U.S. ENVIRONMENTAL PROTECTION AGENCY (EPA)

\+ + + + +

FEDERAL INSECTICIDE, FUNGICIDE AND

RODENTICIDE ACT SCIENTIFIC ADVISORY PANEL
(FIFRA SAP)

\+ + + + +

OPEN MEETING TO CONSIDER AND REVIEW DRAFT
FRAMEWORK AND CASE STUDIES ON ATRAZINE,

HUMAN INCIDENTS AND THE AGRICULTURAL
HEALTH STUDY: INCORPORATION OF EPIDEMIOLOGY
AND HUMAN INCIDENT DATA INTO HUMAN
HEALTH RISK ASSESSMENT

\+ + + + +

DOCKET NO. EPA-HQ-OPP-2009-0851

\+ + + + +

WEDNESDAY,
FEBRUARY 3, 2010

\+ + + + +

The Panel convened at 8:30 a.m. in the Lobby Level Conference Center of the U.S. Environmental Protection Agency, located at One Potomac Yard, 2777 Crystal Drive, Arlington, Virginia, Steven G. Heeringa, Chair, presiding.

FIFRA SAP CHAIR PRESENT:
STEVEN G. HEERINGA, Ph.D.

DESIGNATED FEDERAL OFFICIAL PRESENT:

MYRTA R. CHRISTIAN, M.S.

FIFRA SAP MEMBERS PRESENT:
JOHN R. BUCHER, Ph.D., DABT
JANICE E. CHAMBERS, Ph.D., DABT, Fellow ATS
GERALD A. LeBLANC, Ph.D.
CAREY N. POPE, Ph.D.

KENNETH M. PORTIER, Ph.D.

FQPA SCIENCE REVIEW BOARD MEMBERS PRESENT:

JOHN C. BAILAR, III, M.D., Ph.D.

FRANK J. BOVE, Sc.D.

RICHARD GREENWOOD, Ph.D.

ELLEN B. GOLD, Ph.D.

SHELLEY A. HARRIS, Ph.D.

WILLIAM L. HAYTON, Ph.D.

CHENSHENG LU, Ph.D.

BETTE MEEK, Ph.D.

NU-MAY RUBY REED, Ph.D., DABT

JOHN S. REIF, D.V.M., M.Sc.

T-A-B-L-E O-F C-O-N-T-E-N-T-S

P-R-O-C-E-E-D-I-N-G-S

8:30 a.m.

MS. CHRISTIAN: Good morning. My name is Myrta Christian. I am the Designated Federal Official for this FIFRA Scientific Advisory Panel.

I would like to welcome everyone to today's meeting to continue the discussion on the Draft Framework and Case Studies on Atrazine, Human Incidents and the Agricultural Health Study: Incorporation of Epidemiology and Human Incident Data into Human Health Risk Assessment.

Again, I would like to thank the Panel, the presenters and the public for participating in this meeting. Also, I would like to remind one more time to everyone that the documents related to this SAP meeting are in the docket at regulations.gov, and the presentations from yesterday will be available in a few days.

I look forward to another day

filled with lively discussions and great panel participation.

At this point, I would like to introduce Dr. Steve Heeringa, Chair of the FIFRA Scientific Advisory Panel.

CHAIR HEERINGA: Good morning, everyone, and welcome back to a second day of this meeting of the FIFRA Science Advisory Panel.

Before we get underway with the proceedings, just a little summary of where our agenda will go this morning. We are still in the period of public comment. I mentioned yesterday afternoon that I would leave that open overnight in case there were any additional points of clarification and we did sort of push things along a little bit toward the end to get everybody in.

So, we'll open that up again just for a short period of time, any follow-up questions or issues on the public comment, and then we will turn back to the Environmental

Protection Agency staff for a wrap-up and summary in preparation for the Panel's turn to addressing the charge questions.

And I think I would like again this morning, just to go around. A few new participants have joined on the Panel. Dr. Pope.

But just to remind us of who we are, I am Steve Heeringa, the University of Michigan. I am an applied statistician. I'm the Chair of the FIFRA Science Advisory Panel and here predominantly to help with the meeting itself and running of the meeting.

I'll turn to my colleague on the left, Dr. Portier.

DR. PORTIER: Good morning. I'm Ken Portier, Director of Statistics, the American Cancer Society. I'm a biostatistician and member of the permanent panel.

DR. CHAMBERS: I'm Jan Chambers, a professor in the College of Veterinary

Medicine at Mississippi State University. I'm a pesticide toxicologist and I'm a member of the permanent panel.

DR. BUCHER: I'm John Bucher. I'm the Associate Director of the National Toxicology Program at NIHS. I'm a toxicologist by training and a member of the permanent panel.

DR. POPE: I'm Carey Pope. I'm Professor of Toxicology at Oklahoma State University, Center for Veterinary Health Sciences, and a member of the permanent panel.

DR. BAILAR: John Bailar, retired University of Chicago and now Scholar in Residence at The National Academies in the position of biostatistician/epidemiologist.

DR. MEEK: And I'm Bette Meek, and I'm Associate Director of Chemical Risk Assessment at the McLaughlin Centre, University of Ottawa, an interchange from Health Canada, where I manage several chemical risk assessment programs, and my background is in toxicology risk assessment.

DR. GREENWOOD: I'm Richard Greenwood, Professor of Environmental Science at the University of Portsmouth. My expertise is in the area of mode-of-action of pesticides and in environmental monitoring.

DR. HARRIS: And I'm Shelley Harris. I'm an Associate Professor at University of Toronto and a Scientist at Cancer Care Ontario. And I'm an epidemiologist with a background in exposure assessment and toxicology.

DR. BOVE: My name is Frank Bove. I'm a Senior Epidemiologist in the Division of Health Studies of the Agency for Toxic Substances and Disease Registry, which is part of the Centers for Disease Control.

DR. LU: Good morning. Alex Lu from Harvard School of Public Health. I do pesticide exposure and cause/effect research.

DR. GOLD: Hi. I'm Ellen Gold. I'm Professor and Chair of the Department of

Public Health Sciences at UC Davis and Chief of the Division of Epidemiology there.

DR. HAYTON: Good morning. I'm Bill Hayton, Professor of Pharmacy at Ohio State University with an interest in pharmacokinetics.

DR. REED: Nu-May Ruby Reed, Toxicologist at the California Environmental Protection Agency. I do pesticide risk assessment.

DR. REIF: I'm John Reif. I'm an Environmental Epidemiologist from the Department of Environmental and Radiological Health Sciences at Colorado State University.

DR. LeBLANC: I'm Gerry LeBlanc. I'm Professor and Head of the Department of Environmental and Molecular Toxicology at North Carolina State University.

CHAIR HEERINGA: Thank you, members of the Panel. And before we return to our period of public comment, I'd like to turn to Dr. Steve Bradbury or Dr. Tina Levine if you have any opening comments - oh, Dr. Anna Lowit.

DR. LOWIT: I thought you would start with public comments. We don't have any clarification. We went around the team, and no one had anything we wanted to clarify. We thought yesterday's discussion was excellent.

And just one point we wanted all of you to know that Dr. Michael Alavanja is available today and he's sitting back there, but he will not be available in the room tomorrow.

CHAIR HEERINGA: Thank you very much, Dr. Lowit.

At this point in time, I would like to return to our period of public comment. And we have one more public commenter who has approached Myrta Christian, the Designated Federal Official. It's Dr. Robert Silkin. And I think he can identify his affiliation, but I think he's with the Syngenta Group or -

DR. SILKIN: Thank you, Dr. Heeringa.

CHAIR HEERINGA: Panel members, there is a handout, I think, of Dr. Silkin's slides.

DR. SILKIN: Yes, there is a short handout, and there is a slide set, too, please.

My name is Bob Silkin. I'm a statistician formerly of Texas A&M University, and I'm now a consultant for Syngenta at this particular case. I consult for a lot of people, but I'm here today on behalf of Syngenta.

Dr. Bove yesterday correctly pointed out that the correlation plot that you see up here is really for the 18 CWSs that had SGA prevalence. And it does have one point for each of the CWSs.

In the data set itself, there is the number of births in each of those CWSs. And it's pretty obvious as you glance down that column - I guess that's a pointer, but it doesn't matter.

Second column from the right is the number of births. Fort Wayne is over 16,000. Some of the others have around 200. So, they're not all equal size in terms of number of births.

So, the question might be what would happen if you treat each of those births individually instead of just the CWS, and would that change anything? Obviously, it would re-weight the points.

Since I don't have the individual data, all I can do is basically do a weighted correlation where the weights would be the 233 for Batesville and 255 for Bedford, etcetera.

And if I do that and redo the correlation, I still get a non-positive correlation. The values shown here, they're not really statistically significant, but they're not pointing in a positive direction either.

This plot is what Excel gives me. If I jitter those points a little bit, and I'm not sure whether everybody knows what jittering is, but if I just add a small, normal deviate to each point, I can kind of move them around in Excel and give you an idea that some of those data points have a lot more individuals in them than others.

That big blob there is really Fort Worth - Fort Worth? Fort Wayne. Sorry. I live in Texas now. I went to school in Indiana. I went through Fort Wayne on the way home.

But, anyway, that's Fort Wayne, and you can see there's a couple of others, two or three others that have a much larger sample size than the other.

A little bit of a sensitivity analysis. Since Fort Wayne is so dominant with 68 percent of the births, the question is what happens if you omit the Fort Wayne, which is hugely dominant? What does the rest of them say? Still a non-positive correlation.

Further sensitivity analysis is in this slide you see one point kind of way over to the right. That's Batesville. And it has just a few - about 200 people in it.

If you take that point out, what do you have left? And you have almost just noise, but not a negative - small, negative correlation, not a positive correlation.

So, that was just a clarification.

CHAIR HEERINGA: Dr. Bove.

DR. BOVE: Well, there's still a major problem with this analysis. First of all, I don't understand why you don't read the paper in use and see the analysis there.

It's an individual level analysis. They take into account the season of pregnancy, they have data from the municipality, they do a regression, a logistic regression.

My problem with the logistic regression, by the way, and we'll talk about this later, is they use a continuous variable and they have strong assumptions for how the exposure-response curve looks.

But no matter how you slice the data when you look at it at an individual level with the information they had, with the information on individual potential confounding factors in all and they model it, they have a positive, although very mild, but positive association.

And no matter how many times you do this which uses less of the information in the study, I don't understand the point.

I think that if you read the paper and focus on the analysis done, you can see whether you think it's a strong association or weak association not looking at statistical significance, but looking at the actual point estimates, regression estimates and so on. Then looking at the confidence intervals, of course, and looking if there's an exposure response. And there are problems with the paper on that level.

You don't have to resort to this, which is using less of the information and really a distortion of the information in the paper.

So, I have a hard time understanding why you're doing this exercise when there's plenty of information in the paper for which they'll criticize the paper if you want.

The other thing is that when you look at Fort Wayne, they looked at Fort Wayne separately and they said - and they don't provide enough of the information for me to judge, but they said that the relationship they found with all the data was similar with Fort Wayne. Okay.

So, I think they've dealt with these issues. I think this is a useless exercise. I think that what you need to do is look at the paper and decide whether you think that the effects they saw were important, clinically important, biologically important, whatever, scientifically important or not.

DR. SILKIN: Well, you're quite correct that they had a lot more data than I do to work with. By not having the individual data, not having the individual confounders, all I can go by is what they say and what they put in the paper. I have no way of checking the validity of the confounders, the effects of the confounders.

I have their confidence intervals, and admittedly there are some problems just looking at the paper with what they did. And you're quite correct that this is not as good as what you could do if you had all the data, which they have and I don't have, yes.

CHAIR HEERINGA: Other questions of clarification for Dr. Silkin on his presentation?

Thank you. Thank you very much.

At this point in time, I'd like to turn to the panel. We've heard from a number of public commenters yesterday afternoon. They offered different presentations on different topics. Some of them related to potential methodologies, some of them related to an overview and sort of summary of their interpretation of the charge questions.

Is there any questions of clarification that the panel would like to bring to - if you could just identify to whom and see if we can bring them up.

DR. REIF: John Reif. I have several questions for Dr. Alavanja that I'd like to have him address this morning.

CHAIR HEERINGA: Dr. Alavanja, well, he's actually with the EPA Group, too, so let's hold the questions.

Dr. Alavanja, since you're part of the EPA presentation, let's hold that for a follow-up. For the public commenters from yesterday, going once, going twice. If something comes up during the charge questions, we may be able to negotiate with the EPA to bring somebody up to answer a question.

At this point, then, any final public comments at this point?

Okay. I want to bring the period of public comment to a close and turning to Dr. Lowit, if we can turn to some questions, follow-up questions for your staff, and then any comments that you would have.

DR. LOWIT: Sure.

CHAIR HEERINGA: That includes Dr. Alavanja too. So, Dr. Alavanja, I apologize for the formality, but I just want to make sure we kept everybody in the right balance.

Dr. Reif.

DR. REIF: Sorry I was out of order, but I would like to pursue a few thoughts with Dr. Alavanja regarding the potential for the agricultural health study to shed light on potential health effects of agrochemicals on reproductive outcomes.

So, most of the discussion

yesterday focused on the differences in exposure assessment and the way that exposure assessment was done in he agricultural health study versus the approach that the Agency uses.

But I'd like to ask Dr. Alavanja, if he would, to briefly describe the female members of the cohort with respect to their numbers, their distribution across the two states, and, to the extent that he recalls, the proportion of these women who actually work on a farm or who don't report work on a farm, that is who work off a farm. And then I've got another couple of questions.

DR. ALAVANJA: I'd be happy to answer that question.

I mentioned yesterday that there were three rounds of interviewing that went on, and all of the papers that we have produced thus far have utilized the exposure information from what we call Phase 3. The interviews started in 1993, and they went through 1997.

That exposure information assumed that if you were an unlicensed individual, you didn't apply pesticides on the farm. And that was a mistake when it came to the spouses.

About 45 percent of the spouses do in fact assist their husbands on the farm, because the pesticide applicators are 97 percent male in our study.

So, I made an incorrect assumption in Phase 1 that if you were unlicensed, you wouldn't apply. So, Phase 1 data is not robust with regard to the question of ascertaining exposure among the spouses. And that would be the critical group to look at for the reproductive outcome.

In Phase 2 and 3, I realized that we made a mistake, and we now ask for both the bystander exposure and also the direct occupational exposure. We will now start using Phase 2 and Phase 3 data along with Phase 1 data for exposure assessment, and so we are now able to address that question more rigorously.

So, it's something that we will be doing. We didn't have the capability of doing until we integrated all of the exposure information from Phase 1, 2 and 3 of the questionnaires.

One, I guess, last remark is that for some outcomes we can look at the approximately 1500 farmer/applicators who are women in our study, but it's not going to pan out for the reproductive effects because it's just too small a group to look.

DR. REIF: What are the dimensions of the total number of women in the two states with all exposures?

DR. ALAVANJA: Just under 31,000 individuals.

DR. REIF: So, I'm aware of some of the publications on the website predominantly done by your colleagues at NIHS, if I'm not mistaken.

Has there been an effort to link birth records with the women in the AHS to validate perhaps the self-reporting of events with respect to pregnancy outcome?

DR. ALAVANJA: Yes. There is a manuscript that is now in press that did that in the state of Iowa, and so we were able to validate the responses on births. We also found a few additional births that occurred usually right after the interview was given, and so we can do that in the future.

But there is a potential rub that we're trying to work out in the next few months, and that is if someone was identified - what we did, 89,658 adults in the agricultural health study, when we interviewed the adults, we also asked about children in the family, and so we have the information from the parents.

But for many of those individuals, they're now over the age of 18 and there's a question as to whether or not you can follow those children up without getting informed consent from those individuals directly.

So, that issue is now - well, will soon be in front of our IRB to see if we can do that. It would be a great loss to us if we couldn't do that, and it also would be a great challenge for us with our budget, to go back and try to get informed consent from those - the number is over 35,000 children. So, that would be a real challenge for us to do that.

So, that is a potential problem that we have to address in the next year.

DR. BOVE: Let me ask a quick question on that.

CHAIR HEERINGA: Dr. Bove.

DR. BOVE: But you can see if some of those children had birth defects on the registry.

DR. ALAVANJA: In Iowa particularly since -

DR. BOVE: Yes, yes.

DR. ALAVANJA: - they have a very

well-developed birth defects registry.

DR. REIF: My final question has to do with exposure assessment particularly with respect to the women. And in the discussion yesterday, most of it focused on the pathways that are referable to the applicators, and rightfully so.

But in an environment like what we find in Iowa and probably in North Carolina, the probability of having agrochemicals find their way into groundwater certainly exists, and most farms rely on private wells for their domestic consumption.

So, my question is has there been any monitoring of private wells on the farms that are part of the agricultural health study with respect to agricultural chemicals, and are there any plans to do so if that has not been done to date?

DR. ALAVANJA: Dr. Mary Ward is an environmental epidemiologist who works in our branch and is interested in that topic, has done work in other studies. She is working with us now.

We have the address information for all of the participants in the ag-health study, and most of those addresses have now been geocoded.

And so based on information that we can get from the U.S. Geologic Survey and such with regard to a lot of testing that they've done in the state of Iowa, Mary is doing a modeling exercise to determine where the high exposures were historically.

One would have to bring to bear what was grown in those areas, what was applied to those fields and what were measured in those wells. And she is doing an exercise that will try to incorporate all that, so that that type of research could be done.

As far as making current measurements, some current measurements are planned, but that would only get you so far because of the contaminants in the well would change with time. So, it's really a longer period of time that one would have to characterize the exposure.

DR. REIF: And I think you're responding from the standpoint of a cancer epidemiologist primarily. But from a reproductive standpoint given the finite nature of the relevant exposure period, I would just suggest that one ought to think carefully about current exposure assessment followed by the prospective period of reproduction and the potential for some informative studies to be added to the agricultural health study, which of course is an incredibly rich data source. Thank you.

DR. ALAVANJA: If I could add just one - yes, I am thinking as - I always revert to thinking as a cancer epidemiologist, but I come up for fresh air and think about the question.

The issue still remains, though, the average age of the cohort and the women in the cohort is now 56. So, it's getting very powerful for cancer studies, but it's getting less powerful for our reproductive studies.

Of course there's always a tail, you know, where there are younger women, but that would be a problem. Most of the new births have passed. We're not in that season any longer.

CHAIR HEERINGA: Dr. Harris.

DR. HARRIS: Shelley Harris. Yesterday you talked a little bit about the collection of PPE information and how you did that, I think, by different categories and classifications of pesticides. And I'm interested in whether you collected that for different activities such as spraying or mixing and loading or fixing equipment. And if yes, did you do it - is it chemical-specific information?

I have a couple other questions as well.

DR. ALAVANJA: The specificity to

which we're getting the information is far greater in Phase 2 and Phase 3. We built on our experience from Phase 1, so that is the case that we are getting it in categories.

So, personal protective equipment is now routinely obtained for insecticide use, for herbicide use, fungicides and fumigants. And so we are doing it that way.

DR. HARRIS: So, you don't have a specific measure so farmers might put gloves on or an apron or a respirator, something to mix and load, but not to spray. So, you don't have the data separately for activities.

DR. ALAVANJA: No, we -

DR. HARRIS: No.

DR. ALAVANJA: We have it for two broad categories. We do actually have it for mixing and loading, and then for spraying.

DR. HARRIS: And was there any consideration so you have acre - you probably have a lot of information on acres and different types of crops and incorporating any kind of measures of volume or acres sprayed or active ingredient into the intensity - or into the either exposure scores or do you multiply those times the intensity scores?

DR. ALAVANJA: The difficulty we found was that we couldn't get that information accurately 20 years ago. So, we resorted to those - resorted - we used those variables that we knew in our experience had shown us that we could ascertain those variables reasonably reliable for 30 years ago, 20 years ago.

Active ingredient is a very key variable, but we don't - we couldn't get that information on individuals for historic reasons. So, that's not going to be a part of our algorithm - or is not now, and we don't anticipate it being used.

If it was introduced actually, though, let's say in Phase 4 which we're planning, we could do other sorts of studies that are not cancer-related that might benefit from that.

DR. HARRIS: But do you have some of their pesticide use information or their purchase records over the years historically?

DR. ALAVANJA: We believe that they have those records, but they're not actually interested - well, the vast majority of people sharing those records with us. I can speculate as to why that might be, but it hasn't been a success.

I think it's tied up with the fact that some of that was sort of their tax records. And sort of teasing that apart at an individual home, it wouldn't be something that - we have not been successful with it. There's some technique that one could use, but we haven't discovered it.

CHAIR HEERINGA: Dr. Reed.

DR. REED: Yes, I'm also interested in what kind of information that you use to come up with the numeric coding for some parameters in the exposure intensity level.

I guess sort of give you sort of the back of my head, why am I asking this question, might help. I was a little bit surprised yesterday with the comparison between the two methods, agency's and AHS method. They're so close, 3.7 and three.

And so I'm asking the question from that standpoint that I noticed that. And maybe you covered it yesterday and I was just not catching it right. You have taken into account many sets of data, including the PHED data, when you come out with the codes.

How heavily was PHED considered in coming up with the scoring or how similar your scoring would reflect what PHED data raises?

DR. ALAVANJA: I was part of this paper. But what Mr. Dosemeci who was the lead author on this paper did, he actually started with the PHED data, but the PHED data doesn't make a distinction between chemicals.

And so when you look at the world's literature, and he looked at over a hundred papers that did exposure assessment, and some of those papers were certainly done under more realistic conditions, they were observing farmers at their work and measuring specific chemicals, so he would make adjustments to account for coming up with some type of weighted average of PHED along with the world's literature when it didn't fully agree with the PHED data. And so that's why there would be some differences between the two.

But I would like to point out one thing is that that paper Dosemeci did was published before most of our etiologic papers were done. We really were pushing hard to have a manuscript outlining our methods in the literature prior to what we were doing in the etiologic studies, and then commenting on that.

And we have yet not changed that. But if there are different weighting factors that we should incorporate, our intention is to always do that with a published paper so that everyone will see our methods and make it fully transparent.

CHAIR HEERINGA: Dr. Alavanja, Steve Heeringa. A question about the population in the cohort trajectory over time. These were all licensed applicators. Some of them commercial, but some of them private farm applicators.

In the sequence of measures that you've made over time, have you determined the intensity with which they personally have applied these materials as opposed to the intensity with which they've applied on their particular farm?

The concern that I have is that - and maybe you can correct me - legally you have to be licensed to purchase this material. But I know from experience, you don't necessarily have to be licensed to drive the tractor that puts the spray down on the crop.

And to what extent the applicators

that you are measuring as your cohort may actually over time have reduced exposures personally transferring those exposures to children, to hands, to other people who are - is that sort of context captured in the ag-health study?

DR. ALAVANJA: I think we can always do a better job with that and it has been our concern. But one of the reasons we selected the states that we did, we wanted to go to states where it was a farmer-owner-operated farm. So that when you went to the farmer, in most cases they were the people - that was the person that did these activities.

And so that was sort of the underlying philosophy, and those two states were chosen from a group of many states in the Midwest and the southeast that could have been chosen, and others would meet the bill just as well.

But when we asked the question, we asked whether or not - we always ask did you personally apply? And so we get the information about their personal application.

Over time there is some tendency to have commercial pesticide applicators apply the material to the farm, and so we get some of that information as well. But to this day in the ag-health study, it's still primarily the farmer themselves that are applying the materials.

And with regard to sons and such, and that does happen of course, whether or not that is accurately reported all the time, and there's a question about that, but for farmhands it tends actually not to be the case, seasonal workers and such, not to be the case yet in the agricultural health study.

CHAIR HEERINGA: Thank you. And the concern I guess that I had is to make sure that we capture exposure on the cohort numbers correctly. I know that we can't sort of split off to children and others, but it sounds like that's being done to the extent it can be.

Dr. Reed.

DR. REED: So, since most of the individuals in the database are farmers applying pesticides themselves, I would suppose that the repair frequency would be pretty substantial, right?

DR. ALAVANJA: Yes.

DR. REED: So, the scoring between zero and two, how do you assign the score? Is how frequently the repair makes the difference between zero and two or -

DR. ALAVANJA: If you consider the algorithm, the algorithm asks for the number of total days in a year that a person applies multiplied by years. And then the intensity factor is added onto that.

So, if a person was to repair their equipment and said "yes" to that, it would be a multiple times the number of days of application. So, it would be considered in that way since it's a multiple of that number. And so that's how it would enter into the equation.

DR. REED: Thank you.

CHAIR HEERINGA: Thank you very much, Dr. Alavanja.

Any other areas that the panel would like to pursue clarification before we turn to the charge questions, on the presentations yesterday?

Dr. Reed.

DR. REED: Could I ask Shalu a similar question about -

CHAIR HEERINGA: Shalu Shelat and Jeff Dawson.

DR. REED: About the similarity between the two comparisons, 3.7 and three.

Do you have any take about why this is similar and any sense of speculation on the comparison?

MS. SHELAT: So, for the purposes of the case study at this point, I would just use an arbitrary example. We're not intending to make the conclusion that a lot of the exposure rates will also be that similar.

I'm sure part of the reasoning is as Dr. Alavanja had mentioned, PHED is incorporated into the calculations for exposure rates. But until we look at it in a more holistic fashion, we really can't make a conclusion.

