and David Miller of the Health Effects Division who have a presentation on the proportionality assumption between exposure and the amount of active ingredient handled, critical

assumption in the total exposure assessment.

MR. CROWLEY: Good morning everybody, I'm Matthew Crowley in EPA's Health Effects Division. I knocked over some stuff over here, sorry for the confusion.

Along with me I have David Miller, also in the Health Effects Division and I'd like to thank him for his help in doing the statistical analysis on this presentation along with members of his staff.

So this morning we will be going through the assumption of proportionality between exposure and the amount of active ingredient handled, and that is the unit exposure that has been mentioned frequently in the past couple of days.

First I'll go through some introductory materials, the purpose and background and a brief intro into the case studies. Then I'll follow with data manipulation formatting, the actual analysis and then some discussion points.

So the overall purpose is to discuss the assumed proportional relationship between exposure and the amount of active ingredient handled, for brevity noted as AAIH in this presentation and likely in the task force presentation to follow this afternoon. So that is an example of format there, the x milligrams

exposure per pounds handled that we use in our risk exposure and risk assessment calculations. And that's just an example. Canada may use micrograms per kilogram pounds handled. And we'll be using the 6 PHED case study scenarios that Jeff Dawson went through earlier this week and following that as this follows with the charge to the question and asked to comment on this relationship in light of the historical precedent, its application in risk assessment and subsequent risk management decisions, the analysis shown here and in the background document and the study design objectives of the AG. Handler Task Force.

Briefly I've presented here a plot from Reiner and Sevord in 1985 and this is also taken from our Subdivision U Guidelines presented to the SAP in 1986. This graph is for air blast applications, each point is a mean of a number of exposures, that's why the value there is 23 but they're all the mean. And the authors noted that the r value of 0.7 in the Spearman Rank correlation was remarkable considering the umbrella of uncertainty involved.

On the x axis you have the application rate, pounds AI per acre, and dermal exposure in milligrams per hour. And then the core at the bottom is one of the few that I saw directly relate to the amount

handled or applied as opposed to the application rate or the exposure in milligrams per hour.

So again the 6 case study scenarios, I don't think I need to rehash anything, I think Jeff Dawson did a good job of explaining. If anybody has any thoughts or clarification on what is any of these equipment or formulations we can certainly talk about that.

The next set of slides will be a discussion of how we use the PHED data, how we put it into the, a reasonable format to channel Doctor Johnson's published works it's going to get messy in here.

So again most of the problems and issues related with the PHED data, at least for this analysis was the composite worker methodology, it has other synonyms as well. And again as a reminder this is how the unit exposures are calculated on a body part by body part basis across individuals and you know if you went through the lognorm, you know, the geometric mean and the mean for the distribution that was determined for that set of body parts. So the specific issues I'll be talking about are the grading criteria, again briefly and then the concurrence of measurements across body sections and then patches and whole body dosimetry.

The concurrence of measurements across body sections, actually Curt kind of just, Curt Lunchick kind of just touched on that as well in the discussion of the workers that actually had gloves on at the same time as normal work clothing and that sort of stuff.

So again the PHED grading criteria and our surrogate table, the grading based on laboratory and field recovery and as has been discussed, most of our assessments attempt to use A and B grade data only. And for this analysis we could have simply used the entire A through E data, however we split it up into various combinations and these combinations were not, you know, meant to recommend for a risk assessment of anything of that sort, it was just meant to optimize the data and show a number of plots.

Problem 2, again I mentioned the lack of current measurements of body sections and the lack of concurrent inner and outer dosimetry measurements. For example, and what I mean by inner and outer dosimetry measurements is gloved hands and normal work clothing, in some instances that was not the case. There we'd have, measure their bare hands with normal work clothing on and in that case it was difficult to combine the two to present a measurement of total dermal exposure.

So our solutions to this were just to use the data as it was presented, as we had it and this was by analyzing hands separately, just the body, you know, the torso, arms and legs, chest and back and then total dermal when the measurements were concurrent, either completely outer dosimetry, bare hands and then deposition of residue on the body or total dermal underneath a single layer of clothing, normal work clothing and gloves for the hands. And also analyzing both inner and outer dosimetry.

For this analysis though we may use these, the combination in risk assessment, we did not present the combination of outer, for example, outer head plus inner body for air blast applications. In that case we have just presented the head for the purposes of just, of showing that and the inner dosimetry may be trumped by the outer head patch. And also for inner dosimetry we recognize that the uncertainty that has been discussed previously, the inner, the penetration underneath, you know, the protective clothing or the normal work clothing and the treatment of non-detects, one half the limited detection or quantification.

It should be noted that these hands and body and head and inner and outer dosimetry, these separations are, do have some practicalities in terms

of applying protective equipment and that sort, so it does relate in a sense to risk assessment and PPE practicalities. However it also was beneficial for the analysis just to present proportionality across body sections.

And the third problem being the biggest problem of using the data as a whole, the whole, an attempt to use the whole database and now the problem is combining patches and whole body dosimetry. As we know, in patch dosimetry not all body parts were necessarily measured. Now this is okay for PHED analysis because of the composite work or methodology discussed. And with whole body dosimetry, whether it was a true whole body dosimeter like a union suit or a top and bottom, all the body parts were measured and by, you know, body I'm talking about again the torso and the arms and legs.

And the whole body dosimetry measurements couldn't be separated to match with any of the patches that, you know, we would just choose for the analysis because in the software the whole body dosimetry was basically broken up into body parts and an equal value given to that body part and then adjusted for the surface area of that body part. It was, they were added later than the patch dosimetry data so it was

needed to do that to, for the amenability with the software.

And you may have noticed this going through any of the raw data that we provided that you saw equal measurements for, you know, the forearm and lower leg and that was in most cases because of the whole body dosimetry. So our solution was to use a minimum of critical body parts for the patch data.

And across all scenarios these were, we chose the forearm, chest, back and thighs and according to the surface areas in the PHED reference manual this represented approximately 60% of the total body. And we thought this was a reasonable combination of the data from the patches with the whole body dosimetry. And again that is only talking about the torso, those body parts would then be added to the head or the hands, depending on the scenario

So here is just an example, this is open cab air blast applications, outer dosimetry only. And the table on the left shows what was available in PHED and the table on the right shows what was used in the analysis, the yellow being, the yellow highlight being what was used and minimum again being the forearms, chest, back and thighs.

And you can see we went, as Curt Lunchick

mentioned this morning you go from 77 total monitoring events to 56 that had at least the minimum.

So now we're getting into the analysis starting with some statistical background and mathematical background and then getting into the actual plots.

The method used is a log log linear regression and the directly proportionate relationship between two variables, x and y exists if a plot of log x versus log y produces a straight line with a slope of 1.

This is an involved slide but I will try to go through it fairly quickly. The first box there, the objective is to evaluate whether a proportionate relationship exists and for example if you double the amount of active ingredient handled we assume that the exposure is then doubled. And the first box is the, basically the unit exposure reorganized into the mathematical equation shown.

And throughout, you know, with showing the log on both sides throughout we are focused specifically here on the coefficient B1 and the equation after you log both sides of the original 1, the B1 is then the slope of the line that, and we are looking for the slope of the line to be 1 or not

significantly different than 1. Using a 95% confidence interval of the slope as a significance test and the plots that you will see are ordinarily square regressions, however we do have, since the background document, conducted some hierarchical linear modeling and that will be discussed as well.

So the next set of plots that you're going to see have all these various characteristics that are shown up here, the PHED grading will be noted in the title of the plots, whether it's inner or outer dosimetry, the statistical information will be noted on there, all of them I believe are in the upper left hand quadrant. And we'll be discussing some of those statistical limitations as we go through them.

Also I will be going through some anecdotal data characteristics and I will be just discussing non-detects and inter and intra-individual variability and, you know, some other factors. And I think this is important because it gives another view of the data and, you know, we'll be discussing proportionality but there is value added to some of the anecdotal stuff that I'll be, anecdotal factors I'll be talking about.

And from exhibit C there were 35 plots shown, we're only, I'm only presenting 4 here. I have loaded onto the laptop here all 35 in slide form in case

anybody is particularly interested in any one of them. We can bring them up. Unfortunately I probably won't be able to talk as much at length about those as I will the current ones because there's a whole lot of data.

So the first one shown here is the open cab air blast applications and this shown is only the head exposure and that is the outer dosimetry. PHED grades A through D and you notice on all these plots the exposure is on the y axis in micrograms and the pounds of active ingredient handled is on the x axis. They are both log scale, the solid black line is the regression line.

The dash black line is a slope of 1 line. For comparison the blue dash lines are the 95% confidence interval. And each data point is color and character coded and again they represent, each point represents a single monitoring event. In this case again they are the head only. And we chose this one to show first, I was told to put my best foot forward, and at the beginning we had, we were on, we assumed that if we were going to see anything because of the uncertainties with inner dosimetry, that outer dosimetry would likely probably show the best relationship. So, and in this one it appears that, you know, this is a reasonable situation.

Here's study 435, there were 4 workers, each conducting the activity twice, again open cab air blast applications. And this is the breakdown of the workers showing some of them close together, some of them not very close together and this again may get into the inter and intra-worker varia individual variability that we'll discuss later.

This is the same worker, he has conducted on consecutive days, one in the afternoon and one in the morning and, you know, we're seeing almost a threefold difference with the same person. This is the only D grade data, data point on this plot without this point as you'd see I believe I included it in exhibit C, the A through C, head only exposure for this scenario and the slope of that line is 0.79.

Study 1006, it's one worker conducting the activity 5 times, all 1 hour applications on the same day. And the, it just so happens the one in the morning for an hour was approximately 5 times greater than the one in the afternoon for an hour. And the only non-detect on this plot is this worker from the study 442. It might be worth it to note that this limit of detection was quite a bit higher than the white squares and red squares right next to it. That gets into, you know, again the uncertainty.

We've also provided the conversion of that regression line onto the original scale. On the original scale we would assume that there would be no exponent there, that would be an exponent of 1. And so it's just there for comparison. And again that almost looks like a straight line but it actually will plateau with a slope less than 1, with an exponent less than 1.

The next plot is open mixing/loading liquids. This is a bare hand exposure, only grades A and B. This slope rejects, this plot rejects independence but proportionality may not be reasonable in this case, a slope of 0.16.

In study 434, this is the same worker in Pennsylvania. The collection method was cotton gloves and the difference basically in his exposure was between the morning and the afternoon. And his left hand, the left hand in the morning was a much higher exposure than the left hand in the afternoon.

Study 9003, two workers conducting the open mixing/loading twice at about 0.3 pounds handled and it's, and these are also cotton gloves. And in this case you might think that both of those, that's actually 4 points, both of those at the top might be the same worker and both of those at the bottom might be the same worker, but it's not, it's the reverse.

In this case, study 9001, three workers, each doing it three times. This is perhaps the opposite situation of what I've just showed because worker C had the three lowest while worker A and B were pretty much at the top.

In study 465 the 6 workers each conducting the operation once and again they're all, so far all are shown in the collection method of cotton gloves. And as might be expected, although maybe the minutes aren't that significantly different but I thought it would be worth it to point out the longest in that entire study was 20 minutes and that was the highest exposure as might be expected, but the second highest was the shortest time period of open mixing/loading the liquid.

In study 9010 they used a rinse methodology to collect these on 2 workers each conducting it 5 times, the operation 5 times. And worker A has consistently the highest for those measurements but also has the lowest.

In study 447 again 2 workers, each 3 times collected with the rinse methodology so it shows here that we are comparing as if they are equal. The both collection methodologies. And again the conversion onto the original scale.

The next plot I'm showing is open loading granules. This is actually total exposure underneath the normal work clothing and the hands are underneath gloves. This is one of the few, I believe it was 4 of the 35 that actually wound up with a slope greater than 1. Most of the others had a slope less than 1. And there are only studies in this case from about 7 to I think to 66 pounds handled.

Study 1004, 6 workers, each doing it a couple of times to get to 15 measurements total, these are all A grade. These values here are all non-detects and these values here are, the body value is all non-detects so it brings into the uncertainty of the treatment of those. Again anecdotally these are both, there are 2 data points there, 2 different people. In one person the majority of the exposure came from their body. The other person, the majority of the exposure came from their hand, hands, sorry.

In study 1011 there are no non-detects in this having multiple workers performing the operation multiple times to get 27 measurements, A and B grade. There is what the, on the original scale what a slope greater than 1 would look like. In this case a slope of 2.21.

The last plot I'll be showing is, well you

know, probably the ugliest one that we've seen. Open loading granules, this is bare hand exposure again, however all these data points here are grades D and E, a slope of, I believe the only slope less than, you know, a negative slope.

In study 425 here using cotton gloves again to collect the residue, all D grade, there's 3 different workers. The operation lasted between 3 and 5 minutes.

With study 448 the operation lasted between 2.5 and 4 hours, at least I believe the operation may have lasted 2.5 to 4 hours. The open loading, the loading of the granules may have been like the previous study, only a couple of minutes and they did not collect the residue until half the day and they had done other activities. So these are all E grade.

DR. LU: This is Alex Lu. Just a clarification. You just mentioned those are for bare hands. But the subjects before had gloves.

MR. CROWLEY: The gloves means the residue was collected with cotton gloves as opposed to a hand rinse collection.

DR. LU: And how about the previous case, the inner dosimetry, the previous slide, yes. How about those, their hands? How did you sample their

hands?

MR. CROWLEY: I do not know. I can, that's within the database but I'm not sure. It was difficult to know every specific about everything.

DR. LU: Okay, thank you.

MR. CROWLEY: You're welcome. So this is the case in 448 where there was only, this is the only case where a worker performed the same, the task twice and he was measured twice.

And again this is a note again about the time to residue collection. Basically the same residue, the difference being 22.5 pounds handled versus 750 pounds handled and then 3 minutes versus 2.5 hours. I'm just not sure form the data whether that was a 2.5 hour operation of they just collected the residue following the morning's operation. And again the original scale inset.

So some of the limitations that were discussed following the initial exploratory analyses using ordinary least squares, you saw from, in various cases the clustering of the studies and ordinarily these squares treat each data point independently so that violates the ordinary, that ordinary least square assumption. And the study design may implicate a more appropriate nested or hierarchical linear modeling

method because of the measurements within workers within studies problem that we have.

In this table, sorry it's not very well read, however it compares the ordinary least squares against the results that we have gotten to with hierarchical linear modeling, the difference in the regression slope and the 95% confidence limits. The highlighted ones in red are those that, where before a slope of 1 was not included in the confidence limit, that using the hierarchical linear modeling method, it now does. And I've shown the corresponding slides of the previous 4 plots. Those other 4 had been ones I was going to show as well so we performed the hierarchical linear modeling on those but I can bring those up. As I said I have those, the slides on the computer.

Now again to some discussion points. Again, as has been discussed throughout the past three days, there is data limitations with PHED, the combination of patches, the small range of magnitude, amount active ingredient handled, clothing penetration factors, treatment of non-detects and then the rise methodology and as a result, can we draw reasonable conclusions using the PHED data for the relationship between exposure and the amount of active ingredient handled? And we note that an improved study design could resolve

these issues or, you know, completely control the experimental design where you could, where you would control everything and you could determine the relationship better. But that likely is not the case for us.

Additional points I wanted to bring up is how, you know, in light of how this has historically been used in risk assessment and its practicalities for risk management decisions, we can limit in various ways the amount that a person can handle so it fits well in our risk assessments.

But we also propose there could be alternative predictors of exposure, exposure duration, alternative metrics for application rates in concentration, however, with the last two there might be some, correlating with the pounds of, amount of active ingredient handled.

But this, I believe that decoupling in a closed loading system was brought up previously during this week and that could be a situation that we would, you know, maybe look into where the exposure is better predicted by the amount of times that they lock and load the closed system as opposed to the amount that is actually going through the system.

The next point about conservatism of assuming

proportionality, as I've shown, only 4 of the 35 plots had regression slopes greater than 1 which would not be conservative. However I think for, I'm not sure, two of those, half of those, it had a slope greater than 1 but the confidence limit also included a slope of 1. But there are a couple where it did not include 1 and the slope was greater than 1. So it appears in most cases that the high end will be conservative however that doesn't, we may be underestimating at the lower amount handled.

And we've also discussed are our current scenarios appropriate? Can certain scenarios be combined or conversely do certain scenarios need to be subsetted? For example is there a significant difference between granules and dry flowables? Or should there be a separate unit exposure for high amounts of active ingredient handled versus low amounts of active ingredient handled? And the first example again would get into the problem of how this all fits into risk management, regulating a dry flowable formulation using granule data and how that would work out.

So finally perhaps the understatement of the week, the current data is not designed specifically to address this issue however we feel that with the

improved study design we may be able to do more advanced statistical analysis and it will aid in either reinforcing this assumption or perhaps informing the applicability of another predictor of exposure. And that is all I have. Thank you.

DR. HEERINGA: Thank you very much, Matthew.

MR. CROWLEY: You're welcome.

DR. HEERINGA: Questions from the panel on data presented. I know Alex had a question before. Do you have any follow ups at this point?

DR. LU: No, I don't.

DR. HEERINGA: Doctor Popendorf and then Doctor Bucher.

DR. POPENDORF: That was an interesting presentation and a good explanation. But I just wondered, some very similar plots were in the review document and I wondered if you could just discuss how those that we had earlier differed from what you're presenting today?

MR. CROWLEY: The results were all over the place. In some instances I could have shown all plots that had, all slopes of 1 with tight confidence intervals but I chose to show one. In other cases, there was only one that I showed that had a negative

slope. Hold on one second

DR. POPENDORF: Well then maybe a simple explanation to clarify, that that we were presented in the review is part of what you did, but didn't show, is that true?

MR. CROWLEY: Yes.

DR. POPENDORF: Okay, so

MR. CROWLEY: There were, there were many more, the methodology was basically the same for everything else but again there were 35 of them and I just

DR. POPENDORF: Yeah.

MR. CROWLEY: there were really, I think I could have picked any to show and the same anecdotal characteristics would have come out for each one.

DR. POPENDORF: Okay.

MR. CROWLEY: But, you know, it was kind of a good, bad and ugly show.

DR. POPENDORF: Okay, and the ROB, rest of body in here are the same critical

MR. CROWLEY: Yes.

DR. POPENDORF: body parts? Thanks.

DR. HEERINGA: Doctor Bucher and then Cynthia Hines.

DR. BUCHER: Again from a naive standpoint I was just wondering if the Agency has any explanation for any reality associated with a proportion greater than 1? More active ingredient exposure in relation to the pounds of pesticide used. Is there any situation under which one would realistically envision that occurring?

MR. CROWLEY: I cannot think of anything off the top of my head. David?

MR. MILLER: Yeah, we had talked about that, we can potentially understand a proportionality of essentially less than 1 simply because we talked about, a little bit about the caking affect as you get a certain amount on there. But just it going up, I mean it's almost like the person would be a magnet for it. The more they handle the more it gets attracted to them which just seems strange to us.

DR. HEERINGA: Or the more they handle the sloppier they get. Doctor Popendorf.

DR. POPENDORF: I was looking at that earlier data that we had available and was going to make this comment later, but one explanation if you look at a lot fo the examples in the review at least where the slope was greater than 1, was the inner data and the outer data was 1 or less. And one explanation

would be the inner data, if you're building up dose on the clothing, it takes awhile to penetrate, if you assume a constant rate over time the people that handle more are going to have longer exposures and the delayed effect so they end up getting more proportional dose to the amount handled because they were working over a longer period of time.

MR. MILLER: And some of that might too also be the, remember their default was half LOD and so if you expect lower concentrations on the inner it might also, I mean just the default could have been partly the default values.

DR. HEERINGA: Cynthia Hines.

DR. HINES: Yeah. Just a couple of comments. In your conclusion you stated that in some cases it, you know, this proportionality may be reasonable. But given what we had in our document, the overview document from EPA and some of what you showed today, do you have enough data to state that in some cases this assumption is unreasonable?

MR. CROWLEY: I think you can make that case as well.

DR. HINES: Okay.

MR. CROWLEY: But it's

DR. HINES: You kind of stated the more

positive picture here and I think there is some evidence that in some cases it may be unreasonable. But that needs further exploration.

MR. CROWLEY: And I guess the question is, is our current, are we able to do that with our, make a definitive determination

DR. HINES: Right.

MR. CROWLEY: with the current data set?

DR. HINES: Right, that was why my concern is, do you feel you have enough data yet to make that statement? I might also comment that you very well might find substantial colinearity between the amount of active ingredient handled and exposure duration as well. Those often go hand in hand.

And my last comment is I find the idea of perhaps looking at seeing whether some of these scenarios can be combined is intriguing and maybe that's something the new data collection effort in the Agricultural Handlers Task Database will maybe be able to contribute to that. So that would be an encouraging thing.

DR. HEERINGA: Dallas, Doctor Johnson.

DR. JOHNSON: Would you mind putting up your slide 20? Thank you.

MR. CROWLEY: Would you like me to go to a particular characteristic of the data or is this fine right here?

DR. JOHNSON: Well if you could back up to where you didn't have all the extra stuff on it.

MR. CROWLEY: Okay.

DR. JOHNSON: There, that's fine, thank you. This is probably the only one that I recall in which within a particular study there was a fairly decent range of amount of ingredient handled. The yellow dots, there's a little bit of a range in the amount of ingredient handled and in the red x's there's a little bit of range in the amount of ingredient handled.

MR. CROWLEY: Yep.

DR. JOHNSON: And if you look at the slope, if you were to fit a slope through the yellow lines and you were to fit a slope through the red x's, those slopes would be quite different. And neither one, well, actually the one through the yellow might be closer to having a slope of 1.

MR. CROWLEY: Yeah.

DR. JOHNSON: And the ones through the red x's would have a slope much, much greater than 1. And then if you would also put on the slide 19. In this

one here there's really no range at all within any of the studies in terms of the amount handled. It's basically the same for all workers and at least in the proposed new data there will be, I understand there will be some range that you'll try to achieve which would be very helpful to try to assess whether the slope is equal to 1 or not.

The other point, or a question I guess that I have and maybe we're not, you don't have to answer it yet is, why do we even care whether the slope is greater than, equal to 1 or less than 1? Why not use whatever the slope is? I'm not sure I understand the rationale for looking at the exposure as a ratio of the amount of ingredient handled.

MR. MILLER: Yeah, basically, I mean I'll start and then if Matt wants to add such, this basically the slope of 1 essentially equates to essentially a proportionality of unit exposure. And in essence what we do is we, that's essentially a principal part of how we do risk assessments. Essentially it's you take the unit exposure which essentially is milligrams per pounds AI handled and then multiply it by the pounds AI handled as part of the equations that were showed, that were shown earlier. So the question of whether it's proportional

or not essentially gets to whether that multiplication of pounds AI handled by milligram exposure is a valid way to go.

DR. JOHNSON: And what do you intend to do if the new data shows that it's not proportional? At least not proportional with respect to a slope of 1.

MR. MILLER: What we'd probably do is start looking at other, as Matt had mentioned before, other aspects that might be included. For example the couplings was one of the things. It may be that that's, if you have a close system it's not the pounds throughput that goes through it but rather the number of times you connect it. We would take a look at kind of other things that might explain it. But again I think our thought is that the new proposed data to be collected by the HTF, I mean that is one of their secondary objectives, to look at that and they'll be collecting enough of that ancillary information for us to look at various things like that.

DR. HEERINGA: Questions or comments from other panel members? Yes, Cynthia Hines.

DR. HINES: I just had one quick thought to follow up with what you were just saying. I know in the AHETF's proposal that they're going to try and get good variability on the amount of active ingredient, so

that will help you evaluate this proportionality. If it should be the case that it appears some scenarios that that's not a good normalization unit or measure and say you wanted to look at a number of couplings or a number of tank loads or something like that, the current proposal isn't really designed per se to optimize that variability or maybe that's something that the task force can comment on. So if it's not designed to optimize that, it may limit your ability within that database to evaluate that. It's just a thought.

DR. HEERINGA: Steve Heeringa, I have a, if we could go to slide 18 and this is just an observation. This whole discussion, and I think your approach to test this assumption of proportionality is correct, but it forces a linear model on this data over a wide range of exposures and I think Dallas was sort of getting at that earlier with slide 20 in which you have two nice sets of data, probably the nicest sets of data we have for this exercise.

If you wanted to write a textbook, Dallas might use it. But the real world looks more like slide 18 and did you look at a quadratic term in this over this range of exposure? I mean in my eyeballing of this I would sort of, I would draw a quadratic function

through these data.

Secondly, this limit of detection of limit of the quantification and sort of staking these below LOQ values at a constant and then incorporating them in a regression, you know, economists deal with this problem in income and other types of things where you have censored data and you might want to actually, as you pursue this work here, maybe I'll just ask, have you looked at things like, you know, censored regression, a tobit regression in which you assume this lognormal distribution continues uninterrupted or in a smooth fashion below the limit of quantification, but you don't stake the actual values at a point value at 50% of LOQ. Did you do any of that work, Matthew or David?

MR. MILLER: We haven't done that in relation to the, I mean it's something we plan to do, it's something we've not done in relation to this specific SAP. We do do that however on the residue chemistry side, I've started using it.

DR. HEERINGA: I knew that you might do that.

MR. MILLER: Yeah.

DR. HEERINGA: Do you see it making any difference when you do that in your modeling in the pesticides databases? Residues databases.

MR. MILLER: In the residue databases I, let me think for a second, yeah, I mean you get, we use essentially maximum articulated procedures to do that. In terms of the estimates of the slopes, yeah, I mean it can if you end up with a lot of 20% non-detects or so. But I mean it's something we will probably at this stage kind of look at in terms of this data here, take a look at it. And certainly I think the more important part is we'll be mindful of that with the AG. Handler Exposure Task Force, their proposals. If their, essentially if their detection are low enough that they don't, they're less likely to cause issues or problems.

And then I guess on the first question you asked which was in terms of we did not test quadratic terms on this. We had discussed however that it may be that if the equipment varied, just to look at this one for example, if the equipment one uses at 100 pounds, is that, yeah, at 100 pounds is different from that at 1 pound for example, that might explain essentially kind of a disconnect there and we're talking about doing potentially tests to look at that. I mean one would be to put it in a quadratic term, or in other words a Chow test, things like that. So that's something we're aware of and we're thinking about and as we look more into it it's something we may include.

DR. HEERINGA: Thank you very much. Other questions from the panel at this point in time? Let me ask a question of the Handlers Task Force.

Doctor Landenberger, Bryce, do you expect your presentation to run say 30 minutes or no. Okay well let's, before the break then let's proceed then to the next component of the presentation for this morning and that is a presentation from the Agricultural Handlers Exposure Task Force on the statistical basis for the AHETF data development program. And I'll invite up Doctor Bryce Landenberger, Curt Lunchick and Larry Holden.

DR. LANDENBERGER: Let me clarify.

DR. HEERINGA: Okay.

DR. LANDENBERGER: I misunderstood, there's actually two different components to the presentation. My particular component will not take a half an hour however

DR. HOLDEN: It would be out of sequence.

DR. LANDENBERGER: -- it would be out of sequence in terms of Doctor Holden's presentation

DR. HEERINGA: Okay.

DR. LANDENBERGER: should go first and his is considerably longer than a half hour.

DR. HEERINGA: Is it longer than 45

minutes Doctor Holden?