DR. REED: So in your agency calculation, part of it, and you're using PHED, you're taking the central tendency, not the upper bounds; is that correct?

MS. SHELAT: I'm going to defer that question to Mr. Dawson.

MR. DAWSON: I'm sorry. Could you repeat that? I apologize.

DR. REED: I was wondering in your comparison on the agency calculation site and you raising PHED, you are taking the central tendency, the average or the mean or the geometric mean or anything like that, not the extremes.

MR. DAWSON: Jeff Dawson. This

kind of came up yesterday. Dr. Portier asked a question around this issue. And the last couple years we've kind of very publicly said that there are some areas within PHED that we can improve.

So, we're actually involved in a large effort to essentially refine that database to be able to get at better addressing, for example, what the distributions of exposure look like and ways of ensuring and identifying subjects to monitor that can be related to - allow us to better answer the question, for example, of representiveness.

So, as that new information comes online, we'll be incorporating that additional information into this process. And part of the message I think we were trying to go over yesterday, was that we want to have some intense collaboration with the ag-health folks to refine that as we go and we get better information.

Because, clearly, there are some areas within PHED that we need to improve and that are inherent because of the way that that database was created.

DR. REED: Yes, I guess my question is specific to Shalu's analysis. When you access PHED and came up with the exposure estimate, you are taking it at the average mean or geometric mean and not the extremes, right? Not the bounds?

MR. DAWSON: We're not using the bounds, right. We use the central tendency -

DR. REED: Right.

MR. DAWSON: - value that's inherent in the way that that system was created, right.

DR. REED: I have another question about PPEs.

Have you looked into whether the accounting of PPE in terms of percentage of protection is similar to what was used in the ag-health study?

They have the 50 percent and so forth for various PPEs. Is that the same in the PHED or is it different? I thought some of them, I mean, in PHED would be greater than 50 percent.

MR. DAWSON: Absolutely. Actually, shortly after the initial unveiling, so to speak, of PHED, we actually did quite an extensive analysis where we looked at the variability of different types of personal protective equipment, for example, different - all the data we had with additional layers of clothing, and created an actual document where we looked at this and there's quite a range.

So, we ended up using the values that we use based on that analysis, but it definitely ranges. For coveralls, for example, it ranged from 10 to 90 percent. And gloves as well.

In cases where there are sufficient data, what we've done is segmented the data to just use, for example, here's the folks that wear gloves and here's the folks that do not wear gloves. And we try to avoid, for example, in the way we do it, the use of protection factors.

But certainly the numbers they're using kind of overlay with what we see from the database.

DR. REED: But are different though, right?

I guess what I was trying to get a sense of is still the comparisons in that percentage of protection from PPE in PHED, could be expected to be different or the same compared to the percentage of PPE protection from use in the ag-health study.

MR. DAWSON: Well, they're the same.

DR. REED: They're the same. Exactly the same.

MR. DAWSON: The values that they're using certainly fit on the range of what we see as far as actual performance of

PPE based on the data that we have, right.

DR. REED: And that was used in PHED and - so, in terms of comparison of the two tracks, the PPE are the same - or protection of PPE are the same in terms of the percentage of protection?

MR. DAWSON: Right.

DR. REED: Okay.

CHAIR HEERINGA: Okay. At this point, I think I would like to move on to the charge questions. But before I do that, I'll turn - Dr. Lowit, anything that you would like to add before we turn to the charge questions?

You're free to add, of course, during that process too. Okay. Well, then I think that I would like to turn to the first of the charge questions. In setting this up for us, they pulled the old trick of subparts. So, what looks like four questions is probably more like 11 or 12 questions, but we'll recognize that in advance and move through them.

So, it's actually an organizational strategy, I think. But in any case, we will start with the first question.

And, Dr. Lowit, do you want to read it into the record for us, please?

DR. LOWIT: Yes. Did you want me to read the preamble or just the question?

CHAIR HEERINGA: Just the question is fine.

DR. LOWIT: Okay. Okay. Good.

Section II of the draft framework describes the major types of epidemiology studies along with their strengths and limitations, factors to consider when reviewing epidemiology studies, and ways to use epidemiology in risk assessment. Please comment on the soundness and completeness of these discussions. If appropriate, please include comments on additional factors for OPP to consider when evaluating the quality and weighing the utility of epidemiology studies in risk assessment/risk characterization.

CHAIR HEERINGA: Again for everyone, our process is there is a lead discussant assigned to each of these particular questions, and that individual will lead off followed by an assigned set of associate discussants, and then I'll open it up to the full panel for their comments.

And Dr. Bove is the lead discussant on Question 1.1.

Frank.

DR. BOVE: Okay. I have a whole lot of comments, and I can't go through them all. So, overall I'd say I'd like to see the section rewritten.

On Page 13, there's a list of study characteristics and they're fine. I would just want to add a few more.

One would be that the study have an appropriate interpretation of the findings. Of the studies we reviewed, they focus on statistical significance, findings that are elevated that are not statistically significant are ignored, exposure-response relationships that are not statistically significant are ignored.

If the study does that, that's their problem. But in interpreting the studies, EPA doesn't have to follow what the authors of the study did. So, I wold encourage EPA when they look at these studies, to not let statistical significance trump the magnitude of the association of the dose-response or any other consideration.

A second point to add to the list is appropriate evaluation of the exposure-response relationship. It's sort of tied with the first one, but also it's not sufficient to simply use a continuous variable in a logistic regression or to use tertiles, for that matter. It would be good to have an argument as to why you think that captures the curve.

The third issue is we hear a lot about confounding and it may occur in studies. Although, in my experience and the experience of others, it's rare that there's considerable confounding in any study.

But putting that aside, it's not enough to just say that there may be the presence of confounding. I like to see sensitivity analysis. Just how important is the impact of confounding and other biases, for that matter? And there are methods to do that.

The fourth thing, it was raised yesterday, collinearity of contaminants. In my own work when I look at trichloroethylene in drinking water, oftentimes perchloroethylene is in drinking water. It's hard to distinguish the two.

So, you have to try to find study populations where the exposures vary enough so you can maybe tease out what might be the effective TC and what might be the effective PC or the other contaminants that seem to go together. And that's true of pesticides as well.

So, you may want to encourage researchers to try to choose study populations where there's enough variability that that can get teased out.

Now, on to the types of studies. Again, a lot of statements in here that I disagree with. For example, in a case control study, the controls have to be non-diseased.

There are other control selection methods out there where the controls can have the disease of interest, even, and so I don't think you should make statements like that.

The other thing about case control studies is that the exposures although we were interested in exposures in the past, they may not have to be too distant past for birth defects, for example. But in any case, exposures can be estimated using historical exposure reconstruction methods, which is using environmental monitoring and sophisticated modeling to try to do that.

And so some of the problems that

people raise about case control studies can be - would not be a problem if you estimate exposure in that fashion.

A more serious problem in this section is about cross-sectional studies. Many cross-sectional studies use historical information on exposure so temporality can be established, but the key feature is not - of a cross-sectional study is not that exposure and disease are measured at the same time.

The key feature of a cross-sectional study is it measures prevalence and - prevalence of disease, prevalence of symptoms, prevalence of biomarkers. Okay. So, that really is the key element, not that they're measuring exposure and disease at the same time.

And one of the advantages of the cross-sectional study because it's evaluating prevalence, is it can measure these biomarkers. They're not routinely collected. So, that's a very important advantage of cross-sectional studies that can't be done maybe in other studies unless there's a lot of money put into a prospectus study to do these kinds of measurements over time, but there are drawbacks.

And the key drawbacks of a cross-sectional study are you're studying a survivor population. And that includes births. Births are a cross-section. Okay. You have to survive to birth in order to have a small for gestational age birth or a birth defect, unless you look at spontaneous abortions and other miscarriages and look at birth defects among them.

So, it's a survivor population. And that means that if the exposure affects those who don't survive. If the current employees in a workforce of those who can withstand the exposure and the people who couldn't have already gone and you go in and measure, you're measuring a survivor population. So, there are problems with that.

That's the first drawback with a cross-sectional study.

Second is that prevalence is a function of incidents and duration. So, the question is, is the exposure increasing incidents or is it increasing duration?

So, those are the two key elements of a cross-sectional study not really reflected in the paper.

Ecologic studies, there are good and bad ones. And I think even the bad ones sometimes we can learn something from. I think some of the presentations yesterday showing that there's seasonal effects even in low-atrazine areas is interesting. So, you can even learn something from a not very well-conducted ecologic study. But I think we have to figure out which ones are good and which ones aren't for this purpose. They're not all bad.

But a key element of an ecologic study, and this is where I think people get confused, is that exposure is assigned to a population. We usually have variables such as the percent of the population exposed, the percent of the population who smoke, where we use mean, the average income of the population or average pack years smoking or average smoking sales or something of that sort. That's how exposures are defined in the ecologic study. It stays at the group level.

The Villanueva study the EPA considered an ecologic study, it isn't. Exposures are defined at the individual level. The individuals are characterized by the municipality serving them. The municipality has information to show that the water is uniformly distributed. So, the quality of that water is from one source and it's the same for everybody. So, a hundred percent of the people can be characterized in that population. It's an individual level study.

Now, some studies go from there and then start using ecologic variables to control for confounding. That's a bad move, but that wasn't done in the Villanueva study.

There are other issues around ecologic studies besides the problem with exposure. Ecologic biases also affect how confounders are adjusted for, and it makes it more difficult to adjust for individual level confounders if you try to adjust for them using ecologic confounder information. So, those are some of the issues there.

We'll move on to the next part of the document, which is important scientific factors to consider, and talk about the exposure assessment section which contrasts direct and indirect - so-called indirect measures of exposure.

Direct is biomonitoring and personal monitoring. Indirect is historical records, questionnaires and environmental monitoring. And the problem with direct approaches and the benefits actually of so-called indirect approaches, is that it's difficult to use direct approaches when your interest is in past exposures. Especially distant past exposures.

On the other hand, an approach that uses environmental monitoring and sophisticated modeling, which in my agency we call historical exposure reconstruction, is a very important way of dealing with past exposures. We've used it in drinking water, we used it at Hanford for estimating iodine exposures. We've used it in a number of other areas. Mostly drinking water, but we also used it in air pollution as well.

It's a very good, robust way of estimating past exposures. You can get quantitative information that's useful for any risk assessment you want to do. So, I think that that needs to be emphasized in this section.

Again, historical records and questionnaires can estimate quantitative levels of exposure. It's not just true that they can be qualitative, and it's been done.

Now, some of the problems with direct methods, on the other hand, as I said, they're not very good for past exposures, but they may not even capture the full range of exposures.

Unless you have that personal monitor on you for quite a long time, I mean you change over time and you may miss that. And besides, you're wearing a personal monitor and it may affect your behavior because you're wearing it. So, you may change your behavior because you're wearing that personal monitor, or you may change your day-to-day habits and you need a diary.

So, you need a questionnaire to go along with - so, you need an indirect measurement in order to go along with the direct measurement.

So, these are some of the complications that are not reflected in that section that need to be reflected.

Confounding, substantial confounding occurs rarely. Even when you look at lung cancer and smoking and you look at an occupational exposure to compare workers to the general population, the amount of confounding usually found is less than 20 percent. And that's the most confounding you could probably find in a study.

So, again, people raise the issue all the time. It's important to figure out what the impact of confounding is. And that can be done by sensitivity analyses.

As we heard about the agricultural health study that there is no confounding in that study, that does not surprise me.

The other issues, you have to be careful about what you put into a model. I'm a little concerned about putting seasonality in the model. Seasonality, there are a lot of factors involved in seasonality. In areas where there's low atrazine exposure, seasonality may be affected by such things as air pollutants, trihalomethane and other disinfection byproducts because they're seasonal as well. But in the areas where there's pesticide use, seasonality could be affected by the pesticides themselves. Okay.

So, you put seasonality in a model, you may be adjusting for exposure. And when you do that, you're biasing your point estimate towards null. That's a bias.

Bias one way, you know, everyone focuses on that, but bias the other way is just as important. You want the right answer. Okay. So, you have to be careful about that.

Another study put a variable in there for percent of the land around the house that had crops. Again, I'm worried that they're adjusting for the exposure itself and making an already difficult study, a study that has difficulty estimating the effect, even more difficult. So, they need to be careful about what variables they put in the model.

A few more points and then I'll get out. Effect modification and confounding, totally different ideas. Totally different. Effect modifiers do not have to be confounders, and in fact many aren't.

But most importantly, I think confounding is a bias and we want to minimize that. Effect modification is a hypothesis. You want to design a study to evaluate that hypothesis.

Some studies look at all kinds of effect modifiers and it's a fishing expedition. But really a study if it really wants to evaluate effect modification, has to be designed with enough statistical power to evaluate effect modification. That really needs to be stated here.

There are also issues of statistical analysis that aren't mentioned such as if you put a lot of variables in a model, you may get statistical bias. That means the odds ratio you get, for example, may be inflated because the model just can't handle all those variables, in a nutshell. Or if you use conditional methods, there may not be enough discordant sets or pairs and you get inflated odds ratios.

So, you have to be careful, again, of what you try to put in a model and what you don't.

Finally, two more points. One, interpretation of null studies. Again, the two major causes of null studies are exposure misclassification bias, and the second cause is lack of statistical power especially if you're focusing on only statistically-significant results.

And in birth defect studies when they use birth certificates as mentioned in these studies, you do have under-ascertainment and you can expect they're probably reduced because of that.

Finally, I think that epi studies have been used in all stages of risk assessment, in the TC risk assessments, both the draft and the current. In the draft one, a drinking water study in New Jersey was used, for example. In the recent one that's now being considered, an occupational study of kidney cancer is just one example. But epi studies have been used in quantitative risk assessment, and should be used. And I think that's where I'll stop.

CHAIR HEERINGA: Thank you, Dr. Bove.

Dr. Meek.

DR. MEEK: Thanks very much. It's difficult to add an awful lot after that very extensive discussion, but my sense was that in this context, the text provided rather an overview, generically, the strength and weaknesses of various types of epidemiological studies, but didn't necessarily do so in a context specific to experience on pesticides, with the possible exception of the exposure questions.

And my sense, also, was that while the case studies were helpful, it might have been more informative to include an indication of the extent and nature of reliance upon epidemiological or incidents data for a range of pesticides across the program to get a feel for that in previous assessments.

And based on our experience in industrial chemicals and really following on from what Dr. Bove has said, the use of epidemiological data and risk assessment necessarily varies depending not only on the nature and quality of the studies, but the results, whether or not there is evidence, robust evidence of an effect in humans.

For example, if we have evidence of an increase in a particular effect in humans based on robust epidemiological data in a population to which exposure has been well characterized and for which there is weight of evidence for causality, we would necessarily favor the use of those data in the dose-response characterization.

On the other hand, if we have a negative epidemiological study for an effect which we consider to be relevant to humans based on toxicological and mode-of-action data, we might use this information to bound dose-response estimates from animal studies.

There's different ways to use the epidemiological data depending upon your confidence therein.

I had a couple of comments on the interpretation of null studies. It seems really important to mention in addition to the points that have already been mentioned, that without information on mode-of-action they are exceedingly difficult to interpret, unless, for example - if you take a particular end point, for example, for cancer, there is some understanding in mode-of-action for tumors induced in animals. It's unclear where tumors might manifest in humans and site concordance can't necessarily be assumed.

And, also, the power to detect the effective interest is always critical interpretation of null studies and often never formally addressed. So, that's an issue. And then there's also the publication bias to exclusion of null studies that needs to be taken into consideration.

Another point I think that it seems - and we'll get into this probably a little bit later in responding to some of the other questions, it also seems important to emphasize in relation to biomonitoring, that selection of relevant biomarkers of exposure and effect based on the toxicological database particularly that on mode-of-action, enables much greater likelihood of meaningfully integrating the epidemiological and toxicological databases.

And I had a couple of comments on cross-sectional studies as well, particularly in the context of prevalence versus incidents. But I think they've been adequately covered and I'll stop there.

CHAIR HEERINGA: Thank you, Dr. Meek.

Dr. Gold.

DR. GOLD: Thanks. I have a few things to add. I think the prior two speakers have been pretty complete, but my sense was that the exposure assessment part of this was a little more fleshed out even though it had some issues, than some of the other parts where epidemiologists, I think, focus their attention. And so, I would agree that this section needs some work.

One really minor point is I think in citing some of the references, some of them are old editions of books that ought to be updated. That's a really minor point.

But with regard to the particular studies, I think that's where I focused most of my attention. As I said, the exposure assessment is sort of one part that we focus on in epidemiologic studies, but a major focus is how we select the people that go into the study so that we get a representative sample with regard to exposure.

And so if you want to look at case control studies, you want to make sure that you have a representative group of cases at least with well-defined criteria so you can say something back to whom you can generalize on the basis of these and that they're not bias with regard to exposure, similarly, with controls that you select.

So, we often go into detail when we're writing about how we select people to be in these studies, about what the inclusion and exclusion criteria are so that people who are evaluating the study know to whom they can generalize.

I also thought that under case control studies, some mention of talking about newly diagnosed cases as opposed to using prevalent cases ought to be included. It's a similar issue about whether you're looking at exposures that were related to survival or if you're looking at exposures that are related to the occurrence of the disease.

Also, I thought some attention on how you collect data in terms of making sure that it's similarly collected in cases and controls and exposed and unexposed. And if you're doing cohort studies, that you do it at the same intervals and in the same way. These are, you know, sort of basic fundamental things.

I think paying attention to attrition in cohort studies so that, again, you can generalize from the results and making sure that attrition is minimized and participation bias is avoided.

Also in the context of avoiding bias, that observers are masked as to the case control status, their exposed and unexposed status, they're masked to hypotheses that are being tested.

And also under cohort studies in

the analysis section, just saying that the appropriate analyses are undertaken is sort of a minimalistic approach. But that you're using the maximum amount of information from those studies so that if people are lost or censored, that you use all the possible data that you can in longitudinal studies.

I don't think I have much to add about cross-sectional or ecologic.

One minor - well, it's not a minor point, but it was touched on. I just want to say one more sentence about it that when we're controlling for confounding the ecologic studies, again it's often at the group level, not at the individual level. And so, again, the results that you get may not be applicable.

I also thought it was interesting that I didn't see - maybe I missed it - any mention of nested case control studies or case cohort designs, which are really powerful and useful tools and perhaps haven't been used as fully as they could.

They could certainly be used in the agricultural health study, and there are other cohorts around where specimens have been collected and stored. Lots of cohort studies. I think this is an opportunity, actually, that hasn't been fully explored or used.

I agree with the comment about being able to get quantitative data from historical records, and it's been done many times.

Let's see. I think that's it for now.

CHAIR HEERINGA: Thank you, Dr. Gold.

Dr. Portier.

DR. PORTIER: It's always good to go third or fourth, and they cover all the good stuff. So, then you have to go looking for something different.

A general observation, many of the issues we're discussing relating to the factors to consider when evaluating the quality and utility of epidemiology study results, are the same factors that medical clinicians face when trying to translate epidemiology study results to clinical practice.

And I found an article I'll reference in the paper here, where they used epi data to guide clinical practice in the review of cardiovascular disease and combined oral contraceptives.

And it was interesting because they reviewed 74 epidemiology studies, and concluded that seven of those 74 were relevant for a clinician to use in practice. And actually five of them were directly useful, and the other two might be useful if they re-analyze the data.

And I thought in light of a lot of our discussion, I think the clinicians are doing the same thing you're doing and seeing the same results, that epi studies in general aren't always directly useful.

So, this got me thinking about what questions I might ask when screening epi studies for their relevance to risk assessment, which is kind of the utility part of the question here.

So, I came up with some suggested questions, some of which are implied in the discussion of Section 2, but it might be better to be kind of a little more direct and say here are the kind of questions that EPA would be asking when we look at an epi study in general.

So, was the epi study conducted in a hypothesis generating or hypothesis testing mode?

And I have an aside to myself, we all agree that it's inappropriate and misleading to use data to develop a hypothesis, and then the same data to test it.

But unfortunately when you read the conclusions of many of the hypothesis-generating epi studies, it's like the authors forgot this point. Right?

And so I think the point Dr. Bove has made about really understanding the interpretation of the study, this relates to this. That if it's really an exploratory study, the discussion should be in that context. But often the discussion is very confirmatory sounding, and then there's a disconnect.

I've noticed this in a lot of - when you read a lot of the literature critically for its utility in risk assessment, I try not to read the conclusions. I read the methods and make sure I can understand what they actually did.

Was the method of assessing exposure valid? Has there been some attempt to compare the exposure method to actual exposure? Was the method of assessing exposure reliable, an accurate measure of - was it an accurate measure of actual exposure or is there a bias involved in there? Was the method of assessing health outcomes valid and reliable? On one extreme was it confirmed with histopathology or reading medical records?

It wasn't really discussed that much about the health outcome confirmation part, but that's really another important part of an epi study. Do we really know that they have the condition that they say they have?

Did the study collect appropriate information on related and confounding factors such as cultural, behavioral, dietary and health factors, and related co-morbid conditions?

I'm always reminded that health conditions rarely occur alone. So, we're talking cancer, but they might also have diabetes and heart disease and all kind of other conditions. And those have a lot of impact on the health outcomes.

Factors that are known to impact

the health condition of interest as well as the factors that could impact exposure, we've had a little bit of discussion about that here.

Did the study measure the population or individual it's intended to measure? So, selection bias and generalizability are the issues here.

How does the study population relate to the universe of potentially exposed? So, the section talks about generalizability, but it's really important to make that relationship between what we studied and who we think to infer this to, who we want to be able to - so, it's Iowa and North Carolina, but really it's corn growers in the whole continental U.S. and Canada that we're trying to really make the inference to, and how do those two populations relate?

Did the study examine individuals from a wide range of exposures, including both those with high expected doses and those with low expected doses?

This affects our ability to detect the dose-response and our ability to generalize. So, if we only study a population that has low doses, we're only going to generalize the populations with low doses.

Did the study include populations or individuals not exposed? In my mind, that's kind of the negative control concept in an experimental setting. So, we might do that in a before and after kind of study.

Can we say something about birth defects in the 1920s to the 1940s compared to the 1980s and 1990s? If they have the same seasonal patterns, that weakens the results from the study.

Do the exposures examined in the study relate to past or current situation? This relates to the issue of acute versus chronic exposures, the targeted health end point.

Some of the things that were

brought up yesterday by Dr. Bailar about, you know, is the health effect related to exposures that occurred 20 years ago or exposures that occurred last week? That's a critical issue.

And finally, did the study collect information on sufficient numbers of individuals to have adequate power for preselected health effect differences between the different classes of exposed individuals?

In other words, does the sample size take into account the rareness of the target health effect in the study population?

You rarely see epi studies talk about sample size determination like you'd expect to see in a randomized clinical trial. But that kind of thinking should occur at the design phase, and it should be reflected in the discussion of the methodology of the paper.

And finally, a minor thing. In Section C, the first paragraph on Page 20 of the white paper, there's a statement about how high-quality studies with robust exposure assessment may be used to estimate risk quantitative.

And then that statement is qualified to indicate that most epidemiology studies suffer some limitation in size, scope, exposure assessment or data analysis which prevent their use in quantitative risk assessment. And this is referenced in Caledron (2000).

And I agree with this, but I think you need to support the statement in the document by providing at least one example of an epi study that really provided significantly the quantitative risk assessment.

And I have a reference here to an example of, say, the NIOSH dioxin study where they did heavy-duty biomonitoring and careful study design. And when you get finished, you really have that gold standard-type study.

And having an example of a gold standard study in the white paper helps everybody kind of think this is, you know, this defines what I think is really good, useful epi data. And I think I'll quit on that.

CHAIR HEERINGA: Thank you very much. Comments from other members of the panel in response to this question?

Dr. Bailar.

DR. BAILAR: I have two comments. Both rather brief.

The first is that we tend to think of ecologic studies as being in sharp distinction to studies that have individual measurement of exposures and outcomes, but in fact there's a gradation between these.

You can start with a pure ecologic study, maybe statewide incidents rates, statewide exposure levels, but then you go to municipalities. You come somewhat closer to individual exposures and outcomes.

You could go from there to water sources within the municipality, to households, and finally to individuals. Where do you draw the line?

I don't think it makes sense to actually draw a line there, but rather to treat this as a continuum. And I'd like to see that reflected in the document.

Now, the other thing is just a brief expansion on the point Dr. Portier made about the difference between hypothesis generation and hypothesis testing.

There are times when you really can't do what he suggests. A good example is in animal tests of - a long-term animal test of a new potentially toxic agent where you have the results of the animal test and that's all you've got. That's the only game in town. And you have to use those data in whatever way you can, then, to come up with estimates in testing the hypothesis that there is an effect. Thank you.

CHAIR HEERINGA: Dr. Reif.

DR. REIF: Just a few comments about the general writing of this section on the use of epidemiology in risk assessment.

I do believe that the whole section could be strengthened, it could be more explicit. The definitions of various terms I think in places, could be improved upon using standard references to be very clear about confounding, effect modification, etcetera.

One place where I think this document deserves some substantial exposition is in the issue of exposure misclassification, because this is bound to be a predominant problem in all of the epidemiologic efforts that have been made in the past and they're going to be made in the future.

So, a thorough exposition of the issue of misclassification, differential and non-differential misclassification, the effects of those errors on the risk assessments, including probably a table as also one can find in standard text, would be very, very helpful.