DR. HOLDEN: I could talk faster.

DR. HEERINGA: Well that's what I don't want to do to you but I'll tell you what

DR. HOLDEN: It's hard to say, it might be 45 minutes might be more reasonable yes.

DR. HEERINGA: Let me, let's just go ahead, it is only 20 minutes of 10:00 and if you'll permit, if we feel that there is a breaking point, the presentations and collection will certainly go longer than 45 minutes or an hour so if you'll permit we may find a logical breaking point and come back. I don't want you to rush and I also don't want to interrupt at a point where we would lose track of where we are but I think it may make sense to proceed at this point and we'll find an appropriate breaking point. Thank you very much. This is Doctor Larry Holden I believe.

DR. HOLDEN: Can you hear me, oh great. Oh, thank you. My name is Larry Holden, I'm a statistician working with Silken and Associates and I'm also involved with the Agency Task Force, I mean not the Agency Task Force, I'm sorry, the AHETF Task Force to help with the statistical and both analysis and design issues. I apologize if my voice goes in the middle of this, I'm fighting a cold and so if I start

sounding like the Cookie Monster you just have to listen a little closer.

This talk that I and my colleagues are doing, my talk originally was supposed to be together and would include both the introduction to the design, I'm sorry, to the study and the issue of normalization and then followed by the sampling and sample size issues. However it make it much more logical because of the charge questions on sample sizes on the second to split the talk.

However what I will do and I hope I remember to do this, is that when an issue comes up with regard to normalization or some other aspect that will be talked about again or covered or maybe revisited at least tomorrow I'll try to mention that fact and bring forth any issues that might be relevant. So I hope I'll remember that.

My presentation is really in three sections and so a break could be in any one of these, Doctor Heeringa, if that works out for you as well if it gets too long.

The first two sections are more introductory in a sense and the last section involves more about the issue that was just discussed. The first section of the talk basically describes the target population that

their study, pretty much the target population that our study seeks to address. The second portion is going to address the issue of where the Task Force's study falls within sort of the spectrum between say a purely descriptive versus an experimental approach. You can probably guess from the discussions here it's more towards the former than the latter. And finally we want to discuss or describe briefly how this issue of unit exposure is incorporated into our study and what impact it has. And again some of the panel members have already anticipated me already. They were discussing, the previous speaker.

The pesticide handling scenario has been described before by Doctor Canez and the definition is here, I won't repeat it except to say that it is a composite of many, many different things that are going on but all have some commonality. So it's not as though it's identical like in a controlled experiment, it's a mishmash of many different things but they are tied together in some, by some common framework.

Just to be a little easier, not a little easier, at least a little simpler from a statistical standpoint I'll try to put this in a little more perspective so we can sort of lay the groundwork for the sampling aspect later on. If we think of the whole

universe of all these conditions of which this scenario comprises, what we're talking about is looking at specific conditions or elements of this universe of what's considered a scenario. And so what I've got represented here is just a box representing some arbitrary scenario and three example conditions that are in this scenario. When I just say condition I'm talking about something maybe more than most of you realize, I'm thinking of every possible value for every possible parameter that could affect exposure. So clearly most people aren't going to have the resources to investigate all that but that's what I'm talking about when I'm mentioning a point in this box here.

The, I do want to clarify it because the issue will come up in my colleague, Bryce Landenberg's talk afterwards dealing with inter-person variability, is that what we're, we're not saying that each point here represents an individual, each point in this universe represents a set of conditions of which the individual is only one part of. And so an individual might have in this example here several different, represented many different times depending on the different kinds of conditions he or she can experience. And those conditions might even be differences in behavior.

What I'm showing here is that if you knew all those conditions you might be able to, that would, if you knew some sort of the, if you knew the function or the mechanism perfectly you might be able to predict the exposure that went with it. Even though members of this panel are probably more experienced in getting at this function g than most, I'm sure that even they probably can't specify this exactly. At least I'd be surprised if they can.

There are, and so this looks very deterministic here and I'm saying that there are known factors and unknown factors that determine this condition. But if you knew these and knew how they were related I suppose you could come up with an exposure and so it's not so important here now whether they're talking about actual exposure or measured exposure. But if you want to think of it in terms of measured exposure then think of some of those factors as being what influences the measurement as well. So say an analytical error for example might be part of that.

The reason I bring this up is it's going to enter, it's going to come into play a little bit later. But the diversity of the factors or the conditions within this scenario universe all throughout, and they

are mapping into an exposure is basically going to give you a distribution of exposure and that's because many of the conditions might produce an exposure that is say are closer together than some others. So it's going to produce what I'm calling a distribution. And if you look at those exposures that those conditions generate, that is what our study is attempting to get at.

In other words every monitoring unit, every sample that we collect and monitor hopefully is a member of this universe although the reverse is not necessarily true. What we do is we restrict our target population somewhat from that universe and that's restricted in some of the obvious ways that you see here that Doctor Victor Canez talked about earlier, that we limit our, we're only talking about occupational agricultural handlers thank you very much the, we always talk about workers with some prior experience, at least 18 years old, in good health, English or Spanish speakers, some of these considerations are restrictions for a practical reason, many are for ethical reasons, legal reasons.

There's also restrictions that don't really necessarily apply to the workers, but to the conditions. For example you've heard yesterday or the day before that we are restricting it to, they have to

be at least a half day's worth of work for example before that, they entered into our study. So we aren't talking about someone who works 15 minutes.

We also are, the workers have to follow the Worker Protection Standard. The application rates have to be legal. So there are many restrictions, some of those are implicit, some are explicit.

I think it's important to keep in mind what those are because if one were to, one could always talk about generalizing to the target population but if you generalize to the entire scenario universe you have to use something more than statistics, you have to use subject matter knowledge. I think the gentleman right, Doctor Hughes is it? Yes, you had mentioned the idea of generalizability I think the other day and that's an important concept is that to be able to generalize you need to keep in mind the target population and what it doesn't include. So I think that's very important. And I think some of the panel members mentioned that earlier.

What I'd like to get at quickly now is to explain how our study fits into sort of a rough spectrum between purely descriptive studies and say experimental studies or non-experimental and experimental. These, this isn't a very fine sharp

distinction, it's really more of a spectrum I guess you would say. But nevertheless I'm going to try to describe both extremes and then show where ours fits in and why it fits in there. And that gives, and that means that some things it can do and some things it cannot do. Both of these approaches by the way I consider are valid and very useful but they do have different goals and limitations.

The descriptive viewpoint is one in which this complex function I think of factors that determine exposure can be viewed as something a little novel if you want to call it by something a little simpler, and that is that we're saying exposure might be approximately equal to some measure of central tendency and I mentioned the geometric mean here just for convenience and some components that we're calling random or treating or thinking of as random.

The reason I've got these as multiplicative for those who are worried about that or concerned about that is that you might normally be seeing this additive. Most of the data that we've been working with seems to be at least logsymmetric if not lognormal. And it's more convenient to think of the errors as being, or the random effects being multiplicative. But if it does both you just think of

the log of it and it'll be additive again.

In the descriptive, at least how I'm viewing it, I think the focus is more on describing the expected variance of the worker exposures. And may their location, but certainly the variation and the way in which various statistics, the random effects and I do mean that to be plural because those effects don't, are not going to necessarily seem variable, they're a mishmash of many things that are modeled but necessary, not necessarily modeled independently.

That, the variability of that tends to be in this kind of approach, it tends to be very, very large. And in our studies we're talking about geometric standard deviations for those of you who think in those terms of around 4 which is, for those of you who deal with coefficients of variation we're talking around 250% as opposed to maybe an agriculture where they're dealing with CVs of 20%, 30% maybe as being minimal. For those of you who like to think in terms of orders of magnitude we're talking two, two and a half orders of magnitude between the observations. So a large variability there.

But in other words in this kind of study, at least how, the extreme in which we're talking about this descriptive of the individual observations, it's

the distribution that's of interest and for example the Agency has already brought that they want to look at things like the mean and upper percentiles of this distribution for use in risk assessment.

The experimental viewpoint, or at least how, you know, on the extreme side is one in which we have the same function controlling things but we're approximating it by some simpler function of known factors and also random variability goes along with it as well. However in this case the random variability is more of a nuisance than it is something in which to describe.

It's something we would like to, and ideally get rid of or reduce to as low as possible and explain things in the, in whatever function of some known factors, because in this case it focuses on discovering relationships between known factors in exposure.

Not to say this can't also be done in a descriptive standpoint as well but it's more, oftentimes more a better approach to an experimental viewpoint, especially since in order to get at this function in an ideal situation what you'd like to do is experimentally vary some of those known factors and how as many of the others as possible as constants and unfortunately throw maybe some fo the rest and those

you don't know and treat those as error. But you still want to minimize that. The useful for this is that it's very useful for predicting exposure for a particular set of conditions, useful for discovering the relationship, at least conditional on things you hold constant. But it will destroy, by design, any natural co-occurrence of the factors that may occur in the population and that's a good thing in an experimental study, that's what you want to do.

But our program really is focused more on the descriptive I think than the experimental and I wanted to emphasize this, especially because it's very easy to get off into thinking that this study can look at all different factors and compare things and look at the relationship between this piece of equipment and that piece of equipment, this factor and that factor and you soon run out of sample size very quickly in that approach. And we're not really getting at that approach. I just wanted to specify that, we're not really necessarily concerned with how well the data could undercover relationships, although as you'll see we do a little bit of it.

In general I think it's pretty obvious why you proposed the approach but for more obvious reasons than not, exposure measurements are, if you manipulate

them in situations where that's not necessarily typical and manipulated conditions, when you're trying to use it to describe the distribution it's going to be misleading because you're, you've got conditions that don't necessarily occur very often if at all. The natural distribution however is something that can be used directly for, as you've mentioned, tier 1 assessments. To get at that distribution from an experimental study means you have to throw in assumptions, additional assumptions but the distributions of the things that you used in the experiment.

So it becomes more complicated at the very least. And it's also complicated to design the studies. Remember now, our, the Task Force's study, we're not doing an experiment, doing a study and plan to analyze the data and publish it. We're generating data for others to use. They will be used either in regulatory purpose or perhaps in publications of their own or whatever. So to try to design a study to anticipate all the possible ways in which a, potential users could use data and what kind of relationships they may want to find is going to be not only very complicated but could result in very, very large sample sizes if we try to accommodate everyone, every possible

use.

But, after saying this, and I think Doctor Heeringa, no I'm sorry, Doctor Portier was discussing yesterday, the day before, I can't remember which, about all the factors in which the, all the measurements in which the study would be taken, would be measured, we're going to be measuring a lot of factors that go along, I mean, almost as, very, very many, a large number, so those are going to be available.

So users, there's nothing stopping users of this database from running regressions on as few or as many factors as they wish, but we have a caveat here that many of these factors are very highly correlated and I think you mentioned that earlier on that, today, that they're highly correlated.

Time for example, pounds per AI handled are highly correlated. And certainly in the scenario universe, and since we're trying to collect data that mimics it to some extent, then that, our sample is going to have the same kinds of correlations in it if we're doing a good job. And because of that you've got this multi colinearity issue or confounding, whatever you want to call it, and why you might be able to use other normalization factors for example, you are not

going to be able easily to maybe distinguish one factor from another or perhaps use them jointly because as the statisticians know quite, you run into problems really fast. And Doctor Johnson who taught me messy data analysis I'm sure knows this quite well.

Some of the problem situations, I've already mentioned a couple but some of these factors that are obvious, that amount of AI handled is going to be highly correlated with the number of loads, hours worked, acres treated, spray volume mixed. There are going to be some factors that don't vary that much, are hard to vary and some of these might do with factors that are behavioral that are very, very hard to control anyway. And so they're going to come out to be correlated in the data.

So the bottom line is there's not going to be, we're not going to, the Task Force is not going to be able to guarantee that AHED data can be used successfully to discover factors that influence exposure. The data is going to be available and therefore we're going to try but there's not going to be any guarantee, it's going to be limitations.

But, and I'm always throwing out a but, and here's another one right here. There is one factor that is recognized as being very important, we've been

talking about it today especially, and that is the amount of active ingredient handled by a worker during the daily task. I think it's, I'm not going to beat a dead horse into the ground, but it's obviously important. It obviously, and I say obviously because even some of our data that we've examined shows that it is, it is associated with exposure, not necessarily the same day, it's always proportional.

Although to be fair, every data set that we've examined so far has shown that is proportional within the context of our data, or that our data are incomplete and unable to distinguish whether it's proportional or not, we can't make a conclusion.

So us, the Task Force as of right now can't go out on a limb and say that we have evidence to say that for scenarios it is not true, we have evidence to say that it is true in some cases and we have, our data shows that it's equivocal in others.

But in reality, and I think someone here suggested it, if you know that you go down low enough you're going to be picking up background that's going to flatten out, if you go high enough, pounds per AI handled, stuff, Doctor Popendorf and I were talking earlier this morning, things are going to be falling off left and right. So we know that it's going to

flatten out on both ends if you go far enough, so at the very least it's going to be a sigmoidal at the very least, although it'll be more complicated in between.

In particular though, because the Agency and others are interested in normalized exposure or unit exposure, the two will be synonymous, then we are going to look at that in our study as well. It's going to be like an add on, or it's going to in a sense, even though we're collecting data we're going to orient to try to maximize the information we pick up from this one parameter.

Like exposure, we can also think of normalized exposure having a distribution too. There's nothing, any parameter in that, any measurement, any one of those factors that exist or a combination of factors can be described by data. So you could think of normalized exposure just like exposures having a distribution and then you could look at it to say the mean of normalized exposure multiplied times pounds per AI handled and come up with an estimate of a, the mean of exposure conditional on a given amount of AI handled.

Likewise with the percentile. These are the ways in which the regulators for tier 1 and maybe tier 2 assessments tend to use the data. And as I said it's

always meaningful to describe the distribution or at least doable to describe the distribution of normalized exposure in the data. Nothing wrong with that. But for it to be useful, be useful for predicting exposure at a given level of pounds per AI handled, I'm sorry, amount of AI handled, make s a further assumption, an obvious one, and we've talked about it earlier today, that apart from "linear effects", exposure is proportional to amount of AI handled, excuse me.

Of you write it like I've got down here at the bottom that exposure is approximately equal to amount of AI handled times some constant proportionality, times all your random effects. If you do that, if you write it that way and look at it that way, then this is what we've seen in Doctor Crowley's presentation or Matt's presentation, you end up getting a line, a linear equation that's in terms of log, with log exposure is related to the log of pounds per AI handled with a slope of 1 plus some additive there on this log scale.

So if you believe that that previous relationship is true then of course you've got, you have a relationship like this and do you have a pointer by any chance thanks. How do you work this thing?

SPEAKER: Just push the button.

DR. HOLDEN: Oh, a button, my gosh, technology. I'm used to using my computer mouse.

I wanted to use this slide because it's a great opportunity to comment on some of, some of the conversations I've had about this earlier and also the talk that we just heard from Matt Crowley. Doctor Popendorf and I were just talking earlier today and he pointed out correctly that there's a lot of things assumed in this, even if you don't assume one, but a lot of things that are assumed in this equation that are tricky.

One of the most obvious things or one of the most important things is that you notice that we've got this random hump back there which includes a whole mess of evils and there are many factors in there in which, probably also have exposure as well as pounds per AI handled. We don't know what they are necessarily. And in addition there are many factors in here that probably are in some sense associated with pounds per AI handled in there. So at best what this represents is what I call a marginal relationship. In other words it's the, it's the best all you can hope for is that this relationship is linear or linear to the slope of 1 averaged over everything else that can affect exposure.

So that's the best you could probably hope for.

And that means that we saw a lot of these studies, a lot of the results previously and when you have one study over here at the top say the high pounds per AI handled and another study down here, that there are many, many, many different things that vary between that besides pounds per AI handled.

Some of those things, we know what they are, many of them we have no idea what they are, at least have no idea what affect they may have. And so you've got an extra variability due to those things that the simple regression equation doesn't recognize. And so it's going to declare things as being significant or non-significant if you don't account for those factors. Now, Matt did account for those factors in his high model analysis, at least accounted for them to the extent that you could given the data that he had. And that's a fair statement.

In our studies we are attempting to spread out, and I think as Jan said over here, we are spreading out our pounds per AI handled as much as possible within a study so that when we do have separate they, there's a degree of overlap. And in addition when that's impossible or we have the margin of data that have, already existing that perhaps don't

have such a wide range we try to get as many studies as possible and as many pseudo data points as you possibly can.

And the final thing I want to mention here is that if you focus say in one small area of this plot, in other words you restrict the pounds per AI handled, very small, you can see or maybe it's very difficult to see but if you squint your eyes or put your fingers in front of the plot, that you no longer see that linear trend.

In other words if you focus on too small a range you can't tell whether it's linear or not. As a matter of fact you can even have a reversal. You can look at this points here and these points over here, it might look like you've got a result going this way if you focus on too small a range. And that means that the range of the x variable here is very small relative to the range of the variable, the r variable. So it looks like a shotgun on that level. So it's, you're taking a lot larger sample sizes to be able to revolve that.

So again it's important because that's going to come, that's going to be reflected in what we're talking about tomorrow versus sample size and design with regard to how wide this pounds per AI handled

range might need to be before certain things are possible.

Now that I've said all of the things that could possibly be wrong with data like this, or at least all the complications, if it is true and if this relationship turns out to be true, and this answers Dallas' question I think, or gets to it, then if you look at the exposure per pounds AI, the normalized exposure on the x axis you should have no relationship or no discernable one on the amount of AI handled.

And if that's the case you could sample, take a sample of data and use it to predict a distribution of normalized exposure which would be identical, mathematically equivalent to, as the predicted exposure distribution assuming the amount of AI handled is equal to 1. Then you could use that, superimpose it upon the, this proportional line where the slope equals 1 on the log scale and in a sense get an idea, it says accurately predict, but let me say you'd get a, only to the extent that any sample could accurately predict, you'd get an idea of what the distribution might be.

So the bottom line is if the normalization, the proportionality is true, then it's very easy to see what you can do with the data, you can make some simplifying assumptions. If it's not true then these,

this use is no longer necessarily valid although it might still be approximately true or it might be with assumptions. It could be used conservatively, but nevertheless that's the rationale for this.

Now, to start closing down here, as a result of its importance, this parameter, I said before the monitoring program that we're going to be using will focus, for the purposes of sample size, it's the benchmark adequacy goals on normalized exposure. In other words we're going to use this interest in normalized exposure to a certain extent to help us plan the sample sizes.

I want to emphasize that the study does not assume, we do not assume that normality is true, I'm sorry, normality is true, we do assume that proportional relationship is true, we do not assume that any relationship is true, or any other.

What we are saying however is that we can only design the study to optimize for certain things. And so because there is so much of an interest in using this we are going to use that in helping us design the study.

In one sense we're going to use, and we'll address this detail tomorrow, but right now let me just say that we are going to use the, we are going to ask

that the study be a sufficient number of replicates and clusters and whatever else so that the distribution of normalized exposure, which means that we get information about that distribution in advance as much as we can to help plan sample sizes that that'll be, have some degree of accuracy, some degree of precision accuracy with respect, and we'll discuss in detail what that means tomorrow, but we're going to use that as a benchmark criteria, not to say that that's the only thing we'll be doing with the data, but we will in some sense have assurance or guarantee that the user could use it for that purpose.

And in addition we are designing the study and you've already seen how we're doing it by spreading out the pounds per AI handled in such a way to hopefully ensure that some limited examination of the relationship between AI and exposure can be addressed. When I say limited I doubt seriously whether a user is going to look for a complicated function of pounds per AI handled.

Maybe as Doctor Heeringa said, a quadratic or something may be possible to be able to see something like that or as Matt Crowley talked about, using that data to look at different slopes which effectively you do a simpler spectrum between complete independence.

In complete proportionality there's something in between which has this curved linear aspect to it.

And so certainly the study will enable, the power is going to be, is really going to be limited, the power specification is going to be limited but it will have some criteria for that. And that's pretty much the extent to which we're assuming or we're using the unit exposure concept in our study. So it is designed to get a large number of variables as you saw so users can do any normalization they want and as someone pointed out, because variables like time, say for example, tend to be correlated with pounds per AI handled, the adequacy for normalizing based on pounds per AI handled can probably be done on some other factor like time as well because it's going to be varied in that database as well. So you'll have that but in all cases it's going to be a marginal relationship that can be tested.

And finally let me close with, well, almost, next to last close anyway, that you might, because we've moved our, we've allowed a looking at, or are going to move the study a little bit to try to get an examination of pounds per, amount of AI handled, we are looking at one factor. So if you think of it as being this little, if you think of this spectrum that we

mentioned from complete descriptive where all you're looking at is the variation down to complete experimental where you're looking at a large number of factors, we've moved our study down a bit, not halfway between the two but a little bit towards the experiment in the fact that we're looking at this factor. Although some people in the description of experimental may argue correctly so, that we're, we, you can still have a descriptive study and look for functions but because we're manipulating the factor it's getting closer towards experimental. But I just want to emphasize that that's about the limit of what we're going to be able to do I'm afraid.

In summary then let me say that the target population as I've described for a scenario is a little bit different than the totality of what's in the scenario for many different reasons, and I think it's going to take subject matter expertise to be able to determine whether that can be generalized, if it even needs to be generalized to the larger set of scenarios, but I've heard many comments the last couple of days that expressed opinions on that and all those have been very helpful.

And in addition to the target population I want to mention that it's not just the workers we're

talking about, we're talking about the target population of traditions as well. Our monitoring program as we said is more descriptive than experimental and I think the emphasis, we have an emphasis on exposure normalized by amount of the AI handled but not a dependency. Thank you very much.

DR. HEERINGA: Thank you very much, Doctor Holden. And I think what I'd like to do at this point, we have additional presentations and I'd like to return to questions on your presentation, let you give your voice a rest.

DR. HOLDEN: Thank you.

DR. HEERINGA: Let's take a 20 minute break and come back here at 25 minutes of 11:00 and we'll resume with questions from the panel for Doctor Holden and then continue on with additional presentations, I guess Doctor Landenberger and Curt Lunchick again.

So we'll see everybody back here at 25 to 11:00.

(WHEREUPON, there was a recess).

DR. HEERINGA: We still have a few members of the panel who have yet to reappear so I'll wait another minute before we begin.

Okay, welcome back everyone. I think Doctor

Portier and Doctor Chambers are still to arrive yet. We'll wait just a minute.

While we're doing that I also wanted to acknowledge at the table here, Doctor Tina Levine who is the Director of the Health Effects Division, I normally introduce each day. Welcome.

Okay, let's get underway now that we have a quorum of the permanent panel members here.

Just before our break we heard a presentation by Doctor Larry Holden, describing some statistical aspects of the AG. Handlers Exposure Task Force proposed database or AHEAD, I guess not proposed, the AHED database. And we wrapped up without having any sort of questions of clarification from the panel. There will be several other statistical presentations to follow from the Task Force, but I wanted to make sure that panel members were clear with everything that Doctor Holden presented, or if you have any questions. Doctor Lu, Alex.

DR. LU: I think that Doctor Holden gave an excellent presentation. I think the concept that he provided I would say will be a wonderful doctoral dissertation project. Seriously I think it's, I would take his message with me home.

The question about a selective target

population, if you can go to slide 16, oh, this is different. Well I can ask this question first. When you say manipulating worker activities, could you provide some real example of how we can manipulate worker activities that's kind of deviated from their usual practice?

MR. LUNCHICK: I think an example or examples of this is to dictate specific types of equipment that they would not normally use or to have them, for instance they would normally be using Chemprobes and we're going to want them to use a dry break system which would have a significant impact on the exposure. Our emphasis is to catch what is typically used in the variable that we are going to capture the widest variability on, of course it's the amount of active ingredient handled.

So that's what it is, we're dictating their work habits, we're trying to avoid that to the greatest extent possible.

DR. LU: So in other words this can be called misapplication. It's not applying pesticide according to the label so it would be a misapplication.

MR. LUNCHICK: No, we would never under any circumstance as them to do any application that's not consistent with the label. But the label gives

great leeway. For instance the label may not prohibit an open pour loading and a manipulation would be to dictate in that case to use a closed loading system which would cut out probably the predominant use of a product that does not prohibit open pouring.

DR. LU: All right, thanks. So the target populations, the equation that you presented, there's a random variable term that you feel like you will probably dictate the outcome of the data, right? And I thought, I directly thought maybe that's why you want to select a population to minimize this random variable term as much as possible so you would totally diminish the effect toward the end of the data analysis.

When I look at your criteria of selecting so called target population, I mean what is the, how much do they represent the total pesticide applicator in a workforce? I mean I would say for example one of the criteria is a non-pregnant woman. I don't know who would actually be pregnant and then still go out and spray pesticides. So are you talking about 90% of the total workforce or the 9% of the total workforce? Because I mean if it's really only 9% then that's a lot of data deletion here.

MR. LUNCHICK: This is Curt Lunchick again. This is background for everybody, as we go out

to collect data we very early on will begin to talk to grower groups and others out in the field to make we have a good understanding of the typical use patterns.

We have a subcommittee, AHAP, I don't remember offhand what it stands for, I get acronym overload too, but it's like tree growers, we'll deal with, like if we're going out into the Pacific Northwest to look at the apple growers, we'll get with those people well beforehand to make sure we have a good understanding of the agricultural practices.

In regards to restricting parts of the population, the pregnant women is a necessity. The number of people that are actually limited, probably extremely small. There's no way we're ever going to get an accurate percentage on that.

The English and Spanish, obviously that's the majority of the user population but I think you get into an area like Belle Glade, Florida where you have a lot of Haitians, French of Creole speaking, right now we would not hit that area. But if we found that it was important, there was something going on we needed to in that area, we would amend that type of restriction to make sure we had a Creole speaker. But in general right now that's our plan.

So we're trying to minimize the restrictions

and most of them are based on practicality more than anything else.

DR. HEERINGA: Doctor Chambers.

DR. CHAMBERS: I feel compelled to put my HSRB hat back on and I think it may just be, if you can go back to that same slide, Larry, 16, it may just be a matter of semantics but the term, manipulate, raises all sorts of potential flags I think. And in using that, you know, I understand what you're saying there but again manipulate could mean that you're asking people to be exposed to more than their normal workday would be and I don't know whether that's what you mean or not. As long as you're asking them to be exposed to less than their normal workday I don't think that's going to be a problem but if it's more it surely will be.

DR HOLDEN: Of course in that slide that followed a statement that that's what a experimental approach might do, not what we're doing. Manipulating, when you're, if you're doing an experiment, a designed experiment, not a designed experiment but I also have to go, not us, let's say someone else, they would have to go through HSRB probably if it's done for regulatory issues anyway. And the design, and they also obviously then would not probably do exactly what you say, they

probably would manipulate so they increase exposure, they wouldn't be allowed to.

But if it were an animal study, say, they might choose to do that. An experiment would choose to do things that would not normally be done in practice for the purpose of delineating, maybe testing one treatment versus another treatment or one condition versus another condition.

So in this example I was saying we're not doing that. I mean we're not, we're not doing experimental approach. Yes, we are doing some manipulations like pounds per AI handled obviously.

Is that what you meant or am I misinterpreting it?

DR. CHAMBERS: Well it's just, you know, the whole idea of there being some scripting of these things, again there's going to be some flags raised if this is going to likely expose people to more than their normal work activity.

MR. LUNCHICK: Yeah, I, this is Curt Lunchick again. After two months of preparing for this and having my statistician here restrict me from using words like proportional representative, replicate and other words that I'm used to using because they have statistical meaning that I'm unaware of, this is a case

of you using a term that has meaning, that has implications beyond the experimental. And I think the point is well taken. Because of the requirements of the rule the term, manipulation, has connotations that we have to be careful of too.

And obviously in any experiment, or not even an experiment, any study that we do, they're not experiments, but going out into the field and collecting data, we go through a detailed informed consent process for anybody who participates in our studies.

So, but I think your point is well taken and when we go to the HSRB our statisticians will have to be careful on the terminology also.

DR. HEERINGA: Yes, Cynthia Hines please.

DR. HINES: I just have one comment. One of the objectives that has been described on previous days for the Task Force is to look at some of the more modern handling technologies, both probably at the mixing side and the application side. And it is probably reasonable as was used earlier that there is an association in most cases or in many cases between pounds of AI handled and exposure.

I'm wondering once you start looking at some of these more highly engineered mixing technologies if

that will hold true and if there is going to be an ability within your data sets to get a handle on that because after all, the whole idea behind that is to minimize contact and perhaps you will have a different picture?

MR. LUNCHICK: That's a very good question and actually one we were hoping for, I hope I don't steal Doctor Hamey's thunder with my answer. But actually I wanted to put this in something of a regulatory perspective and I think your question leads to that.

We are using this proportionality between the amount of active ingredient handled and the exposure as an assumption to design our study. When sufficient data come in and probably from multiple studies for a give scenario, we will then analyze the data. And one of the things obviously will be to see if the assumption holds or makes sense. I fully expect that when we get into sophisticated engineering controls, enclosed cab vehicles with air purifying system, et cetera, dry break loading systems where essentially once you connect the hoses you turn the pump on and whether 100 gallons or 10,000 gallons flow through, probably it's going to be a minimal effect.

And yeah, so I think in those cases and in

every case, once we get the data in and start the analysis we will be doing that jointly with EPA, PMRA and DPR in reaching a consensus and a joint recommendation based on the data to users of the database.

DR. HEERINGA: Doctor Johnson.

DR. JOHNSON: Yeah, if you put slide 26 up there. This is a nice slide and you have that distribution whether the slope of that line is equal to 1 or not so I think it's not an assumption that needs to be made and it sounds like from Doctor Lunchick that you're not really going to make that assumption. And I don't think it's an assumption that's necessary for the various kinds of arguments that have been given so far.

DR. HOLDEN: This is Larry Holden again. There is one caveat about that. You are right by the way, we don't assume that the proportionality is true. I know I keep saying this, it must be a force of habit from the old days, we don't assume that the proportionality is true.

We do assume it as a benchmark for the sample size. But in this example that I gave here, this illustration you can flip to the next, the previous slide, yeah, this previous slide, if it's not proportional then what will happen is that instead of

being a, with no relationship there it will actually be higher at the left end and slope downward. And so if one were then taking samples and plot that distribution you would have something that would be bigger than what it shows there because it would be incorporating the variability plus the negative trend because of what you're doing in that case. And so therefore your distribution that you estimate would be bigger there and then if you use that distribution and force it upon the and unfortunately I don't have a slide for this, I should have a slide for that then if you force that distribution then what you'll have is, it'll look, you know, fit exactly on there but now the data point's a reality, it would be flat, they wouldn't follow that line. Assuming it's, or maybe it, I'm assuming it's independent as an example.

But the distribution would be very much bigger so what you'd have then is something that really over predicts the distribution at the high end and under predicts it at the low end. I think David Miller or Matt Crowley mentioned that during their talk I think as well.

So it does make a difference. We're not assuming this of course but in this example, what I said on the usefulness, the usefulness of normalized

exposure, that usefulness has to be expressed with some caveats if that doesn't hold true. Because what happens is you're not estimating, not only are you not estimating the distribution correctly, no, you are, you're estimating the distribution correctly but trying to extrapolate it to future, to regular exposure at a given pounds per AI doesn't hold true. I hope that's clear. In other words it, the distribution will go, let's go back here, instead of looking like this, whoops, suppose it's just flat instead, then if you normalize it you'll get something that looks not like this but higher up here and slopes down like that.

DR. JOHNSON: But the point would be that you would look at the variability around whatever line you fit?

DR. HOLDEN: Yeah and that, right, and you're actually right.

DR. JOHNSON: That variability shouldn't change.

DR. HOLDEN: Yes, I think your point is well taken, I think what you're really saying is, something that you mentioned earlier on, that if you use whatever the data fit, show, and use that distribution then you're home free except for, you know, sampling error, et cetera, et cetera. But yeah,

that's right is that if you, the danger is if you assume something that isn't true.

DR. JOHNSON: I think we're basically in agreement.

DR. HOLDEN: Yeah, we're used to doing this by the way, that's why.

DR. HEERINGA: Additional comments at this point? I think we'll probably get back to the discussion of that issue seeing the look on Dallas' face, we'll probably come back there.

Okay, at this point I want to thank you very much, Doctor Holden and I think there is a, the next presentation from the AG. Handlers Exposure Task Force on the topic of within worker versus, between worker, is that what you're doing?

DR. LANDENBERGER: Yes, that's correct.

DR. HEERINGA: Doctor Bryce Landenberger.

DR. LANDENBERGER: As mentioned my name is Bryce Landenberger and I am the technical leader for the Chronic Risk Assessment and Statistics Group at the Dow Chemical Company. I'm here today representing the AG. Handlers Exposure Task Force but I also wear another hat and I apologize if I get my Task Force mixed up. I'm also part of the Antimicrobials Exposure Assessment Task Force and have been since its

inception. In that capacity I make myself available to the panel if they have any questions concerning design or sample size issues related to the Antimicrobials Task Force and some of the issues surrounding it that differentiate from our AG. Handler colleagues.

With that, in my presentation today I'd like to go over some basic things related to the intra verus inter-worker variability and we're going to talk about this with regards to some basic definitions and concepts and we'll be looking at this in particular by looking at the within worker correlation number. We will look not only at what that general definition is and what its general impact is, but look at what its impact is on long term exposure estimation, how we would get at an estimate of a particular parameter and then some practical considerations for gathering that data to estimate that intra, within worker correlation. And I apologize if I mess up on my terminology, I keep on stumbling over these terms myself all the time.

Let me start out with this first slide which gives us some basic definitions. Up at the top we have a distribution in gray on these slides here. And this is basically the between worker distribution of single exposures and this would be if you were just taking particular individuals as represented in this box down

below the distribution here, and getting their values of exposure over the different particular study and then developing from that the distribution that we see above. If we were to do repeated measures over days or months or years on a particular worker, either those in the green circles here or the red triangles, what you have there is then the within worker variability.

Now, if we were doing just this top part and we had repeated measures on individuals, that whole variability that's considered there is what we are calling the variation between workers, sometimes called the total variation. It has two components to it. One is the between worker component and the other component is the within worker component. But it is important to keep in mind that this is basically a total variation estimate up here because we need that for our definition of the between worker correlation. I already got my terms mixed up.

So what, another thing I want to point out here is these dotted lines would represent the means of a particular individual long term. And this will be important as we start looking at how do you go from the data sets we collect in a single exposure and incorporating that with the within worker correlation to get an estimate of a long term exposure

distribution. Next slide please.

So the concept we want to have here, and I apologize, this particular equation, RW equals 1-W bar over B, I noticed did not print out in the printouts, I noticed that last night and there wasn't much we could do about that. But this is a key equation that we have here, a key concept with the within worker correlation. And what we have if we look at this particular distribution, the same place we started, again you can see this over and over again in the slides, you have the various workers here.

What we're going to do is if we had variation within a particular couple of workers here we might have that variation measured by W1, W2 and in this particular case the example shows these are fairly small within worker variations. I think if you remember what Matt Crowley presented this morning, in some cases that appeared to be the case when they did repeated measures on some individuals, in some cases it was fairly widespread.

This particular example, if you take that average of these variations and divide it by B, which is labeled here, the variation between workers, but remember it's also total variation, this will give you a number between 0 and 1 for the within worker

correlation. And again, since this is small this ratio will be a small number, it will be close to 1. And that's basically what we're doing with this equation here. And this is the concept again, the within worker correlation that we want to deal with. Next slide.

On the other hand if we had a particular individual and their variation was quite wide, W1 is going to be wide and if you had similar results for all of these individuals up here, W bar is going to be fairly large, that average within worker variation. So this ratio is going to be fairly large and RWW is going to approach zero. Again, keep in mind that the variation up here between workers is total variation. Next slide please.

So what is the impact of this particular approach to this and looking at the within worker correlation? Again if we have this distribution up here and you have these individuals here, each one of these workers, if we had repeated measures we're going to be looking at their averages, essentially their averages as we're looking at a long term exposure. This might be something over a lifetime as we have examples down here.

So we would be looking at that particular exposure, the cumulative exposure for that individual

would be the worker mean times the number of exposures. For example if we had the units in milligrams per kilogram per day times the days of exposure over this entire range here, that would give us their cumulative exposure. So as we look at this, RWW is a way in which we can progress from the short between worker distribution of the single exposures and get down to a long term exposure. But we need to keep in mind that this particular distribution at the top is important also for the short term exposure risks that we want to evaluate for particular workers. So if we can move to the next slide please.

So if you think about it, RWW representing various values in terms of within worker correlation relative to the total variation, their average within variation to the total variation. We can sort of understand how we can progress from this single exposure down to a long term exposure. And what we're looking at here is what would be the progression and how might it go depending on where RW is.

If you have a large within worker correlation, remember that means within worker variation is very small, then we're going to have a distribution that's going to be fairly spread out because these individuals are going to have their means

spread out all along this line. However as you start increasing that RWW, in other words the variability, excuse me, as you start decreasing the RWW, that within worker variability is going to be increasing and becoming more similar to the single day distribution for all the individuals. And their averages, because they're spread out across this whole line are going to start to move in towards this arithmetic mean which we've highlighted going all the way down through here. So the distribution in point of fact will actually begin to get tighter and tighter.

If RWW is zero then essentially your long term distribution is going to end up being equivalent to the arithmetic mean. Obviously this is probably not the case, nor is it a case where it's essentially up here, it's somewhere in between. But all these distributions as we see here have the same particular mean exposure. Next slide please.

So what this means, by knowing this within worker correlation we can take and predict from our single day exposures or our single exposures for the between worker distributions, using this within worker correlation and come down to an estimation of the long term mean exposure for the between workers. And in many cases this is really what we're interest in

getting at. Next slide.

So the question is, how do we go about estimating this within worker correlation? There have been some studies and some literature done and which my colleague, Doctor Holden, has gone back and looked at these particular studies and used the data from these in the analysis that was done on these studies to try and develop estimations of this within worker correlation. And there's one case here with Doctor Fromanegerol, there were only 2 air blast applicators and 2 open pour mixer/loaders, there were repeated measures over 6 weeks.

The issues that we have with a study like this is that it tends to underestimate the total scenario variability and it also tends to underestimate as a result, the within worker correlation. We have a limited number of workers and a limited number of applications.

These other studies her, there has been a more or less meta-analysis to try and come up with the within worker correlation as well, I'm sure Larry will correct me if I get this wrong. These are measured actually quite close together as well and they have a tendency to be done under similar conditions which tends to underestimate the long term within worker

variation and overestimate the correlation. Again we also have one other example here where we have some purchased data that the AHETF went through. It was repeated measures, exposures but they were 1 to 6 days apart, very limited in terms of the variability that you might expect to see within an individual worker. And this is going to be an issue with trying to do any kind of an estimation of this is that you're going to have these sorts of problems cropping up in these different studies. What basically these, examination of these studies indicated in the literature was that this within worker correlation is probably between the range of .2 to .5. So if we can go to the next slide please.

So what if we wanted to go ahead and actually get a better estimate of this within worker correlation in the program? Again, we're going to start out with our single day exposure, and again you're going to have these different workers that will have this type of variability on. We need to collect this data. Regardless of anything else this has to be collected because we need to address the short term as well as the long term exposure.

With that emphasis in mind that's why the AEHTF is focused on this particular box in its design

at this point in time. If we start looking at trying to get multiple monitoring units with the same worker, and as Larry pointed out earlier, when we talk about the worker they're not necessarily having all the same conditions relative to their application of a pesticide to the field. They are not the only component that is being varied from one monitoring unit to another.

It could be a change in equipment, it could be a change in the acreage, it could be a number of things that change. So it's not strictly speaking just a worker effect that you would measure. What this would tend to do is obviously you're going to have to at least double the number of required monitoring units to get any kind of an estimation on this within worker correlation.

If you start going to more than that then you start running into obviously more monitoring units. All of these things tend to create some logistical issues, participation, personnel issues, equipment, analytical issues, how far apart is far enough apart, if we do everybody within a couple days is that giving us enough estimate of the within worker exposure, do we need to do it over months, do we need to do it over years? If we have to do it over years we start running into some major logistical issues just trying to

execute the study. Next slide please.

To try and really get a feel for where this is going Doctor Holden did some simulation runs to see how this might be impacted. Using a geometric standard deviation of 4 and 25 workers, there were some simulations done in which a between worker component of the variation was estimated and then a subset of that added to it was within worker variability.

So you start out with 25 workers and if you have to do 2 reps per worker obviously you're up to 50 monitoring units to begin with now and what kind of an impact would it have if our true within worker correlation is .3? And the simulation was set up with this correlation being the actual within worker correlation in the simulation.

So you can see here it ranges from roughly zero to about .61, call it .6. If we go to three replicates per individual we're now having 75 monitoring units and we've gotten a slight improvement on the top end, the bottom end is still essentially going down to zero, we're slightly above 5, maybe 5.3. If we have 5 replications per worker we're now up to 125 monitoring units and so forth.

If we go down to the last line here we're up to 250 monitoring units for the single study with 10

replicates per worker. And what we have succeeded in doing is moving the bar from zero to about .15 call it to about .43. And relative to what we've seen in the literature that was .2 to .5 based on the literature data that we have at this point in time. So as you can see what we're seeing in this simulation is that we're going to have to have a substantial increase in the number of monitoring units to get a true benefit in trying to narrow down the estimate of the within worker correlation. Next slide please.

So in summary what we have at this point in time is that the existing data suggests that the within worker correlation is .2 to .5, there appears to be a large increase in samples that would be required to get a meaningful tightening on these limits. In other words we're going to have to substantially increase the number of monitoring units to get any kind of a substantial decrease in that range.

We do already have within place an existing process to address long term exposure risk. Using the between worker distribution of single exposures and reasonable assumptions about this within worker correlation we can already move in that direction. And it's important to keep in mind that the between worker distribution, that initial box I kept on pointing to at

the top, will be needed for both short term and long term exposure assessments. Thus it's of the greatest regulatory importance as far as we can tell.

With that in mind our basic conclusion that is attempting to measure the within worker correlation in the monitoring program would have a tendency to waste the resources and lead to unnecessary human testing. For that reason we're advocating moving away from doing multiple repeated, excuse me, multiple monitoring units on the same worker and instead trying to maximize the number of workers that we actually measure in the monitoring units. Thank you.

DR. HEERINGA: Questions from the panel? I have one but I'll hold it and wait until Dallas asks it for me. Ken, Doctor Portier.

DR. PORTIER: Could you go back to slide 7. So if I understand this the top distribution is your distribution of between worker, that's your between worker distribution, right? So an individual worker drawn at random would have a value from that distribution which would represent their long term exposure if we repeated them over time a lot, right?

DR. LANDENBERGER: That's correct.

DR. PORTIER: Okay, good. Now, if I assume say a large actually, shouldn't it go the

other way around? If you assume large within worker correlation or you assume the second distribution is that worker's repeated measures distribution, okay, what would happen is that distribution really is shifted to the mean of the value you chose at the top, right? So if I had a worker who was at or near the protective level in their average exposure, when I overlay on that the repeated measures distribution and I start worrying about how often is this worker going to be above some threshold exposure level, it's very important whether I have large worker correlation or small worker correlation.

DR. LANDENBERGER: Let me just clarify, I'm not sure I heard you correctly. You were saying large within worker correlation or large within worker variation, because that was --

DR. PORTIER: Well I really mean, I'm looking at large within worker variation, I'm looking at that second distribution so that's an individual with a lot of variation, right

DR. LANDENBERGER: Yes.

DR. PORTIER: versus the bottom of the

DR. HOLDEN: No, no, that's wrong, it's just the opposite, the largest variation of within worker is actually the bottom.

DR. LANDENBERGER: And what you would have here is if you look at this distribution representing everything your various measurements on that individual are going to be here.

So what's going to happen, there's actually a slide we took out which we probably should have left it, that their averages are going to tend to start to come in towards this arithmetic mean because you'll have, they'll have some high values, you're correct, they'll also have some low values. Now if they're out here in terms of the distribution towards the upper end of the tail and they had tight within worker variation, you're correct that they would probably not move much from that point but it is, you know, questionable about where they are relative to this.

But the other issue is too that that is, they're probably not going to move too far so if they're initially out here, yeah, that's an issue. If they're here it may not be an issue depending on how tight that within worker correlation is. If it's widespread then we would expect them to have values all along here and that mean is going to tend to move in towards this arithmetic mean in the middle.

DR. PORTIER: The point I was trying to make from a regulatory setting, if the workers are very

consistent in their exposure then you have to worry about which workers are exceeding that threshold. If the workers are very variable in their exposure then every worker has a chance of exceeding that threshold and it's a different kind of risk scenario. That was the point I was trying to make. And it ties in with how we look at this within worker variability, within worker correlation issue.

DR. LANDENBERGER: I'm going to say I both agree and disagree. In terms of it being critical and important for understanding a risk assessment mode I would agree. It doesn't change where the distribution goes.

What would be impacted by that is where you set this critical red value here, I'm putting on my risk assessment hat, in terms of protecting those workers. Because that's the issue of protection for the worker. You're never going to guarantee that you protect every worker and if we had a distribution that was really tight at this top end, yes, that would be an issue that needs to be addressed.

But in terms of the within worker variation, I think based on what we have already within the literature, we can sort of develop what that long term exposure would be. Long term I expect people to move

to the arithmetic average because they're going to have some days when they have high exposure and some that they have low. But the combination of those are going to bring them in towards the center. If you're looking at short term exposures you're going to want to use the top distribution between worker anyway and that'll get you to the point where you're wanting to look at what the upper end of the protective level is.

DR. HOLDEN: Doctor Portier, I don't disagree with that. I was just going to say is that, and I may be misunderstanding as well but what we're really saying is that, what you were talking is very important, is absolutely correct I think and that variation is important, but I think what I pointed out in this discussion here is that we can infer some things with reasonable assumptions about the within worker unit measure of distribution from data that would almost be counterproductive in trying to collect for this generic database and I think that's really the point.

Not that the information is within worker variation is unimportant. Actually we think it is important but just that we don't think that we can, with the data that, with the resources that we have, we could do it justice, do it even better than just

reasonable assumptions.

DR. PORTIER: And I follow that, I mean I follow that logic, I'm just trying to, you know, it's one thing to say, we don't think it's important and another thing to think through that issue and say, you know, well, under what conditions might it really be important?

And that's what I'm trying to take the opposite side and argue. From the industry point of view I follow your argument perfectly. But now from the EPA's point of view is that really where they want to go? If they have any say in this is that the best thing for them, thinking about all the scenarios that they have to face which is not just tier 1 and tier 2, they also have other environments where this data may be very useful.

MR. LUNCHICK: Let me address that on behalf of the industry. I don't want to speak for the Agency and I think if they want to add something it would be beneficial for them to come up.

But I'm going to present this from a non-statistical standpoint because I got lost about 15 minutes ago, but what I, when I assess exposure, and let's take this distribution, yeah, I was just thinking of that for you, we're assuming, and I think our

experience shows and what Bryce showed is we have a decent amount of intra-worker variability so at any give time an individual is going to be scattered along this. But if I am looking at one of my products and there is a risk concern out here, regardless of whether I can do any statistical tests, Larry keeps telling me I'll never be able to do a statistical test that's significant, but I would use my experience, we are collecting a huge amount of information when we do our studies.

I would start to look at what conditions are going on here. And let's say it's tank loads or something like that or it's some type of worker property or it could even, you know, whatever, I would begin to focus on that.

I can guarantee you because I'll be working with EPA or DPR or PMRA, we will begin to focus on what conditions may be occurring here that are of concern and not on a task force basis, but on a registrant to regulator basis, we would begin to see if we need more information to delineate something important here that would allow us to hopefully mitigate exposure because the alternative is something a registrant generally doesn't want which is the Agency is going to prohibit any use that would occur beyond this level of concern.

DR. HEERINGA: Cynthia Hines.

DR. HINES: Just a comment on the last slide and it's just a comment on language. I think obviously collecting repeated measurements, and I know from personal experience in an agricultural setting is very sobering and very challenging because of the short time windows you're working with and the applicators may only do a few applications, if that, within a year. So I acknowledge just how difficult that really is.

I would comment though that I don't think a reason to not do it is unnecessary human testing, I think that's, it may be unfortunate language, because if you are looking at the natural variability within workers you're actually trying to capture that variability from day to day or week to week or month to month under their normal conditions or what they're usually doing that you aren't actually manipulating or scripting what they're doing. And so is that just unfortunate language?

DR. LANDENBERGER: Yeah, I think that's a fair comment.

DR. HEERINGA: Good for you, Dallas.

DR. JOHNSON: Yeah, just to play the devil's advocate here a little bit, have you thought about this interclass correlation with respect to

studies and workers within study?

DR. HOLDEN: Yeah, actually I have but, and not only that but you can imagine, the other day you were talking about clustering, we'll talk about that later being temporal as well as spatial if you think about an individual. And we have looked at some of these various components just to see what impacts it has and if you look at it over long periods of time, you know, there's temporal correlation.

With an individual you'd expect that. In other words his or her value you get tomorrow is more likely because of the value today than it would be say three years from now because there are many ways. So you could model that normally in a simplistic way by thinking there's a spatial cluster and also a temporal cluster.

But now when you're talking, but if you forget that and just talk about spatial and within worker correlation, a worker is almost always going to be within the same cluster over repeated times. Until you get long period of times you're going to be in the same, what we're calling a sampling cluster. So that means that the within worker correlation can't be any smaller than the within cluster correlation. It has to be at least that.

And maybe bigger. And so there are some, there's a relationship between this what I'm calling clusters which is just a sampling artifact and I'm visualizing it to model it. And so there are some relationships that have that in there and since it looks like our, a few of our scenarios are in our cluster correlation is around .3, that sort of suggests that the within worker correlation has to be at least that.

So it may be more like .3 to .5 or maybe .6 or something like that. So if you pick some sort of number in the middle that's probably close to what it might be given those kinds of, you know, that logic if you want to call it logic, gut feel, seat of the pants, whatever it is you want to call it.

And of course we sort of inferred this but we didn't really say this that an agency, the EPA or some other regulatory agency, if they chose could pick zero or 1 depending on which is more conservative for their needs to protect, I mean that's the extreme, right? The extreme would be to pick a correlation of, assume a correlation of 1 and then they would always use the between worker distribution for everything which would be super protective in a sense for chronic exposure. So some reasonable assumptions could be made to handle

some of these things.

DR. HEERINGA: David Miller from the Health Effects Division.

MR. MILLER: I'll probably just, I may just end up repeating some of the concepts, but I'll add a few others. If we could just go back to the slide with the four graphs, distributions. Yeah, great.

Basically, I mean if the top one for example and again, this has already come up but I'll just say it, that's the between worker distribution for single, single days and the question that's kind of come up at the Agency and we're talking about is, in essence when we, it's the sloppy versus neat aspect, when you repeat, when you, a worker, this is just for a single day one, when you go out and do it a number of different says, the same worker, if they're consistently sloppy at the upper end they'll essentially remain at the upper end and if they're consistently neat, if they're at the lower end they'll remain at the lower end.

And the worry is, and just to give an example in terms of one of the things that's of interest is, potentially why are they at the upper end and why are they consistently at the upper end? A lot of it may be

due to behavioral practices for example. I just, off the top of my head it might be instead of if you're doing a backpack spraying for example instead of walking backward for example as you spray, if you walk forward and that's your habit you would consistently be at the upper end.

So that's kind of the interest in terms of us in terms of why we think the interclass, the worker variation consistency is important.

Currently what we do is when we look at, we have the, we don't look at the intra within worker variation. Essentially we just use the average is what it is. So there's, it's as if everybody is randomly bouncing around from high end, from the top distribution to the lower end of it or so. So there's no consistency, necessarily consistency within workers in their exposures.

So the question that we're I guess addressing here is, how important is that for example and how much needs to be done in order to kind of clarify some of that? And I think that what the Task Force is bringing up is you would have to do an awful lot of replicates, true replicates within workers in order to get a better idea of that and it may be just the, from the literature .2 to .5 may be sufficient.

And as I think at the tail end it may be just you can do the extremes of those for example and end up with something that's good information, that's better than what we have now.

DR. HEERINGA: Thank you very much Mr. Miller. Steve Heeringa here. This is something I know a little bit about and I think that as I look at what the discussion, the factors that you just mentioned, David, with regard to trying to understand which individuals in the population may be at higher risk or essentially their expected value is up at the tail of that distribution, the factors you pointed to weren't really the individual factors, they were fixed effects of their working conditions, backpacks, wind velocity, type of volume, et cetera.

So those really aren't individual, you know, these measures would really be pure random effects and in a sense you would be better off I'll hold my comments later on but the short of it is that if you want to describe the population distribution of an attribute like unit exposure you wouldn't cluster or multiply measure anyone if there's a positive correlation.

You're just losing statistical information and if we go back to slide 11 these arrow bars, I mean

most of us could calculate. The difference between these obviously is they grow narrow and that's because you're increasing sample size. And I suspect it's not quite a square root function going from A to B to C to D because there is some design affect inefficiency due to the interclass correlation and the cluster size. There's a little simple formula for means that operates there.

But, you know, the picture here, we shouldn't be misled, the reason this picture is changing is because we are adding those observations and the picture would change even more sharply if we, instead of repeated measures, just took 250 observations on independent individuals doing independent things. And that'll sharpen it up.

So I think you've got a critical point here and that is that whether you do repeated measures on individuals that will benefit a certain type of intra-individual analysis if you were really interested in these components of variance.

But with 15 observations on a scenario and multiple scenarios and many other factors compounded with the individual repeated measurements, you will not be able to tease that out. I'll say that again this afternoon.

Yes, Doctor Johnson.

DR. JOHNSON: Yeah, while we're on that particular slide I just wanted to say that you could probably change those arrow bars on RWW a lot faster for the total number of monitoring units if you were to increase the number of workers. So there's two ways to get to the 250, one is to keep it at 25 workers and look at 10 reps per worker but another way would be to have 125 workers and two observations per worker. And you would get a lot narrower bars I think on that true interclass correlation.

So you're looking at the between variability over the total and it's the between variability that you're not measuring very well because you only have 25 workers or 24 degrees of freedom associated with that and you've got lots of degrees of freedom associated with the within worker part.

DR. HEERINGA: Doctor Holden.