Differentiating misclassification of exposure from misclassification of confounders would be helpful. Differentiating non-differential misclassification of a dichotomous variable from a variable with multiple levels of exposure would be helpful because these, again, are very, very prevalent issues in doing any kind of environmental epidemiology, and continue to be one of the major sources of error in whatever direction the error occurs away from the truth so that the document could be improved substantially in that area.

Another general comment in places in the document that the writers used terms like "ecologic" and "retrospective," that is really an improper characterization of epidemiologic research.

The term "retrospective" has a

number of meanings and applications in epidemiologic research. For example, in case control studies which are usually thought of as retrospective, there is sort of a pejorative tone in parts of the document that suggests that case control studies because they are - because exposure is often ascertained retrospectively, are somehow inferior. And that is really an over-simplification of the strengths and weaknesses of case control studies.

Well-done case control studies that pay attention to various forms of bias, selection bias, potential confounding and other forms of bias, are extremely informative and have been used extensively by epidemiologists in a variety of arenas for many years. And, in fact, are probably the most commonly performed form of epidemiologic research.

So, I think that the writers should be very, very careful about the use of terms like "ecologic" and "retrospective." Lumping those two terms is really a distortion that leads to interpretations that are not sound and not based on good epidemiologic practice and research methods.

CHAIR HEERINGA: Dr. Chambers.

DR. CHAMBERS: I'd like to respond to the second point Dr. Bailar made and respond from the standpoint of an experimentalist with respect to hypothesis-generating and hypothesis testing experiments.

In a well-designed animal study, the only thing that should be different should be the chemical of interest if it's a toxicology study. And there shouldn't be a lot of other factors going on, confounders and so forth like that.

So, I understand your point, but I think the epidemiology point that Dr. Portier was trying to make is an entirely different situation when you've got all those other confounders involved in human populations.

CHAIR HEERINGA: Any comments?

DR. REIF: That comment really emphasizes the need to be very explicit about confounding in this document. Because to the non-epidemiologist that thinks about epidemiologic research, confounding is probably almost the first word that comes to mind.

So, a very precise definition of "confounding," the requirements for confounding - actually, both the exposure - relationship to the exposure to the outcome, the causal pathway, are all important elements of confounding.

And what the worry is, is that we have in epidemiology, always the possibility of unrecognized confounding and unmeasured confounding, and those are the really more pressing issues. Because I think the ones that we recognize and that we take pains to incorporate into studies and into analyses are dealt with, can be dealt with effectively.

It's this uncertainty about the residual unmeasured confounding that needs to be also emphasized in the document. Because in many epidemiologic studies, that may be the area that we don't address adequately.

But I think a very precise definition to avoid this tendency to talk about potential confounders as sort of inherently biasing a variety of epidemiologic studies, in fact, all epidemiologic study designs, is a dangerous - perhaps a dangerous misconception that the author should attempt to correct in the document.

CHAIR HEERINGA: Dr. Gold.

DR. GOLD: I should have started my comments, I think, by commending the EPA on undertaking this effort because I really welcome the idea that you're willing to consider epidemiologic studies in your risk assessments. I think it's really important.

So, I think our comments need to be taken - because we've all been really critical, and need to be taken in that context that we really want - I think these documents tend to take on a life of their own after they get finalized. And in the spirit of making it the best document possible, I think that's what we're trying to do.

But I think speaking as an epidemiologist, I'm really happy to see that human studies are going to be considered with all their faults and limitations and so forth. But I do want to expand on one other point that - it's actually two points that have a similar sort of conclusion.

When we try and include - it's almost impossible to include a totally representative sample in our study, but the goal of representativeness is so you can generalize the findings to a larger population.

And this ties into a different point which was made earlier about distinguishing confounding from effect modification, and I specifically want to address effect modification in terms of susceptible populations.

So, as epidemiologic studies are considered, and they're considered in terms of their generalizability, I think they also need to be considered in terms of whether they are representing subgroups of a population that might be at high risk.

So, the example of the AHS is a good one in this regard. I think it's a fantastic study. It's going to answer a lot of questions. However, it's based in two states that are relatively homogenous. I said relatively, not completely.

But I think other studies need to be considered in terms of could you expand what you know about susceptible populations. And, again, using the existing cohorts or developing new ones, even, ought to be considered.

So, I think the point of effect

modification feeds into this generalizability issue a bit. If you want to examine studies that you can generalize to wider populations and look at susceptible subgroups, then additional populations need to be included in some of these studies.

CHAIR HEERINGA: Steve Heeringa. And just to follow up on Dr. Reif's comments in teaching these subjects to many graduate students who are going to be dealing with these issues, this whole issue of confounding, moderation and mediation, terms that are kind of, the latter two, increasingly creeping into the literature, I think it would be very good to delineate specifically in this paper what we're referring to there, because people scramble those things up and they're very, very different.

And so I think it's a fairly small enhancement to this, but to make that clear from the beginning so we don't sort of just lapse into the use of the term "confounding" for things that are actually mediating effects and proxying them.

Dr. Bailar.

DR. BAILAR: I'd like to offer a rule of thumb. I don't know if it's even precise enough to find a place in this document, but it's something to think about.

After a lot of years looking at a lot of studies, it seems to me that almost always if adjustment with a first group of confounders doesn't make much difference, then adjustment with more confounders and better data on the confounders is not likely to make any difference as well.

If your first rough adjustment makes a substantial difference, then getting more data and more confounders is likely to make a further change in your estimate in the same direction.

That might be a way to sort of sort out where you want to put your next efforts. Focus on the ones where it looks like there's really something going on with the confounders.

CHAIR HEERINGA: Dr. Bove.

DR. BOVE: I'll just say one more thing. I agree with Ellen and I really appreciate the EPA is emphasizing epidemiologic studies in risk assessment. And I think examples of where they've been used in other areas like, for example, the TC risk assessment, the dioxin risk assessment, were really helpful.

But I guess I also feel that it would be good to have the same critical approach to the tox literature and tox research as you seem to have with epi research. This is just a thing I have because I have to deal with it at my own agency over and over again, toxicologists interpreting epi studies, but usually not the other way around, epidemiologists interpreting tox studies.

And it would be good for epidemiologists to interpret epidemiologic studies, because I think they can better evaluate those studies than toxicologists, and vice-versa. I think toxicologists often do a better job of evaluating their own research than epidemiologists do. At least that's how it is in my agency.

CHAIR HEERINGA: Okay. I think what I'd like to do at this point, is to call a break for 20 minutes, and we'll reconvene at 10:20. And at that point - oh, Dr. Lowit.

DR. LOWIT: I'd just like to make one point that loud and clear we hear that a certain section needs a major rewrite.

To the extent that many of you, if not all of you, had a lot of very specific points, I mean if this is a long response or even a very bulleted list of all the things, I mean every detail that came up today was, you know, I gave up with the notes a long time ago.

CHAIR HEERINGA: Right. You shouldn't have to do that.

DR. LOWIT: So, if the report can be very detailed and very specific, that would be helpful.

CHAIR HEERINGA: I think that you can assume that that will be the case. And I think as Dr. Bove started his comments out, he didn't want to provide every last comment, but we will provide those bulleted. And if it gets down to Line 5 punctuation, we'll give you that too in the appendix, but we won't do that here.

And so, yes, rest assured that we will do that so that you get not only the general guidance, but also specific guidance.

DR. LOWIT: And a couple of individuals, including yourself, suggested specific places where definitions need clarification or there's some specific nuance that you felt that was important, and really make sure that those points are also -

CHAIR HEERINGA: We'll give citations, too, to help with that. Dr.

Portier is capturing that.

Okay. Let's take a 20-minute break, and we will return and move on to Part 1.2.

And, again, for the panel and for EPA staff if there are other questions or other comments that occur to people as we move through this on 1.1, we will have a chance to return to everything at the end, too.

(Whereupon, the above-entitled matter went off the record at 10:03 a.m. and resumed at 10:22 a.m.)

CHAIR HEERINGA: Okay. Welcome back, everyone, to the second of the morning session of our second day of the meeting of the FIFRA Science Advisory Panel.

We have completed the panel's initial discussion of Charge Question 1.1, and we're turning to Charge Question 1.2 if Dr. Lowit can find her controller.

Jeff Dawson will read Question 1.2 into the record.

MR. DAWSON: Question 1.2. Section III of the draft framework describes the major sources of human incident data along with their strengths and limitations. Section III also describes ways to use human incident data in risk assessment. Please comment on the soundness and completeness of these discussions. Please include comments on additional factors to consider when evaluating the quality and weighing the utility of human incident data in risk assessment and characterization.

CHAIR HEERINGA: Our lead discussant is Dr. Chambers. Jan.

DR. CHAMBERS: Well, this had to be quite a challenge to deal with these isolated incidents.

Incident reports are usually high dose, frequently illegal or accidental exposure incidents. So, they would not be reflective of normal use exposures.

The several sources of incident

data are varied substantially in their completeness, level of description and geographic scope. And in my opinion, EPA has evaluated the utility and reliability of these five sources well, and seems to have made an adequate judgment of the value and usefulness of these sources.

Since the incident data are frequently of limited detail and are largely the observations of non-medically trained individuals, these data are of relatively limited usefulness.

The observations in incident reports are usually short term with only a little amount of follow-up of incidents available. Especially when the risk assessments need to be made on longer term adverse effects, the incident data are probably of relatively little value because I think your regulatory processes have pretty much protected people, ag workers and so forth, from the high-dose, short-term effects.

Exposure estimates in these cases are probably very limited quantitatively and of limited reliability. In addition, the incidents are reported on products, not single compounds. So, the possible interactions or synergies of the main active ingredient with other chemicals are unknown.

In addition, there's probably little, if any, information available in incident reports to indicate what other factors or confounders, and I'm not sure I've used that term correctly now, but what other factors might have been present that might have contributed to the symptoms reported.

Therefore the incident reports, in my opinion, are of very limited value to the risk assessment process.

A special concern to me, anyway, is the report of symptoms by medically untrained individuals. Such descriptions of symptoms may be highly problematic and that those reporting on the incident may not be attune to the types of observations that should be made.

When an incident occurs that is an accident and it's potentially high dose, it would seem very likely that the affected individual would be very scared, in my experience, and report signs and symptoms that are more physiological reactions to fright than to the mechanistic effects of the chemical.

Caution is urged and the conclusions drawn on symptoms that could be attributed to physiologic stress reactions if those are not consistent with the plausible toxicological effects for a chemical.

Also, classic flu-like symptoms are frequently cited as an adverse acute consequence of exposure to some pesticides, and those could also be attributed to some non-chemically induced causes.

The uses cited for incident data, that is the need for changes in risk management, monitoring success in mitigation measures, targeting enforcement activities, are all reasonable uses of the incident data assuming that the incident data are carefully critiqued for reliability.

This is probably their greatest value in the overall risk management process since these typically high, accidental off-label occurrences are of relatively little value in the risk assessment process.

CHAIR HEERINGA: Thank you, Dr. Chambers.

Dr. Reed.

DR. REED: In terms of coverage in this section, the coverage on the five databases and their brief description, I think, is sufficient for presenting their main characteristics in the context of their usefulness or not for risk assessment. And I find Table 3 very useful as a summary.

The description of this section present the toxicity data tally, seem to be focusing on the severity ranking. And I think it's not until the diazinon example that the end points, the importance of end points came out.

And I think it would be good to bring that to the front because as a risk assessor, we would look at it and we're interested in the most sensitive end point. And that turned out to be the lowest ranking in terms of severity, and so it didn't come out until later.

You can either just point the reader to that section and, you know, not to explain too much on this.

The other thing, and I think it was mentioned with the epi study in the same way, that it's good to just point out the importance of looking at human data in the context of what's known about or not about mode-of-action. So, you get a sense of how it fits in into the entire risk assessment process. And mode-of-action, including all the related in vitro data. I think there's in vitro data that comes into play to line all the information up.

There's a couple places where it was mentioned that the data collecting agencies defend their separate data analysis, but I think it would be good to put an example of what kind of analysis that they did and what kind of conclusion they come up with. Sort of get a sense of how other people use these databases. And of course it would be most interesting if it's related to pesticide exposure and risk.

In terms of California data, the PISP, and I have not done that personally, but I always wonder - we're talking about the lack of information in the incidents report. And I was wondering if the agency could look into pulling in the used data to take a look at the incidents report and see if you can get more confirmation/information out of the used report in terms of what they were exposed to and that kind of stuff.

As a risk assessor, this is great. I mean we're looking at human data. But as a risk assessor when I was looking at how to use incidents data, I cannot help but to thinking that if I get a good job, my job is right and risk assessment being predictive, then you're not going to see much of the incidents except for accidental and misuse and intentional.

So, from that standpoint, I am looking at incidents data not just as a group, but I think I get the gist that we're looking at incidents data like the way we look at epi data, which is not, and it's not.

So, I'm more interested in actually picking through the incidents data. I know a majority of them there's no follow through and with all the deficiencies that Dr. Chambers had mentioned. But there might be a few data sets that you can follow up a little bit to see if there's enough information to bring it into consideration in risk assessment, so that you could maybe glean from it either how to modify, how to improve, how to - or to confirm the certainty or uncertainty of risk assessment.

And, again, I'm coming from the standpoint of hopefully these are not to happen. But if it happens, I'm all perked up. I want to know what happens to these cases where there's enough follow-up.

So, It's not necessarily confining ourselves in looking at the incidents data as a whole, as a whole group, but possibly gleaning something out of it that could be useful for risk assessment, sensitive subpopulation, vulnerable population, that kind of information.

CHAIR HEERINGA: Thank you, Ruby.

Dr. Gold.

DR. GOLD: Well, I want to say I agree with virtually everything Dr. Chambers said, but I do have a few things to add just in terms of considerations that might be included in the document in terms of the utility of these reporting systems.

So, one thing in terms of reporting systems in general is if this is mandatory or voluntary. And if it's mandatory, who's required to report, because who reports also affects the under-reporting.

So, I mean what we're concerned about, one concern that comes up in using these systems is the under-reporting, and that's affected by a lot of things. Is it mandatory or is it voluntary?

Also, is there sort of active reporting or passive reporting? In other words, does the agency actually actively go out and seek these reports or do they just passively wait for physicians or whoever, registrants or whatever to report? This will affect under-reporting as well.

And in terms of how the data can be used for trends over time and so forth, I think in terms of weighting the utility, so how useful it is, I think, is how much it's tied to - how much the incident reporting is tied to pesticide use over time, for example, because it could be that you have increased incidents for lots of reasons.

It could be increased usage, it could be better reporting for one reason or another, it could be the population has increased, it could be a sensitive population. There are lots of reasons.

And so to the extent that you can link these data with other pieces of information like that, I think you have the potential for increasing the utility. But I think it's significantly limited, as Dr. Chambers indicated.

CHAIR HEERINGA: Thank you, Dr. Gold.

Dr. Meek.

DR. MEEK: I don't have very much to add. I just wanted to mention that in fact the human incident data, it can be used in risk assessment to a limited extent. The human case reports and surveillance of acute poisonings are quite helpful in considering similarities in site concordance between animals and human in mode-of-action analysis.

So, the concordance table that Dr. Lowit showed yesterday, in fact that information has a place in a qualitative context in terms of looking at site concordance. So, I would suggest that the use of these data really isn't restricted to hazard identification as indicated, but they also play a role in hazard characterization.

So, I think that's all I would add.

CHAIR HEERINGA: Thank you, Dr. Lowit. Other members of the panel?

Dr. Bailar?

DR. BAILAR: I'd like to add just one point. I do not recall that there was anything in this draft document that dealt with reports of clusters.

Three pesticide applicators who happen to work in the same field at different times come up with brain cancer. 17 workers in a pesticide plant had kidney failure when population rates would suggest there shouldn't be more than five.

I think it would be worth adding a paragraph or two about these. Cluster reports are notoriously difficult to interpret, and you might just say something like that. The problems of reporting bias, a cluster doesn't come to attention unless it's outstanding. You never hear about all the places where that didn't happen.

Just as a sideline, I remember many years ago I saw a report of a particular complication with a blood transfusion in a hospital. Somebody calculated the probability of that was one in 7,000.

What struck me at the time was that I happen to know there are about 7,000 general hospitals in the U.S. None of the others ever reported this. They declined it.

So, the one in 7,000 event probably occurred with exactly the expected frequency. The one in a million event does occur with the expected frequency.

So, you might want to add something about clusters.

CHAIR HEERINGA: Dr. Chambers.

DR. CHAMBERS: I guess you mean clusters with respect to epi or with incidents?

DR. BAILAR: I'm talking about clusters that are reported probably through the incidents mechanism, but where you're not dealing with an individual, a report on an individual, but rather a report on a group of people who have something in common like an exposure.

DR. CHAMBERS: But my understanding of these is that it is on individuals that are showing up in Poison Control Center records and that sort of thing on some particular exposure -

DR. BAILAR: Well, what should EPA do if they find a report of a cluster? This falls - in my view, this falls in sort of the same category as incident reports.

DR. CHAMBERS: It wouldn't be - I don't think it would be the same database though, would it? It would show up in some other way.

DR. MANIBUSAN: So, just to clarify, if there are multiple cases of adverse events from the same area, that could show up in our IDS system, which is a voluntary system. It's mandatory for registrants to submit that information to us, but it's voluntary in the sense of the person who's reporting it to the registrant.

DR. CHAMBERS: Would it be something as the example that Dr. Bailar used as brain cancer or something like that? These are more acute reports, aren't they?

DR. MANIBUSAN: Right. So, the

incident information that we often get is more from acute exposure, acute symptomology, not on chronic affects like brain tumors or chronic toxicity.

DR. HEERINGA: Dr. Harris.

Dr. Harris: Shelley Harris. Maybe I'll just jump in here. We just need to make a distinction between clusters or poisonings and clusters of cancer. And typically if you'd have a cancer cluster, someone would call your local health department and they would call the various agencies involved and ask for an investigation.

So, I think that cluster investigations might be more appropriately included in the overview of epidemiologic studies in the first question and more fully described in that area.

CHAIR HEERINGA: Dr. Lowit, and then I'll come back.

DR. LOWIT: Just to add a point to add to what Mary said to clarify, there have been situations where we have in the past, where we have seen clusters where there are acute events. There's an aldicarb in watermelon example. And there was a field example with some female workers who had reported birth defects, I can't remember the chemical, from a few years ago.

So, there have been clusters of events, some of which you may have picked up in an incident because of an acute event, let's say, from a carbamic, for example, but then there's maybe some intermediate term end points, but then obviously if you have a cancer cluster.

So, it fits across different modes of action by duration.

CHAIR HEERINGA: Dr. Portier.

DR. PORTIER: Actually, this discussion is good. And recently I was looking at California's cancer cluster plans, and cancer clusters are handled really well. It's these more acute things.

What I wanted to do is get back to Dr. Gold's issue about vulnerable populations. So, when you're analyzing these things, what if you notice that the people who are calling in who are having these effects are all field workers, minority, you know, kind of the vulnerable, unprotected class?

I mean is there discussion in there about identifying vulnerable populations, which is another part of hazard identification, right, and finding those.

So, that's not geographical clusters, per se. That's more demographic clusters or demographic ID or something of that type.

CHAIR HEERINGA: There was a question, so I'll -

DR. MANIBUSAN: Let me try to take a stab at answering that. We have not yet started to look at using incident data for environmental justice issues. We've tried to do that.

We have some limitations, of course, because many of these databases do not include things like zip codes where we can easily query.

We have to the extent that the data is available to us, looked at the differences between adult versus children's adverse reporting of cases.

And in situations where we notice that children are predominantly reporting adverse symptoms either because they have direct access to the product, we can do things to reduce that risk in terms of special packaging, things like that that we've done in the past with incident information.

CHAIR HEERINGA: Jeff Dawson.

MR. DAWSON: Just another example to follow on with what Mary said in the risk management area related to workers. Historically we've seen clusters of incidents related to, for example, field workers going in too early to harvest certain crops and certain chemicals. So, right away that would be kind of a red flag for us to kind of alter the risk management approach for that. And there have been notable examples over time with that.

CHAIR HEERINGA: Dr. Reed.

DR. REED: Jeff, when I mentioned vulnerable and sensitive subpopulations, age was one other thing. There's greater frequency of showing up in incidents report with children.

But also vulnerable people may be an applicator or a group of people that could be used to both, again, as indicated, go back and look at the exposure assessment, look at the risk assessment, different components of risk assessment.

But also in terms of risk management, you can cater your education, information, dissemination of a group of people who have a tendency, you know, greater tendency to not using the pesticide right or

even not being able to read the label right
because of ounce, gallon, units like that.
So, that's what I'm thinking of
when I say vulnerable sensitivity.
CHAIR HEERINGA: Dr. Gold.
DR. GOLD: I want to address Dr.
Bailar's point. I think clusters are -
they're really difficult, but I think what he
is suggesting is that they get mentioned
somewhere. And whether it's here or in the
epidemiology section, I don't think it
matters.
But I think part of what I have a
little trouble about is you're talking about
clusters that show up in your reports to these
various agencies, and out in California we
have clusters all the time and often they're
not linked to anything in particular.
Often we just hear that there's a
cluster of birth defects or clusters of asthma
or clusters of this or that. And people will
say well, we think it's the pesticides that

are in the fields adjacent. It's a non-specific. So, it's not related to a particular product or agent.

And so I think adding a paragraph or two about things like this because they do occur, and how you should handle them and with great skepticism and care, but it is something that the public is concerned about.

So, I think not mentioning it is somewhat of an omission. And so mentioning it delicately and properly and not always oriented to a particular product. Okay? Because that's not how they always occur.

I mean sometimes the community is just concerned and they just see a lot of something, and they're not relating it to a specific agent in particular.

CHAIR HEERINGA: There's a nice paper, and I believe it's by Persi Diaconis, on this whole issue of how humans sort of over-interpret clustering of events when in fact - the randomness of those cluster occurrences. And I'll try to throw that citation in. It's pretty accessible because I think it was published in Scientific American or something like that.

So, it's just a citation to throw in there on this whole issue of how to address clustering.

Dr. Reif.

DR. REIF: Yes. There's also an issue of the American Journal of Epidemiology in 1992 that devoted a whole issue to analysis of clusters and cluster busters, and that also would be a useful reference.

CHAIR HEERINGA: Dr. Bove.

DR. BOVE: Let me just say one thing about clusters. First of all in this database, you will be able to link particular products with the cluster. So, actually this is a little different than the usual circumstance where people say there's a lot of cancer on my block and we're not sure what's around.

But there have been cluster investigations that have ended up into full-fledged epi studies, Woburn, for example, using very sophisticated water modeling to determine exposure.

So, cluster investigations can move towards a full-fledged epi study and that needs to be - that will be mentioned in that.

CHAIR HEERINGA: Okay. Mary.

DR. MANIBUSAN: I just want to ask maybe for some clarification, it sounds like there is a recommendation for us to do more active surveillance. So, realtime information.

Right now we don't currently have access to realtime incident data inputting, so we purchase PCC data, for example, in two-year increments, and there's a lag of about three years for data quality. Our incident data from our registrants are submitted perhaps quarterly. It's not on a day-to-day basis. We don't have that active surveillance system.

And I think there are articulated very clearly right now, our incident information is reported to the agency in paper copies.

CHAIR HEERINGA: The question -

DR. MANIBUSAN: So, some suggestions about that.

CHAIR HEERINGA: The question to the panel, I think, is should this be done or not. I would add is it feasible both in terms of its timeliness and the cost that it would take for the benefit that's reaped.

Dr. Bailar.

DR. BAILAR: I was not meaning to suggest that there should be any expansion of that particular effort. I think you should continue doing what you're doing. It's potentially useful. You will certainly come under criticism if you don't do it, but I wouldn't look to that for big answers to big questions.

CHAIR HEERINGA: Other thoughts on

that particular issue? Hopefully that clarifies. I think there seems to be consensus.

Yes, Dr. Hayton.

DR. HAYTON: Well, this discussion makes me think about the fact that the systemic exposure to a chemical is not only the external exposure, the applied dose, but also the clearance of the chemical.

And that when you talk about susceptible populations, people who are missing the gene that makes the enzyme that metabolizes, say, the chemical, how do you - I don't know what the answer is, but how do you pick that up? And it's sort of a characteristic of the population looking for sensitive individuals.

For example, cytochrome P450 2D6 is a classic example. And if you're using a drug in therapy and somebody is a poor metabolizer, about ten percent of Caucasians I think is how it works, are deficient in that enzyme, then - in therapeutics, we're trying to identify those people beforehand and adjust the dose accordingly.

But is there a role in monitoring for that?

DR. MANIBUSAN: Right. Now, you bring up a very good point.

I think it's very challenging to think that we can use incident data to identify susceptible populations on molecular basis.

I think it calls back to the framework analysis. The framework really provides an opportunity to integrate what we know about a chemical through its mode-of-action, through its human relevancy. We can make some determinations about what we're seeing in the animal and how relevant it is to humans.

In our risk assessment, our risk assessors around the table, we have uncertainty factors to account for some variability that we might not be able to anticipate in our end points that we're seeing in our animal studies.

We have an intra species uncertainty factor to account for that, but it is a very big challenge for risk assessment to try to identify sensitive subpopulations with differences in CYP induction, with differences in metabolism.

CHAIR HEERINGA: Thank you, Dr. Manibusan.