DR. HOLDEN: Actually that's correct, I mean, yeah, that's true, you can, it's the, it's the combination of both of them. But there is some sort of a tradeoff because we're talking about estimating the correlation, not estimating the mean or something like that.

But nevertheless you're right, the general

picture is the same is that it's the total resources. I don't know what the exact balance is, I imagine that there is some, one's going to be more efficient than the other. Certainly for measuring, if one wanted to do both you'd put more information into more workers than you would into repeated measures.

But then I probably would want to, almost insist in that case that you separate these workers by a bit more longer in time than you might, you know, I think some of these studies where you measure a worker in the morning and then you do him again in the afternoon or the next day, you're probably underestimate or overestimating the correlation, underestimating the variability. That might be more relevant.

DR. HEERINGA: Thank you very much. A clarification on my comments too. I was looking at this as arrow bars on mean. This same affect would be observed, that's a good point. This is essentially arrow bars on the estimated interclass correlation.

DR. LANDENBERGER: Just one clarification on these bars, these are actually empirical ranges from a simulation, 10,000 runs on each one of these.

DR. HEERINGA: Very good. Other questions on this presentation? Okay, am I correct that we have

one more presentation this morning in this group?

DR. LANDENBERGER: We're done.

DR. HEERINGA: We're done, okay. What I'd like to do then is to move on to at least an initial discussion of the question, charge question number 4. But before I do that I indicated to the panel members and to the participants here that I would allow individuals an opportunity at the start of each day's charge question session to go back to add additional comments on prior charge questions that had been covered.

Is there anything for the panel members that has been presented or discussed here this morning or any thoughts that you've had over the past sixteen or so hours that would cause you to modify or amend any of your comments or extend your comments from yesterday on any of the charge questions? And if so if you would, when you speak just say, this is in relation to charge question x and I have this comment. Any members of the panel? Doctor Popendorf.

DR. POPENDORF: Yes, I was just trying to think back on the question but as Mr. Evans points out this morning I commented yesterday with regard to biomonitoring, simultaneous biomonitoring and passive dosimetry, I didn't recommend that but there was a

misunderstanding in terms of the goal of the biomonitoring.

So I would amend what I said in that regard. And I thought a little bit about that and I can see some benefit to doing the biomonitoring as a, if you're doing a good job of passive dosimetry your biomonitoring would essentially be zero. And if you get any detects then you would be able to get some estimate of breakthrough or inefficiency of your passive dosimetry.

On the other hand I think I wouldn't recommend it be a requirement because the practical aspects of being able to find someone with no prior exposure and probably no subsequent exposure without the passive dosimetry would, any biomonitoring would be for that particular day.

That would be a big restriction on trying to collect data. If it works it's a good backup but it, I couldn't see it being a requirement, I wouldn't recommend that. It would just make it very difficult to get representative users.

DR. HEERINGA: Thank you very much Doctor Popendorf. Other members of the panel? Doctor Barr, are you okay, it looks like Jeff is going to leave the building so I won't ask him to read the charge

question so I'll tell you what, Doctor Portier has a good suggestion, it's 25 to 12:00, let's take an early lunch and if we could have everyone back here at 12:45, let's do that.

And I think that makes sense and then we'll start on charge question 4, we'll be 15 minutes ahead of where we need to be in the agenda.

Just a little foresight for those of you who I know are making travel plans for tomorrow and many people in the audience, I do not believe that we will try to accelerate things to finish today, there's just no way to do that.

The agenda is sort of set with an interspersion of presentations and questions. I do not anticipate to go overtime on tomorrow's session by any means but I think we need to maintain the agenda roughly as it is.

And I think the chances of us ending early today or finishing up today with the public part of the meeting or the general meeting of the SAP are small. So we'll stay with the general agenda so would expect to be here again tomorrow morning for those of you.

But I do expect to be finished by the time that we are scheduled to be finished at noon.

(WHEREUPON, the meeting was adjourned for lunch.)

FIFRA SCIENTIFIC ADVISORY PANEL (SAP)
REVIEW OF WORKER EXPOSURE ASSESSMENT METHODS
January 11, 2007
Afternoon Session

DR. HEERINGA: Welcome back everyone to the afternoon session of the third day of our four day meeting of the FIFRA Science Advisory Panel meeting on the topic of a review of Worker Exposure Assessment Methods.

This morning we heard several presentations on statistical aspects of the problem of exposure measurement and also on the design of the Agricultural Handlers Exposure Task Force AHED program, sort of motivations and design considerations for that.

At this pint in the agenda we have, we're going to return to charge questions to the panel. And before we begin I guess I'd like to offer one more opportunity for panel members. Is there anything you would like to revisit on questions number 1 and 3 at this point?

Not seeing anything then I guess, Mr. Miller if, or Matthew if you would want to read the charge question number 4 into the record please.

MR. CROWLEY: This is Matthew Crowley, Health Effects Division, EPA.

The normalization of exposure by amount of active ingredient handled, the unit exposure has since the mid-1980s been a principle of the relationship underlying the use of exposure data in the Agency's pesticide handler exposure assessments. It is based on the assumption that the two variables are proportional. That is, if one doubles the amount of pesticide they handled or applied the resultant exposure will be doubled as well.

The Agency is unsure whether the results of our exploratory work showing that proportionality between exposure and amount of active ingredient handled is reasonable in some, but not all cases as a function of the limitations of the data within PHED or whether this relationship is in fact not a reasonable assumption for all scenarios.

It may be the case that an additional ancillary variable, for example, boom length, number of tank mixes or number of decouplings in a closed loading system in addition to or in place of the amount of active handled may improve the predictive capabilities of our exposure model.

Though it is recognized that neither the studies in our current database nor the proposed studies by the Agricultural Handlers Exposure Task

Force were designed for the primary purpose of examining proportionality between exposure and amount of active ingredient handled, or to determine the extent to which other parameters influence exposure compared with our current database.

The Agency believes that the proposed AHETF studies will generate data that will reinforce the assumption of proportionality between exposure and amount of active ingredient handled, or alternatively inform the applicability of another variable as a more appropriate predictor of exposure.

Based on the themes presented on this topic, including its historical precedent, it's application in risk assessment and subsequent risk management decisions, the Agency's exploratory work using the six PHED scenarios in the case study and the study design and objectives of the AHETF, please comment on the assumption of proportionality between exposure and amount of active ingredient handled as a default.

Also, please provide comments on whether the proposed AHETF study design is adequate to evaluate proportionality between exposure and amount of active ingredient handled. What other parameters should AHETF study designs measure in order to improve the prediction capabilities of our exposure model?

DR. HEERINGA: Thank you very much. Doctor Lu is the lead discussant of this particular question. Alex.

DR. LU: Good afternoon. I think this question can be answered in a different direction. I think the panel with their different background expertise, I'm sure they're going to attribute it from their perspective as well.

I'm going to look at this from just form the exposure assessment perspective and pharmacokinetics. I think in theory absorbed dose increases proportionally with the exposures. In these discussions the matter is whether the Agency is using the right or statistically significant surrogate for dose and exposure respectively, or the Agency is simply just trying to extend its exposure and dose paradigm to add another component which is the amount of active ingredient handled before the exposure term.

So if I can show the data, I just, it just kind of struck me this morning that I do have some data that can kind of validate this statement here. If I can have the pointer.

Anyway, look at this graph, this is animal data, it's a controlled dosing animal data, this is for Atrazine and this is published data so I feel

comfortable presenting it here.

Look at this, I'm sorry, I should point to there. Look at this black square. It represents the how come the figure, the legend is not showing, anyway, this is the plasma concentration of the Atrazine that are dosed to the rats. And those are the saliva concentrations. Okay, so look at, this is one milligram per kilogram of Atrazine and this is 10. Okay, as the dose increases the exposure increases or vice versa. So that's very clear. Again this is an animal controlled study it may not be applicable to the field of human data. But again in theory this proportionality exists. Next slide please.

Okay, do you want to scroll it down a little bit. Okay, again this is animal data, this is for Diazinon, they're similar studies. Again this is the plasma concentration, you can ignore the saliva, it's for different purposes. The concentration was obtained from the 10 microgram, no, I'm sorry, 10 milligram per kilogram of dosage. Scroll down a little bit. So the highest concentration is somewhere around 800 ppb. Going down, going down. So the next graph shows essentially the similar data except the high, the peak concentration is somewhere around 100 which reflects the lower dose that were given to the rats which is 1

milligram per kilogram. So this data, and those are all published data, okay? The reason I want to show those data is to just kind of prove in theory this proportionality exists.

The question is that, what is your definition of exposure? What is your definition of dose? Or simply you just want to create another component in front of exposure. So here is the amount of active ingredient handled. Here is exposures, here is the dose. And you try to validate these two boxes the same way that, you know, that is shown between exposure and the dose.

The study that was presented by the Task Force provided actually a more realistic approach to assess these problems, these issues. Besides the exposures and other factors, many amount of active ingredients handled, they're not known or can be quantitatively or qualitatively assessment. A random variable, the term that is added to the algorithms and which may or may not affect the exposures. However, according to the presentation made by the Agency using the PHED data, this random variable may likely not only exist, but affect the result, the resulting exposure substantially as well.

This is evident by the PHED case study

scenario number 2 in which the relationship between the exposures and the amount of active ingredient handled is dramatically different in terms of the slope of the linear line, 2.1 versus -0.26 although the tasks of this application is the same which is the open loading granule. So in these two cases the random variables that are proposed by the Tash Force study actually affect the results substantially. So one of the charge questions is that, do we need to add another term or look at other factors that affect, may affect this proportionality. I think that's probably a wise approach.

The Agency made a presentation about their future efforts in facing in this the desire of investigating the proportionality between exposure and the amount of active ingredient handled. In the future studies it's absolutely needed however, if the objective of doing this exercise is to use the data in the risk assessment paradigms then the agencies as well as the task force group should actually incorporate the absorbed dose into the whole exercise as well. Because, for example if you end up, if you can validate this proportionality by adding new data or whatever you want to do, the next step is to incorporate this exposure to the risk assessment and you're going to

transfer the metrics from the exposure to dose and there's another black box that you have to work on. And so instead of doing it later you might want to do it now.

Assuming the proportionality between exposure and the amount of active ingredient handled is proven valid for any reason, the same proportionality should also be proven valid before it can be used in the risk assessment, meaning exposure and the dose. So the task force people propose that at the end they're going to normalize the exposure with the active ingredient handled to come up with the best estimate in terms of this proportionality. I thought that might be, that might not be the right approach. What should be done is that using the amount of active ingredient handled as a modifying factor. For example, in the morning discussion there was a phrase that came out that the sloppy people tend to be sloppier and that's why they get higher exposures.

So, think about this scenario, if the person is really sloppy, the more the pesticide this person applied the high exposure will likely happen to this worker. But if you normalize the amount of active ingredient handled you dilute the affect, the true affect that this sloppy guy actually has a very high

exposure.

On the other hand if a person as a pesticide applicator is very careful, follows all the standards, guidelines and the labels, so no matter how much pesticide that he handled or applied, his exposure will still be very low.

So this modifying factor will not affect the final exposure and probably not the dose as well. So I would not say you should normalize, that you should use this amount of exposure as a controlling factor, modifying factor that would kind of compensate for good reasons as well as for bad reasons.

That's all I have for this question.

DR. HEERINGA: Doctor Appleton is our first associate discussant on this question.

DR. APPLETON: Thank you. Well since this is totally out of my field for a change and I couldn't steer everything back to biomonitoring, I guess we got close here, I'll be the one that says the nice things.

I'd like to commend both the EPA and the industry task forces for all of their work and more rigorously defining or trying to explain the potential sources of variability that we see in the measurement of one of the key assumptions, that is the proportionality of exposure with the amount of active

ingredient handled. This is a key assumption that was accepted without much debate 25 years ago at the EPA, I was a fledgling scientist in the exposure group back then, but kind of sat as a fly on the wall watching it all happen.

And I accepted along with most other folks that there was a proportionality in the relationship, it's very intuitive and you can find examples in other kinetic examples as we just heard to support that. And to a small extent my branch chief suggested that I did accept that notion.

I'm not sure where all of this discussion will lead in terms of being a user of either the Pesticide Handler Database or the Agricultural Task Force's database, so until I hear differently I suppose I will operate business as usual and assume proportionality.

My only real tangible recommendation I can make on this right now is that as a PHED user for the Forest Service I would really strongly and somewhat selfishly recommend that EPA not give up on the PHED. Stick with it and do try to find out as much as you can with the resources that you have to use the data. And I say that not only for my organization but for many other organizations that deal extensively with more

minor use pesticide issues that may or may not be addressed by the agricultural database. We grow tress, we don't grow, you know, a country full of winter wheat, so things of that nature, understanding that the industry has its own agricultural regulatory needs. So those are really the only comments I have so I'll rest with that.

DR. HEERINGA: Thank you Doctor Appleton. Paul Hamey, doctor Hamey.

DR. HAMEY: Thank you. I think just sort of thinking about the issue in advance of looking at the data I think we can sort of hypothesize that for example where exposure arises through contact with airborne material evolved at a fairly uniform rate, there might be sort of a strong association between amount handled and applied. And we might find this for example when measuring exposure during the handling and applying granules where airborne dust is generated during hopper filling. And I've seen at least one modern regulatory study where there was quite a nice relationship between the amount of exposure and the amount of granules handled.

And so I think, you know, under some circumstances we can imagine there will be a proportional relationship.

But as already suggested this morning I think we can also imagine alternative situations where exposure results mainly through contact with contaminated surfaces and where the residues on such surfaces may be at a equilibrium plateau or a sort of constant level. And for example with mechanical transfer devices for loading liquids were mentioned and here the exposure is probably limited to contacting the concentrated residue left on the dry break coupling which is actually independent of the amount transferred. It is affected by the concentration of the active ingredient in formulation so even in that situation under some circumstances you might find some relationship.

Other sort of situations where the user is protected in a closed cab when making ground boom applications, an exposure might occur there when handling the contaminated boom or nozzles and the outside of the cab and particularly residues on the doorhandle of the machine.

So exposure could, in those two scenarios be proportional to the number of decouplings or the number of times the driver leaves and reenters the cab, both of which may not be directly related to the amounts or indeed any other useful parameter that we can

determine.

So for some scenarios we may not, well I do not expect a strong relationship will be found. However I think the assumption probably serves as a reasonable default but I would caution that this may be true within a closely defined scenario but may be more challenged when we have wide scenarios of work tasks and practices and equipment within those scenarios.

For example if we consider ground boom sprayers these may range from small antique machines mentioned yesterday to state of the art large modern self-propelled machines with induction bowls, clean water supplies for washing gloves before removal, glove lockers to put your gloves in after you've taken them off before you get in the cab, automatic folding booms and with the operator and the equipment controls positioned in closed air conditioned cabs. Such machines, the latter machines will achieve much higher work rates and consequently the user will handle more pesticide but because of the technical controls the unit exposures will be much lower than on the old antique machines. And if we have all the ground boom machines in one category then we're misleading ourselves.

I think it's also reasonable to consider the

data used to derive the specific unit exposures in more detail. And if the unit exposures are derived from the antique machines then we would expect the unit exposure to be overprotective in the case of modern equipment, but if it's the other way around and the data would derive with the modern equipment then we may underestimate what's happening with the antique machines. So that's two additional points to consider.

And I think I've also got some brief comments to make on the figures presented in the EPA's documentary evidence which I thought was, you know, the evidence I thought was a really good job, I was really impressed with the amount of effort that had gone into that. Having recently done something similar myself with the European data I know how difficult it was so I was very impressed with what was presented this morning.

On figure 4.1 which I think this morning turned up as slide 21, is it possible to see that again? Yeah, this was a very interesting graph which, you know, shows a relationship that none of us expected. But I went back and looked at the data here. Study 425 was performed in 1977. The other data set is nearly as old. So similar vintage there, but when we look at these they were very short replicas, 3 to 5

minutes long and in 425 they loaded about half the product from a 50 kilogram bag. And in the other studies they were handling several complete bags. And maybe the work tasking using, measuring out the product from part of a bag was very different to using complete bags so those sort of aspects we need to consider as well.

I think another comment I would make was using data from about 30 years ago to do risk assessments now and the composition of modern granular formulations is probably much better, probably a lot less dusty so these data might be quite conservative anyway.

I think it's also worth noting that the 3 to 5 minutes duration in these studies illustrates why you need to describe the relationship so you can estimate exposure form whole day work tasks. And it's probably important to recall in our deliberations that the new data will be more representative of the full working day and in those cases extrapolation is necessary. It won't be so much of going from short replicates to try to estimate what happens, you know, a representative working day but it will be looking at extrapolation to different application rates or different product concentrations which may be, have slightly less

uncertainty than this situation has.

One of the other figures in the document was figure 4.2, I don't think we need to get the graph up for that, but it was, it looks at the rest of body outer dosimeter relationship for studies 1003 and 448, that they didn't seem to have the same sort of relationship that the overall regression had. And it seemed to be influenced by, very much by a single point from Study 428. I just think we need to look at those sort of aspects as well.

Slide 20 from this morning which was figure 4.4 I think in the original evidence, Doctor Johnson made one of the points that I was going to make, looking at the yellow dots from Study 1011, if you took those in isolation maybe the slope is closer to 1, that's what I thought when I was looking at those before. I think further inspection of the data here is also interesting. 1004 and 1011, they are similar in age, they're both from, well one's from 1992, the other's from 1994, both products contained 5% active ingredient, both were done outdoors, one in the U.S., one in Canada, don't expect any difference from there, both involved workers who were separately monitored over short periods again so each of those dots represents about a 10 minute work period so there's

some issue with that. But they're similar in that respect. There were differences in the environmental conditions, 1004 was done under winds speeds of, some of the replicates with wind speeds recorded as 6 to 8 miles per hour and others of them 10 but Study 1011 was done under much windier conditions and all the wind speeds were recorded as above, being above 10 miles per hour. That's might have had a partial reason for this sort of relatively higher exposures in the second study. There were also differences in the product packaging. 1004 used 10 pound bags and 1011 used 50 pound bags. Again this could have contributed to the differences observed between the two studies with the larger bags being much more difficult to handle. Another difference was the product. In 1004 it was an insecticide and that in 1011 is a herbicide. This suggest that to me that there may be some differences in the equipment being filled and to continue this comparison it would be helpful to have descriptive details of the tasks done. Unfortunately we don't have those in the database and I looked for the details of the tank and the hopper size to gain some idea of the sort or equipment being used. For Study 1011 it's given as 450 to 1,150 and I assume the units are gallons but I can't remember what they were specified,

but we don't have any information for that parameter for the other study so we can't really compare. But these are sorts of the issues that I think are important.

And I, you know, the point I'm trying to make is that when we're doing these analyses we're not comparing like with like and care is needed when looking at the relationships within studies.

Just briefly, would additional variables help to explain some of this variation we see? Well I think it, maybe they would but I think there are issues because a lot of the additional parameters recorded in the studies, if we start including them we do have problems with multi-colinearity which we, when we start adding a lot of related variables into the analysis so that we have to be very careful of that. And the other sort of point of caution I think is that when we're looking at the variables to see if they helped us give a better description of the determinants of exposure, then they have to be sensible in terms of, from the regulatory perspective and one can imagine that some of the parameters are not very helpful from that point of view. Statistically they may be interesting but from a practical point of view they're not.

I expect the new data will possibly help to

inform on these sorts of relationships because there's obviously a lot of issues when looking at the PHED data. It would be nice to design sort of an experimental approach but I am persuaded by the arguments that the new data should really be more descriptive and that would really meet the requirements if the other objective that such studies really should involve applications representative of what workers' typical practices are rather to have some artificial element built into them. And I think that's all I have to say for the moment.

DR. HEERINGA: Thank you very much. Doctor Kim.

DR. KIM: Being from a school of public health when I look at this data, the data sets or the charts that EPA presented, what comes across my mind is that assuming proportionality between exposures and the amount of active ingredient handled is like the problem ecological fallacy. The ecological fallacy is the assumption that all members of a group have the same characteristics of the group at large. So something like, you know, all students from Harvard have higher SAT scores than students from Cornell but if you take a random sample you find that there's a student from Cornell who has a higher SAT score than a student from

Harvard.

And in the figure shown in section 4 of the background document provided by EPA we see evidence of this. In other words, across studies we see a linear trend but if we look at each individual study separately the assumption of linearity doesn't seem to hold. And Doctor Johnson discussed this in his comments. However if unit exposures are going to be applied in a risk assessment management context then it should hold across studies and scenarios. In other words it should be generalizable. I think that the AHETF, the task force is moving in the right direction by collecting data for a wide range of pounds applied, a variety of locations, MUs and diversity of workers and equipment can be used to better assess this proportionality, the assumption of proportionality. However a controlled study may be much more informative for teasing out this relationship.

My final recommendation is that strict adherence to linearity might not be the right approach. In PPPA models when we construct them we use scaling laws for cardiac output, ventilation, using body weight to the power of .75 and this can, when we look at volume of tissue we don't necessarily use those, the power law. However these are, these relationships are

established from empirical measurements and are determined using log or regression analysis. These laws hold across and within species and is a generalizable law that has been determined empirically and used widely. Something similar may be helpful here.

DR. HEERINGA: Thank you very much Doctor Kim. And Doctor Popendorf.

DR. POPENDORF: Well, let me begin with pointing out a couple of weaknesses in the analyses that are, have been done. I mean given the data they were a good job but I think the problem lies with the existing data. For instance the first figure in the presentation was that was slide number 4 from Reinert and Severin, good correlations but again that really wasn't active ingredient handled if you look at the x and y. It really was a prediction of dose rate as a function of concentration of what was being sprayed. So, you know, good data, different, obviously a different correlation.

The second point is the weakness of a lot of the other data. Again, what was there, for instance figure 4.2, you're looking at a very narrow range. You know, looking ahead to some of the discussion I think in question 6 where they're designing the study and how

big a range do you need to answer some of these questions? We don't really have an obvious measure of the variability of the individual points that I was talking about yesterday. And if you consider how variable those points are you're looking at a, that is the y value, you're looking a the x value over a range of 10x and the variability in the y's could be, you know, 5, 10, the same kind of magnitude, you really can't test the validity of that test, you know, you're just looking a those points but those points aren't as accurate or as precise as they appear on paper.

So, you know, we definitely would need to look for opportunities within PHED to look over a wider range of x's and that looks like it'll be designed into the new database.

I think as a point, you know, and some people kind of question the idea, do we really need to use that active ingredient handled? And I think in order to extrapolate to other pesticides that may be at different concentrations, if you change formulation you're going to end up with different amounts or perhaps the actual amount applied on per acre and if some user applies a given number of acres you need to be able to do some kind of extrapolation to those other settings. And I think the logic that was pointed out,

I mean it's intuitive that it should be related, it's been used, it's sort of that default value that we kind of have to go with.

But I can fully agree with the comments in terms of the issues of equipment. That's got to make some differences and I think you need to really consider the equipment within the categories that are being proposed. Perhaps look at that and narrow those definitions of what you're, what is being used within a category so that, you know, you might end up with a few more categories but you'll be able to make better predictions because of the influence and very easy to get a 10x difference just due to equipment that has been pointed out.

You know, I was going to suggest, you know, the rate of exposure issue kind of like the Severin paper, that seems to be something to look at in the future, whatever that design is, the critical event type of thing, the handling of some number of bags or valves or that type of experience is certainly important. Perhaps the speed of what is being used and it might be the rate at which you're actually putting the material on. You know, that could differ if you're certainly in a tractor environment, the fast you go the, no wind at least and average in various

directions, the further away that mist is going to be so you could predict speed having an affect. If you go to a manual type operation, biocide type operation and speed could certainly again sort of dilute the affect of exposure. There's a lot of possibilities that could be looked at. I think you just sort of consider at this point that amount of active ingredient handled as the default consideration.

I think I mentioned earlier that, my observation about the, when the slope is greater than 1 so that's sort of on the record, given what I just said about how reliable is the slope, I don't know how good that conclusion is. It's just strictly an observation.

I think completes my comments.

DR. HEERINGA: Thank you Doctor Popendorf. Comments on this particular question of proportionality. Doctor Landers.

DR. LANDERS: I'd like to support the comments that were made earlier by Paul Hamey. As I mentioned yesterday, not all sprayers are the same and so we discussed that.

I would like to support Doctor Kim in his ecological fallacy. In engineering we have engineering fallacies. I'll give you a quick example, nothing to do with the SAT scores but as a fellow academic we all

know that Princeton is in New Jersey and we leave them out of the debate. We have to be fair.

Anyway, closed transfer systems appear to be heralded as the answer to everything, you know, just stick one of these things on and we've protected the man. I would like to point out that in California for example where they have such legislation, removing a dripping probe from a pesticide container puts the operator at greater risk than removing a sheathed probe. And so not all closed transfers are the same.

So what is to be done about this? Well my recommendation to the two groups is, when they consider closed transfer system that, or any other engineering control, that some form of matrix system be developed, either looking at high risk/low risk components or looking at ancient versus modern, some simple system so that you can create a better system. Thank you.

DR. HEERINGA: Doctor Landers made a very interesting point. When you talk about sort of a qualification or some quantification of relative sort of risk level across technologies within these, even these exposure scenarios, the 30 that are proposed for the AHETF. So you would just have some sort of graded system that might be part of the database that could be used like the data quality measures are currently.

DR. LANDERS: Yes, indeed.

DR. HEERINGA: That's a good suggestion. Thank you. Other comments form other presenters? Yeah, Doctor MacDonald, Peter.

DR. MACDONALD: I'd just like to ask Doctor Hamey to give some examples. You referred to some variables that were more of statistical and less practical interest and I was just wondering what you had in mind there. It seemed that everything we've talked about is practical.

DR. HAMEY: It's just when you look at the database there are about 80 or 90 different variables for each application that recorded crop height, boom height, wind speed, boom width, number of nozzles, boom pressure and some of those things. I don't think as a regulator we have much understanding as to what are the different practices in reality and if we want, we decide it's a parameter that is useful and we want to have some control over it, then we have very difficult, great difficulty in making recommendations and enforcing recommendations that people can follow.

DR. HEERINGA: Cynthia Hines.

DR. HINES: Yes, I would just like to actually commend the task force for recognizing that it is important to test this assumption of proportionality

and to try to work that into their data collection efforts and recognizing that you do need to improve the variability in the amount of active ingredient handled.

I agree with Doctor Kin that this is something critical that needs to be looked at because we are normalizing everything by pounds of active ingredient handled and it may not always be appropriate. And so it's time and it's something that's concerned me over the years to look at this to the extent that we can.

DR. HEERINGA: David Miller.

MR. MILLER: Yeah, just to add on I guess to Doctor Hamey's and Doctor MacDonald's, in terms of things that are of regulatory interest, a 60 foot boom width versus a 20 foot boom width or so, that's not something that is, can be, well, I mean a farmer's equipment is a farmer's equipment so that's a lot less manageable in terms of by label changes than for example a variety of other things.

DR. HEERINGA: But in that context I think Doctor Landers' suggestion that across the whole array of equipment within a scenario that there may be some method based on expert judgement to sort of qualify the relative possibilities. Within that we'd still have enormous variability but that might go a long way to

sharpening this out.

Let me ask the panel, in the context of this generic system, database and its use in the regulatory framework that you've heard about, do we agree that this proportionality assumption is probably the conservative assumption, the straw man against which everything else would be tested? Are there consensus on that? Dallas you'd better

DR. JOHNSON: I'm always bothered by this word proportional because it's not really being used as proportional. If something's half of something else they're proportional. But it's more than being proportional, it's the slope is equal to 1. And I think the planned data will allow one to test that assumption and I guess that, I think it's something that should be looked at and then I guess people will use it the way they want to use it.

DR. HEERINGA: Doctor Popendorf.

DR. POPENDORF: Yeah, think it might help to support that default if the Agency were to develop a sort of physical model or a rationale based on physical mechanisms of exposure that would be the background document if you will. It doesn't have to be extensive but, you know, why would the amount handled be a predictor? Some of the critical events that

Doctor Hamey mentioned, for instance in some scenarios there's no good reason to think that it would in some exposure scenarios. But if you had a rationale for each scenario then I think it would help to justify that, whatever assumption you do go into, active ingredient is your probably most common category.

DR. HEERINGA: Doctor Bucher.

DR. BUCHER: In thinking about how this might be utilized by the Agency I think it might be worthwhile for the Agency to prospectively consider what kind of information you could generate from this kind of database that would convince you that you should depart from this assumption of linear proportionality, because I'm afraid that if there are not really compelling data generated and there's lot of data points that are sort of randomly scattered around on the map, that it's going to be difficult for you to really decide where you would depart from this assumption and where you wouldn't.

DR. HEERINGA: Doctor Lu.

DR. LU: I think this proportionality issue may have something to do with how the pesticide is being mixed in the spray. So it definitely is not a default assumption. It's probably a case by case scenario. I look at all the graphs that the Agency

provided in the background documents. It kind of strikes me that if you pay close attention, the cases or scenarios that are close to the linear proportionality will be for the application I remember as open cab air blast. I mean the r square is close to 1 and so on and so forth. And I think there's a reason for that because again I have a very limited field experience and all the field experience that I have was associated with fruit tree growers. They use air blast in the field and what happens is the active ingredient is OP or other pesticides they use, even if mixed in a tank and then being blown out by the big engine as the tractor hauls the thing along the orchard.

So if you think about an exposure scenario it's close to uniform exposure because it's become a smog of aerosols. And probably you can assume that each aerosol if you were able to collect each droplet they'd probably have an even concentration across the air.

So in that scenario if you look at the graph that EPA produced they're close to the r equal to 1 slope. That's the only case that I can see. Again it has something to do with the way the pesticide has been applied but it definitely is not a default assumption.

DR. HEERINGA: Other comments from panel

members? We'll turn to David Miller and Matthew. do you fee that we've addressed this question? Are there any concerns or is there any information that you feel you were looking for but you haven't received?

MR. MILLER: Maybe just one, I've heard kind of some different things about definitely not a default assumption and others that it seemed not an unreasonable assumption. Obviously we'll be looking at it in terms of the, some of the data we have but of course the AHETF, they will also, as you'll hear more about tomorrow, part of their secondary objective for that is to determine, well essentially get better information about whether proportionality with a certain power. So I think we'll probably be visiting that some as well

DR. HEERINGA: Okay.

MR. MILLER: tomorrow.

DR. HEERINGA: Okay. Any additional comments at this point?

DR. JOHNSON: Just one, Steve. The good news is that it doesn't really matter as long as, I mean as long as it's proportional with respect to some slope.

DR. HEERINGA: Right.

DR. JOHNSON: It doesn't matter that the

slop, whether the slope is equal to 1 or not, you can still predict risk, or still predict exposure.

DR. HEERINGA: It just changes your threshold of risk.

DR. JOHNSON: Right.

DR. HEERINGA: Yeah.

MR. MILLER: I guess one of things too that is kind of key on that in terms of was if most of them tend to have essentially, when you do the log log the x values, the exponents are less than 1, that's tending to be protective at the upper. So maybe a default of 1 it tends to be protective in that aspect if we do agree that it doesn't curve up like that.

DR. HEERINGA: Doctor Popendorf.

DR. POPENDORF: As you say, if we agree. And you know I made my point earlier about it might but can we really believe the data that's there? We don't really know.

DR. HEERINGA: I'd like to do a little group thinking here because we are a little past 1:30, I'm going to make a liar out of myself, I think we would have the prospect of finishing today if the presenters who are scheduled for tomorrow morning are ready. And I want to, I don't, wouldn't want to push that ahead without them but I'd like you to be thinking

about that between now and the time that we finish our discussion of the next question because that will be a decision point.

And so I think that the representatives who will be speaking on behalf of the Agricultural Handlers Exposure Task Force and also I think, David, you have a short presentation too which you said is very short, it's more of a lead in, thinking about whether we might do that and then finally address that last question this afternoon. We won't make that decision now but be thinking about that and obviously we won't, I won't make that decision on my own, I'll confer with you at our break.

But let's, at this point move on to the next question which is question number 5, having to deal with within worker and between worker variability. And Matthew, if you would read that into the record please.

MR. CROWLEY: This is Matthew Crowley, Health Effects Division, EPA.

The proposed AHETF study design does not include true worker replicates and is not intended to examine the issue of variability within workers. The AHETF notes that to appropriately investigate this issue would require significantly more sampling and resources. They propose however that their single dat

exposure distribution results can be used to evaluate longer term multiple day exposures by placing reasonable limits on expected interclass correlation coefficients, the ICC. They indicate that, from their own research and review of the literature the ICC is likely to be between 0.3 and 0.5 over relatively short periods of time, for example seasonal, and likely to be even lower over longer periods of time.

Please comment on the AHETF's approach to estimating the number of samples needed to determine within worker variability and their inclusion on the importance of measuring such variability in their proposed studies.

DR. HEERINGA: Thank you very much. And Ken Portier is the lead discussant on this question.

MR. PRESENT: I'm going to preface my comments by saying, you know, the question asked is a two part question. One is on the sources of variability on these observational exposure studies and whether the within worker, what's the importance of the within worker or within handler variability term. And then the second question you're asking us to comment on sample size determination as it relates to clustering or studies.

And I think the second question on sampling

really belongs in our discussion on question 6. And my comments on the clustering sampling, the sample size issue really needs to follow the presentation that was planned for tomorrow morning. After talking with the panel and Doctor Holden it's clear we kind of need to see that presentation before I jump into my comments.

So I'm just going to concentrate on the first part, on the within worker variability issue and maybe let the other members of the panel comment on that and then we can have our discussion on whether we want to move forward with the Agency and the AHETF's presentation before we really get into a kind of a full discussion of sample size. And I think my panel members probably agree to that.

So with that I'm going to start with my discussions on sources of variability and observation exposure assessment.

After listening to the presentations and discussions it's clear that exposure data collect from observational studies has the potential to address three sources of variation. First there's among handler variability, that is variation among different individuals doing the same or similar task, typically under differing environmental or other conditions. Everyone agrees that this variability is necessary to

measure and is a basic component along with the estimate average in almost any risk assessment.

The second component is the within handler variability. That is, variation among different measurements on the same individual doing the same or similar task, almost certainly under different environmental or other conditions. We say different environmental or other conditions to stress that these are not true repeated measures in the sense of an experimental study since study settings are never completely under our control, and in any case, since repeats have to be done at different times. We expect other factors that could affect exposure to have change during that time period.

The concern with repeated measurements is that because all the measurements are taken on one handler then the resulting values should display a positive correlation. The AHETF proposal argues fairly strongly that this source of variation is unimportant and/or too expensive to measure given the objective of the resulting data to support benchmark or minimal adequacy requirements for tier 1 and tier 2 risk assessments.

The third component is the within study variability. That is, variation among measurements

taken on different handlers within a specified study. This is referred to as the cluster affect and the concern is that handlers doing the same or similar task at one site may produce similar exposure values because the measurements are taken under common environmental or other conditions. The measure of similarity used for within study variability is typically the interclass correlation or the ICC, and the range of interclass correlation is estimated to be 0.3 and 0.5 for measurements taken over a short period of time. I needed to get this, yeah it's on there, sorry, I have to get slide out because for me to read this equation, you know, it's nice but, and statisticians can't do anything without equations right? I could wave my hands, would that help? This is under EPA SAP, EPA, January '07 and it's the PowerPoint slide. That was good. The folder, EPA January '07 in the PowerPoint slide. This is one of three memory disks, right, next slide.

Okay, thus the true model for MU exposures is a modification of equation 1 in the procedures for determining the required number of clusters and monitoring units per cluster to achieve benchmark adequacy, which is one of the background documents that was provided to the panel.

So here we have an equation for the adjusted, what's called the normalized exposure as a function of an overall mean, in this case the log geometric mean, plus a variance component for cluster, plus a variance component for workers within a cluster and then a variance component for repeated measures within workers within a cluster. Just click it one more time so you don't have to read all that stuff, it's just the same thing that was in the document. But the basic idea is that if we, you know, the kind of the overall conceptual model involves all three of these and what the proposal basically says is, throw out the last one, my r, i, j, k since we're not really interested in that particular variance component. Okay, so I wanted to have that just so we'd know that and that's going to be useful tomorrow or later today when we talk about sample size.

The AHETF arguments to de-emphasize within handler variability in Section 5.3 are clear and compelling. In particular they argue that repeated measurements on a single handler are difficult to do, typically are done without strong controls on other factors that can impact measured exposure and the resulting repeated measurements would be expected to demonstrated low correlation, that is a RWW term

between 0.2 and 0.4. The RWW term essentially relates to the variability of the repeated measure term, okay? This is not the ICC term, that relates to workers within a cluster. This is the repeated measures correlation of variability term. The between handler data which will populate the AHED database is expected to support tier 1 and tier 2 risk assessments. For such assessment the focus is on cumulative exposure over long time periods. The distribution of individual long term cumulative exposures will be best described by the between handler distribution regardless of whether this RWW term is 0 or 1, and I think they make a very good and clear argument and description of why that's the case. So I don't have any problem with that.

The also argue that the between handler data distribution could be used to simulate both within handler and between handler variability in any probabilistic Monte Carlo risk assessment by specifying and drawing from some RWW distribution. So, we don't necessarily have to know what this repeated measure correlation is, if we've got some idea of what its distribution looks like in the population we could kind of simulate correlated repeated measurements in a Monte Carlo simulation and I think that's a clever way to

actually utilize the between handler data to handle within handler variability.

All of the above I understand, I can see the industry viewpoint on the issue of within handler variability. But myself and some of the others who have talked about this issue can't help but feel that EPA could be missing an opportunity by not pressing or investing in some limited repeated measurements. And our justification for allocating time and resources to limited repeated measurements is as follows. So, you know, there's always a 'but' in these things, right? And here's the 'but'.

Current literature on which within handler correlation is, the current literature on within handler correlation, or on which the correlation's values are estimated is small and problematic. This doesn't, this literature does not provide strong justification for limiting the range of RWW, something which might be needed if indeed the AHETF approach is used to incorporate within handler variability in future assessments. So it's kind of one thing to say, well, we have these four or five studies and we can kind of guesstimate some values from this, but my experience on this panel is that kind of a statement usually doesn't cut it when you're trying to justify a

fairly complex risk assessment. And we're thinking that, you know, real data would make an easier argument.

Any probabilistic risk assessments, we want to, we want, will want to incorporate both within handler and between handler variability. The AHETF approach is always going to be weaker than an approach which is founded on estimates that are backed by actual data. The actually data case is always an easier case to make than kind of an assumption based case.

While the AHETF states little interest in examining potential determinants for factors of exposure, EPA and other researchers will definitely be looking to use these data for exactly this reason. The best data for identifying exposure factors is to have measurements taken under different levels of the suspected factor on the same individual using that individual as a block. So any matched measurement study is always going to be more powerful than any other kinds of study. So if I've got repeated measures of an individual doing the same task under different conditions it helps me determine whether that condition really affects exposure or not.

So from a study point of view for other

users, and I understand where AHETF is coming from on this, but I'm saying other users who are going to look at this database as a gold mine, they'd be happy to see some limited repeated measures in there.

So some limited repeated measurements on handlers in the database could result in more powerful studies of the importance of exposure factors. Adding limited repeated measurements should be relatively cheap, especially when compared to the cost of starting a new study, say bringing a new location and/or recruiting new handlers. Once the study side is established and handlers are recruited and documented it's relatively easy to get permission to take measurements from multiple times. I see no reason to examine repeats that occur at very separate periods of time, which would really increase the cost significantly. But same day or next day repeats would provide kind of the most information for the least additional investment.

Fifthly, I don't see repeated measurements as creating an ethical or other issue with the Human Subjects Review Board review. To get one measurement on the handler will require approval by that handler. Since these are work tasks that the handler is typically engaged in, there's little extra about

getting a repeated measurement that would raise a flag in front of a review board. You've already aksed the person to participate, you've gotten their approval, it's a, I just don't see them really getting all upset with that.

For these reasons we think the AHETF should reconsider their strict stand on repeated measurements. In the long run a better understanding of factors that affect exposure can lead to better understanding of how to mitigate risk and lead to better risk assessment.

And I think I'll stop it at that point.

DR. HEERINGA: Thank you very much Doctor Portier for the introduction. At this point Cynthia, Cynthia Hines will be our next associate discussant.

DR. HINES: Thank you very much. I think Ken covered a lot of major points so I might just expand on a few.

I have a lot of mixed feelings about this because in my own research I have very much valued doing repeated measurement designs. And I kind of can relate to Dana over here who had to kind of give up on the biomonitoring when it's near and dear to her heart for the database. I feel your pain.

I also have, because of my field experience I really am very familiar with the challenges of trying

to do these studies in agriculture. And at a minimum, trying to do a repeated measure study within one of these scenarios is going to mean at least doubling, you know, of course the number of measurements you make. So if you have say 25 persons in your scenario that's 50. You know, it would even be better if you could, you know, triple that. And there are going to be a lot of agricultural situations where getting even 50 people is going to be challenging and even getting two measurements within your study period.

I think some of the other challenges that are posed, that have been mentioned in terms of the frequency of collecting these measurements is a very good question. You don't have a lot of control over that sometimes in agriculture in terms of how often these applications are made. So you may be waiting a couple of weeks for the next one, you may be, you know, it's just very challenging.

And also of course analytic costs go up and, you know, and the task force here hasn't really spoke a whole lot about that but I would imagine that's one of the considerations when you're doing as much monitoring as they are in terms of whole body dosimetry and face wipes and hand washes.

So I can really see that trying to do

repeated measurements on all the scenarios may not be a feasible proposition, even if it's data that would be very valuable and of interest to all of us.

What I am wondering is if there is, and I haven't thoroughly looked at this, but is there a possibility of looking at maybe one or two scenarios that might be of interest to EPA or, and the Task Force, where that scenario or those couple of scenarios could be looked at from a repeated measures study design where the conditions in the field and the number and frequency of applications might lend itself to that kind of study design? And so I might encourage that you think about that before totally giving in on the whole idea of losing information on our within worker variability.

I agree with Ken that once, and from my own experience, that once a person has consented to a study, that getting a second or even a third measurement is usually not a problem. Usually your barrier is getting them to agree to the study and accepting the procedures. I think once they're comfortable with that, unless you're doing something very onerous, I don't think that will be a big issue for you in a repeated measures study frankly. And I've never had any problem from a Human Subjects Board as

long as the burden is not excessive.

There, in reading through the Task Force plan, and you may have changed some of the parameters since you wrote the document, but at one place I read that if a person were to, a consented person were to withdraw from the study that you would consider taking a person who'd already participated once and having them in a sense do a repeated, have them participate a second time, like a repeated measurement on just a few of the people in your study. And I guess I would kind of want to suggest that you just, discourage you from doing that, that if you're not really going to commit yourself to a repeated measures study design, then just get unique individuals and not end up with a hybrid, because I've seen those situations and then they get really messy in terms of, do we treat this data independently, do we not? That kind of thing. And then sometimes observations get thrown out. So you might think about that.

And then a couple of my other, or some of my other thoughts really, like Ken said, I think relate more to sample size issues. So maybe we should defer on that. I will only just mention a couple of things because maybe you can address them in your, the Task Force could address them. And there was something in

your document about biasing toward conditions that might yield higher exposures and I don't know if you'll be talking about that but if you could go into that a little bit. And I'd also like to hear a little bit more about the evidence supporting the, the importance of the geographic distance of the clusters and why you feel strongly about that.

And I think that's it.

DR. HEERINGA: Thank you very much Cynthia.

MR. MILLER: I wonder if I might

DR. HEERINGA: David Miller.

MR. MILLER: just a couple of questions. At some point it might be a good idea to see if the AG. Handler Task Force might be willing to come up and just give some of their thoughts on it.

I did have a few questions if I could just in relation to what's kind of been said before. It that

DR. HEERINGA: I wonder if we should hear from the remaining associate discussants. Can we do that and then go back? Let's, if you don't mind, Doctor Kim.

DR. KIM: My comments will be short. The point made by the Task Force regarding the focus of their monitoring program and the intended use of the AHED is well taken. Their statistical analysis is convincing and I appreciate the simulation exercise that was presented with the technical summary.

I commend the detailed work of the AHETF, the Task Force and the presentation of this material. However their data does not address within and between worker variability. Their conclusions are based on work done by Doctor Crimalt's group and Doctor Fenske in response stated that it would be helpful if EPA would conduct its own analysis to determine a range of values. I don't know if anyone has already done this but I'm assuming not because it was not reported in the EPA's background document.

So, I calculated the within worker correlation coefficient using the repeated measures data presented in figure 5.1 or the EPA's background document. I estimated the variance components using a one way random affects model for 10 individuals with a number of repeated measurements ranging from 2 to 6. The total number of observations was 39 and I obtained a within worker variance component of 0.4 and a between worker variance component of 2.5 and the within subject, or within worker correlation coefficient of 0.9. Thought this is a different finding from the assumptions made by the Task Force, it lies outside the

range of within worker correlation coefficients.

So there are two things here. One, the assumption is wrong. Secondly there's not much information gained from repeated sampling within an individual so your conclusion is supported with this quick analysis. However, before making final recommendations it would be prudent to examine the within worker correlation coefficient for various exposure scenarios.

On the flip side, repeated exposure studies are, I think needed to better understand biomonitoring data. As discussed yesterday, biological differences between individuals is important for interpreting biomonitoring data. Because of the physiological differences between individuals, because they vary much more than within an individual, such as body weight, metabolism, height, et cetera, a repeated exposure study would be very helpful for understand the relationship between passive dosimetry and biological monitoring data.

DR. HEERINGA: Thank you very much Doctor Kim. Doctor Johnson.

DR. JOHNSON: Yes, I, initially I didn't feel like I had much to contribute to this discussion and I probably still don't but I'm not smart enough to

keep my mouth shut.

The, I thought the Task Force made a good argument and I could live with that. I thought Doctor Portier also made a good argument and I could live with that. And so I'm not really quite sure where I come down.

One thing that might be of some interest is the relative cost associated with collecting the data. I guess I assume that a lot of the cost is associated with measuring the amount of chemical that are in the, whatever garments you collect, that that's probably the major source of expense. There may be some expense in getting to a particular study location but it might be, I don't know how these costs compare to one another.

I thought Doctor Portier was suggesting that you didn't need to collect the repeat data, you didn't have to add the same number of repeats on every subject and you didn't have to actually repeat every subject, that you could repeat fewer subjects. But then that's in contradiction with Doctor Hines who wants to have every subject done the same number of times. I guess given those two options I would, if the major objective is to measure this interclass correlation I would be inclined to go with Doctor Portier's recommendation. I think the, you've got these three variance components

to measure and there are certain numbers of degrees of freedom associated with those and so in trying to make a decision I guess I would try to do something that would sort of give me equal numbers of degrees of freedom associated with each variance component. And the only way you could do that is to not have the same number of samples for each individual and/or the same number of samples of workers within each study. So it may not be feasible to do that but I do think you can, if you do replicate the individual workers and if the major source of cost is associated with analyzing the data that you collected for the chemicals deposited, then it might not be necessary to measure each individual the same number of times within each study and within each worker.

DR. HEERINGA: Yes, Cynthia Hines.

DR. HINES: I just wanted to clarify. I agree, you don't have to measure the same person the same number of times. I was more concerned about the situation where you have just a few people that you've done repeated measures on and it's too small to do anything with. But you certainly don't have to have a perfectly balanced situation.

DR. HEERINGA: Okay, your statement is, if you're going to do it, do it right so you'll get

utility from the data that's there?

DR. HINES: Yeah, that's right.

DR. HEERINGA: Right. Doctor MaDonald.

DR. MACDONALD: Yeah, I also didn't think I was going to have anything to say but I'm going to give some preliminary remarks on some ideas I'm trying to put together for the discussion of the sample size question.

But we've got really three different types of design going on here. If you're trying to sample a total population economically you use cluster sampling because once you get into one situation you can pick up a few more subjects fairly cheaply. So that's really the purpose of cluster sampling, to lower the total costs. And that assumes that the cost of going to a location and getting the workers signed up then it's relatively cheap to do more workers or even to do repeated measures on one worker.

If however you want more detailed precise information on specific scenarios then you do what's more commonly called stratified sampling and there you pick the scenarios that you're most interested in and you make sure you get enough observations within each scenario. And that's a quite different objective from cluster sampling.

Then on the third level a number of people here in the room, on the panel and on the Task Force seem to have an interest in the different levels of variability in their own rite, within worker, between worker, between scenario, between machinery. And that's a different kind of question yet again and requires a different approach to the sampling design and I think some of the confusion is that we're looking at the one database and the one design to answer all three types of questions, and that the sort of recommendations you make on sample size depends really on how important you think the three different questions are.

DR. HEERINGA: Thank you very much Peter, I appreciate that comment. I think it's just everybody recognizes that simply because we are not measuring them to estimate them, these components of variance aren't disappearing, they're present in the overall set of measurements, we just can't differentiate them and put that information in the differentiated components to different uses. And there clearly are different uses as you point out.

David Miller, you had a, you wanted a follow on question for the panel?

MR. MILLER: Just one of the things that

the HSRB brought up during June was in terms of they brought up this specific issue in terms of consistently neat or consistently sloppy workers. As you know, the way we, in PHED right now there are true replicates but they aren't counted when we do our analysis that way.

I guess I had a question for Doctor Chambers just in terms of kind from what you've heard here in terms of some of the recommendations or thoughts or suggestions, in terms of repeat measurements, maybe a limited number, et cetera, do you have any thoughts on the HSRB perspective on that in terms of, if we go with, one, if it goes one way, how is the HSRB going to feel versus another way for example?

DR. HEERINGA: Doctor Chambers, are you comfortable answering that?

DR. CHAMBERS: Thank you for giving me a segue because I was going to ask for a little bit of time to philosophize on it anyway.

There are two things that have been brought up at this meeting that I think are going to raise some significant flags with the HSRB. One is the scenario, excuse me, the scripted business that came up yesterday and then it's repeated measures.

I've been working with the HSRB for its first year of existence and I'm on the science side obviously

and I'm speaking now for the ethicists which I'm not one of. But their main concern seems to be that we are putting people at a higher level of potentially risk than they would be in our, their normal occupations. And even though they may agree to an informed consent, Ken, there are still concerns I think that the people are being exposed to more than they normally would be. And I'm really not, I can't answer your question, David, I really don't know how they're going to come down on that but that's clearly the types of discussions that have occurred over the past year.

What I was going to suggest and I don't know whether this is out of line, Bill Jordan, or not but, you know, if the HSRB could be thinking about those in a highly generic way beforehand, the scripted idea, the scripted exposure and repeated measures without having a particular proposal in front of them, we did that kind of, you know, when we first reviewed our, the previously done studies we came up with a list of criteria by which we would judge whether or not these were scientifically valid studies or not.

If we divorce the consideration from any particular proposals and come up with the philosophy of the group on scripted exposures and on repeated exposures, I'm wondering if that might be a little bit

useful. And I would say if that is considered a useful suggestion it should be done in all haste really before any of these proposals get brought up. That's just something that occurred to me over the last couple of days here.

DR. HEERINGA: Doctor Portier.

DR. POPENDORF: I want to kind of reply on one thing. On the repeated measures I guess if this were an experimental study I'd agree with you, because if we were doing a scenario where we're asking the worker to repeat for the purposes of measurement only a task over and over again, you'd be exactly right, we'd be increasing their exposure by forcing them to redo an exposure event.

But this is an observational study, we're not asking them to do anything other than what they would normally be doing. So there, you know, whether we turn our back and measure them or not they're going to still do that next load. And so I see no, actually I see decreased exposure because they've got this stupid whole body thing underneath that's intercepting the dose. So if anything we're mitigating exposure on these events. So I guess I would kind of argue that. I mean I could see the experimental setting, it's a mouse, you're jabbing it and you're aggravating it and

it's unethical but with a person who's going to be normally doing the task they're going to do it whether we measure it or not.

DR. CHAMBERS: This is Jan Chambers again. My understanding and you all correct me if I'm wrong, but my understanding is if it's a strictly observational study it will not come to HSRB, it will only be brought to us John Carley is nodding his head that's correct. And it's only if it's something beyond observational and semi-experimental that it will be brought to us. And these scripted things are not going to be absolutely observational, regular routine things is my understanding.

DR. POPENDORF: That may be true for the biofoulant study, the other task force, but I think for this one, I don't think they're planning to do anything scripted, they're going to do things, I mean that was my understanding and maybe we need that clarified. But for the handler/loader kind of scenario that's not scripted. For the other task force that is going to be scripted and I agree with you on that because there you are asking them to do something in a scenario that they might not normally be doing it.

DR. HEERINGA: Cynthia Hines.

DR. HINES: Yeah, I think this is a major

source of confusion right now, is how much of the Task Force current plans are scripted. Because when you read through the proposal there are definitely scripted elements in there. And while I agree in a purely observational study, repeated measurements should not have an HSRB issue, if you do throw in that scripted element and you are intentionally affecting or changing that exposure, that's a different story. So if we could get some clarification on this whole issue of scripting that would help.

DR. HEERINGA: Yes, Bill Jordan.

DR. JORDAN: I'm Bill Jordan from the Pesticide Office at EPA. There is a continuum between, in the ways that one can design research. At one extreme research could be for exposure purposes could be designed in a very scripted way where, for example one of the studies that was discussed the other day was a jazzercise study in which people wore whole body dosimeters and they were told exactly what body movements to do.

They were to spend so much time crawling around on the floor, then they were to roll over, then they were to do jumping jacks or something like that. They were doing that on a floor where a pesticide residue was present and then the whole body dosimeters

were removed and the quantity of the residue that had moved from the floor surface with which they had had contact into the dosimeters was used as a basis for estimating potential exposure. That's at one extreme. I think we'd all agree that's a scripted study.

At the other extreme people who are engaged in research could go out, find a field that had been treated and ask the workers, hey, may we have your t-shirt at the end of the day? May we have your pair of pants? Step into the booth over here and we'll trade new pants for old ones and there would be simply a collection of those items of clothing and the same kind of analysis might be done on the clothing residues measured. And there might be an interview with worker to say, how long did you work there and there might be an interview with the farmer to collect information about how much was in the spray tank and so on and so forth. That latter one would be observational.

What the AG. Handler Exposure Task Force is doing falls sort of in between and each one of them will be looked at on a case by case basis to figure out whether as one of you mentioned, the scripted elements dominate to an extent that we would consider it intentional exposure, i.e., that the exposure that the subjects in the research receive would not have

occurred but for their participation in the research.