Dr. Bailar.

DR. BAILAR: Dr. Hayton mentioned the deactivation genes. There also should be some concern about activation genes with this vinyl chloride and some other things. And it's worth noting that each of those bends the dose response curve in opposite directions, but their source is non-linearity, which might be important.

CHAIR HEERINGA: Dr. Chambers.

DR. CHAMBERS: I think both those

points are very important. But this question is about incident data, and I don't think you're going to get anywhere near enough information to make any judgments about that. Perhaps not in some epidemiological type studies.

But with respect to this particular question, I don't think you're going to get information like that.

I would have to question how much - you're going to have to do a time benefit analysis. Certainly I know you're stretched to the limit on an awful lot of activities and whether or not putting a lot more effort into this activity is going to get you information that's of value where you could put energy into other activities, you're going to have to make that judgment.

CHAIR HEERINGA: Okay. Dr. Manibusan, Dr. Lowit?

Let's move on then to Part 1.3.

DR. LOWIT: Section IV of the draft

framework describes a proposed WOE approach for evaluating human and experimental animal data from in vitro and in vivo studies. This proposed approach makes use of the "source to outcome pathway" and the modified Bradford Hill criteria like that in the MOA Framework, as tools for organizing, evaluating and describing the human health consequence of a particular chemical based on available data. Please comment on the proposed use of the modified Hill criteria in the context of the source to adverse outcome pathway for integrating a variety of types of data at different levels of biological organization including human incident and epidemiologic data in risk assessment.

CHAIR HEERINGA: And Dr. Meek is the lead discussant here, and I think also for 1.4.

DR. MEEK: Right. They're a bit similar, 1.3 and 1.4, so we may cover most of it currently.

CHAIR HEERINGA: We'll just stick with 1.3 and -

DR. MEEK: Okay. Well, in my view, use of the source to adverse outcome pathway and the modified Bradford Hill criteria is extremely helpful not only as a basis for organizing, evaluating, describing the human health consequence of a particular chemical based on the available data, but also in identifying critical data gaps.

And so I really think that the agency is to be commended on their pioneering work on the framework, its contribution to transparency and risk assessment generally and I really am very supportive.

I also think that thinking in this context is important in transitioning our focus in toxicology and risk assessment from delayed adverse effects that we normally consider to earlier biomarkers of exposure and effect so as to collect more informative human data at relevant dose levels. So I think it, again, it serves really a dual purpose.

The framework is also helpful in directing attention very early in the assessment available data to dose-response relationships for early key events, and it is these dose-response relationships that are critical in the subsequent risk characterization.

So, I believe that the source to adverse effect pathway and framework offer significant potential that transparently and appropriately integrate human and toxicological data as proposed in the documentation.

That said, and there's always a but, in my view, however, there is clear benefit to be gained in more clearly distinguishing the qualitative and quantitative aspects of mode-of-action analysis and human relevance as a basis for integration of human data and subsequent dose-response characterization.

While preexisting epidemiological and incident reporting can be helpful in hazard characterization, unless we're able to robustly access exposure or more appropriately incorporate biomarkers of exposure and effect based on identification of key events in a mode-of-action context, the contribution to dose-response characterization will necessarily be more limited.

I think that's very nicely characterized in the source documentation in Figure 1. And so the kinds of data that we're looking for in an epidemiological context are pretty clearly highlighted there.

I also think that consideration of the human and toxicological data and the context of the framework contribute to conservation of resources. For example, lack of adequate characterization of exposure-response relationships in epidemiological studies may preclude the need to do an extensive weight of evidence analysis for these data since they cannot contribute to the risk characterization.

So, you would actually have to change the order of the way that you look at different types of data. And I think I'll come back to that at the end of my comments, because in large measure I think this relates to how you might meaningfully use problem formulation.

So, another point I want to make, the value of framework analysis and coordinating assessment in research has not been emphasized in the documentation.

For example, there is repeated reference to problem formulation, but without indication of how the broader toxicological and epidemiological databases might be considered at this stage in integrated fashion as a basis to identify critical data gaps to inform the assessment.

This would be the appropriate step in my view, for example, to identify limitations in available human data in the context of the overall database, as a basis either to focus additional research or at least to increase understanding of the likely contribution of the existing human data in the context of the overall database, and that's critical.

The appropriate human data might include in vitro studies in human tissues or cell lines and perhaps very focused epidemiological studies to address the specific questions and identified subgroups through consideration of early biomarkers of effect.

A couple of points as well, these are more specific about the criteria used in the framework specifically, it's important to recognize that the criteria used in the framework here are those that relate principally to weight of evidence rather than consideration of individual studies. And that's the difference between the Bradford

Hill criteria as applied to the consideration of causality and epidemiological studies and their consideration in this framework.

And that's appropriate, because what you're doing is looking at the weight of evidence.

Based on increasing experience and application of the mode-of-action human relevance framework and to avoid confusion that this addresses exposure in any way, it's suggested to consider revising reference to dose-response relationships to concordance of dose-response relationships between the key and end events. It's not just that there's a dose-response relationship.

I also think that there's confusion with exposure, which is considered in a different part of the risk assessment paradigm. So, I think it's very important to be explicit in describing what we're actually doing in that step.

Also, we don't expect here to have

earlier key events occurring at lower doses where the data don't support the hypothesized mode-of-action. The document was not very clear on that point.

We also expect the incidents of earlier key events to be greater than or equal to that for the end toxic effect where the weight of evidence doesn't support the hypothesized mode-of-action either. So, those are more specific points to the actual documentation.

Another point is based on increasing experience with the mode-of-action human relevance framework, I think we would normally consider potential alternatives for hypothesized modes of action at the outset of a framework analysis as a basis to distinguish relevant pathways and key events in an integrated fashion. So, I think it was just something to take into consideration.

So, I have a number of more specific comments as well to the documentation, but I won't present them here because there's probably no need to.

CHAIR HEERINGA: Thank you, Dr. Meek.

Dr. Hayton.

DR. HAYTON: Yes, I agree with Dr. Meek. I thought the Bradford Hill criteria were highly appropriate and they're well accepted. And since I don't have a lot of expertise in epidemiology anyway, I didn't want to say other than it made sense to me.

One comment that resonated with me is that the criteria shouldn't be viewed as a checklist, but rather just as a group of characteristics that taken together provide a systematic way to aggregate observations.

And then an issue that came up yesterday that not all the criteria deserve equal weight, I think there's something to that. So, weighting among the criteria would be an issue.

In Section IV, I didn't see very

much about extrapolation of dose amongst species and extrapolation of high-dose toxicity in experimental animals and human incident cases to environmental exposure. In humans, it typically occurs at much lower doses, and I'm just wondering how the extrapolation would be done.

Would it be a linear, low-dose extrapolation would be the standard approach, or other approaches and issues surrounding the phenomenon known as hormesis where low doses actually seemed to provide a protective effect, and it just occurred to me that some thought ought to be given to including that issue in the framework.

And maybe it has and maybe you decided just not to put it in, but that's what stuck out to me.

CHAIR HEERINGA: Thank you, Dr. Hayton. I think the point that you raise has been a point of considerable debate and also a guideline formation within the agency.

Dr. Bucher.

DR. BUCHER: John Bucher. So, the draft framework for integration of in vitro/in vivo animal and human incident in epidemiology studies has many advantages and the agency should be congratulated for their efforts. There's really no question that establishing steps to ensure rigorous and consistent evaluation of studies of any type result in a better risk assessment.

The mode-of-action framework certainly helps in the organization and evaluation of the data. That said, it's important to remember the historical context and purposes for developing the MOA frameworks. These originally focused on cancer outcomes, and only more recently have been extended to non-cancer end points.

For decades, positive findings from animal cancer studies were assumed to be relevant for human hazard identification. In the late `70s and early `80s, research programs were begun to systematically examine the biological events that appeared to correlate with and perhaps account for the induction of cancer and the number of common sites for tumor responses in rodent cancer studies.

The original mode-of-action framework for experimental animal tumor sites and types was established to ensure that the many modes of action hypotheses were being offered up to, frankly, explain away problematic results were in fact based on solid, scientific foundation.

The second mode-of-action framework, which is Bette's framework, affectionately known as Bette's framework, was established to specifically examine the claims that certain animal tumor mode-of-actions were in fact not relevant for humans.

And this proved to be an illuminating exercise for those of us who participated, and highlighted just how much some folks were relying on assumptions that certain animal tumor types were not relevant for humans based only on the perception that humans would not be exposed at sufficient levels to get these tumors rather than on the fact that there were true differences in physiology or biology.

The most important aspect of the framework for assessing the human relevance of animal cancer findings was that the failure to establish that an animal cancer mode-of-action could not occur in humans, resulted in the default assumption that the animal cancer finding was in fact relevant for human health assessment.

So, why is this fact important? In the current proposed framework for incorporation of data from in vitro/in vivo human incident in epidemiology data by its very nature, it creates to me the expectation that inconsistent findings in any one area could lead to inaction on the part of the agency.

Consider, if you will, that uncertainty in any particular area, for example, the relevancy of a particular end point in an in vitro assay, say, in a high through-put screening assay, to the toxicity pathway that one thinks he or she is probing, for example, we know in fact as we go into the HTS programs, that we're tying to pick out targeted enzymes or targeted parts of toxicity pathways and probe those in ways that are meaningful. We're not sure, in fact, if those probes are hitting critical parts of those pathways or parts of those pathways that have a lot of play in them, for example.

Or if you look, for example, at human - or at animal cancer data, there are a number of end points that could show up as clearly positive, strong outcomes in animal cancer studies in, say, the Harderian gland or the Zymbal's gland or the forestomach, targets that do not have a clear human counterpart.

So, one might consider that in fact these kinds of things are equivalent to the confounding in an epidemiology study. And I enter into that based again on the same fear that you had after talking - your earlier comments.

But confounding as we all know in epidemiology studies, is often used to explain away findings, when in fact we know that it can also mean that the true signal is stronger than it appears because it's basically fighting through the fog.

So, this is the main issue that I think EPA needs to be aware of and guard against when attempting to bring to bear all these different types of data in reaching public health decisions.

By bringing all of the relevant data to the table, the EPA cannot raise the bar so high that nothing is recognized as a threat to public health.

We heard yesterday assurances that

strong epidemiology signals wouldn't be ignored, but having more data and having more data of different types with different associations and strength of associations with the actual outcome that we're evaluating, can as easily lead to confusion as lead to clarity.

It's still going to come down, I think, to professional judgment of the strength of the data in the separate areas before a decision can be reached on the collective cohesiveness or biological plausibility of the data set that you're looking at predicting a human health outcome.

CHAIR HEERINGA: Thank you, Dr. Bucher.

Dr. Chambers.

DR. CHAMBERS: I don't have much to add and it certainly won't be as articulate as that was.

The opinions I had on this to add anything to it is that I think that in most cases the epidemiology and the incident data will be mostly apical end points and will not look at the intermediate steps like the key events and the dose-response for the key events and the pathway. So, I think the information you'll be getting from the epi studies would be confirmatory for some of the animal research on mechanisms and action and so forth.

For the most part, I don't think you're going to get - I think the Bradford Hill are very good criteria, but I don't think you're going to get a lot of that information out of the epi studies that you don't get out of the animal studies.

CHAIR HEERINGA: Thank you, Dr. Chambers.

Dr. LeBlanc.

DR. LeBLANC: Most of my points have been covered. So, I'm going to try and avoid redundancy here, but I do want to echo Dr. Hayton's point that while I feel, as everyone else dose, that the approach is very, very appropriate, I think that the various parameters within the Bradford Hill criteria have different weights and should be - these different weights should be applied when organizing the information and making judgements as to how the individual pieces of information should be used.

With reference to a couple of the criteria, one being consistency, it was noted in the framework that human and animal responses may not be consistent, and that certainly is the case.

And as I read it, the resolution to those situations would be to identify the most sensitive end point in the animal models, and to use that end point in decision making to assure protection of humans.

And I would suggest that perhaps as an alternative or in addition to that, one should simply look at the chain of events that occur that's pretty much identified in establishing key events, and taking one step back on the rung.

So, for example, a chemical is inhibiting an enzyme. And in the rodent models, it's causing cardiac arrhythmia. And in humans, it's causing really nasty headaches, and you can't measure really nasty headaches in the rodent models.

Cardiac arrhythmia itself may not be appropriate, but you can take a step back and look at concentrations when that information is available, as to what level of enzyme inhibition causes the effects in the rodent models and what levels of enzyme inhibition occur in the human models, and make decisions that then sort of titrate with the results derived from the animal studies for the protection of humans.

I recognize that's difficult to do simply due to the lack of information most often in the epi studies.

Another point relates to the fact

that the epi studies may indicate other modes of action. You set up this paradigm, you have biological plausibility and everything works really well, and you have some well-designed epi studies that show effects, but they're not consistent. They question the - there's no biological plausibility to the effects that are observed.

And I think that's very important information. It may be of limited information from a risk assessment standpoint, but it certainly can't be ignored. I think what it's doing is just opening up new areas of investigation.

The epi studies perhaps could be categorized as - well, they could be categorized many ways, but there are certainly epi studies that are not good epi studies. And presumably, they would be triaged in the process early on.

There are other epi studies that are good studies. They're just not predictable. They gave results that were unanticipated. A hypothesis was set, and the hypothesis wasn't supported, but good information is derived and I think that information is important.

It may suggest that based upon the Bradford Hill criteria, that certain box of confidence was developed and that this information falls beyond the box. It may indicate, then, perhaps that box needs to be widened.

So, from a regulatory standpoint I recognize that it is limited, but certainly can't be ignored. I think it has to provide guidance to new hypotheses, to new studies, to new analyses of existing studies. That's all I have.

CHAIR HEERINGA: Dr. Bailar.

DR. BAILAR: I don't have much to add. One thing is to emphasize something I mentioned yesterday, is the original purpose of these criteria was to bring some order and rationality and interpretability to a very difficult art, and it is an art, of trying to interpret observational data.

Now, Dr. Weed has pointed out that even that Surgeon General's report was not the first in the field. He has written about it in his textbook, the prior history, but it is a very difficult kind of thing to do.

Hill himself did not stick rigorously to this criteria. He modified them in response to specific issues that came up in things he was looking at. They have to be interpreted flexibly. And as Dr. Hayton said, it should not be taken as a checklist.

I really emphasize the need for flexibility in applying these. Add things, subtract things, modify them further as needed to suit a particular problem.

Another point is that on Page 28 you refer to the literature search, but you do not say how that search will be organized and conducted. And it does not refer to the necessary screening of papers to find those that have some merit for whatever your present purpose is.

I've written somewhere else about the characteristics of data that I've noticed in big, public problems. One is that the data tends to be vast. Just enormous amounts of stuff available.

The second is that they tend to be highly complex in that there are all kinds of issues. They often involve many, many different kinds of scientific and technical expertise. You have to have a way to deal with that complexity.

The third is that almost everything you'll find is of poor quality. It has to be screened out or at least weighed so that it's severely down weighted.

And the fourth is that it's often not what you want anyway.

In the literature search, I think you might want to deal with this and bring out points, not necessarily exactly these, but how the literature search should be conducted and how the results should be interpreted and dealt with.

The third point is that you refer to the postulated mechanism of action, but you don't really deal with the difficulties of determining the mechanism of action with reasonable certainty, or how the residual uncertainty should be dealt with in the analysis.

A further problem is that there may be special difficulties when there are two or more outcomes of concern that seem to have different mechanisms of action and how you're going to deal with them. Thank you.

CHAIR HEERINGA: We turn now to other members of the panel for any comments.

Yes, Dr. Greenwood, please.

DR. GREENWOOD: I think the use of the Bradford Hill, as he called them, I think, review points rather than criteria, it is a very sound way of approaching what is a very complicated set of systems.

But one of the things that hit me yesterday listening to the input that we received, was that if you were to look at plausibility before you looked objectively at the outcomes of your assessment of studies which you are looking at, that it could actually change your objectivity.

And that is a danger, I think. And it was brought home to me by the way that people were looking at the atrazine data and the responses in that.

I think that it's very important to look at the plausibility, the biological plausibility, but I think that probably needs to be looked at after an objective assessment of the value of the study rather than before.

Because none of us wants to do work that we don't really have to do, and it's very easy, actually, to throw something out if you've got a good reason for not looking at it too carefully because it didn't seem plausible. And I think you might miss important factors if you were to look at the plausibility and let that influence your assessment of the data that you're looking at.

CHAIR HEERINGA: Thank you, Dr. Greenwood.

Dr. Reed.

DR. REED: I want to say thank you, Dr. Greenwood, for bringing this up. I think this is the area that we are most afraid of. It's predetermining what is biologically plausible, what is not.

And then in the next step we have the other mode-of-action, and so the two might come working against each other. You already write it off on the plausibility, and then you have a new mode-of-action of possibly the new manifestation of the same mode-of-action that may be different between animals and humans, and you write that off too.

So, I really appreciate that

comment. I think it's important to be objective about plausibility.

CHAIR HEERINGA: Dr. Reif.

DR. REIF: Just not a disagreement, but a note of caution because in epidemiology, I think most of us try to do hypothesis-based research. And part of our - the formation of epidemiologic hypotheses does in fact rest on biological plausibility.

So, we find ourselves in a bit of a conundrum if we say to ourselves well, I'm just going to go and do an exploratory data dredging without having the recognition of at least some awareness of what happens in animal systems.

So, I just throw that out as a note of caution on the biological plausibility issue.

CHAIR HEERINGA: Dr. Reif, on that topic, and also Dr. Reed, Dr. Reed mentioned there is the sequence. There is the biological plausibility criterion or element.

And then follow that, I assume if you reject, then you move on to is there another mode-of-action which would explain, and cycle back.

Is that -

DR. REED: Mostly I think it's important to have a placeholder for something that you're looking, that you're puzzling about. It's more of a benefit of doubt kind of way of looking at data and I think it's important.

CHAIR HEERINGA: Is that appropriately placed at the end? In other words - Dr. Bove, you look like you -

DR. BOVE: I was just going to say that another word of caution is that the biological plausibility evolves over time and changes so that what we think is not biologically plausible today, becomes biologically plausible.

So, it's true we develop our hypotheses that way, but we also leave it open sometimes to evaluate the data. It's not really a dredging exercise. We still have a hypothesis. We just don't have biological plausibility for it.

For example, in the early days we were looking at disinfection byproducts. None of us thought that there was an association with birth defects or with small for gestational age. There was no biological plausibility at the time. And the research developed because of findings that way.

These were hypothesis driven, but there wasn't much biological plausibility. Since then, you know, research has developed since then.

CHAIR HEERINGA: Dr. Lu.

DR. LU: I just want to comment. I might be the least qualified person to comment on this, but I do believe that we have to lose this co-called modified Bradford Hill criteria. Because in my opinion, the principle of epidemiology was developed to study incidents like infectious disease, like which wells that contain cholera, the bacteria that cause cholera and so on and so forth.

When we deal with these issues that involving chemical exposures, especially the chemical that come and goes and to create an environment, I mean nothing really fit into the criteria.

So, if the agency is bounded to these modified criteria that, you know, it can really go out of the box and seek for other evidence.

And as Dr. Bove just suggested, and we actually talked about it yesterday, that the biological plausibility evolved.

I still remember we talk about melamine case. If you ignore the incident data, then melamine would still be able to added to the plot, but it's the incident data that actually raised the red flag and then lead to a lot of regulatory action.

So, I mean in this case it doesn't fit into the Bradford Hill criteria, and you will pretty much ignore the incidents of the melamine data.

So, I mean I'm going to talk about a little bit more about this when we get to Question 2, but my position here is that it's a criteria that for the reference. But if you can bind yourself into this, I don't think it's a wise move.

CHAIR HEERINGA: Dr. Meek.

DR. MEEK: Could someone define "biological plausibility" for me? I think there's a real issue in terms of how people are interpreting biological plausibility around the table and I think we're talking at cross-purposes.

DR. LU: Right. I mean my definition is a disease caused by certain chemical exposures.

DR. MEEK: I think it's the issue that Dr. Gold raised yesterday, and I think we used biological plausibility in an epidemiological context completely differently than we would use it in either a toxicological or a mode-of-action context.

And the criteria with which I'm most familiar in terms of the Bradford Hill criteria being applied in a mode-of-action context, really means do the data support. And in the more generic biological data, do they make, you know, does what you're seeing make sense?

So, I think we're really talking at cross-purposes in terms of biological plausibility.

I'm not entirely sure how biological plausibility could ever be used as a barrier for not considering any fact, so I didn't really understand much of that discussion.

CHAIR HEERINGA: Well, how do we sort this out?

Dr. Gold.

DR. GOLD: Well, I wanted to make a comment about that, and also about using the criteria, but they relate to each other.

I agree with the comments that were made that you don't weight all the criteria necessarily equally. And in terms of biologic plausibility, I think when these criteria were developed and published and so forth, a lot of it was around the smoking and cancer relationship.

And we didn't have the pathophysiologic mechanisms by which smoking caused lung cancer, but that didn't prohibit action in terms of public health education and smoking cessation efforts and so forth.

And by the way, you could still maybe even do trials to see, you know, preventive trials. That might shore up your causal argument. So, I think we need to be careful about how we weight them.

One other comment, and then I'm going to come back to the biologic - so, I think even some of these criteria have fallen out of favor, for example, the specificity of the association of the exposure-disease association, because a lot of exposures that we're talking about have systemic effects.

And so we don't see specificity of the association, and I think that one in particular has - I'm not saying that it shouldn't be in the list, but has brought - should get considerably less weight in most circumstances.

And then with regard to the biologic plausibility, I just think that when the toxicologists talk about it, I think they're talking about it differently than perhaps when the epidemiologists or the biologists are talking about it.

And maybe there needs to be recognition in the document that different disciplines are speaking here and they come at this with different viewpoints and interpretations, and maybe there needs to be some clarification from those different viewpoints.

But I also think that lack of - apparent lack of biologic plausibility does not necessarily mean that you don't have enough evidence to take action.

CHAIR HEERINGA: We'll go to Dr. Lowit, and then Dr. Bove.

DR. LOWIT: I was glad to see Dr. Meek speak up, because I was having trouble following the line of some of the discussion. And I'm quite uncomfortable with the idea that biological plausibility is a yes or no answer that really thrives in the face of the idea of evaluating the totality of the information across multiple lines of evidence and to understand the strengths and the weaknesses.

And sometimes those weaknesses can tell you as much as the strengths do in that there's - at least as we see it, that there is nothing about the use of the framework as written that precludes action when you have a robust data set for which you have maybe a mismatch across the humans and animals that you take a public health eye and you take action in that appropriate way.

And that's actually the power of - the transparency of the framework is that you can lay that evidence out and talk about your uncertainties and your strengths and how you came to a particular conclusion.

So, I hope we continue this discussion about the plausibility issue.

CHAIR HEERINGA: Dr. Bove, and then Dr. Bailar.

DR. BOVE: Yes, remember they're viewpoints. And as viewpoints for having a discussion about causality, they're important to take into consideration, but Hill never thought of them as something that you rule out.

I think the only thing that rules out is temporality. That's the only one. The other ones don't rule out anything. They're just issues that should be raised. Some of them, anyway, and some shouldn't be raised because they're not relevant.

And when you say it makes sense, there's enough uncertainty throughout toxicology and it's hard for me to say well, what makes sense. And, again, that evolves as well.

So, I don't think we're talking at cross-purposes. I think what we're saying is that, you know, just what you said. It's not a yes, no, that biological plausibility is something to think about, but that we also are aware there's plenty of uncertainty about biological plausibility and that the knowledge evolves.

CHAIR HEERINGA: Dr. Bailar.

DR. BAILAR: I think the biologic plausibility in terms of health - how do you evaluate the likelihood that this is going to be a real effect?

In terms of the, you know, the statisticians would deal with what's the prior probability based on what you already know?

Things like possible mechanisms of actions, what you know about related compounds, what you know about potential confounders, try to integrate all that to get at least a general sense of is this a place where you would be surprised by finding an effect or is it a place where you really might sort of expect an effect?

I don't know that you could really be any more precise than that in terms of plausibility.

CHAIR HEERINGA: Dr. Reed.

DR. REED: I think this discussion is very helpful. I think when we look back at the document, maybe we can - because some of it captures some of this discussion a little bit on Page 31 on the bullet about biological plausibility, because it was more specific to a known mode-of-action.

And so that might be part of the confusion. So, that would be what I would recommend.

CHAIR HEERINGA: I'd like to move on to Question 1.4, there is a relationship here, before the noon hour. I don't want to lose track of this line of discussion.

And so if there are additional thoughts on the part of the panel members on not only the criterion, but the sequencing and the action steps.

One thing that I heard, and I'm somewhat of a naive listener here, is that there's some emphasis on differential weighting of these criteria. There's some emphasis on qualitative interpretation.

I know from having sat here, that there will never be enough data to decisively address any one of these criteria in most of these investigations. So, there's a missing data problem.

And I think that one thing we ought to think about is that this is a framework. So, is it organizing or is there more of a decision rule theoretic-type approach? I suspect that answer will be rejected.

So, where do we draw the line stepping back where this essentially sort of devolves into yes, we followed these steps, they proved inadequate, so we're going to fall back on sort of an arbitrary decision.

It's just a thought that I had listening to the process here. So, I think we need to be at least in terms of each of these steps and the criteria and the potential weighting and interpretation, as clear as we can be in our reporting.