In thinking about the AG. Handlers Exposure Task Force protocols, at least the ones that I'm familiar with, what they are doing is that they decide on a particular scenario that they want to evaluate, they decide on a particular chemical that they want to evaluate, they go out and find a farmer who's willing to be a cooperator in the research and agrees to hire a crew or use employees working with that farmer, to, or even a research facility, to apply that pesticide to that crop at that rate and allow the researchers then to work with the crew who would be engaged in the handling activities to collect the necessary measurements of exposure, hand rinses, whole body dosimeters, face wipes and so forth.

Now the farmer is not going to agree to participate in that research if they don't need to do the pesticide treatment or if the pesticide treatment is the wrong one or somehow or other it's not appropriate for the particular site. But the decision of the researchers about what they need is driving to some extent the decisions about what pesticide is used and therefore influencing the potential exposures received by the subjects in the research. That characteristic is in my view, and I think in the view

of the other folks at EPA, sufficient for us to say that the subjects are participating in research that involves intentional exposure.

But I will happily concede to anybody who wants to discuss it that, yes, indeed there is a range of research designs and this one falls in the middle, closer to that line between observational and intentional exposure. But as a policy matter we are choosing to apply our regulation in such a way as to call this research that involves intentional exposure.

DR. HEERINGA: Thank you very much Doctor Jordan. Doctor Landers, did you have a comment? Any additional comments from panel members at this point in time on the issue of within worker and between worker variability and I guess we led off into the notion of replication. Yes, please identify yourself.

DR. LEIGHTON: Yes, this is Tim Leighton form the Antimicrobials Division. I want to clarify the Antimicrobial Task Force.

The scope of some of the studies in the antimicrobial perhaps will be observational. For the antimicrobials we do have industries such as metal working fluid where the workers are there every day and we can actually go in and monitor them. So some of the ones in the antimicrobial will actually be potentially

observational.

DR. HEERINGA: Okay, any followup on that? I'm interested in this distinction and and I won't drag it out here on an IRB2 between something that's observational as opposed to experimental. I thought if it were research it would be reviewed and you maybe exempt it.

DR. LEIGHTON: And another good example in the antimicrobial is the pressure treatment of wood where there is an existing facility, the worker is there every day, they're using a compound that we actually have analytical techniques for so we can go in and monitor those workers without changing what they're doing that day and get the measurements. So there are some example of what we hope will be observational studies.

DR. HEERINGA: Thank you very much. Additional comments form the panel on this? We, it sort of led us off into another area. David Miller.

MR. MILLER: Yeah, I wonder if I could just follow up on a couple of

DR. HEERINGA: Absolutely.

MR. MILLER: Okay. One of the suggestions that came, or thoughts that came up on the panel which we hadn't thought of before, but it might be possible

to have repeated measurements on only a number of scenarios, or not all of them in other words.

And the question I guess I had, I'm not a subject matter expert, but are there certain scenarios where you think it might be more likely to get repeated measures because you get that sloppy versus neat worker affect? In other words, I'm just making it up but it might be backpack sprayers, you might be kind of more likely to get high correlations from day to day within a worker versus, and I'm just making this up, aerial applicators, they may be much more less likely. So one of things I'm wondering is is it possible that if there could be repeated, or at least thoughts about repeated measures on scenarios which you think are most useful?

DR. HEERINGA: Paul, Paul Hamey?

DR. HAMEY: Well I haven't got an answer in terms of the behavior of workers. But it might be useful to try and relate it to the actual use patterns. So if you have scenarios where workers will not be repeatedly exposed to the same chemical then perhaps it's of low interest. But if you have a scenario where they repeatedly use the same chemical then it's of more interest.

And something that would come to mind, again not knowing the use patterns in the U.S., but if I was

looking at this in the U.K. I'd be particularly interested in orchard applications where there are spray programs that have repeated applications every sort of 10 or 7 days of the same chemicals.

So that might be a consideration.

DR. HEERINGA: Thank you. Doctor Lu.

DR. LU: It sounds like Mr. Miller has some sort of fear factor that are kind of associated with your study design, mainly because of the Human Subject Review.

But I remember the protocol for the IRB protocol, there is a component that, yes, you need to assess the risks that are imposed on the human subject, but also there's a component of benefit. So what happens is that we all value repeated measurements and that's why a lot of people that conduct this type of research, we kind of move from a cross section of study design to longitudinal, longitudinal meaning that you are going to repeat taking samples from the same participant over a period of time.

And we recognize that's the important step forward because we will get a lot of information from this type of a repeated measurement study that will never exist in the cross sectional. So that's a huge benefit and I will argue that, yes, we do put

additional risks on the participant because instead of giving us one urine sample they're going to give us 24 urine samples over a one year period of time.

But is the risk reasonable because the benefit will outweigh the risk? I don't know whether the HSRB ever can see the benefit component when they look at not just this study but all studies.

DR. HEERINGA: Doctor Barr.

DR. BARR: I just want to clarify something. As a sitting member on one of CDC's IRB boards, there's a distinction between risk and burden to the participant. Collecting urine is a minimal risk, especially if it's urine samples collected as a part of their daily duty. Now you may be imposing more than minimal risk if you're asking them to do an application where they wouldn't otherwise be doing it or you're asking them to apply a chemical that they wouldn't otherwise be applying. Then you might be putting them at more risk but it still might be less than minimal risk.

So by adding additional sample collections you're not putting more risk into the study but you are placing more burden on the participant. And so the compensation to the participant would actually have to equal the burden that you're placing on the

participant.

DR. HEERINGA: Doctor Landers.

DR. LANDERS: Am I right in thinking yesterday we decided that the reason for these tests was for the registration purposes of a new product? And therefore if it was, or if it is, sorry, then if it's a new product then if they're going spraying for example, then the application of that product would be for experimental purposes, the crop would have to be destroyed and therefore we would be conducting an experiment by its very nature rather than an observation.

Would you like to comment on that?

DR. PORTIER: I think the thing is that they're not going to do this with new chemicals, they're going to do this with a common one that they understand a lot about the properties and that is in current wide use and they're avoiding the new chemical issues and worrying about the physical issues. So I don't think this database gets into that. Its utility will be toward the new chemical but not its measurement.

DR. HEERINGA: Doctor Curwin.

DR. CURWIN: I mean this whole subject seems a bit off topic and I wasn't going to bring this

up because it is off this variability topic. But on the human subjects side of it, it's my understanding of the new data that would collected would be on existing chemicals which I think was just clarified. And what I understand the manipulation part of it would be is the amount of active ingredient handled and that's what they're trying to get at. So the way you would manipulate this is either by manipulating the application rates or the acres applied. But my understanding is that this would still be within the labeled rates, the label use rates.

So in terms of maybe increasing the potential subject's exposure that may be the case, they may be asking the farmer to apply a little bit more than they normally would have but it's still within the acceptable risk level that's been established by the EPA already. So if we're to believe the risk assessments that the EPA has done on these existing chemicals then their actual risk shouldn't be increased.

DR. HEERINGA: Doctor Chambers.

DR. CHAMBERS: Just real, real briefly here, let me just clarify that the HSRB is not an IRB, the HSRB looks at the IRB approvals that have come down and everything. The HSRB has almost no experience

right now at looking at protocols. All we've been doing most of the time is looking at completed studies and so we're really kind of going into protocols right now. And that's why I suggested that perhaps now is a very, very good time to try to get some philosophical tenets down for HSRB on why they would approve repeated studies, scripted studies and that sort of thing before they have to look at an individual study and judge it in the context of a new proposal.

DR. HEERINGA: Thank you very much Doctor Chambers. At this point in time Doctor Portier.

DR. PORTIER: I just wanted to get back to one last clarifying statement which I think Doctor Dallas said. The statisticians really don't care if things are not balanced in this data set. We just want measurement somewhere so if you have the measurements we can estimate it.

And so to re-emphasize the balance issue is not really a big concern for us. It's an existence proof than an amount proof, right? We'd just like to see some there so we could estimate those values. And I think Doctor Kim's little study that he did on the side kind of makes the point that I made, that you have literature that says it's this, and yet your own data says it's that. You're going to need something that's

going to tie this down a little bit firmer if you're going to use it at any point in the process. And we can, I can at least envision some situations where you're going to want that source of variability.

MR. MILLER: I know Mr. Villanueva has a quick question but just an immediate followup on that, would one possibility be to go deeper into PHED and essentially look for essentially the, that kind of information? There are repeated measures there and would that help inform I guess, instead of that range, .2 to ..5 or something like that? Would that assist in

DR. PORTIER: Yeah, yeah, I thought of that and the reason I didn't make that recommendation is because the general feeling we has is the level of control that you've had within the PHED data is fairly low. You either have control or you have a lot of measurement, is really what you need to be able to do this. And you don't have a lot of measurement on external factors.

You didn't have a lot of control in the generation of the data and so I'm not sure that exercise is going to produce a lot of useful information. The AHED on the side is going to have a lot of control and a lot of exposure factor

measurement. So you've got two good things going for you there and all we're asking you to do is think a little harder about what we're going, we're going 85% of the way or 90% of the way, another 10% and you're going to have a real jewel of a database that's going to allow you to look at all three sources of variability.

That's all I'm kind of, trying to say here is that we're, you're almost there and we don't want you to falter at that point. At least that's my feeling and I think that's supported by the rest of the panel.

MR. MILLER: Can I do a quick followup on that? Just, you had mentioned before one possibility was essentially to do a limited number of replicates, I'm making the number up, 10% or something like that. Do you see kind of in terms of do you have any thoughts in terms of the balance if they increased the number of replicates by 10% but decreased by a certain percentage the number of different individuals? I mean I guess it's a cost/benefit type thing. Is that something that, do you have any idea in terms of numbers?

DR. PORTIER: I think we'll get into that in the discussion of sample size.

MR. MILLER: Okay.

DR. PORTIER: I mean I, I didn't really

plan to address repeated measures in the sample size question, I was going to go on the assumption that you weren't going to be interested in it and we'll, we would discuss, but I think given a little bit of time we can probably give you some recommendation on that. At least some approach recommendation on how to handle it.

DR. HEERINGA: Mr. Villanueva, do you have question for the group?

MR. VILLANUEVA: Yes. Actually Doctor Kim's exercise got me thinking and I was, with the ICC I'm being bounded for 0, between 0 and 1, if we have 1 then repeated measures doesn't really matter at all, we might as well just have sampled another individual, 0 meaning that there's no correlation between the measurements.

And I guess what I'm thinking about is the variance components that Doctor Kim mentioned, I think they were about equal worker to worker and then within worker. I think, or were they similar?

DR. KIM: If I recall it looks like .4 for within and about 3 for between.

MR. VILLANUEVA: Okay, yeah, well, that's even, I think that's probably even better but the real question is, is I guess thinking about the benefits of

repeated measures. If you just have, it really gets down to estimating that variance component within worker and the question I'm thinking is, is there any reason for the panel members to think that the within worker variation is actually going to be larger than the worker to worker variation? And if not we can just use as a surrogate the worker to worker variation to represent the within worker, being a conservative estimate. So saying that, you know, they're not correlated at all and you're just as likely to measure another person.

DR. PORTIER: You know, in looking at the data on the graphs I mean I got the feeling that the within worker variability is going to be less than the between worker variability. But that's, that's hard because everything we see is confounded by study condition too and so it's a little bit, and it's a little bit hard to be able to make a real firm statement on that.

But I wouldn't be surprised if the within worker and between worker variability are very much the same in this situation. I mean my feeling is that the RWW is probably close to 0, much closer to 0 than it is to 1, in the sense that, you know, repeated measurements on the same person are very much like

repeated measurements on other workers. And that you haven't lost a lot by assuming that these repeated measurements were independent in the previous risk assessments. I haven't seen anything that really makes me worry about that too much. Maybe Doctor MacDonald can comment on that too.

DR. HEERINGA: Doctor MacDonald.

DR. MACDONALD: Yeah, Doctor MacDonald, I'm just thinking there's got to be some compromise between us sitting around and trying to decide on whether a certain scientific fact is true or not and insisting that the new database have 25 repetitions of every worked and every situation. There's got to be some sort of compromise and certainly a limited number of directed studies just might be enough to establish whether this is going to be a reasonable assumption or not.

But I am concerned by what Doctor Portier said, that of course that anything you do is going to be, the results will depend on the conditions of that study and you always worry about generalizing to others. But at least it would be better than us just trying to guess the answer.

DR. HEERINGA: Doctor Kim.

DR. KIM: Just a followup on that. Yeah,

I agree that the whole purpose of that analysis was to demonstrate that, okay, let's just not go, we'll proceed forward based on studies that were done by other people but actually do a few studies ourselves, or EPA, to estimate the within subject variability or a correlation coefficient.

DR. HEERINGA: Steve Heeringa here, I want to follow up on Doctor MacDonald's comment too. And I think that in terms of the model that Doctor Portier put up, that's a pure random affect for the individual and I don't think we can ever measure that in any sort, because we are always confounding the fixed affects of the measurement, the boom length, the temperature, the wind, with those individual measurements. We can never sort that out so unless you have a utility for that pure individual affect or you're willing to work with those individual affects contaminated by a particular application scenario you need to think carefully about what you're getting out of that. And I know Doctor Hines, you probably have to do that all the time in your work. But I think that we have to realize that everything we've seen, those random affects, even if you estimated them with models are probably going to be cross contaminated by the fixed affects in that geometric mean.

So that's just a, that'll come up I think in the sample size discussion.

What I'd like to do at this point is I think I'd like to call a break for 15 minutes. And if I could have a little pow-wow with the EPA scientific staff and the AG. Handler Task Force, the Microbial Task Force who are here this afternoon and are going to be doing presentations. Just to see, I want to make a quick decision as to whether we have a green light to go ahead today or whether we would be putting anyone at a disadvantage in terms of preparation and coverage.

So if we could just, those people can meet up here.

We'll come back at 10 minutes to 3:00 please.

(WHEREUPON, there was a recess).

DR. HEERINGA: Okay, let's please have our seats and we'll get ready to start again.

Let me first explain where we have come down in terms of the progress of the agenda. As Myrta said at the start of the meetings it is a floating agenda so we are able to move things around. But I had a conversation just prior or just after the start of the break with sort of the principals from the Task Force and from the EPA and from a few representatives of the panel.

And where we have come is that we would like to use the balance of the afternoon to actually begin to have two presentations, one of them a short one by David Miller and Matthew Crowley and then following that up, that was actually scheduled for late today, and then following that up we'll have another major presentation by the Agricultural Handlers Exposure Task Force, Larry Holden, which gets into some more of the statistical issues related to the exposure assessment and the research plan development of the AHETF.

That'll leave us probably I suspect about an hour and then we will have time for questions of clarification. If we do begin our discussion of the sample size charge question this afternoon I would expect to return to that tomorrow morning. Everyone was in agreement that it would be beneficial what the case, to return here tomorrow morning after everyone's had a time to digest the information covered today, to think up any new questions, develop any new sort of responses or clarifications. So we would have that session tomorrow morning. We would finish our discussion of the sample size question tomorrow morning, have a general wrap up set of questions so that everybody leaves here comfortable that they've had a chance to present their positions and we've had the

kind of exchange we've had this afternoon.

After three days we have to get a life, you know, and we're still talking here about sample size issues after three days. But it's been great and I really appreciate everybody's contribution.

So at this point in time then what I'd like to do is I'd like to ask David Miller and, or Jeff Evans to give their short presentation.

MR. MILLER: Yeah, this is just a real short one or so, really kind of geared more toward an intro to the next one which you'll hear from industry. It just kind of recaps a little bit about what we've talked about before and what the industry is going to be presenting next.

Just to begin on that slide number 2, just it's kind of a quick recap of the exposure topics discussed to date. It began with a historical perspective on handler exposures, it went through some of the developmental, development of PHED as well as its use as a generic database, its function, kind of where it came from, the data that went into it, why it was generated to begin with, et cetera. There was also a discussion of passive dosimetry and biological monitoring and then also hand washing methods. And then kind of the EPA presentations ended with a

discussion of the proportionality between pounds AI handled and exposure and discussed generally the kind of the inter exposure concept and how that's central to a lot of kind of what we've done.

Just the next presentation is introduced, to introduce is the Agency's asked the panel to discuss the limitations of the present data and how additional data which is proposed to be collected by the Task Force can improve our ability to assess worker exposure.

What we're not asking is the panel to review the plan proposed by the Task Force for addressing these issues.

The panel received as part of its package, two specific documents that are going to relate to the Task Force presentation next and again it'll blend in a little bit to some of the other documents, in particular the intra versus inter-worker variability. But the two documents, I won't read them, one was just a technical summary and the other the specific document on clusters and monitoring units. So basically a statistical simulation and a general plan for what the Agency is, what the Task Force is proposing to do.

The primary goal is the collection of worker exposure monitoring data and its incorporation into a

generic database that will define exposure distributions. That's stated in the Task Force documents.

The AHED you've heard about and the idea behind it is to have the ability to estimate individual exposures for a single workday given two pieces of information, mixer/loader/applicator pesticide handling scenario as well as the amount of active ingredient handled.

The purpose of the presentation you'll hear next from the Task Force is to discuss their data development plan that's consisting of both kind of their overall plan which relates to the first document that went through. And then the other issues related to sample size and its adequacy. It deals with the number of sites, the number of subjects per scenario and activity and also kind of how that relates to the previous discussions in terms of the repeated measures.

DR. HEERINGA: Thank you very much Mr. Miller. At this point in time, unless there are any questions from the panel, Jeff, did you have anything to add at this point?

MR. EVANS: No, I mean only to add that the sample size was probably one of the biggest issues raised by the HSRB and it's a constant battle with the

agencies and data doers so this is an extremely important part of the program for us.

DR. HEERINGA: Thank you. We'll devote adequate time to it. Okay, at this point then I'd like to request that the presenters for the AHETF, if they were willing to come forward. Do you have a microphone there Doctor Holden?

DR. HOLDEN: Thank you Doctor Heeinga. Let me get this set up here, okay.

This is the second part of a two part statistical presentation I guess that I'm sure you all were looking forward to. That wasn't supposed to be funny but I guess it is.

DR. HEERINGA: Doctor Handwerger before he left asked us to videotape this.

DR. HOLDEN: Oh no. I'm sorry, my mistake, yeah, I'm technologically ignorant here, there we go.

In this presentation I was going to cover the sort of the last three, not the last three issues but the last three issues that we wanted to discuss today. The sampling methods and the description how we're actually going to sampling or proposed. Doctor Portier has assured he's got the solution to all these problems so we're all set. And secondly, a discussion of the, I

would say, in a sense an example or an example of how we propose to at least start to address the issue of the goals to help determine the sample sizes and I think that you'll see that we feel strongly that there ought to be some goal other than just collect data. And then finally just a few slides outlining the general, if you want to call it kinds of things we're going to be talking about in the scenario summary analyses that the Task Force is planning to provide along with the documentation for AHED if you will.

Here we go with my little population diagrams again. What we're starting off is with this term monitoring unit which I'm sure that people are scratching their heads and wondering where that term came from. And it's basically what we are referring to as not necessarily the individual, but the set of conditions that we select from this target population and the resulting exposure measurements and everything else we measure.

So a monitoring event is actually probably a better term. I think the Agency or someone threw that up but I actually like that term better. This term monitoring unit was sort of patterned after the main sampling unit which was just to try to find something a little more neutral that replicate which has sort of a

different meaning, depending on who you talk to and the context.

So we are obviously going to be taking samples from this target population but these samples are not going to be what's called a probability sample, or not a true probability sample anyway. In other words we're not saying, unfortunately, not saying that the, that we have enumerated either implicitly or explicitly all the possible conditions and are sampling from them with some sort of known probabilities, be it completely simple random or otherwise.

We sort of feel that in order to do that we might have to have some approach, or one approach possibly would be to do something like a multi state sampling that's typical for so many surveys where one, and it's important to describe this, where the, we might start off by making a list of all the counties let's say in which this scenario was, is being used and maybe time periods in some sense, maybe weeks, months, whatever it is, some unit of time, some increment of time. And then randomly select from those as the first stage. We might first, we might do that by first enumerating or getting some idea about the, how many workers that might be inside that and select the size. We might use some sort of stratification, I think

Doctor MacDonald mentioned that, some sort of stratification approach, whatever might be appropriate to increase precision.

Once we've got the sample of county/periods, I'll use that term, then one would proceed probably within that county and conduct a large number of activities designed to ultimately get down to enumerating the number of potential monitoring events inside this county. This may involve, and probably will involved doing it maybe several more stages of sampling, but in some sense we'd get down in some sort of organized probabilistic way into a smaller set of potential monitoring units and then randomly select those.

What you'd end up with is a sample of monitoring units when then you would use to conduct your study. These monitoring units, or at least the exposure results from the monitoring units because this probability sampling is unequal, would each have associated sampling rates. In other words there would be some rate such as, that would represent in a sense the relative numbers of individuals in the target population that each of those monitoring events represents. And it might not be, if it was a simple random sample the rates would all be the same so it

falls out and then that's why, I mean, that's what you oftentimes see in your statistics 101 textbooks.

So we, this approach is not the only approach. I've done surveys like this and they're complicated, expensive and there may be simpler ways to do things. But in this, this approach is quite costly as you can imagine and most of the costs are overhead costs. Going from the county/periods, the first stage is very simple, it's not usually, it doesn't take very much effort but the real costs come in the lower stages within the counties in this example here. And so we've got a lot of costs.

Now for larger surveys like the ones in which I've been involved with we have like 1,500, 2,000 sampling units that you end up with eventually. It's worthwhile for small precision. I think we're talking, and hopefully we're talking about sample sizes a little bit less than that per scenario in which case the overhead cost might really swamp out the monitoring costs so it might be quite expensive. And what that, when those kind of costs come that means that something has to give and usually it means that there's fewer scenarios or fewer replications per scenario or something.

There's also issues that go along with many

surveys or anything else that, the timing of activities and whether you do it randomly in some sense and such and the selection of workers and travel expenses. When you start selecting things with some degree of randomness it becomes a bit more difficult.

There's delay in monitoring activities and in setting this up, and it would delay the start of additional collecting but these are things that can be worked around given enough time, effort and money.

The AHETF studies, however we do have studies in the database already that were acquired. Some were already, there are some in there that were generated from the methodology we're, I'll be discussing that were collected previously. So whatever is done, it might a little bit like a, not very cost effective to worry too much about a real complicated survey sampling design, say multi stage design and then throw it in with some other data that were collected under different ways. It makes it, it's almost your, you have apples and oranges in here, very expensive apples and a little bit cheaper oranges but you still have, you're mixing and matching there. You might be able, it may be difficult.

So what we're, have currently done, and for the studies that have already been collected and what

we're proposing to continue, barring Doctor Portier's solutions, are what I'm going to call purpose and diversity sampling which is actually something that exists and I just didn't make it up. And in a sense this is a, the monitoring units that we use are actually selected based on judgement and I'd like to state here ultimate availability, I mean sometimes you just can't find something and you take what you get. But ultimately with judgement. And the selection of process, so that's where the purpose of aspect comes from, not selected randomly but purposely. And in addition the purpose is to increase the diversity of some or various factors. And the factors that we're most interested in decreasing the diversity of are amount of AI handled which you've already discussed previously, location and a date, what I'm going to call cluster and we'll get to that in a second, and workers which you've already discussed, was discussed in the last session and it was also, I mean there's some issues there about whether we continue that process or not but that's currently what we're doing.

And then in most other respects were some caveats which I'll describe later, the monitoring of the workers, will perform their tasks pretty much in the normal fashion.

Let me just briefly describe the methodology that's used, at least the target that's used for diversifying the amount of AI handled. First the Task Force determines the, a range of AI, I'm sorry, amount of active ingredient handled under the scenario based on expert opinion and other information they have available. This range is truncated a bit. On the upper end it's my understanding that there is some, it's so difficult to find situations, even though theoretically something that high pounds per active ingredient can be handled it's hard to find something that, an example of that actually being done and so that might tend to truncate the upper end a bit and but we're talking levels of about a ton per day per user. So we're still talking pretty large amounts that we do include in the database.

The, on the lower end oftentimes it has to be restricted, at least for the surrogate compounds that we're using because we want to stay above the limit of detection. So if you get too low then the exposures are just too low to measure, reliably anyway.

So what's left in the middle we're calling the practical range of AI handled and the, well we, basically what we do is try to space out the pounds per AI handled pretty much equally spaced on a logarithmic

scale so I'll call that logarithmic spacing for a short cut. That's pretty much what we do there with the understanding that it's not always possible to get that exact but that's sort of a target.

Now we mentioned the, this issue has been brought up time and again over the last several days so it's probably no surprise to many of you, the, we found, both in our data and analysis and it's very, it seems very common in a lot of data that we will do an exposure monitoring study and you have individuals that are being used that are within the same general area, the same general time and perhaps more importantly, being done in the same study with the same study monitoring personnel, that there tends to be, and even when the pounds per active ingredient handled are exactly the same, you tend to have study effects if you want to call each one of those things a study, you have these study effects or cluster effects that might be different.

These may be due to known factors, equipment was a suggestion. An example of these study effects to me seemed to be very evident in some of the data the Agency presented earlier on on the PHED data. And I think Doctor Popendorf referred to that quite often so that's great.

So this, this kind of, but it's also factors that we really don't know what they're due to. I mean there's so many factors that differ, I mean any two, any two studies you get on locations you're going to find hundreds of variables which differ. Some of those with your subject matter expertise you might be able to say are reasonable candidates for the cause. But there's really no easy way to prove it without doing an experiment, trying it under those conditions and those, that's the need for future, for designed experiments in my opinion. But nevertheless we find these effects, it's if, given, I'm sorry, given that that's the case it seems like what the ideal thing to do would be to diversify or to not sample any more than one person within one location at one time. In other words, one person per study would seem to be a great way to go to reduce this artifact of sampling.

But that's not very cost effective as I think Doctor Portier mentioned earlier in the last session. It's oftentimes cheaper, there's a cost, there's an advantage to once you've found a location there's a lot of effort involved, a lot of overhead into getting a study set up. The additional cost per unit is substantial but it's all, but it's a heck of a lot cheaper than going ou9t and doing another study and

getting one person. So there is a cost savings involved in doing multiple people per study. And I'm sorry I gave Doctor Portier credit for that and I think it was Doctor MacDonald who mentioned that.

In any event our clusters or our monitoring units tend to come in clusters and these clusters typically represent a study. Although in a few cases we have a single GLP study that might have multiple locations spread out over a period of time in which case that study would have multiple locations but in many cases a study and location are synonymous.

And because of this fact we tried to, what we're trying to do is when we sample a scenario is try to have multiple clusters at least per study, we don't want to get studies with only one cluster or two. That just is just is asking for trouble because the cluster effects make life very difficult. And so we always, there's a balance then between number of clusters and number of units per cluster that's going to be involved in the sample size.

And finally, I know this issue's been hammered to death already in the last session, but we are, have been in the past been attempting to not get repeated measures or multiple measurements per worker, certainly not within a small space of time or space

because we feel that that information is just not as, we're being really redundant because there, it just may be correlation between those values so we have been attempting, our goal had been taking only one individual once. I think there was an allusion, someone questioned in the protocol that is there a situation where we would use another person, the same person again if somebody dropped out of the study. And I think if I recall correctly that happened once maybe in the entire study, although I could be wrong about that.

And in addition I think that the requirement, if that ever had happened, is that that person, if they are used again could not be used for a certain period of time. They had to be spaced out in terms of time a little bit. I mean it be several days to try to minimize their affect. But never, so the bottom line is we're not really interested in individuals being measured close together in time.

So in general the other, the other monitoring conditions are pretty much left up to the worker and the attempt in here is to improve representativeness as much as you can with a non-random design.

And what goes along with that, that second bullet is pretty much saying, it follows along, it

permits the natural correlation among the factors to sort of come out, with the exception of pounds per AI handled, it's occurrence is going to be, is obviously mandated but it's, but the correlation, anything that happens to go along with it from a causal standpoint will be, will still go along.

So if large pounds per AI are associated with longer times, then longer times will tend to be spaced out in such a way as the pounds per AI handled is.

But there are some and this is described in Appendix B of the technical, there are situations where some, what I call minor onsite diversification of other miscellaneous conditions occur at the cluster level. This is pretty much an informal attempt, not so much formal as the other approaches were, but more of an informal attempt to diversify. And you get an example of maybe some equipment types, crops rates, the dates of monitoring by location, things that sort of fall within the purview of what would normally be used but you have some sort of choices in the matter. So you try to say, well, instead of these two people doing the same thing they'll do something a little differently.

And the reason for this was to take what looked to be a cluster affect that you see here in A, B and C and spread things out just a little bit, force

them to be a little bit more diverse. Again there's a difference between diversity and representativeness as we freely admit here. We're trying to increase the diversity, that does not necessarily mean it's more representative because I think I talked to some folks earlier that one could sample, get a sample of every tie that's sold here in Washington, D.C., you get one representative and you've got great diversity but that doesn't represent the number of ties that are in use. So it's really important to distinguish those two concepts.

But diversity was seen as ours, as the second, the second best thing from a random sample. It was the next choice that we could have here.

As all the statisticians here know, and as lot of people forget, that the, that when you do non-random sampling that means that generalizing to the target population becomes, is not a statistical exercise anymore. It's a matter of, to be nice you say subject matter expertise, you could also, some people would say maybe faith is probably a better term to use in that case. But nevertheless it's something that you can't really, technically use statistics to do without making further assumptions. And one of the assumptions that is very common, maybe more common than it ought to

be, but it would be very common, is that this, the purpose of this non-random sample can be approximated, or can be approximated by a, some kind of a probability sample from the target population. And if you make that assumption, then of course you can go ahead and use your statistics to make issues.

But you always need to keep in mind however that even though this is common, that when you do that you're running the risk that one runs is that the probability statements that you make at best are only approximate. And it's very difficult to judge that. Maybe it's a function of subject matter expertise.

So going with non-random sampling isn't the best choice but sometimes it's the only choice left available to you.

Data adequacy and necessary sample sizes. Assuming, we're going to assume that there is some sort of a surrogate model that applies here just so we can proceed forward. The AHETF program, in order to do sample size there has to be some goals. I mean the goal really shouldn't be just to collect data, to develop data or whatever the terminology happens to be. And then likewise to my knowledge there's no universal optimal sample size that one could use, you know, be that number 20, 30, 40, 50 or whatever. It, you really

need some, to establish some sort of purpose, some goal and then maybe it's only approximate, maybe you're, maybe it's based on assumptions but nevertheless you then use that goal to establish some sort of sample size that will address that goal.

And so we're going to do that. We, we're doing this by setting what we're going to call these benchmark objectives and benchmark objectives are basically saying that these are objectives that we want the data to accomplish for sure. If it doesn't it doesn't mean that users of this data will have some assurance that those, that the data used for those purposes will meet certain objectives. It does not mean that the users could only use the data for those purposes, nor does it mean that the data are inadequate for the other purposes. It just means that we cannot specify as in a sense developers of a database providing data, that every possible conceivable use of the data by anybody, we can cover. Even if we could we probably couldn't afford the sample sizes that it would take to do that. But nevertheless it's very difficult to do it so we have to settle on something, whether it's the ones we're proposing or whether it's others that the panel comes up with, you know, that's great.

But nevertheless we have to settle on

something so to get things started for this panel and for the project we've proposed several things that we, that get at things that, that get at goals, benchmark goals which members of the Task Force and the Agency seem to thing, certainly the Task Force for sure, seem to feel were important or are objectives which the, which were de facto, in other words, these are things which people were using without actually explicitly stating.

The two benchmark objectives that we're talking about here are that the, and I'm going to express them first in a general sense, that the first, the primary objective is that selected measures of a distribution, you know, some things like means percentiles in sense is a feeling should be 'accurate', and I'm putting that in quotes because accuracy here is really going to depend on the surrogate distribution assumption. But accurate to within K-fold when exposures are normalized by amount of AI handled.

That last phrase might seem as though we're saying that this database is designed to be a database of normalized exposures and that's not the case. This is going to be a database of exposures and many, many factors, several of which are potential normalizing factors. So it's not a database of normalized

exposure, it's a database of exposure and supporting data.

However, we're saying that for the purposes of sample sizes that the distribution of normalized exposure will be, well we're trying to make it accurate to a certain, we're going to have enough clusters and numbers and individuals per cluster to achieve the sample size, to achieve this objective.

It's also true however that it's going to turn out that it doesn't really matter whether we're talking about normalized exposure or some other measure for this objective, as long as the probability model that's used applies to some other measure, the results will apply to that other measure too. So whether it's just pure raw exposures or exposures normalized by time or whatever.

The secondary objective in that suffix is important. It's not second objective, it's secondary objective. And it's in some sense subordinate. And it's only to be met for some scenarios and those scenarios are going to be those for which the amount of AI handled is sufficiently broad so that it's possible to do this, another sense, reasonably. And that is that the users of the data should be able in some sense distinguish between the complete proportionality case

and the complete independence case.

As you recall on the talk earlier today I said that there would be, we were going to as a goal, one of the goals was to provide some limited ability to investigate pounds per AI handled. So in a sense we might be guaranteeing this, not to say that users cannot use it and be able to use it for more complicated analyses, but we have to guarantee something.

And to be useful these objectives really need to be more precise. I'm not, I've only stated them in a broad sense so they're not really very useful right now. We have to give a little, a few little preliminaries here. And one of the preliminaries is that this is the model. Doctor Portier actually may put a version of this up here later on when discussing the variability with an additional variance component, but this is the model we're sort of visualizing as a surrogate for the purposes of sample size.

And what we're saying is that the normalized exposure on a log scale is the log of the geometric mean exposure plus a couple of random affects and those random affects we're assuming are normal with the variances, V sub C, V sub W, which is just a fancy statistical way of saying that they're lognormally

distributed. And with the geometric mean of GM and the geometric standard deviation, GSD, which is defined in the terms of those variances as of here, as the back transformed variance or standard deviation.

And the correlation between monitoring units, monitoring events in the same cluster is described by interclass correlation or ICC, where ICC is defined as the ratio of these two, as the cluster variance divided by the total variance.

You might wonder why I use the term ICC for here and use the term RWW for the within worker correlation. And mainly the reason was because people got it confused when I talked about the interclass correlation or ICC in one case and interclass correlation for the other case and I find it just simpler just to, just to use two different letters because, even though I still confuse people I didn't confuse them as often. And ICC is used for this case because this is the first one done. And it was too late to change it afterwards.

So determining the sample sizes, we've already said what the model was. But in order to do anything with this, to do anything useful we're going to have need some reasonable values for the geometric standard deviation and the interclass correlation. You

don't need in this case you're not going to need the geometric mean because we're going to be talking about relative variability so that's going to fall out. So that's not really important in this case.

It is important when we analyze the data but it's not going to be important for doing sample size.

We don't have values for that you might think, so we examined some of the existing AHETF data that's in the database now. These data consisted of primarily the data that was acquired, purchased data, and with some data then that was done prior to this, the HSRB. And so there's a number data sets involved in that as well.

In most cases these data had as you can, here's the list of the scenarios for which we had data. The one that says air blast is really one scenario, is really sort of two manifestations of the same scenario and one is with headgear and one's without headgear. So that's really the same scenario there. And for the purposes of this exercise I treat as one, because you can see the results are very similar there in terms of variability. Not in terms of the geometric mean by the way. Without headgear it's a much lot higher because of the head exposure.

The number of monitoring units in these

different scenarios that currently exist in the data set I analyzed is give there in the table. And you can see on the second column, those are the number of clusters that are in the data. And you can see that a few closed mixer/loader granule has one cluster and hopper box seed treatment has one maybe. There's a, it's a situation where we had two locations but we're not really quite sure whether that's one or two clusters. It's really, one falls on the, it falls on the dividing line because it's separated spatially a little, but the same individuals were used. Not the monitoring units, the same personnel were being used in both cases. There might be some association there so for right now we're just calling it one.

And in that case the geometric standard deviation we estimate, I'm going to just treat as that's lower down from the estimate. In other words the estimate, if I would, knew the cluster, knew it had more than one cluster it would probably be higher because I'm just measuring intra-cluster correlation now.

Nevertheless these are the estimates of geometric standard deviation. You have this data in your, in the technical document, the second document that was quoted, along with confidence intervals and

you also have the data for not just dermal but for inhalation. So you've got a more complete data set in the document. This is just shown up here for illustrative purposes.

And you also have the interclass correlation that was measured here so I used that model, that variance component model that you saw a few slides back on these data. And these are the results that we get.

One thing I wanted to point out, and it should be very clear form the data you have in your, in the report with a, from the table where the confidence intervals are given, that the confidence intervals on geometric standard deviation are certainly large. But the confidence intervals on the interclass correlation are ungodly large.

In many, for example I'm not sure of the exact number but as I recall, I don't have it right in front of me, but as I recall these ones in which the interclass correlation is 0, I think the upper bound could be up to around .6, in the vicinity of that. So there's a large variation in interclass correlation. In other words these kinds of measures are hard to estimate, especially when you don't have a large number of clusters. You need, as Dallas Johnson said earlier, correctly so, it's not just the number of monitoring

units, it's the number of clusters that's important as well to getting a good estimate there. And that's true for anything else.

Nevertheless it looks as though these data have quite a range. To give you a feel for these they seem to range from say 5 right here for this one scenario to I think the lowest was 2.5 for a dry flowable mixer/loader, which is one of the better studies. We got more complete data for that scenario than perhaps many of the others. So it ranged from 2.5 to 5.

Well a geometric standard deviation of about 3 is about what, a coefficient variation of about 150, something like that? Or maybe about an order of magnitude variability, no, almost two orders of magnitude variability. A geometric standard deviation of 4 is about a 250 coefficient of variability, a 200% coefficient of variability or about 2.5 orders of magnitude. So in other words 95% of values range within about 2.5 orders of magnitude. That's a pretty substantial variability there.

If I look at the, if I look at the, not saying that every one of those scenarios has the exact same geometric standard deviation in interclass correlation, but to get some sort of typical value to

use let's look at the summary statistics, the mean is about 3.8, medium for geometric mean of maybe 3.3., this is for the geometric standard deviation. And if I do a combined model of all the scenarios, allowing, using, forcing the variation components to be the same, but the geometric means or the means on the log scale to be different, then you get something like 3.8 as summary statistics. So think of that as just a glorified summary.

And the ICCs are a tad, somewhere between 2.5 to, I'm sorry, .25 to 3. So if I were going to pick a number, rounding it off suitably to use for maybe looking at sample sizes, at least as a first pass, it seems like something like 4 and 0.3 might be reasonable. Especially since I think if you look at, you have the table for inhalation which gives summary statistics a tad higher than that, maybe just as far on the upper side, it seems like 4 and 3 might be something that's useable for both. So I picked 4 and 0.3 as something that's, something to use at least as a start to get estimates for, for sample size or investigating these goals.

Now we're ready to look at the goals a little more precisely. I'll narrow this done a bit by saying we know that the number of clusters and monitoring

units per cluster and they should be sufficient so that these estimates of the geometric, arithmetic mean in the 95th percentile from the surrogate logarithm of distribution with specific, you know, yada, yada, yada geometric standard deviation and ICC should be within K-fold of the true values in this 95th percentile.

Now I'm putting a little bit more precision on this just to give us something that we can work with.

The reason I'm picking the geometric, the arithmetic mean and 95th percentile is mainly because the geometric mean is easy to measure and there's actually an analytical solution for that. I don't have to resort to simulation to work with that. Arithmetic mean and 95th percentiles are very, are statistics in the distribution that are quite of interest in the regulatory arena. So those seem to be reasonable statistics to focus on.

K-fold, the, it's not going to take rocket scientist to figure out that the more stringent requirements you put on K, the larger the sample size is going to be and the more relaxed it is, the lower the sample size is going to be. And it's very difficult to figure out, to ask the question, what kind of error or relative variability then to tolerate on

the results? And it's very difficult to get that.

Through numerous with people a tthe AHETF and getting a feel by saying things like, well, if you got a result like this, how far off would it have to be before you'd say things are useless? So using questions and answers like we sort of honed in on something that 3 at least is a start. It might be something that seems kind of reasonable.

Much more than that we're starting to say, well, you know, I'm not so sure that it's giving us very useful information and you know what's going to happen is much more precise, much more lower than that.

So as an example, not to say this is hard, this isn't a hard and fast rule and it certainly is not going to necessarily be true for every scenario, but just as, for an example so that we can get something to work with, let's just work with 3 and that's an example I used in the document. And I want to just emphasize that that should be treated as an illustration, not as the agent, not as the Task Force's recommendation that that's what should be used. This is a, we're trying to come up with something, a process as opposed to necessarily a specific number. And this is what we saw as an approach to address the process.

Nevertheless, 3 is what we'll use as an

example.

You can go through and do, and I have described the simulations and I think actually the panel has received the SASS programs that were used, either ours or the Agency's, I'm not sure which version you got. But the Agency did a validation in a sense or let's say, I mean validation is too strong a word, they tried to reproduce the results themselves and so I'm assuming they sent mine, but if they did you call tell because I'll have my name on it somewhere.

In any event, you can use the simulations and as I said the, in some cases the geometric, for the geometric mean you don't need simulation to address the result. But it was just as easy to include it for generalizing, for generalizability.

But nevertheless what we do is we generate, we take the number of clusters, say 5 in this example and the number of monitoring units per cluster, in this case it's 5 as well. We look at the geometric standard deviation and the interclass correlation, generate data that correspond to that model, then calculate the statistics of interest. And there are many different ways of calculating them but in the, and I think I presented at least one way here, calculate an estimate of geometric mean, the arithmetic mean, 95th

percentile, look to see how far off they are form the values you started with and from the lognormal distribution for the 'true values'.

And look at that in terms of a ration and that's what we call the, and the 95th percentile, I'm sorry, the 95% bounds, the 97.5% and the 2.5% whichever bound is bigger, that's basically what we're using for this and looking at the relative variability. So in other words what we're seeing is an example of the geometric mean is that it says 2.2., that should be interpreted as saying the error in the geometric mean is going to be less than or equal to 2.2 with a probability of 95% apart from simulation error. And I think in this case I used 10,000 replicate simulations in this instance.

And these are the results we get and what I'm showing you is the result of several different simulations. I'm picking the one here that gave a result that was, that seemed to be close to a threefold, it seemed to be that those boundaries at least were within threefold. And it's always going to be the case that the higher percentiles are going to be more uncertain than the lower percentiles, I'm sorry, that the percentile is more towards the center.

This is a lognormal distribution so the

arithmetic mean is often around in this case the 70th or 80th percentile anyway. So an arithmetic mean in a lognormal distribution is up in the higher percentile so it's going to also tend to have, and that depends on how the, what the geometric standard deviation is. But it's going to be very close in terms of the geometric mean and the precision will be higher as well.

So these last two parameters, 95th percentile and arithmetic mean are driving the results. If you go to the 99th percentile or 99.999 or whatever you want to go, the error is even bigger.

So this is an example. We, I looked at many different examples, you've got those example I think, both from sort of a sensitivity analysis approach and also looking at some of their values in the document. I don't think it's in the document but just if you're curious, with this scheme if you wanted to go down to say 2, for twofold, instead of having numbers like, and let's just say 5 monitoring units per cluster just to stabilize it, you'd be getting about, you'd have to have 12 clusters with 5 per cluster so that would be in, that would be like 60 monitoring units.

So you'd more than double the sample size to go from 3 to 2. To go to, from, to get down to say 1.5, a relative error of 1.5, it's like saying the

estimates up there are plus or minus 50%. I think as I recall, I'm thinking it's around 200 total observations. So whatever that would be 40, I think 40 clusters times 5. Is that right? Yeah, or something like that, 40 or 50, it's very, it's close to 150 or 200 so it's a, the sample size, it's not surprising, the sample size is, as Doctor Heeringa mentioned earlier when talking about the interclass correlations, the sample sizes are going to start going up like the, go up by a factor of 4 every time something else goes by a factor of 2. It just, it tends to get worse and worse as you get tighter and tighter.

You're never going to get anywhere near the kind of precision or accuracy you might expect of studies, like plus or minus 10%. I mean that, if you're wishing to have errors plus or minus 10% or 20% or 30% I think it's wishful thinking. Not with this kind of variability.

It may be that if you, as someone suggested, if you nail the scenarios down or some scenarios down to real, real narrow and focus on scenarios like that, that if a lot of the variability tends to decrease, and I expect that there would be some, then you might gain this. But just keep in mind that the, that even for a scenario or even within a study oftentimes we've seen,

I think it was the, I can't think of which one it was now, the study that's on the very bottom, let me skip back here, there we go, open pour mixer/loader, I think that the, that there was one study in which the individuals were applying separately, or not applying, this is a mixer/loading, were doing it separately, they were scattered out over several days but it was all in the same study so you've got a cluster. Within that one cluster there were three orders of magnitude variability.

So we're talking about, so if I narrowed that scenario down to that single, that single cluster, that one location, that one, I think it would have had to France, I'm not really sure about that but it was single, it was a purchased study, a single, a single cluster that was done over the course of say two or three weeks, you had three orders of magnitude variability in terms of the exposure.

So maybe the, maybe that's an exception to the rule but it seems like it might be not as obvious that cutting down the scenarios is necessarily going to cut down the variability always.

Did I do something wrong here? Oh, that's great, that was a feature I think, not a bug.

The, anyway overall the scenarios seem to me

anyway that the accuracy, and you'll see this, I'm not going over it but the accuracy is, the information is in your document, that the accuracy seemed to be influenced more by the number of clusters than the monitoring units per cluster. That's not, it should be not a surprise, it's the same issue you saw with the interclass correlation for within worker and between worker. If there's a correlation within a cluster then it's to your benefit to take more clusters, not waste your effort to take more individuals per cluster.

The only reason, and somebody over here mentioned it earlier, that it's cost considerations oftentimes while you do cluster sampling. Cluster sampling, it's sort of like they say democracy is the worst form of government except for all others. Cluster sampling sometimes is bad but it may be the only choice you have in some cases.

Give a large variation exposure this geometric standard deviation of 4 or 3 or even 5, there are pretty much practical limits to how much the accuracy is going to be increased. I mean like I said before, going down to 10% is wishful thinking.

In this example we gave here a threefold accuracy could be achieved with 5 clusters and 5 monitoring units per cluster as long as the surrogate

model holds. With that caveat there.

And the result is not exceptionally sensitive to small changes in the geometric mean in the ICC unless they both happen to vary, increase in the same direction or you underestimate them in the same direction.

And just to say again that we as the Task Force are not saying that even if we pick a value like threefold we would necessarily use that for every scenario. So whatever value is picked is not necessarily going to be used in every scenario. It might be that the variation we know for, in some cases the variation in some scenarios is lower then others. Not just due to random variability but we have some evidence other than just that from just sample data, it's something a little more stronger perhaps.

In that case if there's lower, I mean it's not, it doesn't take rocket science to tell you that if the variability is lower you can use a smaller sample size, if the variability is higher you need a larger sample size.

Increased accuracy is desired if that's the case. If you need twofold in some cases and say fourfold in others, an example given might be that if it's a very important and high risk scenario in some

sense you might want more accuracy, especially in the upper end. If it's a scenario where the exposures typically are low it might not be quite as important to have the same level of accuracy. That's really something that's not for a statistician or for me to decide. It's really something that people who use the data or are planning to generate the data need to make the decisions on.

And I guess because, this is just an artifact here, but the purchased data in the scenario, the purchased data available, some of those have very, very large numbers of monitoring units per cluster. And in that case, if you have say 16 monitoring units per cluster or 10, then those don't contribute too much to the cluster, you going to still going to need say, if I say that 5 and 5 are needed I'm still going to have to have pretty much 5 for all the others besides that one. Because you don't really, you don't gain a whole lot of information by that 16, that 16 doesn't give three times as much information as a cluster of 5.

Okay, closing down this, the surrogate, let's look at the secondary objective real quickly. The secondary objective, I can sort of rearrange and redefine the surrogate sampling model in terms of the log exposure is equal to some linear function of the

amount of AI handled and it also has those same two variance components that we talked about before.

And expressed in this form, we've already talked about this many times so it shouldn't be too different from what you've already seen, that if the slope is equal to 1 then exposure is proportional to amount of AI handled. If the slope is equal to 0 then the whole turn with the log AI drops out and you've basically a simple variance component model for exposure and pounds per AI handled is unrelated.

So there are two extremes. Granted, there are many things in between and things that aren't even described by this model. But nevertheless we can reduce this secondary objective to a situation where we're just testing whether the slope is equal to 1 or 0.

So to make this objective more specific I'm going to say that for scenarios with at least a 10x practical range of AI handled, and maybe that should be even bigger than 10x given the variability that we see in normal I think Doctor Popendorf mentioned that earlier that we should be, might be wanting to make that even bigger, it probably needs to be bigger. Even where the data should be adequate and I'm going to say sort of the power to reject the hypothesis that this

slope is 0 and the slope really is 1 is at least 80% and symmetrically you can reverse that in this case under certain assumptions and say that is also the same as saying that you have the same power to test the other way.

In other words you're going to reject the slope being 1 when it really is 0. That clearly does not take all possibly hypotheses you can test but it does at least get at something.

So what you need is a surrogate cluster sampling model, we already mentioned that, that's in the previous slide, you need reasonable values for the residual, geometric standard deviation and the ICC. I think the values of 4 and 0.3 are probably reasonable for this as well.

The power in this case however really depends also on how those amounts of AI handled are allocated among the clusters, both what the levels are and how they are allocated. And that's kind of tricky.

We've already said that we're trying to spread the AI levels out sort of logarithmically and I think if we assume that that's the case then we've addressed the first part. The second part is how are we going to allocate those AI levels, I'm sorry, the amount of AI handled out among the clusters becomes

kind of problematic.

What I've done in this example is present two alternatives. One is which, most of the AI values are in the, the AI values are maximum overlap, that's the bottom example there. So we achieve maximum overlap within each, each individual cluster has a wide, the widest range possible of amount of AI handled and given the spacing. And the other cluster has the, those are just the opposite, it scrunches as many of the AI levels and minimizes or allows no overlap at all.

What you would expect in this situation if there's a cluster affects is that the top one would be worse because now the cluster affects are being confounded with the affects of amount of AI handled.

Whereas in the bottom case you would expect that to be a bit better because now the cluster affects are sort of like blocks in a way and your, it's not contributing to the additional variability of that estimation of the slop.

I'm not going to go over all of the different things that we've examined but just to give an example, you also use simulation to address this and basically this means joining the data under one of those two assumptions and then plugging the results, the simulated results into a mixed model for variance

components, testing for the slope, being from 0 or 1, depending on which way you want to go and look to see what proportion of time that actually is rejected. That's you power.

So for example on the second line down where it says 8x, if the range in the amount of AI handled is eightfold, then in that case of minimum overlap between the two, within the clusters, only 37% of the time would you reject the hypothesis when it's false.

In the case of the maximum overlap you do it 82% of the time. Again the caveat being that there's simulation error, whatever is associated with that which is rather small.

So you can see that the intuitive assumption is true that the, that that, when you have maximum overlap between the clusters you're actually doing better in terms of this association.

You can bump the power up by increasing the pounds of AI handled range. So when you're up around 50x you can see that you're almost at the 80% power. And certainly at 100x. So that works and it's also shown in the simulation that you've got in your report, that you can increase the number of clusters to compensate for this and to some extent even the number of individuals per cluster, but that's probably

inefficient.

But in that second case there the only solution is to use, really to use more clusters because in that case what you don't have is, you don't have, say if you have 20, if you have 25 observations in 5 clusters you don't have 25 observations, you've got sort of 5 fuzzy observations there. And you need more of those fuzzy observations in order to compensate for that.

So there is, that's the kind of issues that you have to work with.

In examining several simulations it seemed to me is that if we're going to assume that things are logarithmically spaced and the amount of AI handled is at or greater than an order of magnitude and there is some strong overlap between the AI levels and the clusters, that it seemed that in this specific example anyway, that when you met the first objective, the second came along pretty much for free.

That may not be true in all cases but it's certainly something that needs to be, would have to be investigated. But in this example that seemed to be what happened.

Okay, only a few more slides and we're home free. You are, I'm not.

The statistical analysis of the data, I just want to say, and this is something that's very easy to be confusing here because it's, I just want to remind you that the Task Force, our purpose isn't to analyze the data really. Our purpose is to generate the data that others are going to analyze. And so we're not really here in a conventional experimental study that we're going to take the data, analyze it, write a report, write a paper or two and get tenure or whatever.

What we're going to do is provide data for others to analyze so, however what we want to do is we want to make sure that others have data that is going to be useful to them but we cannot anticipate every kind of analyses that's done and that's why the generic database is going to provide many different features for people to do things on their own, something that may not even have been anticipated. You can, obviously you're able to query the data, pull out data based on any of these factors that are in the database.

You're going to be able to calculate exposure pretty much normalized or transformed pretty much in many, many different ways. And if the database software itself does not provide the analyses that you need to do, you need something more specialized, all

the data can be easily exported I think as I recall to, probably I think into an Excel format which then can be put into some other format. Any, most any data, any statistical analysis program will handle that.

So users of the database can do pretty much anything they want. Our analysis is pretty much going to be limited to attempting to verify that whatever requirements that we say we meet, we try to make an attempt to verify that level of adequacy. And also to provide some guidance to users regarding proper use of the data. For example, the issues of clusters and things like that to provide as much, it's mostly like documentation from a statistical point of view.

Those, that information is going to be provided in something called a scenario monograph which also will include many, many other pieces of information and I think that's described in the technical document that was provided. Not the one on the sample size but the other one I think in Section 9 or 10 or something like that.

The main focus of our analysis is basically going to be things like confidence intervals for the distribution parameter, the parameter estimates, some post hoc power information, that sort of stuff.

And I wanted to say that if the, if it turns

out that the data are somehow grossly inadequate from what the goals were, the Task Force is certainly going to consider augmenting the data with additional clusters. You almost have to talk about additional clusters because you can't go back into a cluster once you've already done it. A new study is going to be a new cluster.

So certainly there's going to be some guarantee that the, whatever adequacy requirements can be set, no matter how we do it, are either met or determined not to be a problem, one of those two things.