So, let's turn to Question Number 1.4.

MR. DAWSON: Question 1.4, OPP has extensive experience applying the mode-of-action framework to experimental animal data. However, OPP has not yet completed a WOE approach that also includes epidemiology or human incident data like that proposed in Section IV of the draft framework. Please include in your comments what, if any, additional scientific considerations not discussed in the draft framework OPP should take into account when conducting such WOE analyses.

CHAIR HEERINGA: Dr. Meek.

DR. MEEK: Thank you. I think we've probably largely addressed the content of 1.4 in the previous discussion, but one other thought comes to mind when I was listening to the discussion on biological plausibility.

One of the aspects that seems relatively important here is to be characterizing relative degrees of uncertainty in the various components of the database and how we relate them one to another in an overall framework analysis.

We can really only characterize that uncertainty relative to other data sets that we know, and there are no absolutes. So, there are no yes/no answers.

So, I think it's probably really important to do that to the extent that we can in doing these framework analyses. So, what's the extent of the data in each of the areas and what's our overall confidence in our conclusions?

CHAIR HEERINGA: Thank you, Dr. Meek. That's a better way of saying what I was trying to say before.

Dr. Bucher.

DR. BUCHER: No, I agree and I don't really have anything to add to what Bette said.

I think the important thing is to try to get across maybe in the document, a little better idea of how one would handle situations where different types of data that either agreed or disagreed would be handled, how you would try to weight different types of information against one another.

That's very difficult to do without some examples of exactly what one might do in different cases.

CHAIR HEERINGA: Dr. Reif.

DR. REIF: Well, the question asks for additional scientific considerations not considered, and let me just address the issue of individual susceptibility first from an epidemiologic perspective.

Of course this notion has been around for a long, long time if one goes back to the smoking question. Critics of the association between cigarette smoking and lung cancer pointed to people who smoked two packs a day for 50 years and died of a broken heart, as I sometimes describe it to students.

But in a more modern context, the issue of susceptibility and genetic variation, I think, is an important one that hasn't been addressed in the document. And no doubt as you heard yesterday, is being incorporated into good epidemiologic studies like the agricultural health study.

So, the addition of some

discussion of genetic variability in terms of SNPs or in terms of genome-wide association studies from a human side and the incorporation of knockouts.

And I'm not aware of the extent to which those sorts of studies are done today, for example, at NTP, but I do believe that we would be remiss if we didn't anticipate within a very short period of time, or actually currently the appearance of studies, both human and animal, that incorporate genetic analyses using the contemporary techniques.

So, that's my suggestion for additional scientific considerations that might be helpful.

CHAIR HEERINGA: Dr. Hayton.

DR. HAYTON: the only thing I thought might be missing from Section 3 was an explicit statement of intention to estimate or quantify exposure or the dose.

CHAIR HEERINGA: Comments from other panel members? And feel free to return to the previous topic, too.

Dr. Portier.

DR. PORTIER: I just wanted to reiterate what Dr. Reif said. I was sitting here thinking that the American Cancer Society has done cohort studies for over 40 years now, and we have a new one in place. And the new cohort concentrates highly on collecting blood samples for exactly the reasons he's talking about.

Even though it's a prospective study, you can't do a modern, expensive, long-term epidemiology study without collecting biospecimens for genetic analysis. So, that data is coming down the line.

CHAIR HEERINGA: Comments on these issues? Question 1.4, anything missing scientifically from the discussion of the frameworks?

Turn to Dr. Lowit, Jeff Dawson, Dr. Manibusan on this. Any points of clarification, something that confused you on?

DR. LOWIT: Not to re-open, but as you work through your report, this issue, the biological plausibility issue, I think Dr. Meek's point was right on that people coming from different discipline areas think we're thinking about that concept in a very different way.

And to the extent that there's discussion around those issues, would be helpful.

CHAIR HEERINGA: Dr. Bailar.

DR. BAILAR: That comment is right on target. I talked earlier about prior probability in relation to biologic plausibility. Different people will have different priors. That's just a fact of life.

CHAIR HEERINGA: Dr. Portier.

DR. PORTIER: Something that came to mind in researching the Bradford Hill criteria, and there's one other thing that kind of came up that we haven't really talked about, which is the concept of analogy, using in the risk assessment, information on similar chemicals, not specifically the one being studied, but ones where we've already studied and have a lot of information in kind of bringing that information into the decision-making process.

And in the whole weight of evidence approach, I didn't see any discussion on using chemical analogies, bringing in analogous chemicals into the discussion process.

So, I think you might want to add that as well.

CHAIR HEERINGA: Dr. Hayton.

DR. HAYTON: That makes me think it could also be pathway disruption. I mean it wouldn't have to be chemically similar, but what if it were - you could also look at your categorize on the basis of pathway effects rather than chemical class. Just another way.

CHAIR HEERINGA: You mean like endocrine systems or things like that?

DR. HAYTON: Yes, like endocrine disruptors which might not be chemically similar, but act on a common pathway.

CHAIR HEERINGA: I think that we have returned almost to the agenda schedule. Not that that's that important, but I think everybody deserves a long lunch. And so let's plan to reconvene at 1:00 p.m.

Thank you, everyone.

(Whereupon, the above-entitled matter went off the record at 11:38 a.m. and resumed at 1:00 p.m.)

A-F-T-E-R-N-O-O-N S-E-S-S-I-O-N

1:00 p.m.

CHAIR HEERINGA: Good afternoon, everyone. Welcome back to our afternoon session, second day of the FIFRA Science Advisory Panel meeting on the topic of the draft framework and case studies on atrazine, human incidents and the agricultural health study.

We are in the process of the response by the panel to the charge questions. And we have through this morning's session, covered Charge Questions 1 and its four parts. And we're ready now, I think, to turn to Charge Question 2A, which relates to the first of the case studies.

So, either Jeff or Dr. Lowit.

DR. LOWIT: Let's see. As discussed in Question 1.1, the draft framework provides general descriptions of the strengths and limitations of ecologic and retrospective epidemiology studies with respect to human health risk assessment. Please describe what you consider to be characteristics of robust, well-designed ecologic and retrospective epidemiology studies.

CHAIR HEERINGA: And Dr. Gold is our lead discussant.

DR. GOLD: I'm afraid I have quite a bit to say. Some of it we touched on this morning, so I apologize in advance that it's long.

And the other thing is that in addition to answering what we need for robust and well-designed studies, I'm also going to say something about what I think is a little bit missing in the parts here. So, that's part of the reason for the length.

So, the first thing which won't directly answer that question is that I think we need clarity in the terms that we're using. So, the term "retrospective epidemiology studies," I think we need greater clarification because many epidemiology investigators use this term to describe case control studies, as was mentioned this morning, because the information about exposure is gained retrospectively in these types of studies.

However, the term "retrospective" is also used in the framework for retrospective cohort studies, which is a different design, but could be considered retrospective because the exposure cohorts are assessed retrospectively.

So, case control studies and retrospective cohort studies can share some of the same challenges of accuracy and completeness of retrospectively ascertained exposure information, but they sometimes determine exposures differently or use different methods.

For example, in case control studies, frequently participants are asked about their prior exposures, which ,ay suffer from inaccuracy in recall. In retrospective cohort studies, the exposed and unexposed cohorts can, not always, but are often identified by existing records about prior exposure.

For example, in occupational studies, this is frequently the case. And so records may have the potential to be more accurate, not necessarily, and complete than participant recall of exposures.

Also, it should be noted that nested case control or case cohort designs would provide less potential for bias in ascertainment of exposures than would case control or retrospective cohort studies that depend on recall of exposures.

And just a really minor point that I quibble with the use of the term "predictors" in this section, because in ecologic or cross-sectional studies it's not always clear that we're talking about things that preceded - exposures that preceded disease occurrence. So, it's not necessarily a predictor.

So, and then in the portion where ecologic studies are summarized on Page 41, much of the data on occurrence of birth defects, pre-term delivery and small for gestational age are derived from birth defects registries, birth records and national data sets.

So, important considerations in ecologic studies using such data sources, I have several of these, include; number one, whether reporting to the registry or on the birth record is mandatory as would - this would tend to make these sources of information more complete, and reporting from areas where it is not mandatory could be influenced by factors that might also be related to exposure, for example, socioeconomic status or may be related both to likelihood of reporting and exposure to pesticide.

Second, whether the registry is

actively identifying birth defects where it depends on passive reporting will affect how complete the ascertainment of cases is.

Third, reporting to the database depends on who reports. Because the more the reporting from different sources, the greater likelihood of more complete ascertainment.

Fourth, whether the criteria in definitions of birth defects, pre-term delivery and small for gestational age has been explicit and consistently used so they're comparable across years and regions.

And finally, what was the length of follow-up for birth defects? For example, was it just at birth or for one year? This will greatly affect how complete the ascertainment is.

So, to answer the question, robust, well-designed ecologic or retrospective studies should derive reproductive outcome data either from registries with mandatory reporting and active surveillance with explicit and consistently used criteria and definitions of outcomes for ecologic studies or routine regular screening for outcomes in exposed and unexposed cohorts with explicit and consistently used criteria and definitions of those outcomes in these retrospective studies.

I have some additional considerations for information on confounding variables, whether they were obtained and controlled in statistical analyses.

So, this is important when you're comparing rates across geographic areas or between a state and the United States as a whole, because differences in the population distributions with regard to such factors, and I have a whole list of them, could affect the rates. They may be small, but need to be examined. And, thus, influence the determination of the difference in rates between the areas being considered.

In the case of ecologic studies,

this is further influenced by the fact that adjustment for confounding at the population level may not sufficiently remove confounding effects, and, thus, result in differences from studies in which confounding factors are adjusted on an individual basis to obtain summary statistical results for comparing groups.

Also, in some of the examples that were cited, the CDC natality database was used and they do adjust for confounding factors, but some of these factors are missing in certain states.

And so they're not comparable over years or they may not be comparable - the results when you adjust for the confounders may not be comparable across different states.

So robust, well-designed ecologic or retrospective studies should obtain complete information on as many potentially confounding variables as possible from all groups in the case of ecologic studies, or all individuals in the case of retrospective studies, and evaluate them as to their relation to exposures and outcomes and their modification of the exposure/outcome relationships. So, it's not just confounding, but effect modification as well.

Also mentioned in this section, referring to Page 41 of the framework, generally involved surrogate measures of exposure, for example, levels in the drinking water, proximity to fields and so forth, rather than measures of actual exposures. Don't measure how much women actually drink or measures at the tap, for example.

So robust, well-designed ecologic or retrospective studies should use the best, possible measures of exposure that are most likely to relate directly to the outcome, for example, drinking water measures instead of ground or surface water measures, as individual amounts of tap water and bottled water consumed as well.

Another issue related to the use of surrogate measures of exposures is that studies of reproductive outcomes often use the place of residence of the mother at time of birth to relate to environmental exposures by area, and studies of chronic diseases often use the address at the time of diagnosis.

So, use of these addresses, I don't think we did discuss this earlier today. The use of these addresses can result in misclassification of the relevant area of residence. For example, in the reproductive studies, residence at conception depending on which outcome you're looking at, might be more relevant than where the mother is living at the time of birth.

And so in robust, well-designed ecologic or retrospective studies, we should acquire residential histories whenever possible. That's very difficult to do, I realize. More likely to be done in retrospective studies than ecologic studies, and even in retrospective studies it's difficult.

So, the purpose would be to use the most relevant residential location for assessing exposure, although it would be noted that even obtaining relevant residential location may misclassify individual exposures because individuals spend substantial portions of their lives at work or otherwise away from their residences and the exposures in these locations are often not considered.

I think this was mentioned yesterday, but not today: Care should also be taken to consider different health outcomes separately. For example, different kinds of birth defects may have different etiology or a different relationship to the agent under study. And the birth defects will have different relationships, potentially, than birth weight, pre-term delivery or small for gestational age, because the etiologic mechanisms may differ.

So robust, well-designed ecologic or retrospective studies should provide sufficient sample size to have adequate statistical power to examine the relation of exposure to different outcomes separately as they may have different etiologic mechanisms.

Let's see. Some additional considerations in ecologic studies, and this is highlighted, actually, in Table A2 on Page 43, that there are some quite large sample sizes there.

So, sometimes when you're comparing populations, you can end up with huge populations. I think there were over 30 million births in one of the studies.

So, this can result in very small differences being statistically significant. And what needs to be decided is if they're really meaningful or clinically or public health-wise important.

So in robust, well-designed ecologic or retrospective studies, it would be important to consider whether the magnitudes of the differences are truly meaningful, and whether the differences could be due to uncontrolled confounding and whether the differences are internally consistent.

I have a couple more - no, one more. So, another important consideration for ecologic studies, and sometimes for prospective and retrospective cohort studies, is that if multiple outcomes are examined, for example, different types of birth defects or multiple associations, for example, using different timing and measures of exposure, if you're testing those multiple outcomes, multiple associations in relation to an exposure, then some associations are going to be statistically significant by chance alone due to multiple testing.

So, in such circumstances for the ecologic or retrospective studies to be considered robust and well-designed, investigators should account or adjust for multiple testing unless a strong hypothesis indicates the likelihood of the exposure being related to more than one outcome.

That's all I have.

CHAIR HEERINGA: Thank you, Dr. Gold.

Dr. Bove.

DR. BOVE: I have a few things. The first thing is that if you do a birth defect study in this day and age, you should use a population-based birth defect registry. And if you're going to look at the birth certificates, look at your state's birth certificates. I mean I don't understand why you need to go elsewhere or national databases when these are available, certainly, the health departments, and that might help with the ascertainment issue.

There are a couple of other issues. Again, a lot of criticisms at these studies are that there was confounding bias or selection bias or exposure, misclassification bias, and sometimes they're there, and sometimes they're not. But it would be helpful if a case is made as to whether they exist, and some effort made to determine what the impact might be of unmeasured confounders, for example.

As I said earlier today, a significant confounding is a rare event, actually. And it would be good to get a handle on just what impact confounders will have and whether it would really change the interpretation of the data.

Speaking of interpretation of the data, I mentioned earlier, too, that the reliance on statistical significance to determine whether a finding is worth looking at or not is a bad approach. And that also part of the problem with multiple comparisons is just that, that you're busy looking at statistically significant findings, and even worse you penalize findings because you worry about multiple comparisons, and really that's

not what you should be evaluating.
You miss quite a number of good, important findings if you're focused only on statistical significance. Studies often are under-powered for birth defects.
So, if you're going to focus on statistical significance and determine whether you're going to take something seriously or not, then you know the answer before you do the study. If we don't have enough power, none of these things will be statistically significant, so why bother even doing the study in the first place.
The reason you do a study in the first place is because that's not what you're supposed to focus on. You're supposed to focus on the magnitude of the association. That's the odds ration, the relevant risk, point estimate, the coefficient regression and so on so forth.
Certainly a confidence interval is important, but if you look at just one end of

the confidence interval, you're wasting your time. You might as well not calculate it at all.

There are two ends to a confidence interval, and I think that there are other issues that make you want to look at a finding and take it seriously or not, and statistical significance shouldn't be one of them.

Okay. And I think that part of the problem with some of the evaluations of these studies by the EPA and by the researchers themselves, fall into this trap of using statistical significance as a yes/no, whether I'm going to take it seriously and failure to look at exposure-response relationships that aren't statistically significant, but actually indicate that there may be something there. Enough on that.

So, some of the studies actually didn't report findings because they weren't statistically significant. That's unfortunate as well.

I do think that drinking water studies are difficult to do, but I do think that if it's possible to use the monitoring data and do some sophisticated modeling, we can actually get down to monthly estimates of contamination, which fit well in evaluating trimester or exposure.

Granted, we don't have information oftentimes on bottled water use. But if the contaminants are volatile, there's a showering exposure and dermal exposure at least as important as well.

So, I do think drinking water studies are extremely valuable, and so are air pollution studies as well, if you can use the monitoring data and use sophisticated modeling to estimate what those exposures are. And I think for past exposures, that's pretty much what you have to do in order to estimate exposures.

I do take exposure-response relationships very seriously. I think that researchers should too. I think that they should look at smoothing methods to see if the categorization made sense. I don't think they should just throw a continuous variable into a regression model, especially an exponential model. I think they need to see if the way they've categorized it or the way they're portraying exposure-response relationship actually is nearing the actual curve itself.

At the end of the day, I think it's important for the study to have a study design that matches what the objectives of the research are.

If a study really wants to look at effect modification, then it needs to be appropriately designed so it can look at effect modification, because that is a hypothesis and just as important maybe as the usual hypothesis we look at when we look at exposure-deceased relationships. So, that's about it.

CHAIR HEERINGA: Thank you, Dr.

Bove.

Dr. Reif.

DR. REIF: I think my colleagues have done a very thorough job in answering the specific question here under 2.1. And what I'd like to do for just a second, is to step back and just add a little bit to the general framework of reproductive epidemiology especially in the first - in the early stages of exploring exposure-outcome relationships.

So, going back to what we discussed this morning however you'd like to define "biological plausibility," supposing that there's evidence from animal studies that there are reproductive effects in laboratory animals, but that the likelihood is that the epidemiologic database for a specific chemical or pesticide is going to be sparse.

So, if one begins to think about the potential effects in humans to evaluate concordance, then one should take a broad approach to the various end points that reproductive epidemiologists use.

So, for example, going to some things we haven't talked about, assessment of fertility in any of these designs, not in ecologic, but certainly in case control studies, done by examining time to pregnancy in conditions of unprotected intercourse is a metric that's been used in a variety of epidemiologic studies.

There are examples of studies of spontaneous abortion, of course, from the drinking water. Family, as well as from - actually, from the pesticide research. And they may also add important information with respect to phenotoxicity that might be seen in a laboratory animal.

Just to point out along the way that there are issues with many of these end points. For example, the well-recognized phenomenon that about 20 to 25 percent of spontaneous abortions are not clinically recognized because they occur so early in gestation that they can't be counted.

So, once you go through the thought process here of examining the end points, then moving to stillbirth, to neonatal deaths, and then that issue of growth retardation that is shown in some of these examples, is a convenient metric for these analyses because the information is generally available from the birth certificate. And that's the reason I ask the question about linkage, data linkage with the AHS, because it would be a convenient and rather efficient way to examine at the early stages of exploration, potential associations between specific agrochemicals and outcomes.

So, you could use birth weight as a continuous variable, you can use the definitions of "low birth weight" and "very low birth weight," you can use intrauterine growth retardation or small for gestational age directly from birth certificate data, and pre-term birth defined as delivery before 37 weeks of gestation, again, are data sources and data points that are available to use in this general category of studies.

It's not, obviously, a longitudinal prospective study, but the more efficient and easier studies to do in epidemiology can in fact address those end points.

The birth defects work is hazard - it's a little tricky. It is very tricky, actually, for reasons that others have mentioned particularly due to the heterogeneity of probable pathways and causes of specific defects or groups of defects.

So, for example, it's very common for people to look at cardiovascular defects when, in fact, this represents a heterogeneity of lesions that may, in fact, have different etiologic mechanisms.

So, I think just thinking about these things in a more generic way can help guide where we should direct our efforts.

And there are examples, in fact, in the atrazine literature of menstrual cycle changes with respect to the characteristics and length of menstrual cycles. Of course, under-control of the hypothalamic pituitary access and the gonad, I think, are important end points to look at, as is age at menarche in younger women, and age at menopause in older women, because I think all of those are informative, again, linked back to the biological plausibility question and the mechanism of action issues that have been raised by the agency, which I concur heartily are absolutely the right direction to go.

And finally, when thinking about reproductive end points, one shouldn't ignore the male. I have to say that as probably about 50 percent of the equation. And, in fact, I think the recognized or the usually quoted statistic is that about half of the infertility that occurs in human couples is the male side of the partnership as opposed to the female side.

And there are even, for example, studies that have explored the relationship between atrazine and semen characteristics that are important to bring into a weight of the evidence analysis.

CHAIR HEERINGA: Thank you, Dr. Reif.

Dr. Bailar.

DR. BAILAR: I don't have much to add. This will be relatively short.

CHAIR HEERINGA: Get your mic there, John.

DR. BAILAR: The EPA report should first acknowledge that ecologic studies, and to some extent retrospective epidemiologic studies, are inherently weak vehicles for quantitative estimation.

This is not a criticism of such data which may have considerable strengths in other ways such as generating hypotheses, supporting smaller and inconclusive data of stronger inherent character providing floors to the size of some effect in support of legislation or regulation and so forth, but it is important that EPA not overstate the strength of such work in the interpretation and analysis of problems.

That said, there is a spectrum of strengths. Not all such studies are equal. Look for some of the following things: A defined population base, whether the population base has grown as narrowly as compatible with substantial exposure, that is, don't dilute possible evidence of real problems by including persons with little exposure, except maybe in dose-response analyses when dose groups are compared, look for estimated proportion of problems reported, accuracy and completeness of diagnosis, the quality and completeness of data known on unknown or suspected confounders and effect modifiers, the quality of exposure data.

I'm not saying here as much as it

might appear. I want to know whether the investigators have well-defined categories and whether exposed persons are classified correctly under whatever scheme the investigators have chosen, and do not try to reduce this to a checklist.

There will be too much variation from problem to problem so that thoughtful interpretation by the best epidemiologists available will remain necessary.

CHAIR HEERINGA: Thank you, Dr. Bailar.

Dr. Portier.

DR. PORTIER: Knowing that my colleagues would do a much better job of listing a lot of the details, I'm going to come back up and I'm going to give you some general criteria that I've extracted from the literature. And there will be some references in there as well.

I'll start with a quote I came across which I thought is pretty good, a pretty good summary. It says good, quality epidemiological studies are those with sound methodology, lack of bias, long enough follow-up times to observe a health effect - this was a carcinogenic study. So, I've replaced "health effect" with "carcinogenic" - a health effect response, adequate exposure information and dose-response information.

Before a lack of health effect can be inferred, it's essential that the exposures be of substantial duration and intensity, and that the number of exposed persons be reasonably large.

One of the things that this brings to mind is that a good epi study - most epi studies aren't good at looking at very rare events. They're much better for events that you can get cases on. So, it's not going to be looking for those really rare health conditions.

And then I listed a set of nine general criteria, which a lot of which have been talked here. And I'm just going to read these, because it summarizes it. Again, a lot of this was extracted from a paper by Swaen in 2006 in Human and Experimental Toxicology, but I've kind of added and enhanced it.

So, the study design should be appropriate to the study objectives, and should take into account the time frame of the exposure to health effect relationship. That's number one.

Number two, and appropriate comparison groups should be used. Appropriate matching. We've talked about that.

If applicable, measures of exposure should be corrected for known confounding factors, including concomitant exposures.

Now, a big one. The sample size of a study should be such that if in reality there is an association between a certain exposure and health effect, the study will be capable of distinguishing this effect from a no-effect situation on a statistical significance level.

In other words, it has to have sufficient statistical power for identifying meaningful differences. A negative study can't be interpreted if one doesn't know the probability the study can detect an effect if it is present.

In other words, we talked a little bit about negative studies, but the key component about understanding a negative study is understanding the sample size that's involved, because that tells me whether I missed a big difference or not simply because I didn't look at enough people.

The fifth thing is the appropriate statistical analysis is used. Use statistical estimation and testing methods which account for multiple comparisons when multiple health outcomes are examined.

I think all of you know the dangers of exploring without taking into account the fact that every time you do a test, there's a chance you're going to be wrong and that it's not significant.

The quality and reliability exposure measurements must be assessed. Three types of exposure data may be used. Actual data, and there's kind of two forms of those; external, where the external dose is taken from measurements in the individual's microenvironment, or an internal dose through biomonitoring, that's actual data. There's analogous data where an exposure situation is used as a surrogate for the actual data. And then there's, which we haven't talked about, is predicted exposure data where we're getting an exposure value from some kind of validated modeling technique.

All of these things have to be assessed against some kind of actual data to ensure their predictive ability.

We want to avoid the use of exposure indices that have poor predictive ability. And if an exposure index is used, it must be validated to the actual data.

The exposure measurement must provide adequate, discriminating power to detect exposure-related hazard. At a minimum, that means we need reliable gradations of relative exposure amounts.

So, if you don't have individuals in exposure categories that relate to some kind of a scale, you're not going to be able to get toward the dose-response relationships.

Exposure metrics can represent dose values, for example, average daily dose, cigarettes per day or some peak dose, it can represent a duration value like length of exposure, example, years smoked, combined together into a cumulative exposure metric. And I'd like to think of cumulative exposure metrics kind of like area under the curve statistics that integrate duration and intensity, which you talked about in your document.

A few last things. The quality and reliability of the health effects data must be assessed. Medical records typically have higher value than self-reported health effects, more recent events of more value than events that happened a long time ago. And we just mentioned using state birth and death registries where there's a quality assessment of the health record data that comes along.

Temporal variability, spatial variability and variability due to individual behavior as it relates to exposure, should be accounted for in the study. It should have a sufficiently long observation period with respect to the expected latency health effects.

And then in case control retrospective studies, completeness of case ascertainment should be the same between the exposed and the non-exposed group, that relates to high power and fair comparisons.

And I think those criteria really

work for both retrospective studies and ecological studies. They're kind of general things. Although, most of the ecological studies are going to be unable to meet some of these criteria.

Now, I can't let Dr. Bove's statement about statistical significance go unchallenged. So, my question is how do we know something is significant and needs focusing on if we don't pay attention to the statistical significance or confidence intervals.

So, we constantly say things like an odds ratio of four, and you say oh, that's great, but the confidence interval starts at .5 and goes to 27. And I don't really know what that means anymore unless I look at the statistics and take into account the sample size and the power of the study that we started with.

One can, though, compute and use the probability that the odds ratio is greater than a particular value in the decision process. So, kind of the - if you, you know, statisticians kind of less emphasize the point estimate and more emphasize the interval estimate or the probability of an event occurring.

So, it might be that when you're factoring this into the risk assessment, the chance that the relationship could be high, that chance is significantly high, you can factor that into the decision process.

There's still a good chance that you could be wrong if the confidence interval includes one, but I'll go part of the way with Dr. Bove on this, but not all of the way.

I think I'll stop at that point.

CHAIR HEERINGA: Thank you, Dr. Portier.

Dr. Bove, I think I was interested, too, in these points of view. And this is, I think, very important. I and others here have been on other advisory panels and advisory boards that were chemical specific, and the wide array of epidemiologic data - I'm thinking of arsenic now. And in the end, there was a clear difference of opinion on the expert panel as to how to utilize some of these data that clearly the power wasn't adequate to necessarily detect true values of interest, but the indications from the data themselves or of trends, of comparable odds ratios across studies.

Dr. Bove, is that sort of the line of thinking you were saying that if you see trends and results across studies even though individual studies may not be adequately powered, as Dr. Portier suggests?

DR. BOVE: Well, that's true, too, if you look at several studies and they're all seeming the same direction, but none of the studies had the statistically significant finding. That's one issue.

But yes, you know, I think that - but that is part of why I'm saying what I'm saying. In any particular study, you can always do a whole p-value function curve and see what alternative hypotheses are probable based on that or at least more likely than others.

I mean if you're interested in doing that, that's fine. But if you use several confidence intervals, for example, nested confidence intervals, the p-value function really interpret the data and that's fine.

If you're just going to say if it's in or out of the lower limit of a 95 percent confidence interval, to me that takes that art away from interpretation. We've heard that conversation this morning.

And it leads you to do things that would actually lead you astray such as ignoring findings that after several studies seem to pan out. I mean there are examples actually I can point to today from the studies under review if we want to go through, you know, and I'll do that later maybe.

But this is a big debate in the field. The issue about multiple comparisons is a big debate. A lot of epidemiologists come on both sides of that issue. And some say that instead of worrying about the p-value, if there are ways to adjust the point estimate itself through some kind of Bayesian method, that's fine. But, again, I don't know if we can resolve this here.

Again, I think that what's important is that statistical significance is not one of Hill's viewpoints, although he does refer to it in the paper. And I think that of course someone can make that claim that if it's not statistically significant, I'm going to ignore it, but I don't think that that's a very good method of interpreting data.

CHAIR HEERINGA: Dr. Portier.

DR. PORTIER: I do agree with you that when you read the literature, oftentimes the researcher excludes from the discussion things that they thought were not statistically significant. And I really think they should put it all in there.

I'd rather see all of them, even the non-significant ones, because you - you're absolutely right.

The other thing that EPA hasn't asked us about in this document is a meta-analysis when we start putting these things all together. And that's when you start seeing those things that are near significant, near significant, near significant. But then when we put a number of studies together to get a broader range of exposures, the trend all of a sudden shows up. Right?

They may be insignificant in the smaller studies. But in the broader view, all of a sudden it becomes significant. And you won't notice that if those things are not provided, and I agree with you on that.

I think non-significant is not a reason to not talk about it, to not put it into the - I would agree on that.

CHAIR HEERINGA: Thanks. Dr. Bailar.

DR. BAILAR: A couple of footnotes. Generally, I'd like to encourage EPA to minimize the use of p-values and go for confidence bounds wherever you can calculate them or pick them out of the literature. They're just so much more informative.

Among other things, they take care of the problem of small samples. Confidence bounds turn out to be much wider, and everybody can see that you didn't have much chance of binding an effect with this kind of study, this kind of design, this kind of sample.

Furthermore, we're used to seeing confidence bounds at a particular probability level, but it's not difficult to put in more than one level. You can show ten percent bounds, five percent, one percent on the same figure, the same set of bars, and sometimes that can be informative.

The other footnote, Dr. Portier went through a very nice catalog of the sources, kind of exposure information. To that I would central monitors. That may not be a big issue in pesticides, it might come up once in a while if you're dealing with very broad, airborne contaminants, but it is important in some other kinds of analyses that EPA is interested in.

CHAIR HEERINGA: Thank you, Dr. Bailar.

And I saw a number of people nodding at the comment on the display and use confidence intervals to reflect uncertainty.

In my view, that's the world that I live in. A simple one star, two star, three star, no star doesn't provide you the type of information that you need to make an art out of this.

So, I think we're converging a little bit here. Does anyone else have thoughts on that?

But I think that's important because it reflects how people present results in scientific papers and in presentations. And I think particularly in this domain seeing those confidence bounds on relevant statistics is very, very important.

Dr. Gold.

DR. GOLD: I would support that you get more information by the confidence intervals, but I'd also point out that there's nothing magical about .05 either. It was totally arbitrarily picked.

And so also looking at 90 percent confidence intervals, for example, it's often done in occupational studies. And as, in my training, one of the statisticians point out most of us make decisions based on values much greater than five percent.

So, I think that's, I think, where the comment about not adhering so stringently to statistical significance comes up, because you make decisions in life based on much higher p-values.

CHAIR HEERINGA: We should all have been born with a table of T statistics in our heads, and we would have walked back and forth between the two.

Other comments on this particular item?

I guess we'll move on. I'll turn to Dr. Lowit. Any questions? Lieutenant Niman, we would like to read the next question into the record, please.

LTJG NIMAN: Question 2.2, ecologic and retrospective epidemiology studies are particularly useful in identifying new hypotheses about the human health effects of pesticide exposure and may confirm the human relevance of findings from experimental animal studies. However, these types of studies do not typically include robust characterization of exposure and they do not address confounding factors as well as prospective studies. Although there may be exceptions, generally, ecologic and retrospective epidemiology studies are generally not sufficiently robust for use in quantitative risk assessment, i.e., for use in deriving point of departure or in quantitatively informing extrapolation factors. In light of the strengths and limitations of the ecologic and retrospective studies, please comment on appropriate ways to use these types of epidemiology studies in risk assessment/risk characterization or their utility in problem formulation.

CHAIR HEERINGA: Dr. Greenwood is our lead discussant.

DR. GREENWOOD: Well, I looked at this against the background of the proposed changes towards the National Research Council move towards looking at pathways and so on, and I think it's going to be a long transitional period. And I think that you're going to need to continue looking at any sort of information that you can get in the move towards the new paradigm for risk assessment.

And I think the epidemiological studies certainly have potential to provide important information in looking at your assessment, the assessment of other people.

Certainly they could inform experimental toxicological end points, and it could also be useful in making people aware of possible lesions that they haven't predicted or taken into account in the absence of, say, of a mode-of-action study.

But like all information, they really do need to be scrutinized very carefully just in the same way that toxicological data need to be scrutinized very carefully to make sure that the design and analysis and so on, the way they do the methodology, are appropriate.

There is a difference, though, because for a lot of toxicological assays, there are standards available. There are definite and well-defined methods for demonstrating the validity of the assays, and they can only be done by people who are accredited to do some of those assays.

So, a lot of the routine assays, toxicological assays, have been validated. And I think there's a real need for looking at developing some sort of framework, and I think already Dr. Bove has already hinted at or given some ideas for looking at the validity of epidemiological studies.

It makes an enormous difference, and I always call it the field of analytical chemistry where at 20 years ago the analyses you got from laboratories were by and large very unreliable. And that was shown by the early interlaboratory trials.

And over the last 20 years, validation protocols have been developed and the reliability of analytical data now is very, very much improved. It's a totally different field from what it was 20 years ago.

And I think that in your document, you've identified the main difficulties. And one of them is obviously in the assessment of exposure and identifying the other factors which might be correlated with that exposure or associated with it.

It's very, very difficult to estimate exposure even under good conditions. If you look at some of the studies where - biomonitoring studies, people have had great difficulties. Even when it's on an individual basis, they've taken the urine, all blood samples, done the analysis, and then try to link that to the exposure scenario, and it's not as easy as you might think. And often there are large uncertainties associated with it. The confidence intervals are very wide.

I think that the estimates of exposure that we've seen in some of the studies really would not pass any validation mechanism. We'll come up to that perhaps in the next question.

And the other problem is that yesterday I think someone was saying well, why not stick to the high dose with worker exposure, but actually that doesn't cover a huge proportion of the population. And that certainly wouldn't cover, for instance, spouses and the offspring of workers, let alone the general public.

And certainly that source of contamination through people taking home contaminated clothing and so on is well-documented for the asbestos cases, for instance, where spouses were exposed to asbestos from the clothes of their spouse.

So, none of this sort of should happen, really, if there's good practice, but contamination - there isn't always good practice. There are failures and contamination can occur, which adds an extra uncertainty when we're dealing with epidemiological studies.

I think rigorous estimates of

biomonitoring are difficult to achieve. And I do feel that there's this scope in the way that epidemiological studies are looked at in future and maybe trying to do some validation of the exposure or maybe insisting on some validation, external validation of the exposure data.

But it's not easy even when you have urine samples, because - there's a pilot study by Bartlett I think in 2007, where people traditionally look for atrazine capture rates in urine, and use that as a measure of exposure. But actually what he found was that something like 70 to 80 percent of the exposure was due to a couple of metabolites, the diamino chlorotriazine and deethyl atrazine, which were the predominant compounds in the urine and a much better indicator of exposure as those compounds have been thought to be active in their own right.

So, the atrazine recapture is actually underestimating the exposure. So, it's very, very difficult to get a good fix on that. And I think that if we're to take epidemiological studies seriously, this aspect really needs some sort of validation.

I think the other problem is that as with experimental toxicology, it's a big problem for the field, that there's a tendency to look at each compound in isolation, because that's the way we did the toxicological testing, that's the way we think about it.

Actually, that is a chemical soup that anybody's who's into water monitoring knows just how many compounds there are in that soup, and it's more than we can measure realistically.

So, it is difficult with any study, to actually understand whether that one compound on its own is the problem or whether it becomes a bigger problem or a smaller problem in the presence of other contaminants.

And somehow or another I think this needs to be taken into account when assessing the exposures which are used in epidemiological studies.

And it's not just the pesticides. You've got to think in terms of the huge numbers of industrial chemicals, pharmaceuticals, household product components, personal care products and components. And some of them even vary on a seasonal basis. Suncscreen components in lakes in Switzerland, big problem in summer. Not in winter, because it's too cold to swim and not too much sun, but they are a problem.

And these sorts of things often are forgotten about when people are concentrating on one group of compounds which are of known biological activity.

Given all of these reservations, I think that epidemiological studies do have the potential to make a significant contribution to particularly risk characterization, and in some cases risk assessment, in a number of areas.

And I think that one of the areas is the identification of potential health problems which may not have been previously considered as being associated with exposure to pesticides.

And I think vigilance is something which needs to be maintained. And it could help in that case, to prioritize research efforts. But there may be opportunities to look at the validity of some studies in areas where there have been changes in practice.

So, for instance, in some areas of Europe now, the use of atrazine is severely curtailed and it's prescribed for many of the applications for which it was used previously. Particularly for maintaining road surfaces and weed maintenance on those railway embankments and so on, which really are a very large - provide a large component of environmental contamination.

And it might be possible there to look at an effect, if it's a seasonal effect, see whether it was there before the removal of the compound, and after the removal of the compound not just for atrazine, but for the compound.

But I think the big contribution that epidemiological studies might be able to make is at the problem formulation stage. And I think as you move towards the new paradigm, I think it's going to be even more important that the people in different areas, the toxicologists, the epidemiological toxicologists, all actually speak to each other, the analytical chemists, all people actually have an input at that stage.

And I think certainly it could inform prioritization of research and maybe help to inform what sort of internal exposures people should be looking for on the basis of observed external exposures.

So, I think that providing we can in the future move towards better collaboration between disciplines, I think it would be easier to actually provide validation of some of the steps in epidemiological studies.

I'll leave that there and hand it over to my epidemiologic colleague.

CHAIR HEERINGA: Thank you, Dr. Greenwood.

Dr. Reif.

DR. REIF: Yes, thanks. I commented earlier today about the inherent dangers of kind of lumping ecologic studies and what are called retrospective epidemiology studies. And I think that, again, and I'll just reiterate that that's probably not a useful collapse of study designs.

I think the question as it's framed, probably applies pretty well to ecologic studies. It doesn't apply equally well to case control studies or historical cohort studies.

I'd just like to give a couple of examples of, first, one of a historical cohort study that is, I think, a useful discussion point although it's not about atrazine. It's about DDT. And then I'd like to talk about one of the other designs for a moment that's in the case example.

Going back about 20 years ago there was a small body of evidence that suggested that accumulation of liquiphilic organochlorine pesticides, DDT and others, was related to breast cancer risk. And those original observations were based on biopsies of fat taken from women with tumors and women without when fat was available from other means. And some small case control studies using that biomonitoring approach found differences in the concentrations of DDT and other organochlorines. And that began quite an effort, actually, to evaluate the role of organochlorines in breast cancer risk.

One of the answers came from a study done in which sera had been banked from women about 20 years previously. And I believe it was Nancy Krieger who did this, but my memory may be faulty, and I'll check on it.

But the beauty of it was, and this is a retrospective design, this is an historical cohort study, that these sera available from a fairly large sample of California women taken over 20 years ago, could now be used in exposure assessment and by using registry methods - I forgot what the ascertainment was specifically, but to then ascertain those women today. And that is at the point when the investigators did the study.

So, they had historically faced exposure data, quantitative using for the day, sophisticated measurements with levels of detection that could measure these things in sera which was kind of new, because prior to that they had to have a fat sample prior to the resolution improving.

So, it's just an example of here's a retrospective study, it's just one of those phenomena that we're going to see more and more because it is becoming more and more commonplace for epidemiologists to bank tissue as they do studies. To bank buccal swabs for DNA, to bank sera for who knows what down the road as new hypotheses develop.

So, it really isn't fair to because the study is retrospective, as is that historical cohort study, to sort of say it has these inherent limitations that are not going to make it useful for risk assessment.

It could be extremely useful, and that's the example that I'd offer in that context.

The other thing that the question sort of raises in my mind, has to do with this surrogate exposure for proximity to fields where pesticides have been applied.

Actually, I've been involved in a study like this, and I think they are - this kind of spatial epidemiology using GIS tools and using data, for example, that's available for specific pesticides in specific locations where one knows what the application dates are because of the emergence of the weeds that are being killed, there are relatively specific windows of exposure when certain pesticides will apply.

There are also very specific windows of susceptibility in the development of the fetus so that it at least conceptually is somewhat attractive to explore the potential relationship between the timing of application of pesticides and specific events that may be occurring in utero in the pregnant female. So, this has some - at least some theoretically appeal.

What hasn't happened, which I think is a worthwhile objective, is for a validation step to occur to determine whether or not the women, for example, who live within 500 meters of the edge of a field containing corn or sorghum at a specific point in time in the early spring when the pesticides are applied, whether one can find evidence first of environmental exposure by house dust sample.

So, the whole technology, as you know from the industrial hygiene field of exposure assessment, has really improved recently. And these kinds of validation steps which are really not that difficult in a study where one would collect house dust samples and analyze it for the persistent kinds of pesticides that are environmentally persistent, has then the potential of adding validation to the spatial GIS-based analysis.

We went a step further and took a sample of farms where you had proximity to fields and you had house dust samples, and then did human biomonitoring. You would now bring the exposure assessment closer to the objective, which is really to get a handle on what human exposure is.

So, I see even the study that's in the case analysis, is rather prude and certainly can be criticized. That doesn't mean that the design of considering spatial analyses using GIS techniques should be dismissed.

In fact, my belief is that it should be strengthened by enhanced exposure assessment methods done by industrial hygienists.

So, those are the comments I wanted to offer. They're not terribly germane to the question, but the question itself has some features that are difficult to conceptualize for me.

CHAIR HEERINGA: Thank you, Dr. Reif.

Dr. LeBlanc.

DR. LeBLANC: Well, first off I would agree with the agency that these types of studies have limited quantitative value in the risk assessment process. But having said that, I think they still offer significant qualitative value that should be considered.

I think that the greatest value to these studies as it relates to the risk assessment process, is both at the front end of the process as well as at the end of the process.

And I'm not an epidemiologist. And I'm proud of myself, I suppose, because the points that I listed as being relevant to the answer to this question were voiced precisely by Dr. Greenwood. So, thank you.

What I would like to do is just touch upon some of these points in the risk assessment framework just to sort of draw attention as to where the strengths are and where the weaknesses are to the retrospective ecological-type studies in this process.

As stated by many now, I think problem formulation is certainly, I think personally, the best place to place these studies, the results generated from these studies.

In my mind, they seem largely to

be exploratory in nature, but the observations that are derived from these studies, nonetheless, can be used to direct hypothesis setting, as well as directing the analysis towards testing these hypotheses. And I think all of that would fit up front in the risk assessment process.

Another thorn in the side of certainly toxicologists as it relates to risk assessment is evaluating the risk of toxicity associated - or hazard of toxicity associated with chemical mixtures.

And the reason it's a thorn is because there's so much complexity associated with the process. What chemicals do we use in our assessment, what concentrations of each chemical do we use in our assessments? And oftentimes toxicologists will simply throw their hands up and leave it to someone else to worry about.

But I think there's certainly potential that these ecological retrospective studies can provide insight into exposure to environmentally relevant mixtures of chemicals particularly in an agrochemical setting.

And as such, effects that are discerned under those particular situations, might be again used in a hypothesis setting that the results could perhaps run through the weight of evidence framework to establish the degree to which causality can be associated with the mixture, and then provide guidance.

The information itself can't be used, I don't think, in the risk assessment. They can provide guidance to animal studies that would be directed towards relevant exposures to chemical mixtures.

Moving on to the exposure characterization process, it seems that exposure characterization is inherently a weakness of these studies. So, I would suggest that perhaps the utility of these types of studies in exposure characterization is limited, but I'll get back to that in a second.

Moving on to hazard characterization, there are certainly limitations. And the limitations relate to the effect that we have a lack of understanding of exposure in these studies.

And without quite knowing what the exposure is, it's hard to ascribe hazard of toxicity associated with the exposure, but nonetheless they can be cooperative in nature. They can be used to look back at the animal studies and see if there's concordance, if there's consistency between effects observed in human populations that have been exposed and the animal studies.

And also, and we touched upon this many times, they can be used to identify hazards that are unique to humans that perhaps we missed in the animal studies.

Exposure characterization and hazard I would argue are sort of both weak areas that these studies may have limited value. However, I don't think we can - I think we still need to keep a perspective of the fact that consistency in observation among these studies may be very informative to the exposure characterization and hazard characterization process.

That is, we may see something in an ecological study and say well, based on that study alone because of the lack of control or recognition of mitigating factors or lack of understanding of exposure, we can't make any judgments. But if we consistently see that response, that effect among many of these studies, then I think that is very informative and that should be taken into the risk assessment process.

And then lastly the risk assessment, I think that these studies can be used to assess the validity of the risk assessment, that is, sort of looking back and seeing whether the judgments that have been made in the risk assessment are reasonable based upon what we're seeing in exposed populations.

And they also can be used to provide guidance in risk management, that is, in situations where the use of material has been curtailed or perhaps regulations have been lessened.

One can look at responses and see if there's concordance, if the expectation is reached and can provide confidence to the risk management, or alternatively can provide guidance to appropriate risk management. That's all.

CHAIR HEERINGA: Dr. Bove.

DR. BOVE: Looking at this question, I thought it was rather negative and pessimistic. And I think that it would help if the EPA would look at actual risk assessments that have been performed by your agency both in draft form and in final form, and see how human data were used.

And I'll give you a for instance.

An early draft of the trichloroethylene risk assessment utilized a couple of occupational studies and a New Jersey drinking water study which I was involved in and actually did the exposure assessment for. And I can tell you that the water data in the New Jersey study, and the water data in the Indiana study, roughly about the same.

In the Indiana study, you have a lot of measurements during the growing season and hardly any the rest of the year. In New Jersey you had - if you have a contaminated system in any system that was a little bit above the MCL, the New Jersey Department of Environmental Protection went after them and made them test quite regularly so that for the contaminated systems you actually had quite a bit of data, sometimes monthly, over a three, four-year period.

Anyway, the point I want to make is that the early draft of the TC risk assessment used both occupational studies and this New Jersey drinking water study, calculated the cancer potency for that, compared it to the animal data, and all the potency ranges sort of lined up nicely.

So, even with data this poor - I hear all this about how poor this exposure data is, you know. It's not that poor. Okay? It's not perfect, it's not great, but it can be used.

My feeling is that instead of asking the question this way - or better yet, the best thing the EPA could do is actually go and look at how health data is being used, because it's being used. There's no question about it. It's being used in every part of risk assessment, including risk characterization.

And maybe looking at those studies that are used for risk characterization, that is used for problem formulation and so on and so forth, you get a sense of what quality data is being used and how far you can push epi data, because you can push it quite a great deal.

We push animal data, so we can push epi data as well. So, that would be my recommendation.

CHAIR HEERINGA: Dr. Gold.

DR. GOLD: I have very little to add. I'm just going to make two really quick points.

I think ecologic studies have significant limitations and we've voiced those. So, I think they are largely useful and one could say arguably more useful than incident data, for suggesting hypotheses for future well-designed studies so that they can help drive the research agenda, they can help in the problem formulation, and they can also help in examining the consistency of findings across studies, including ecologic studies.

And then in terms of retrospective studies, I too find this sort of a very pejorative way of framing the question. I think they really can be quite useful in many circumstances that my colleagues have already outlined.

So, I would just say that they can help provide insights into future analyses as well. And, again, help identify gaps, help drive the research agenda, help in the problem formulation and help generate better research in the future to overcome some of the shortcomings of previous studies. That's all I want to say.

CHAIR HEERINGA: Thank you, Dr. Gold.

Other members of the panel that would like to weigh in on this particular questions?

Dr. Chambers.

DR. CHAMBERS: I'd like to pick up on a comment that was made by a couple of the panelists a minute ago. Dr. Greenwood and Dr. LeBlanc both mentioned the mixtures issues.

This is a generic question not

toward any particular risk assessment, and it kind of brings to mind the chlorpyrifos discussions from several months ago in an SAP where there was some human epidemiology data on chlorpyrifos, but there were also two other anti-cholinesterases in the people at the time or in the households and everything.

And it troubles me if compounds are known to have the same mechanism of action, and yet the effects are attributed to just one of a mixture of several that have the same mechanism of action. That doesn't seem valid to me.

That was sort of brought up a few months ago when we talked about chlorpyrifos, but it would be valid for any case where the mechanism of action is known and you've got a mixture of several compounds that have the same mechanism of action.

Trying to sort out in a quantitative sense, the effects due to any one component of the mixture, I think, would be very problematical.

CHAIR HEERINGA: Dr. Hayton.

DR. HAYTON: Yes, maybe just a small point, and maybe this is obvious to many, but Dr. Greenwood mentioned external versus internal dose. And that since exposure is a central issue, I think we need to keep it in mind that the external exposure is really just a surrogate for the internal exposure which we often don't know.

And I'd just like to point out that even when we know within a population and think you know about people taking drugs, if you give a group of people all the same dose rate, you will see internal exposures that vary by, say, a factor of ten pretty commonly.

So it introduces, I guess, a foggy lens between what we think we're measuring as exposure and what's really going on.

CHAIR HEERINGA: Dr. Portier.

DR. PORTIER: Ken Portier. Actually, looking at this question and listening to the discussion, I thought to myself I wonder if what we're talking about is actual value, first, versus potential value for these different study designs.

Is it the fact that we like - we think that perspective studies are more valuable mainly because they're more complex and more expensive to do, and therefore we invest more intellectual capital in those studies, more time and effort. And, hence, we get better data in the long run.

Whereas if we were able to kind of invest the same intellectual/capital in a retrospective study, we could get almost the same thing.

There's this perception that ecological and retrospective studies are, quote, cheap. So, they're easy to do. But maybe we should try to change that and say these things could be very useful if we spend the time on it, because we have examples of retrospective studies that have very useful information. Those are the ones where they spent the time and put the effort in.

So, I don't know if it's the study design itself that gives us the value for the risk assessment. It's something else. It's what scientists invest in those things.

Which for the agency means you need to be thinking about how to encourage the kind of retrospective studies that garner those resources to give you those kind of answers. That's a tough one.

CHAIR HEERINGA: Dr. Lowit, are you satisfied at this point to move on?

Okay. I'd like to move on to Question Part 2.3. That could be read into the record, Lieutenant Niman.

LTJG NIMAN: Question 2.3, the atrazine case study, Case study A, provides specific examples of ecologic and retrospective epidemiology studies. Please comment on OPP's reviews of the studies discussed in Case study A. In your comments, please provide specific feedback on the OPP's descriptions of each study design, exposure assessment, use of appropriate statistical methods, and ability to address bias and confounding in addition to other factors that may be important in the interpretation of these studies.

CHAIR HEERINGA: Dr. Greenwood is our lead discussant again on this one.