But I guess I should say, and this is where the confusion comes, when you're looking at some of these, and this is what people I think thought we were, our purpose of a database was to generate, to basically provide estimates of normalized exposure. It turns out that in order to get the confidence limits for the parameters of the normalized distribution, we're going to have to calculate the parameters of the normal, estimate the parameters of the normalized distribution and do confidence intervals.

And so as part of testing for the adequacy we're also going to be in effect making that calculation. So we'll be providing some descriptive

information, some estimates of the distribution of normalized exposure just because we need it in order to do other things.

And in addition in doing like I say, a power comparison for the regression, the test of effective amount of AI handled, they almost have to do in order to get some sort of estimates you almost have to do the regression in the process so you end up getting those results anyway. To what form we'll actually put those in the scenario monographs hasn't really been decided yet.

So this is where the confusion comes out, it seems as though people sometimes are mistaken that we, this is what our purpose of the data, our purpose is to somehow collect data so that we can analyze and provide these estimates. Our purpose is to collect data so that others can do things with it but we also want to establish some sort of benchmark.

And I guess in addition as I said we'll provide information like, things like degree of confounding, variance factors, and description of the clustering so that people will know what the clusters were and this sort of information so that users who really are sophisticated can factor these things in.

And so in summary then let me just say that

the, we started off saying that we're talking about our purpose of sample and monitoring units with certain objectives that we're sort of thinking about right now, and proposing actually about using.

Threefold, there's nothing magical abut that, that seems to be something that's been tossed around alot.

It that level is there it sort of translates into 5 clusters and 5 monitoring units per cluster, given the data, the assumptions we used in the sample size estimation.

And just to say again that the AHETF statistical analyses are going to be limited to really focused on evaluating the adequacy, not doing analyses for our own purpose.

Thank you very much.

DR. HEERINGA: Thank you very much Doctor Holden for that presentation. A lot of material again, I think very nicely organized, systematically presented.

I want to turn to the panel now. I think certainly we would like to cover some initial questions of clarifications on this presentation and some of the sample size issues. Doctor MacDonald.

DR. MACDONALD: Yeah, there were two

numbers that you've used that I don't see a justification for yet. Maybe you could explain where they came from. First of all the 5 MUs per cluster and secondly the 95th percentile of exposure is the highest you've consideration and both of those choices have real implications for the sample size.

Can you explain where those two numbers came from please?

DR. HOLDEN: Yeah, the 95th percentile is one where there seems to be an interest in the Task Force. And again these are, I must say these are examples. So there's nothing that says that these have to be, and we're not necessarily, you know, wedded to these numbers. But the justification for the 95th percentile is these happen to be numbers in which, at least that I was informed, were very, very commonly used regulatory estimates from the distribution. These were used in many cases and the 95th percentile and the arithmetic mean, the geometric mean just came along for free so that's just provided.

Granted, I've heard that there are other, I'm sure there are other people who have used like the 99th percentile, like food consumption and this sort of stuff like that. And so those issues, and we knew that the higher one, percentiles you use, the relative

accuracy gets worse and worse and worse. The 99.99999th percentile is a heck of a lot worse than 99th.

And so, and so those, so there's a point at which we stop. We just picked these saying that we're going to guarantee something and those are the two numbers that we say we're going to guarantee.

If it turns out that there is a very great need to have adequacy for a much higher percentile, then we'll use that number.

So, assuming that this process is used, not some other process but if that process is used then a higher percentile will be used as a benchmark and the now the relative accuracy, it may be that given resources and needs, it may be that the relative accuracy required for the 99th percentile might be less stringent than for the 95th let's say or the 90th or the 75th given certain considerations.

Because the sample size implication is such that we aren't going to do, let's say, I'm just making this up, let's say that it's impossible to do the study or cut back on a number of scenarios, you know, there's cost considerations, if required that the 99th percentile be accurate within 50%. In which case something's going to have to give, it's going to be the

number of scenarios, the number of, how scenario is defined, the number of replicates or something. Or it's going to be, what's going to give is the precision requirements.

But what we do want to ensure, I think it's really important that whatever requirement is establish and however it's done, we don't want to say that we're going to generate something that's going to have and objective that when push comes to shove people say, I'm really not happy with.

Because if it, you know, we don't want to just to do just because we can do it. It doesn't seem to make a lot of sense to me at least. So, so whatever happens, so that's the 95th percentile.

I hope that's an answer. Maybe not the answer you wanted but it's an answer.

The other question was the

DR. MACDONALD: The 5 per cluster.

DR. HOLDEN: Oh, 5 per cluster. I did look, and there are other choices and I think what's, what happens is that that again was examples and I didn't say, I didn't mean to imply that 5 per cluster oh, no, a target of around 5 per cluster appeared to meet the benchmark requirements.

I'm not saying that there aren't other

configurations that are going to also meet the benchmark requirements. In fact, using, I think an example I gave you could use fewer, fewer numbers per clusters, fewer monitoring units per cluster with a large number of clusters and actually come out with actually a smaller total sample size.

What's happens is that there's cost implications of those and before a decision, a final decision will be made I think that might have to, a balance might have to be established between the cost. Now the cost, the cost of, given the same number of observations, putting more into more clusters is going to give you, changing the balance to more clusters gives you a better precision, at least in this case. But also it might be more costly.

And so there's a balance, I think, you're the one who mentioned, earlier the clusters, the effect of doing cluster sampling? I think it was, wasn't it?

DR. MACDONALD: Yeah, yes and that

DR. HOLDEN: Right and this is the issue that came up

DR. MACDONALD: that's a fairly

DR. HOLDEN: in the clustering sampling.

DR. MACDONALD: there's fairly we known formulas for working through that.

DR. HOLDEN: Yes, that's right. That's right. Well, there are cost allocation formulas for doing that that can be used. And if those formulas don't work it may be that the specific, and I'm really a fan of not using a conventional sample size formula.

If the formula that you use is based on some end point that you're not interested in I'd much rather use simulation say for end points you are interested in, than an exact or quick calculation for one you're not.

But nevertheless the principle is still the same, the allocation of cost principles are the same. You're right, absolutely right.

DR. HEERINGA: Doctor Holden, someone raised the question earlier and I'm interested too.

I know you can't put specific or exact costs on the different components of these data collection activities, but in general, and I think the big question is, you know, is the cost in getting the data collection teams to the location or is the cost in the assays of the actual dosimeters?

What is sort of the relative order of magnitude? I mean is the cost in the analytic component of that? Is it five times the cost of flying somebody to the Central Valley to oversee this

administration and pick up these and bag these clothes?

DR. CANEZ: This is Victor Canez. I'd say it varies, but among all the monitoring units we've done we look at 5 monitoring units in one cluster as kind of the economic threshold that we want to go out there.

And it turns out I'd say over all the monitoring units we've done, it averages to about $18,000 per monitoring unit. And those are divided up into, obviously if we do more monitoring units at one location it comes up cheaper.

The analytical costs are pretty much the same but there is some time to set up to go there, find the workers, find the sites, do all that up front coordination. And then there is some time and cost involved in camping out there. If you have rain you're out there longer. And so depending on that period that could impact it also.

And so that's pretty much how they're divided up into those three areas, the initial cost, the study conduct and the analytical costs.

DR. HEERINGA: And roughly what ratio, did you have any sense of what those, say if you had to go is it

DR. CANEZ: I'd say the analytical costs

are a little bit under 50%.

DR. HEERINGA: Very good. And the others we can, that's really the main element. And I think your own sense too that, you know, your experience, you've sort of done this cost optimization in your head if not on paper.

The one question that I have is that if you design for studies, or your recommended protocol is for 5 measurement units per cluster, does that limit the type of applications that you might consider?

Let's say somebody has 40 acres of orchards and that's what they have and that might well be a good half day or full day's work for one person, or it might be two half days for two people, but it certainly wouldn't , you wouldn't want to necessarily split it up.

Do you have to look for large scale operations in order to make this work?

DR. CANEZ: It depends a lot on the scenario we're looking at. And with some scenarios where we're looking at let's say orchard, for the example you used we're looking for open cab or orchard applications, those are generally, in some areas they're smaller farms and so you might have only 40 acres. In cases like that we're going to do, we're

going to go to one, go to a place and on one day we may use that all 40 acres. The next day we may pack up and go to another location where there's another farmer with another 40 acres and those sorts of things.

If we're lucky we find two farmers in one place and each of them have 40 acres.

So it just varies on the situation, the equipment, the number of acreage we need, depending on the application rate. So it just, it's one of those coordination efforts that has to be done before we get to the place.

DR. HEERINGA: Yeah, but getting back to that, the reason I raised that is I was getting back to Doctor Landers' concern of, you know, small operators, association with equipment types, et cetera.

The reasons I asked was to think, you know, is there any sort of prejudice in this choice against going to smaller operations?

And what you're saying is that if you looked at say small here, we're not looking at Ma and Pa orchards but, you know, say 40 acres that you might find multiple locations in the same physical vicinity to distribute your observations in.

DR. CANEZ: Yes, you're right. In some places there's corporate farms that have many acres and

they have many 40 acre blocks.

When you start looking at different types of equipment, closed cab ground boom application, those are usually large operations, these guys move through the fields and you'll have farms that'll, that have all the acreage you need.

But if you go to an open cab ground boom application you're going to be mostly looking at small farmers.

So it depends on the operation. But you will find some small farmers with closed cabs so, you know, you can fill those gaps in also.

DR. HEERINGA: But it sounds like the scale bias that I was concerned about is really pretty much taken into account when you break out the different scenarios?

DR. CANEZ: Yes, yes.

DR. HEERINGA: Thank you. Members of the panel, Doctor Johnson.

DR. JOHNSON: Thank you. Can we put slide 21 up there for a second?

The one question I had on this particular slide is that in the monitoring units there that are within each cluster, are those unique individuals or are there some individuals that have been measured more

than once?

DR. HOLDEN: It's a good question. For the most part the, there's the combination of studies there. There's studies that were acquired previously and then there's studies that were, new studies that the Task Force has augmented to that, that were done before the HSRB.

And the ones the Task Force did, every one is an individual study. The ones that were acquired, there are a few studies and I, the only one I know for sure, the aerial one, I know that some of those in the aerial application, and Victor probably knows this much better than I do, I'll let him comment in detail, but i know those were repeat, some of those were repeated, the same individual repeated.

Although I think almost all always, not always, not almost always, in many case there were like separate days but I think there is actually was one case inthe aerial where they were two per day with the same individual, he did two per day.

And so that's, but that's something that we're avoiding in our, in the data we are collecting. That's something we don't want at all.

DR. JOHNSON: Right. The second question on that same table, you'd think after as many years as

I've been involved with analyzing data, I have to admit I never looked at much in terms of where I had to look at geometric means. I always made things not so messy that I couldn't call it normal.

And so if I had lognormal data I'd generally log it and then think of it in terms of normal stuff anyway and getting my confidence intervals and everything and then I'd just exponentiate the end points to get back to the raw units on which the data were analyzed.

So I'm having, what I want to ask you I guess is if I think about this then, when I look at this geometric standard deviation, mostly I'm thinking about when I look at a regular standard deviation I sort of think of a mean plus or minus so many standard deviations.

So here when I look at a geometric mean and I don't really want to add and subtract the geometric standard deviation, the I want to divide and multiply by it, is that right?

DR. HOLDEN: Yeah, that's right. I think I'm going to have to run it myself but the geometric standard deviation, you could think of it as, oftentimes when people use the geometric mean, when you multiply it times the geometric standard deviation, and

that gives you, the sort of equivalent of one standard geometric deviation above the mean, if you do it with the log scale.

Or then you divide by it and you get, so if you take those two numbers, the geometric mean times the geometric standard deviation and take the log of it, and you take the, or you take the geometric standard deviation, I'm sorry, the geometric mean divided by the standard, geometric standard deviation, take a log of that, that would be equivalent on the log scale to the mean plus or minus one standard deviation.

And likewise I think if you, what people oftentimes do is take the geometric standard deviation and take it to the, to some normal quantile and that gives you a sort of confidence balance on it.

But yet you put it in the exponent I think as I, if I recall correctly. Is that right?

DR. PORTIER: Just take a log, do the standard plus or minus and then re-exponentiate it and you've got the

DR. HOLDEN: Yeah, that's what I do too, I was telling you that other people do it. Yeah, I'm no fool.

DR. JOHNSON: That's what I say, I'm an old dog.

But then the other thing I was trying to do is then try to figure out what the meaning of this K-fold measure of accuracy is.

And so the K-fold would be like if I had a geometric standard deviation of 4 then a K-fold would be, well, if I take the log of the, of 3, I get 1.1 approximately so a K-fold, I guess a K-fold measure of accuracy would be equivalent to being within 1.1 units on the log scale. Is that right?

DR. HOLDEN: I'm not

DR. PORTIER: Sorry.

DR. HOLDEN: That's fine.

DR. PORTIER: A 3 is pretty tight, right, on a log scale? That's a

DR. HOLDEN: Yeah, so it's about a .7 standard deviation.

DR. PORTIER: I have a question.

DR. HEERINGA: Okay, Doctor Portier.

DR. PORTIER: I was looking at the same graph, you know. One of the points that I see here is that your GSDs and ICCs really are estimates, right? Based on the data.

And so you don't have it on this chart but on Table 2 in the document that we have you have upper and, you have confidence intervals, upper and lower

bound

DR. HOLDEN: That's correct.

DR. PORTIER: and, you know, normal risk assessors would say, well, I'm not going to base my sample size on this estimate, I'm going to use the upper 95% CI to base my sample size on because that's my more conservative mean concentration level.

And I was trying to figure out for the ICC whether I want the lower bound or the upper bound, but I think I want the upper bound to be more conservative.

In which case your 5.5 becomes something more like a 7 or a 10.3, right?

DR. HOLDEN: Oh yeah, absolutely, absolutely. Yeah, if you take the worst case, basically saying the data I have, what's the most, what's the highest it possibly could be, you know, you could think that way.

And basically if I design my study around that, then I can guarantee you that no study would get done in that case.

DR. PORTIER: I'm just trying to get back to the

DR. HOLDEN: Yeah.

DR. PORTIER: 30.

DR. HOLDEN: Yeah.

DR. PORTIER: That's we had two days ago, right? So I'm saying

DR. HOLDEN: Yeah, that's right, that's right.

DR. PORTIER: 30. So where did that come from?

DR. HOLDEN: It may turn out to be 30 but I think there ought to be a reason for it.

DR. PORTIER: Yeah.

DR. HEERINGA: Doctor Johnson and then Cynthia Hines.

DR. JOHNSON: When, I'm still not sure that I've got it stated, that I've seen it stated anymore, exactly what the sample size determination is going to be, that we should review and comment on.

And also with respect to the picture on page 30, you've indicated the affects of choosing the spread of amount ingredient handled under the two different scenarios, one where they're spread out within each cluster and one where they're spread out between clusters.

And have, does the Task Force have an indication of what they're planning to do with respect to those? Both the sample size issue and the spreading of active ingredient within each of the scenarios.

DR. HOLDEN: Yeah, what may not be clear, I think we're, in this document we were trying to propose a, like I say, I'd say a process.

And I think it's important to us that there be some sort of goal and a goal that is of interest. And from that point on whatever is reasonable, you know, we're okay with.

I think that we needed to propose something and we've, I've proposed something that, with checking with the Task Force, that was something that seemed to be reasonable to them and seemed to be something, I'm not saying they can live with, but something that was getting at the goal that they were interested in.

That's not saying there are better goals and likewise the issue of the, you mentioned about the cluster configuration being important, what we do in that case, in reality I think I might have said that in the document. But in reality we're never going to have one of those configurations or the other.

The studies done in the past tended to be more like this first configuration. Let me put that on that I forgot what it was but there it is, whoops, there it is.

Yeah, studies done in the past tended to be like the first configuration with maybe these pounds

per AI handled more closely together, not spread out.

Studies that the AHETF has done, at least some of them so far are more like this.

So, but there are many situations where for many different reasons they had to, they fall, the fall somewhere in between. And so I sort of gave two extremes and looked at the power there.

So whatever is done, again, this is going to depend on Doctor Portier's solution to all the errors of the design. But given that it's something analogous then configurations like this will probably have to be investigated and we would then gear our sampling to try to pick the optimal configuration within bounds, within the restrictions of the logistics of course.

DR. HEERINGA: Yes, Mr. Lunchick.

MR. LUNCHICK: I just wanted to add also, what we're really looking for from this panel is, are the processes that have been proposed by the statisticians to help us in determining the sample size proper?

As we get into looking at different scenarios we'll be sitting with the regulatory agencies and knowing the magnitude, or having a good understanding of the magnitude of exposure, what patterns are very critical in agriculture, others which are kind of

unusual, things like that that will go into a consensus on the accuracy that we would need and things like that to determine the actual sample size and the clustering.

But we wanted to make sure as we do that that whatever process we're using, people agree with and not at the end of the process tell us, you shouldn't have done that.

DR. HEERINGA: I think that the panel will certainly be in a position to do that.

I think Doctor Johnson's question really gets at how much error is tolerable? Because, I mean that's where we as statisticians start and work backwards. And I think you've gone through that, Doctor Holden, with the Task Force.

And I think where you're at looks, I mean it's very reasonable in some respects and we'll hear more details from other people.

But when we factor this all into the regulatory process where it ultimately goes, you are going to want to have some level of accuracy or precision in these data for generalized purposes. That really makes the data not the determinant in whether you get registered, I mean the data itself, but not the error in the data, determinant to whether you get registered or not on a product.

So we've got MOEs of 100, two orders of magnitude. Would a third of an order of magnitude be acceptable level of error for Doctor Johnson to work with?

I mean that means that your errors in estimation are a third of the, say one, you know, one component in the total margin of exposure.

Are those, that sort of thinking, that's the way I think about it naively but is that the way where you're really sort of headed when you

DR. HOLDEN: Yeah, that's the idea that we were headed with that.

I can't comment on the, like MOEs and things like that but when I discussed with the Task Force these are the way, these are exactly the, what you're saying is exactly the approaches that they were thinking.

And that's where I came up with the idea of the fold accuracy with leaving the, that specific number up to the people who really know better than me.

DR. HEERINGA: All right. I wonder if I could ask Jeff Dawson or Jeff Evans too, maybe to comment on this issue, or someone who is familiar with that process.

I mean as you think about it what we're

trying to do here is to put this sample design in a perspective where the data should not be, at least errors in the data should not be the determinant of the decision, that we can see through the errors to see enough data to make decisions incorporating your other uncertainty factors.

MR. DAWSON: Jeff Dawson. I think we'll probably need to clarify this a little bit more because this is a fairly, a question with many levels and it really, to focus everybody on this issue I'd like to remind everyone that FIFRA related to occupational risk assessments and the risk management decisions that we make, is really a risk/benefit statute.

So essentially the answer can vary in a practical sense, depending upon, you know, the situations that we're considering. And certainly we consider these in the risk context relative to let's say the severity of the toxic effect, the relative steepness of the dose response issues that are associated with the chemical.

And also as specified under the statute the, you know, the associated benefits.

As far as, so those are the factors that we consider when we make decisions relative to what MOE is going to be the number that we can live with or we

can't live for risk management.

As far as the relative uncertainties or the error that we're talking about now, I think we need to think about it some and, you know, maybe talk a little bit more about it in the morning.

DR. HEERINGA: Again I didn't want to necessarily suggest that there is one number.

MR. DAWSON: Right.

DR. HEERINGA: But I think some notion of scale as to how, because when, if you go one to probabilistic work later on or if you work deterministically with the model that we've seen today you will want to put some uncertainty bounds on that.

And if those uncertainty bounds are excessively large due to variability and uncertainty in the actual input data, and we've seen that a lot over the years, I think here's a chance to sort of say, you know, we want to be practical, we want to be cost effective, but we also want to make sure that what we deliver has essentially a small enough uncertainty in it that it's not going to ultimately be the determinant as to whether we can or cannot make a risk decision or a registration decision on a project, a product.

MR. DAWSON: Uh-huh. David Miller wants to answer.

MR. MILLER: I just, in terms of, we, when we first looked at this we didn't think the goals established on that were unreasonable.

But we did note that there are, and it's, I mean the percentile, for example the K-fold. I mean we can, we may alter those later, it's essentially kind of, as much a risk management decision as anything else.

So I think what we're asking more of the panel is to the extent that, in terms of the methodology being used by the Task Force in order to determine the number of clusters, et cetera.

We can decide at some later point we want a higher percentile or we want a different K-fold and that should be relatively easy I think just to go back and kind of recalculate some of that information.

And that's the first aspect and the second aspect, just keep in mind kind of where we're at now, which in essence looking at a central tendency, 50th percentile geometric mean type thing.

So just kind of, it looks a lot better than where we're at now and just in terms of specifics, in terms of the K and the percentile that you're trying to cover with a certain degree of confidence, that seems to be, I mean something we can decide at the Agency in

terms of any discussion with risk managers.

DR. HEERINGA: One other thing just to set us up for tomorrow too. Doctor MacDonald pointed out that it will make a difference at which quantile of this distribution, the median or the central tendency of 1, but if you were to jump out to the 95th percentile

MR. MILLER: Uh-huh.

DR. HEERINGA: it could affect our response on this.

How, I don't want to put you on the spot but in the larger context of thinking about this database and its utility for five or ten years

MR. MILLER: Uh-huh.

DR. HEERINGA: is that something that should be consider on an equal basis with the historical use of the

MR. MILLER: For example up to the 95th

DR. HEERINGA: Yeah.

MR. MILLER: for example the 95th percentile?

DR. HEERINGA: Yes.

MR. MILLER: I'm thinking, just thinking into the future that would not be an unreasonable thing for us to consider.

DR. HEERINGA: And that's just in terms of planning for this data set and what it ultimately might be used to look at.

MR. MILLER: Yes.

DR. HEERINGA: Okay. Doctor Popendorf.

DR. POPENDORF: I guess I'm not quite sure what your question was to the degree that, I mean, like Table 4 in the report or somewhere here they have looked at the 95, 95th percentile.

Are you just saying, should we look at that or

DR. HEERINGA: I was trying to get at some sense of whether that's there for demonstration or really should we be looking at the results that have been, we'd be primarily focusing on results that look at, at means and say, medians as opposed to something that is out a little further in the distribution?

And I think the answer is, we should look at all of them equally in terms of our comments tomorrow.

Cynthia Hines.

DR. HINES: Just a clarification. I understand that our primary objective at this point is to look at the process. And this might be more directed at the Agency.

In their document, Table 5.2 on 102, page

102, there are proposed AHETF sample sizes, so I mean is that kind of just a, is that really in the works or are we really further up the stream than I'm confused.

MR. MILLER: You're talking about the Agency's document on that.

I think there's a sentence in there that says that these are essentially kind of a first cut at looking at things. And depending on what the panel kind of discusses and the conclusions and the thoughts, that obviously could be changed depending on where the Agency wants to go.

But that's a kind of a first cut at looking at things.

DR. HEERINGA: Yes.

DR. COLLIER: I'd like to respond to it

DR. HEERINGA: Doctor Collier.

DR. COLLIER: from the perspective of the Task Force, Richard Collier with the AHETF.

That information was provided by the Task Force to EPA. It was preliminary information, information we used initially in our scoping process and in setting an overall budget.

However the process for determining the sample size for individual scenarios that's been

described to you this afternoon is the process that we expect to use going forward.

So that table had many early uses but shouldn't be considered what we consider a roadmap at this point.

DR. HEERINGA: I think that we're going to plan to reconvene tomorrow morning for a continuation and a discussion of this charge.

Before we break for the afternoon though, is there any member of the panel that really has sort of a different twist on this, something that you'd like to sort of lay out there quickly before the evening so that people would have a chance to think about it and potentially prepare to react to it?

I guess what I, you know, if anybody is expecting to throw a curve ball, throw it now.

DR. LEIGHTON: I've got one.

DR. HEERINGA: Okay, introduce yourself.

DR. LEIGHTON: I'm Tim Leighton from the Antimicrobial Division.

The document that we provided you also has a sampling plan for the Antimicrobial Task Force. That showed I think 19 studies, 15 replicates each. That 15 replicates was based on our guidelines for the minimum

recommendation.

So any comments along those lines would be helpful for us also.

DR. HEERINGA: Thank you very much, Doctor Leighton. Doctor Popendorf.

DR. POPENDORF: Yeah I'm, not exactly a curve I guess but I was going to ask the question yesterday, I think that was yesterday, to the person from the Aerial Applicators Association.

And it kind of goes back to question 4 I think and also maybe applies to this, that maybe you could, I don't if tonight, today or tomorrow, but the statement was made that in that study there was no correlation with active ingredient.

And I was wondering two things. One, were there any correlations with anything else and, you know, how, that would seem like it might be a good example to talk about, both in terms of the implications of study design here in this clustering, it's like a case study?

And also going back to question 4 for an example of something that doesn't correlate with active ingredient handled.

DR. HEERINGA: I'm not sure that the AHETF is directly involved there, that was

MR. LUNCHICK: Yes we are.

DR. HEERINGA: Andrew Moore I think from the

MR. LUNCHICK: Yeah but I could, this is Curt Lunchick from the AHETF.

That is a study we cooperated with the USDA APHIS on, and we have not completed the analysis of the data and at this point haven't made recommendations, interpreted everything it means, so I can't really answer your question except for those are things that we will be looking at too.

I mean there's other issues with the aerial data, the ULV applications versus normal spray volumes that we need to address and that will be done in conjunction again with EPA, DPR and PMRA to make sense of it and make recommendations on whether we have sufficient data or if we need to obtain more.

DR. HEERINGA: Thank you very much Mr. Lunchick.

Any other comments or inputs now, things that you'd like to put out there that the, all of us may want to think about some more this evening in preparation for tomorrow?

DR. JOHNSON: Steve, this is kind of minor but I wonder if it would be possible to get electronic

versions of the PowerPoint slides that have been presented?

DR. HEERINGA: Is that possible to provide those? I think they would be part of the docket, right?

MS. CHRISTIAN: Yes.

DR. HEERINGA: I, Mr. Lunchick or Doctor Collier, is it possible to get copies of the electronic presentations from the AHETF?

DR. COLLIER: They have been provided for use in the docket. I guess I'm not sure we have the capability to make copies of those overnight though. Perhaps a copy or two might be able to be generated. If you

DR. JOHNSON: I'm not saying

DR. COLLIER: specifically

DR. JOHNSON: yeah, I'm not thinking about overnight, I'm thinking about it as we write up our report and review it and so on.

DR. COLLIER: Oh, certainly.

DR. JOHNSON: It gets terribly, I have a terrible time searching through the documents to find things and it would be much nicer to be able to search through the PowerPoints there.

DR. COLLIER: We will certainly cooperate

with the Agency and

DR. JOHNSON: And so you could just email them or whatever, it would be fine.

DR. HEERINGA: Actually if they have been provided for the docket we'll take care of it from the SAP, so you should not worry. I just want to make sure of permission on that.

DR. COLLIER: Yes.

DR. HEERINGA: Okay then, unless someone else has something else for this afternoon, I'd like to thank everybody for their participation today. It's been a very productive session. I think we've had some very interesting discussions.

And again our plan for tomorrow will be to convene at 8:30. At that time we will pick up on a review of some points from the previous few days and then return to the charge question on sample size.

So we'll see everybody tomorrow morning at 10:30. Have a good evening.

(WHEREUPON, the meeting was adjourned for the day.)

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 11, 2007

N