DR. GREENWOOD: well, when I look at this, I guess the question is really looking for a general approach, looking at the general approach taken to the analysis of these various studies.

I think looking at this, the general approach to evaluation seems to provide a very useful framework, I think, and covers the important factors that need to be dealt with.

And I think that the descriptions are good. The descriptions provided of the designs, I found those very easy to evaluate.

And I think looking at the case studies that are presented, most of the main weaknesses have been identified, but I think maybe more attention could be paid with maybe looking at them in a little bit more detail about a number of things.

One is definition of the outcomes and the number of outcomes being studied. I won't cover the ground again, because colleagues have already been through that, that we're looking at large numbers of outcomes. We have one in 20, anyway, to be significant by chance.

And I think the area that the agency has identified, and I think everybody around here has identified as one of the major problem, is looking in maybe a little more detail at the exposure and exactly how valid those exposure measurements are made.

I think in the evaluation the agency has made, they've made a good assessment, but I think this is one area which really does deserve extra attention because it's got to be right.

And I think one of the things that people tend to do because analytical chemistry is so precise these days and you can depend on it, people tend to take the analyses, this quality of them for granted.

But actually if you're looking, for instance, at surface water, even drinking water and you look at the analytical chemistry, it can provide you with precisely the wrong answer. And that's nothing to do with the analytical step, it's to do with the sampling step.

And often, for instance, with surface waters, the quality of a river can be very strong. 12 bottles of water a year. So, how long does it take to fill a bottle? 30 seconds? So, you're looking at about six minutes out of a year that are being sampled.

And I know from work that we've done and certainly work from people in

Switzerland looking at rivers, that when we get a rain event, levels of pesticides can change from next to nothing up through into tens or 20 micrograms or higher per liter. And just as quickly, it can fall.

So, depending when you take your sample, you can see that there's a very low exposure or a very high exposure, and neither is actually representative of the real situation.

So, I think you need to bear these sort of factors in mind to look a little carefully at some of the data. Drinking water levels tend to change more slowly because just of the volumes that are collected and treated and stored and the time of the flow through the system.

But, again, I think with some of the exposures with drinking water, you could in fact end up with some quite flawed information depending on whether people filter through carbon filter systems, some people do, the water before they drink it in the house. So, the filtration and so on, because that's likely as well to be correlated with associated demographic factors.

So, there are lots of problems with exposure data, and I think that's probably one area where I think particular attention, maybe more attention than was paid here.

But overall I think that the approach that you took was very reasonable. And to be fair to some of the authors, they did actually point out the problems with their own data.

So, I'll leave it there again and pass it over to colleagues, I think, who may have some more detailed examples.

CHAIR HEERINGA: Dr. Bove.

DR. BOVE: I don't know if this is worthwhile doing or not, because, first of all, these aren't all the studies that have been done. There are two Iowa studies, for example. One that's published and one that's not.

But the one that's not published on birth defects was given to the EPA back in 2000. I personally gave it to you at one of these panels.

So, you don't have the whole universe. So, that's one thing. So, maybe I shouldn't even talk about the birth defect study in this packet because there are some similarities.

Min was also elevated in the Iowa study. So was heart. If you look at the table here, there's slight effects for heart certainly not statistically significant.

There are problems with this study in terms of using this metric of distance. And it's the only one out there, so I don't know what else to say. I have no problems with the way it was interpreted.

Why don't we move to the small for gestational age, because we see two studies that you did have in the packet, plus one that's referenced by one of the studies, the Munger study in Iowa which should have been in the packet, but isn't, and all three of them have somewhat similar findings.

The effect is small, but the effect is somewhere - in the Iowa state, it's like 1.8, but the Iowa study evaluated individual level study and turned it into an ecologic study, unfortunately.

It's not the individual level study which, you know, if I could have gotten it and did it, I would have done it for them. It would have been more informative. But they didn't, so good luck with that.

The other two studies that are in the packet are individual level studies. The Villanueva study is not an ecologic study. It's not analyzed, it's an ecologic study, exposure assessment is not done at an ecologic level, it's an individual study, and it's not negative either.

Actually, the stronger effect if you look at Table 2 in that study, was for pre-term birth which they don't even evaluate at all on their paper. The researchers don't, and the EPA doesn't.

You have actually an exposure response. You have a medium exposure. You get a 1.22 odd ratio. And for the high, you get 1.93. And the 1.93 is pretty high compared to all the other odds ratios you see in the paper.

Then you turn to the - another table in the paper. Let me see if I can find it myself quickly here. What you see is actually ver similar between pre-term birth and small for gestational age.

For pre-term birth looking at the first trimester and during the growing season, they have an odds ratio of 1.36. For small for gestational age in the third trimester during the growing season, odds ratio of 1.37. You couldn't get that much closer with the two, but the difference is one is statistically significant and one is not. And that's just a numbers problem, okay, but pretty much the same effect.

What's interesting is we don't see much in the next study with pre-term birth. And to tell you the truth when I've looked at pre-term birth and the kinds of exposures I've looked at, granted they're not pesticide exposure, I usually don't see much with pre-term birth either.

But still, I think if you want to interpret the Villanueva study, first of all it's an individual level study. Secondly, there is an effect both with pre-term birth and small for gestational age. And third, it's not a big effect, is the three things.

Now, if you go to the next study, which is the Indiana study, which has a better drinking water exposure assessment because they had more data, okay, and there what we see - do I have all the tables in here? I'm not sure if all the tables are in this thing or not.

But, anyway, when they looked at first month for pre-term birth, they see nothing, basically. So, that doesn't jive with the previous study.

If you look at small for gestational age, the findings there aren't very strong either. And if there is a exposure response, it's very slight.

So, how do you interpret that? Well, we've seen, as I said, two other studies, the Iowa study and the Villanueva study, showing slight effects for small for gestational age. We're talking about pretty low exposures for the most part. And so actually those three studies kind of agree with each other, that is, small effects, small increases in small for gestational age around the realm of 1.1 to 1.2 in terms of odds ratio, prevalence ratio, whatever you want to calculate here. So, maybe that's how they should be interpreted.

Again, my problem with doing this exercise is that there are - you don't have Munger's paper here that I would study in the packet.

The amount of work that's been done in reproductive end points with atrazine is limited. Okay. It's not as robust as some of the other chemicals.

What is more robust and what I would like to have seen in this packet, we're going to look at epidemiology and the role epidemiology plays in risk assessment.

I would have been occupational studies because there are a number of them. We've asked EPA to look at them in the past, and in particular non-Hodgkin's lymphoma studies. And so, again, I'm blessed with - I'm not sure the value of this exercise - this doesn't necessarily - it doesn't capture even the entire realm of reproductive end points and atrazine. There's some papers that Dr.

Reif knows about and put forward to the panel. And it's not certainly the universe of studies of human data and atrazine, including the occupational studies.

So, I'm left with just what I said, that this is a kind of funny exercise, but I do think it's important again. I'm not going to say it anymore after this, but when you look at these studies, certainly look at the confidence interval. But look at the point estimate too, and look at the exposure response that you see in front of you.

And don't say that an exposure response where you see 1.2 and then 1.9 and say there's nothing there. That doesn't make any sense. You can say it's a very weak finding, you can say it's based on small numbers, but you can't say it's not there.

CHAIR HEERINGA: Dr. Reif.

DR. REIF: I have very little to add. I also found it actually quite frustrating to consider of the six studies that were included in the case study for a couple of reasons.

I understand that the logic was that these studies have been published since an interim decision was made by the agency. I accept that.

But on the other hand, the selection of these six studies certainly doesn't fit into a weight of evidence scenario, because the weight of evidence scenario would dictate that we should consider all relevant studies and then do triage to decide which ones might be informative and which are not suitable because of study quality issues.

So, I had some frustration with the exercise also, and in particular because there were no case control studies in the suite, and there was a preponderance of the ecologic hypothesis-generating studies.

It was just difficult, actually, to develop in my own mind any sort of a consensus about well, where are the data taking us?

So, I didn't go to the extent that Dr. Bove did to look for small differences and very mild increases in the risk assessments, because the whole strategy here to me was incomplete.

So, that was my frustration with the case study, and I just think in the next iteration of course for future meetings if one is actually going to look at the potential risks associated with exposure to atrazine, that this has to be a true weight of the evidence analysis in which all the literature whether it's from 1997 or 2007, is brought to the table in a comprehensive manner.

That goes back to the comment that was made this morning about how do you do a literature search, what are the parameters of the literature search, what are the protocols for the literature search?

The same considerations that are

used when we select studies for meta-analysis, they're equally applicable here because that kind of rigidity and rigor is important, as you well know.

And I know that there was - that the agency know that I was torn between the sort of bi-partite mission here of trying to get a handle on the use of epidemiology in risk assessment. And while we're at it, let's take a look at some atrazine data. But it's so fragmented that it's really difficult to come away with, in my mind at least, with any kind of clear understanding of risk.

CHAIR HEERINGA: Dr. Lowit.

DR. REIF: Let me just add one thing. I have very little - other than the point Dr. Bove made about the Villanueva study, I do think that the reviews of these studies by the agency are generally quite good, accurate and complete.

So, as far as those six studies go, I think the reviews are adequate for certainly the next step.

DR. LOWIT: And just to respond to the comments about a small subset of studies, we certainly understand it's dissatisfying of the scientists to see a small slice of something that you know in your head and in your heart is very large, complex database not only the epidemiology side, but there's a very complex, rich animal database.

And our view, you've heard us talk about the two hats that we're trying to balance here. And we felt that those six studies, as Aaron described yesterday, really encompassed our goal for this case study as it is for today.

We have a need to make September, which will involve a more complete evaluation of all of the epi whether it's reproductive outcomes, birth outcomes, cancer outcomes in context with the animal database which is very large and very rich. And those need to be done in combination as we're proposing the weight of the evidence.

But those six studies provide a sense of some of the things that we're going to struggle with in the atrazine, but what they also provide is a spectrum of the things that we struggle with in these kinds of studies, period.

I mean it's very common for a study of this six variety to be published whether it's atrazine or another robust chemical or something we don't know much about where we struggle with how to think about the exposure assessment and the design and where it fits in characterization versus quantitation and how that works. And we want to do it in the most robust way.

So, we understand it's dissatisfying, but we are working very hard in the background to complete that picture with respect to atrazine.

We're also working very hard in the background on a lot of other chemicals.

And the feedback that we get on both fronts is important in both of those goals.

CHAIR HEERINGA: Thank you, Dr. Lowit.

Continuing with our associate discussants that have been assigned here, Dr. Bailar.

DR. BAILAR: I did go through each of these separately and made notes. There are three common problems before I get to the individual papers.

One is that what I see here is a reflection of a very common problem in data on what you could call big problems, substantial problems. And that is concerns about bias dominate concerns about randomness.

P-values and confidence bounds deal only with randomness, and my feeling is that here the p-values and confidence bounds are of limited significance because of this underlying concern about bias.

The second is that many of these

are subject to problems with multiple comparisons, and I don't think that any of them really dealt with that head on.

The third is that the effects they're reporting are relatively small, that is, relative to the size of the background effect you're looking at small wiggles in bigger numbers.

The first paper by Winchester, Huskins and Ying first - the size of the effect is not at all striking. Maybe six percent variation from low to high with a lot of possible season related confounders.

The peak incidents in terms of LMP is May to June. The data would be more convincing if the authors had found a lack of such a pattern in mothers who had been drinking groundwater, and they did not look separately at surface water and groundwater.

Also, I could not find any evidence in what was here, but the authors suggested for other seasonally changing chemical exposures, nor did they look at concurrent data from other states where atrazine exposures are much lower.

The data in Tables A2 and A3 I found to be somewhat troubling. All but one of the birth defect types was more common in April to July than in other months. And the exception, that is nervous - what was just called nervous, barely fell below our ratio of unity.

About half of the differences were statistically significant, but what I know about chemical teratogens, which is not extensive, most of them simply don't work that way. They tend to be much more specific and I take this broad pattern to be some evidence of a pervasive bias related to some other type of seasonally changing factor.

The authors and EPA here, I think I mentioned this yesterday, have missed a potentially useful analysis related to this and other papers in this group as they've ignored the older literature.

The data were not of the same quality we have now, but state departments of health, state vital statistics offices have for decades collected information on birth certificates about birth defects. And that might be of some relevance, because a lot of those data could be used to find out what was going on before atrazine was in use.

The second paper by Mattix, Winchester and Scherer, first a couple of minor points. They had a gap in the data from 1990 to 1995 to 2002. That was not explained in this draft report.

Also, they cite some CDC data and some Indiana data, but I'd like to know how the CDC data for Indiana correlate with what Indiana reports. Did they show very much the same thing? If there is a serious discrepancy, that needs to be explained.

The authors note that the elevated Indiana rate they report was statistically significant only in three of the years, but a critical question here is whether statistical power was great enough to say that an effect was present or greater in some years than in others or are we just looking at the effects of having small numbers of AWDs in each year. And were there any special features of atrazine use during the higher incidents years?

It appears that they didn't figure A2 were not adjusted for nitrates, and I wonder if that can be done.

The third paper, I'm again concerned about multiple comparisons especially because the confidence bounds on the adjusted rates for fields of corn with an odds ratio of 1.22, which was barely statistically significant, does not match the ratio for soybean fields, which does not suggest an effect.

And it's hard for me to see why atrazine in one kind of field would not have the same overall effect as atrazine in another field. And rough agreement with the paper just above the dates of conception were pretty much the same.

The next paper is focused on low birth weight, pre-term delivery and small for gestational age rather than birth defects. These outcomes are not independent, so the three sets of results may not provide much more information than any one of them. It would be worth checking the correlations if that is possible.

They use the geometric mean. It's not sort of why. If this was because of skewness in the distribution, that was simply the wrong thing to do. It's the high exposure points that concern us, and it's counter-productive to reduce their impact on the analysis by using a geometric mean.

Only one year is examined, so possibly year to year patterns could not be studied. It's an ecologic study with the strengths and weaknesses of such work.

The critical number of data points on exposure is the number of water distribution units which was not given in the materials that I have at hand, nor is there any analysis of possible co-variates correlated with distribution units such as ground versus surface water or local contamination by known sources of toxic chemicals.

The high point is again May to September, but now this is in terms of third trimester which puts the peak six months out of phase with the data on birth defects and no explanation for this is altered.

Table A7 summarizes the results. Only one of the nine odds ratios was statistically significant and barely made it. And further, we do not know what else the investigators may have looked at or worked on, so the problems of multiple comparisons come up again.

And the next paper, I think it was the last one, roughly 70 percent of the birth records available to the investigators came from one community, which raises questions about selective effects on reporting, how implied, was Fort Wayne dominant in the data and were there local confounders.

There's nothing in this draft report that explains that. It might be in the original paper.

The exposure data seems to be quite weak. Estimates are constructed from sparse data, especially sparse in the winter months. But the winter months are critical because in a sense they're basically the control period.

Weak signals of an effect were detected for SGA with exposures in the third trimester and the entire pregnancy, but that was not found for pre-term delivery, and low birth weight was not reported in this part of the analysis.

The range of confidence bounds is smaller for SGA than for pre-term delivery, though each is comparing roughly a three month versus a nine-month period, which may be the result of having substantially more sample for the entire pregnancy than for the third trimester, but this different is not explained in the materials I have here, and the last paper was not presented.

Overall, I would say that what is presented in this draft report, and I'm looking only at what's in this draft, seems to me to be entire compatible with no effect.

CHAIR HEERINGA: Thank you, Dr. Bailar, for the careful review of each of the papers. And, again, I think in conjunction with Dr Reif's comments about the quality of what's presented in the actual document in describing these studies, I think there are some things that have been pointed here that might be added to what you already have there.

I'm interested during the break in

speaking to someone - the corn/bean thing is a bit of a puzzlement. And I think it may be due to economic practices, no-till versus Roundup Ready beans and things like that when the atrazine goes on. I don't know if somebody can provide me a little background, but I'd like to be educated.

The next discussant is Dr. Hayton.

DR. HAYTON: Thank you. I read the Case study A descriptions and the papers of interest here, and I thought the case study fairly described the study design. I was satisfied with that.

I thought the assessment of exposure was acceptable.

The third question we were asked was whether the case study indicated appropriate statistical methods, and I thought there that there was an issue that really there was no judgment call on whether or not the statistical methods that were in those papers were adequate or acceptable.

So, if in fact the case study should have addressed appropriateness, I thought it did not.

The fourth question had to do with bias and confounding and other factors, and I thought those were reasonably well addressed. One thing that popped out to me, and that's the - I think it's the second paper. The Mattix paper where it's mentioned in that paper that the abdominal wall defect incidents occurring in the Riley Hospital - it says 279 over the 1990 to 2002 paper - fewer than half of those were simultaneously identified by the state registry.

I thought that was kind of disturbing that the concordance there was so low. So, I think maybe that needs some comment. I don't know what to make of that. Anyway, that was my response to the questions.

DR. BAILAR: Could I comment?

CHAIR HEERINGA: Dr. Bailar, then Dr. Portier and Dr. Bove.

DR. BAILAR: Abdominal wall defect is sort of a yes/no diagnosis. Probably somebody was poking at their abdomen, found a gap between the muscle on both sides that may be of no clinical significance, whatever. But one obstetrician doing that consistently with a high sensitivity to abdominal wall defect, could account for the whole thing.

CHAIR HEERINGA: Dr. Bove.

DR. BOVE: That's true. But in general, birth certificates and birth defect registry data doe not jive very well at all. And that's because the birth certificate data is just not a good source of information on birth defects. It never has been, never will be. That's why you need to use population-based birth defect registries.

And Iowa does use it in their studies. These other studies did not. That's a flaw. That means they're both under-ascertaining birth defects, and then there's also disease misclassification of the defects they do have. They may have the disease, they might have something else, they may not have it at all.

So, this is the problem when you use birth certificates for these kinds of studies. They should not be used for birth defect research. That's why we have birth defect registries.

DR. HAYTON: So I understand why there's no - or poor concordance there, but - so, what's the bottom line there?

DR. BOVE: Well, in a birth defect registry, they have to verify the diagnosis. They get medical records. They confirm it. Even the passive systems do that. And the difference sometimes between passive and active is not that big a difference.

One major difference would be some of the active systems actually go out to a year, sometimes they go out to several years, a good birth defect registry, for example, California's or New Jersey's or some of the other ones that have been around for a long time. They go out and capture defects that occur after the child, the infant comes home.

Birth certificates, it's just whatever a person puts down on the certificate, whatever is recorded at the hospital. It's a mishmash.

I've done this comparison in New Jersey, for things that you would think that you would - that would jive. I mean a neural tube defect, how could you miss it? And yet, they don't agree and the birth certificate is often wrong.

So, to explain it, I can't explain it other than the birth certificate is not meant for that purpose. The birth defect registry is meant for that purpose, and that may be the difference right there.

CHAIR HEERINGA: Thank you, Dr. Bove. That's a good observation.

Any more comments from members of the panel on this particular question?

Dr. Portier.

DR. PORTIER: So, I agree that the EPA's review of the atrazine case study examples, I think, was adequate.

For the EPA summary, I would have liked to have seen a comparison of how LOD observations were handled in each study. We know that how the LOD observations are handled can have a major impact on the summary statistics, on the associated confidence intervals, and on any of the statistical testing and modeling that's done.

So, they used half detection or did they estimate the missing data or did they set them equal to zero, all of the above, none of the above?

What I also found not adequate was the background for evaluating the study, so the context into which we went into these six studies.

It's almost assumed that we were reading these studies kind of with a blank slate, and in my case it was a true blank slate, and that the studies themselves would provide us all a background on the health, the target health effects, that they would provide us information on the reproductive health effects from atrazine, that they would provide information on temproal and spatial aspects of the reproductive health effects discussed in the studies.

So, a lot of these were not kind of provided. So, when I read the papers, I mean you read the papers cold, but then you want to know is this reasonable or not, because I didn't have the background context.

So, I would have liked to have known something about low birth weight, SGA, pre-term births and the general population and what kind of trends we've seen nationally and within these target states over the last hundred years.

I mean, we've had birth records for a long time. Although they're not perfect, some of these states have had birth registries for 50 years, maybe.

And at the same time as I was reading this, I wondered about things like well, does the body mass index or the mother have an impact on any of these outcomes? In none of these studies did they talk about the mother's body mass or the mother's weight.

They talked a little bit about health condition, but I didn't know how that was assessed. And I would think for these kind of birth defects, things like that, especially the mother's, quote, condition to have birth, would have a big effect on these kinds of outcomes. And I just didn't get that in the papers, and then I didn't have the context in the EPA case study.

So, I think as you move forward, you really need to think about the wrapping of the studies as well.

CHAIR HEERINGA: Additional comments?

Dr. Reed.

DR. REED: Yes, I'm still curious about the possibility of exposure to other triazines and their breakdown products that are supposed to have mode-of-action on some of the end points. So, that would be good to be addressed.

CHAIR HEERINGA: Dr. Bove.

DR. BOVE: the Iowa research did look at cyanazine as well as atrazine. So, they had some data. There's Rathburn Reservoir. They had some levels of all those in that reservoir, and they just compared to people who aren't on that reservoir, is basically how they did that study.

As for your comment, birth certificates have changed over time. So, what's recorded on them changes over time too. So, for example, in trying to do a study of birth weight at Hanford in the `40s, I couldn't do it. There was no birth weight information. There was gestations age though, so we looked at that.

Birth certificates also change in terms of what kind of maternal risk factors are there. In the early days, they wouldn't have smoking and alcohol. Now, they do. How useful that data is, is oftentimes questionable, but sometimes it captures some of the smoking, but the alcohol information usually is not very good at all.

There are also other maternal conditions in the birth certificate, so that's where all this information is coming from, right from the birth certificate, and birth weight is useful.

Small for gestational age is a useful end point because low birth weight sort of mixes together pre-term birth and - wait. I'm sorry.

Low birth weight is a mixture of small for gestational age and pre-term births. I guess it's getting late. And so by looking at small for gestational age and pre-terms births, we're separating two different outcomes.

Although, I do think that in the future, they should also look at the fifth percentile, not just the tenth percentile. In my own work, the fifth percentile seems to show a stronger effect than tenth percentile. I think tenth percentile is too broad or a term low birth weight, which is even more narrow.

But these are useful end points. The data is there. Oftentimes people look at these end points not because necessarily they're biologically plausible, but because you can look at them. The data is readily available, and they're looked at for that reason.

CHAIR HEERINGA: Additional comments?

Dr. Bailar.

DR. BAILAR: I would not dismiss the data on birth certificates too quickly.

They are full of errors. No question about it. Birth defects are grossly under-reported. No question about that.

But the critical question is not whether there are errors, but whether those errors are differential. Are they more or less the same in different places? Are they more or less stable over time?

My guess is that even with the changes in birth certification, that the errors in the them, the pattern of errors, the size, magnitude, direction have not changed rapidly over, say, a period of ten years, and I am even a bit more accepting of comparing patterns in different areas.

If you see something in one place, but not in another in the birth certificate data, I would give that some consideration.

DR. BAILAR: Didn't mean to denigrate birth certificate information. I use it all the time. I'm just saying that for birth defects, there's a much richer, better source of data, and that should be used in studies.

CHAIR HEERINGA: Okay. We have one more remaining part to Question 2, but I want to make sure everybody is fresh for that one. So, let's take a 15-minute break and return at 10 after 3:00.

(Whereupon, the above-entitled matter went off the record at 2:54 p.m. and resumed at 3:14 p.m.)

CHAIR HEERINGA: Question 2.4, can you read that into the record?

MR. DAWSON: Question 2.4, in light of scientific issues discussed in Questions 2.1 to 2.3, OPP requests input from the SAP on factors to consider when integrating these studies in the atrazine WOE analysis currently under development.

DR. REIF: I'm Dr. Reif, and I would like to ask -

CHAIR HEERINGA: We had some feedback from the audience, that they're not able to hear us. So, please pull your mics right up tight and speak loudly.

Dr. Bailar doesn't need to pull his quite as close, because he's got a big, booming voice, but everybody else speak clearly and closely into the mic so that everybody can hear.

DR. REIF: This last question in the series that deal with Case study A is, I think, an overview question of how these particular studies identified by the agency can be used in the weight of evidence analysis.

And I believe that with the probable exception of the Indiana study, that these studies have significant limitations that are going to make it difficult to do much substantively to incorporate these particular epidemiology studies in a weight of evidence approach to risk assessment.

And that's why I was tempted, I guess, on my own to look for other studies that were relevant to the question, not to answer the question of whether atrazine causes or is associated with adverse reproductive outcomes, but to look for examples of epidemiology studies relevant to the question that used other study designs or that used other methods of exposure assessment that are more informative when discussing the broad issue of incorporating epidemiology studies into risk assessment.

So, I went personally outside the charge a bit and identified a number of studies, all of which contain risk estimates for atrazine for a variety of outcomes.

And I'll put these of course into the report with the appropriate references, but again it wasn't to do risk assessment, and epidemiologic evaluation of risk for atrazine. It was to explore the diversity of epidemiologic approaches, and then to say to the agency now, if you were to consider, for example, this cross-sectional study, you could see what the strengths and weaknesses of a cross-sectional study are and how the cross-sectional data might be integrated into the risk assessment.

So, the example of a cross-sectional study is a study by Farr, et al., from the American Journal of Epidemiology published in 2004, which is built on the agricultural health study. So, it is actually nested within the AHS, which brings it to an additional level of relevance. And the outcome of interest here is menstrual cycle activity, length, irregularity, etcetera.

So, it isn't - the point is not whether atrazine was associated with aberrant menstrual cycle activity. The point is to say here is an example of an epidemiologic study well done integrating data from the AHS, which is using the cross-sectional approach, and now what can we learn from the study with respect to epidemiology's contribution to risk assessment.

So, that was the way that I thought through this after I got through the six studies and was somewhat disappointed in the lack of quality for most of the studies that's been described by other commenters.

There's another I think that bears discussion, and that is just as an example again, not as evidence for or against the health effect. And that's a study of male men from two states conducted by Shanna Swan and published in Environmental Health Perspectives.

What Shanna Swan did was to take urine samples from this group of men who had evidence of abnormal semen characteristics, i.e., they were cases, and another sample of men who had normal semen characteristics, and assess their exposures to atrazine using a single urine sample measuring atrazine metabolites.

So, again, it's another approach that I think adds to the breadth and depth of our understanding of how epidemiology can contribute to the central question that will be discussed, as you pointed out, in September.

So, I think these other sorts of approaches have actually extreme relevance to the question about the use of epidemiology in risk assessment. And that's why, in particular, I understand your response. And I understand also that your task in September is very large, because this initial step of discussing various study designs and how these particular study designs can be incorporated into risk assessment, I think, is extremely important to the central question.

So, that's one point that I wanted to make, and I won't go through the examples that I selected, but they're well-known and they're published in the epidemiology literature.

The other point I want to make about incorporating these studies in the weight of evidence approach, has to do with the potential shape of a dose-response curve.

That hasn't really been discussed here. So, we look at epidemiology studies, and many of us do the categorization of exposure using the quartile approach or using a tertile approach depending on the number of subjects that are in the study. And we, as has been described, let the exposure data, for example, in a case control study, let the exposure data for the controls drive the cut points for the analysis and apply those cut points to the cases.

So, that approach is obviously different from looking for a linear dose-response relationship that one might do with a regression analysis or other tools. And I'm not a biostatistician and others may wish to comment on this, but to me the issue is important because it goes back to the punitive mode-of-action of any chemical, that is, is there a linear dose-response relationship or is there a threshold.

And we haven't really talked about thresholds and linear responses here and it's slightly outside the questions, but I believe it's relevant. Because if you were to take the epidemiology data, most of which is examined as Dr. Alavanja described, in terms of quartiles of exposure, you get one answer that might be referable to a mode-of-action that involves a threshold.

Whereas if you take the approach of using statistical tools that look for a linear dose-response relationship, you're looking at a different approach that answers a somewhat different question.

So, I just want to raise that because I think it's relevant to toxicologists and the people who do risk assessment, as another important consideration in beginning to understand what the epidemiology findings bring to the weight of evidence.

CHAIR HEERINGA: Thank you, Dr.

Reif.

Our next discussant is Dr. Lu.

Alex.

DR. LU: Since I'm assigned to address this charge question, so I can kind of aggregate my comments for various question to here, so I sort of hope to facilitate as proceeded here. It won't be long, anyway.

And also, I'm going to use some of the slide that presented in yesterday and today's public comment section, because I just realized that I learned a great deal from the previous presentation. I think they are useful in my address, in my response to these questions.

So overall, this is my opinion: That those ecological study that's cited by the agency may not suffer from so-called ecological fallacy to the level that EPA has acknowledged.

The evidence that presented in front of me, and including the data analyses that present yesterday and today, actually suggests some possible link of atrazine exposure to birth defect.

For example, one of the slide I presented yesterday shows that if we were able to separate states from highest, medium and lowest atrazine use, we actually see a nice, seasonal effect. Meaning that during the early April, I'm talking about a highest atrazine state here, the increase of the birth defect is quite obvious and then that state assuming atrazine is the guilty party.

Take into account atrazine's half life in the water and the outcome of the birth defect measured monthly, they actually tell you something about the possible link.

How about those lowest atrazine state? Well, there is a signal trend, but not relevant to atrazine use. And keep in mind the atrazine is not the only teratogen or endocrine disrupting chemical that is being proposed right now.

So, from here it's clear that there is a link. The question is that whether this is a true link or a false link, and this is where the ecological fallacy come into play.

But it's obvious from this data analysis that we can rule out there's not generic variation associated with birth defect and associated with atrazine exposure. Otherwise, you will not see these type of a trend.

So, again utilizing the data like this nature will kind of rule out some most likely not being part of the game plan.

So, if we look at - now, say we have the national birth defect data which is not tied to individual state or individual regions. But what happen is that if we're able to link those incident data through data satellite in this case, atrazine concentration in the surface water versus number of tornado that hit in this area. I mean you start thinking about maybe those are possible linkage, right?

And it's up to the agencies or people that are interested to prove whether a tornado is a likely cause or the atrazine.

So, again, based just on those data, I would like to address the question the agency posed in terms of so what should be incorporated in the overall weight of evidence analysis and the risk characterization for atrazine?

Well, I would look at, first of all, the window of susceptibility. The differences between this graph and the next graph is - well, the difference is obviously one shows some relationship, some show there's no correlation. But another difference is, is that this data analysis does not take into account window of accessibility.

One of the public comment presenter used the average or median concentration of atrazine in the water bimonthly, and then tabulate - and then correlate it with the small gestational age.

We know or based just on the paper that presented in front of the panel, we know that the time which is defined as last metrical period, is critical for birth defect. And if there is a significant amount of atrazine in the environment, this probably you will see.

So, I think window of susceptibility is very important in the weight of evidence analysis. Especially if the data has no such component, I think the agency should actually re-evaluate part of the data in a way that makes sure we do not introduce a virus.

The second important factor, I would say, is the longitudinal or temporal variation of exposures and the correspondence to the disease outcome. In this case, birth defect.

Again, we see clearly there is the

temporal variation. Now, the temporal variation of the atrazine correspond to the birth defect recorded. In this case it would be national level.

So, is this important? Maybe. But if, say, for example, there is no atrazine variation month by month, whereas in the meantime you see a spike of birth defect reported in May, July, and you probably can rule out atrazine may not be an important player, or vice versa.

If the incidents of birth defect is distributed throughout the year, no matter how fluctuate atrazine concentration are in the drinking water or surface water, it doesn't matter. It's not an important player.

So, that's how I look at it in terms of you look at what should be incorporated into the weight of evidence analysis. I would put this two factors - I would weight these two factors heavily.

So, the next question is that

well, how are you going to address this epidemiology data?

Well, it's - again, my position here is that there's never a perfect epi study to address certain issues. So, we have to kind of think about what you have. And yesterday we spend some time to discuss this framework, and I found it very useful.

The reason because they tell you a lot about the evolution about biological plausibility in a sense.

For example, when we start using a lot of cell phones, we claimed that electronic magnetic field had something to do with brain tumors. There are animal data that strongly suggest that that's the case.

But as we go, there are some good human data - or epi data suggests that's not a case. So, we can move this box from here to here.

And yesterday we talked about melamine, which is the opposite case, right?

So, the question is that can we put the box of atrazine exposure and birth defect in one of the four boxes?

Obviously, it's not to the level we can put a box here, and neither here either.

So, the question is which one, where should we put it? You weight the evidence. Weight of evidence will help you.

The question to answer these two questions in terms of why did this pregnant woman living in - and I use this from the paper that we're assigned to read.

So, why does a pregnant woman living in Fort Wayne County, Indiana have a birth defect baby?

The answer to this question is not necessarily the same as the answer to this question. Why does pregnant women as a group living in Fort Wayne County, Indiana have so many birth defect babies?

So, the last factor I want to say

about the weight of evidence analysis approach, is that EPA needs to take into account the protection of public health. In my opinion, those ecological result actually - based on those ecological result, it is proven that agencies should leave these parts here.

There is some evidence for the associations, and leave for future datas. So-called good, quality epi data. Or in this melamine case, some incident data that dictate where this box that's temporally a part here, should go this direction or this direction.

Leaving the box here will actually safeguard public health with this part here. Will actually encourage more data because I do believe that one way or the other all the boxes here should be moved either to the Number 4 or Number 1 box.

So, that's just my comment. Thank you.

CHAIR HEERINGA: Thank you, Dr. Lu.

Dr. LeBlanc.

DR. LeBLANC: In gathering my thoughts to answer this particular charge question, I found a lot of repetitiveness in my thoughts. And I wanted to avoid that particularly as this day is coming to an end.

I was able to titrate my comments down to four points I'd like to make. And as I look at them now, I see there's still some repetitiveness in there, but I'll try and manage that as best I can.

Now, the first point that I have here is that the agency needs to give serious consideration to study selection, that difficult decision, perhaps, of what to use and what not to use.

In my experience, I could be in a room with a group of colleagues who are arguing that the EPA excludes relevant data in their risk assessment of a certain chemical. I could leave that room and go into an adjacent room, and hear another group of colleagues talking about the same chemical and arguing that the EPA includes junk data in their risk assessment of that chemical.

And I suppose if you're making no one happy, maybe you're doing the right thing. But I think there is certainly a challenge there, and I think part of the answer at least is - well, there are two answers. Two parts to the answer.

One is making some good judgments as to how you go about selecting data, and then the other is being transparent in that decision making.

And I can only assume you thought a lot about this and maybe you've even had an SAP meeting about it, I'm not sure, but - I'm not suggesting another SAP meeting, by any means.

(Laughter.)

DR. LeBLANC: But I think there are considerations that go into that judgment and there can be quantitative approaches to data selection. And you could take that approach, and you might be criticized by scientists saying that scientific judgment is involved.

And, again, I don't know the answer, and there are two sides to every coin. But there are - I think the EPA needs to give that a lot of thought.

I think it's an incredibly important point in making decisions as to how to use and incorporate this epidemiological data into the risk assessment of atrazine.

The second point is managing potentially new information. And this ties in with the points that I just made.

Say, for example, you have an epidemiological study where a novel observation is made with respect to potential effect, but that the decision had been made that the study wasn't going to be used.

If you simply file the study in the trash, I think you're setting yourself up for a lot of criticism that is they're not using relevant data in their risk assessments.

But accordingly if you use it, then you're setting yourself up for they're using junk data in their risk assessment.

And I think that maybe what you need to do is when you find yourself in that kind of situation, that the information needs to be filed away not in the trash, and it's not used in the risk assessment at this point in time, but it's filed, it's put away pending further investigation so that everybody knows you're aware of the data and you haven't forgotten about it. And you just need some corroboration and you need greater information before you can actually use that information.

The third point I have is that consideration should be given to reproducibility of observations among studies. And certainly that is repetitive with a lot of things we've been talking about.

But the only point that I want to make here is that we need to be thinking about reproducibility or consistency not only in effects that are observed in epidemiological studies, but where possible the concentrations at which those effects occur.

In the ecological world, I don't hear a lot of controversy with respect to the effects that atrazine causes among exposed amphibian populations. The controversy that I hear relates to the concentrations of atrazine at which these effects occur.

And I don't think you can separate the two. I don't think it's fair to say atrazine does this. The question is does atrazine do that at an environmentally-relevant exposure concentration?

And the last point I have is that the agency needs to continuously pose the question. He question being to what degree do the epidemiological studies decrease uncertainty with extrapolation from animal studies to the protection of human health?

Certainly in my mind, that's the

big question. We've discussed a lot of points, a lot of items in the past could of days, but I think it all titrates down to whether or not the epidemiological studies allow us to reduce the uncertainty associated with extrapolations.

And whatever the answer to that question is, the agency needs to use that answer in adjusting uncertainties accordingly in determining what uncertainty factors might be applied to the risk assessment of atrazine or other chemicals.

CHAIR HEERINGA: Thank you, Dr. LeBlanc.

Dr. Bove.

DR. BOVE: Well, I think we've been over and over all these issues, but let em just say a few things.

One, I do think that study selection is important, and I would want EPA to error on the side of being totally comprehensive as they can, including going through the gray literature if necessary.

Cast the net widely, and then give reasons why you're going to exclude studies. Good reasons.

But that gets me to another point. And that is that I think that when EPA is evaluating the epi research, they need to have epidemiologists review it. If they need to get help from outside, get help from outside.

So, if you're looking at drinking water studies, have expertise not only in - not only bring people in who have done these kind of studies, but also people who have done drinking water exposure assessment, done water distribution system modeling, groundwater fate and transport modeling, whatever. So, you have the right expertise evaluating these studies.

Similarly for occupational studies. Bring in the epidemiologists who have done these studies. Bring in the IH people who have done the exposure assessments, and have the right expertise there to evaluate these studies as you probably do with the tox studies.

So, I don't think I'm asking for something a whole lot different. Just bringing in the right expertise for these studies, I think that would help.

And, again, I think that when I hear that we want to see if the human data help in the extrapolation from animal data, again I want to get away from thinking that there's one set of data that's much better than another set of data.

Again, the epi data is looking at the right species, it's looking at the right exposures, it's looking at the right ways that people get exposed. So, granted there are advantages to tox studies and animal studies, I'm not going to dispute that, but there are also advantages of epi studies.

And I think the better idea is to look at both sets of data and see what it tells us and not assume that we'll get the end point from one and see if the other data set agrees with it or not. Look for the most sensitive end point that is being told to you by both data sets, and move forward in that way.

CHAIR HEERINGA: Thank you, Dr. Bove.

Comments from other members of the panel in response to this? Dr. Portier.

DR. PORTIER: I just had a clarification question, Dr. Bove. When you said bring in these outside experts, are you brining in the people who did the studies of the concern or are you brining in people who have done those kind of studies, but not the ones being, you know, is it the owners of the studies or people who know how to do those studies, but who could be freer to be critical of the studies?

DR. BOVE: The EPA could make that decision. Certainly you'd want people there who know how to do these studies, who know the pitfalls, who know how difficult it is and what these studies can actually tell you.

So, if the only people around are the people who actually did those studies, I guess you're going to have to include them. But I'm sure there are people out there, epidemiologists out there who have done drinking water studies on other end points.

For example, I've never done a drinking water study on atrazine. I've done drinking water studies on TC, PC and so on, for example.

There are water modelers who maybe haven't done groundwater fate and transport of accuracy, but have done floor solvents or gasoline or whatever.

So, there is that expertise out there. Okay. So, I think you can find it. I think you can find it.

CHAIR HEERINGA: Dr. Chambers.

DR. CHAMBERS: I'm getting a little uncomfortable with a few of the answers that have come up with this afternoon, this question, and earlier on some conclusions about atrazine.

I think the point of all of your questions at this point in time is generic methodological types of questions. And that this is - it's premature at this point, I think from your standpoint, and from our standpoint, to make any conclusions about atrazine's effects as such.

I'll be very uncomfortable if our report starts making some judgments at this point, because I think that's the point of the September meeting. And I would urge the panel to refrain from putting some conclusions about atrazine's effect at this point, because the data sets that you're providing us are not complete at this point.

CHAIR HEERINGA: Thank you, Dr. Chambers.

Dr. Lowit.

DR. LOWIT: Just to respond to Dr.

Chambers and not to cut off a lot of productive discussion, but I think there's certainly elements of what we thought we would hear in 2.4, we actually heard in 2.3, the cautions about individual studies and the way to think about those individual studies in a way that maybe we haven't yet in both the generic and the specifics of it.

I'm not encouraging you to cutting it off. I just -

CHAIR HEERINGA: No, I understand what you're saying with respect to the discussion in 2.3 has in fact identified a lot of the character of the individual studies that would be relevant to bring forward to this assessment.

Dr. Portier had a comment.

DR. PORTIER: This is just coming from my experience. When you bring in experts to review these things, the first thing you have to do is get them to all agree on what are the criteria of what they, as a group, are going to call a good and a bad study for that objective.

I've done these things where you get in and you immediately start reading the studies and you're trying to develop these criteria as you go along, and it changes. After you've read a whole bunch of bad studies, a kind of good study looks really good, right? And it's better if you can - and it's even better if EPA kind of lays down some general ground rules before you even bring them into that evaluation.

So, I totally agree that's the way to go for a lot of this stuff in terms of especially looking at utility and trying to put weights on utility. You're not going to get that statistically, but you're going to get that from a consensus assessment of the expert.

So, there's got to be a box around that. Otherwise, it becomes a moving target.

CHAIR HEERINGA: Thank you, Dr.

Portier.

I think what I would like to do at this point is to move on to Charge Question 3. And just looking ahead for this afternoon, it would be my intent to try to wrap up Charge Question 3 per our agenda this afternoon. And we will try to take care of Charge Question 4 tomorrow morning.

And I think that we should probably be fairly close to the agenda. Whether we finish - I don't want to force it to finish before noon. But I think if we have three-and-a-half hours tomorrow morning on Charge Question 4 and wrap up, we should be pretty close to finish just to sort of give you a forward look at this.

But if we could read Charge Question Number 3.1 into the record?

DR. LOWIT: We're going to do some quick musical chairs if you give us a second.

CHAIR HEERINGA: Sure.

CHAIR HEERINGA: Sarah Winfield.

MS. WINFIELD: Case study C describes various analyses and evaluations that can be conducted when evaluating human incident data. Please comment on ability to use incident data for the following types of analyses: trend of incidents over time, frequency of reported symptoms, common product clusters, frequency of repeated exposure scenarios, and assessment of children versus adult symptom profiles, which is in the diazinon case study, and please suggest alternative and/or additional analyses, if appropriate.

CHAIR HEERINGA: Our lead discussant is Dr. Lu.

DR. LU: I guess I'm going to start by saying that the human incident data for diazinon is quite unique in the way that those are acute toxicity regardless of how the data were gathered by different agencies.

And also the dose acute toxicity or the report of symptoms are very unique to all the member in the OP families.

So, I mean it's unlikely that we discuss in this case of diazinon would be attributable to other pesticide group such as triazine herbicide which does not have significant acute toxicity or very apparent symptoms that people can report it to.

So, I mean this alone would pose a significant limitation for future utilizations in this weight of evidence analysis and risk characterization.

The other limitation which the agency has acknowledged is that those incident report data are in terms of the quality, those incident report data are varies to a great extent. And the trouble is we don't even know how to quantify those variations. We don't know which one is good, which one is bad, and so on and so forth.

So, there's no doubt that if we are going to - if the agency is going to incorporate those acute incident symptoms or data, it will introduce unwanted bias and likely uncertainty to the future analysis. So, just in my opinion, those are my disclaimers.

Speaking of trend of incident over time, diazinon data obviously shows that once the use of certain pesticides is limited, the reported symptoms in terms of the number and the frequency were reduced as well. And I think that's the intention of restricting diazinon and other opiate pesticide use.

However, it's not clear that at the population level the exposure to diazinon, that will not trigger acute symptoms also reduced.

So, I'm talking about the emphasis on acute toxicity versus chronic health effect. Although we see a great reduction of self-report symptom data throughout this country, but there are exposure that has actually triggered no effect, no apparent poison symptoms at all. So, how would you account for those chronics?

This is even worse for pesticides that does not have apparent or dramatic acute toxicities like triazine herbicide, for example. So, I will suspect that the incident report for those pesticides will be sparse, inconclusive.

In other words, are their report, the incident report related to atrazine, related to one of the pyrethroids and so on and so forth.

So, having said that, there is important value for the incident data over time for use in risk characterization and risk assessment. So, the apparent decline of diazinon OP is related to a restriction of use, right?

So, what I just mentioned in Question 2.4 is that if - this is hypothetical. If we were able to remove certain pesticides, in this case the topic of discussion, atrazine, if we were able to remove atrazine from the water bottle in the critical month of women get into pregnancies, will we see a reduction in birth defects nationwide?

That's the value of looking at a trend of incident over time by taking away a pesticide or compound and we know was related to the health effect, and see whether that health effect will disappear as well. And that will actually enhance the hypotheses of certain chemical cause certain health effect.

And that could be done, but it would take a while. And that's what I'm suggesting.

If you think about a two-by-two box, it's prudent to put something in evidence against or evidence for and waiting for the new data to move the boxes around.

And data like incident trend over time manipulate by restriction of pesticide use or complete removal from the market, will help you to see whether this is the case.

And this is the second question in terms of frequency report symptoms. I'm going to lump this with the frequency of report exposure scenario, because I do think they are - you are asking the same questions.

In report symptoms, especially the similar symptoms or exposure by different individual within a defined time period might raise some concern not only for the use of patterns of that specific product, but also the potency of acute toxicities that has not yet been discovered or disclosed.

In the circumstances in which incident data reveal a health outcome that is not previously observed in the toxicological studies, right, those human incident data will be quite valuable in terms of exploring unfounded biological plausibilities associated with the specific exposures.

So, move on to the next question is common product clusters. We spent some time talking about clusters. There are confusing points, there are some misclassification, but cluster also provides some valuable information.

For example, the cluster actually occurred, right? But it's more of an acute public health concern instead of a risk assessment purpose.

For example, many years ago there is an incident on methyl pyrithione in ten or 13 states in the southeast regions. And we finally found out through ATSDR, it was caused by misapplication.

So, again, those information are critical for acute public health mitigations, but what is the value of the risk assessment?

It's an application error. It's almost a misapplication. It's supposed to not happen.

So, gain, I actually raise the question about cluster data. And sometimes it's manipulated, it can happen because of misapplication, it's totally irrelevant to regulatory risk assessment.

And the last question is assessment of children versus adult symptom profiles. This is my knowledge that the known acute toxicity of OP or diazinon is not differential between adults and the kids.

The dose that will trigger the acute toxicity might be different between the adult and the kids, but self-report incident data really contain exposure dose information.

So, again, I don't see how you can use those for assessing children versus adult exposure.

So, in conclusions, I will say incident data like diazinon has some value for risk characterization, especially for pesticide other than OP. But its value for risk assessment, in my opinion, is highly limited.

I'll stop here.

CHAIR HEERINGA: Thank you, Dr. Lu.

Dr. Chambers. Jan.

DR. CHAMBERS: I'm not really familiar with incident data. So, seeing this was my first pass at looking at that kind of compilation and I found a fair amount of it pretty confusing.

With respect to trends of incidents over time, I think this is, as Alex pointed out, this is a very good example since the numbers of incidents decreased with the decrease of approved uses of diazinon.

I think you have a pretty unique case here that will probably not be duplicated with other pesticides. And so how generalizable this is for the use of incident data is kind of hard to say. Probably not very generalizable.

With respect to frequency of reported symptoms, these can be tallied, but the tallies do not discriminate, as near as I can tell, between high and low exposure in Table B1.

Also, with the low numbers of

incidents such as the PISP had two incidents, the percent response don't make any sense when you have very, very low numbers. So, I don't know how you interpret that meaningfully.

Table B5 shows absolute numbers, but wasn't clear to me whether the symptoms in the generic categories are the same. And so if they're blending a lot of different symptoms into this generic category, then that may not tell you a whole lot either.

Another thing you asked about was common product clusters. And I guess I saw that in Table B, too, but I wasn't sure of what that actually was telling us. So, I don't know how to interpret that either.

Frequency of repeated exposure scenarios, I couldn't find where that was compiled. So, maybe that was in there and I just missed it, but I couldn't find it.

Assessment of children versus adult symptom profiles. These are compiled, but again it's unclear for similar symptoms between children and adults at least with this presentation.

So, maybe there's a lot of information out there that really got compiled very, very briefly here, but I couldn't discriminate a lot of the stuff that was there that maybe is useful.

You asked for potential alternative and additional analysis. About the only thing I can think of that might be worthwhile is to separate out the suicides and abuses of intentional exposures as opposed to just accidental exposures.

The accidental exposures will tell you a little bit more - or will tell you something about potential risk management issues or if the intentional exposures are just totally random. So, they would not tell you much about how your risk management processes are going ahead.

Any confounders or other factors that were present that may have been responsible in whole or in part for the symptoms reported should be determined. And if significant, then the reliability of that incident report should be questioned.

But then kind of back to the earlier question this morning, is it worth the time and energy of your staff to really dig into these data when they're going to be sort of isolated bits of data and maybe not really contribute to the quantitative risk assessment in any meaningful way.

CHAIR HEERINGA: Thank you, Jan.

Dr. Gold.

DR. GOLD: I took this question sort of generically. I figured the case study was just kind of an example and that you wanted more generic input on use of incident data. So, that's how I answered it, and I have a few points.

I think the advantage of these sources for evaluating incidents, trends and so forth, is that the data are collected in a relatively uniform manner with regard to product information or severity rankings and symptoms. So, that enhances the comparability of the data sort of over time.

The disadvantage is, I think, though, in using these kinds of data for the kinds of scenarios that you've outlined, are several. One is the lack of mandatory reporting by anyone other than registrants, so that you are likely to have under-reporting.

Second, the potential lack of concomitant information on trends in the amount of pesticide use so that it's not possible to determine if there are really more incidents or more usage.

And third, that you're largely only capturing - and this was stated previously - only capturing an acute event and not events that have long latent periods or are associated with long-term exposures.

So, I think the human incident data may be useful in the problem formulation stage in suggesting future research that should be performed and data that should be analyzed to assess better the magnitude of the relationship of specific types, doses and amount of exposure to specific health outcomes.

Also, providing evidence for mitigation efforts and providing information on whether human effects are consistent with those observed in toxicologic studies of animals.

In terms of the comparison of the distribution of symptoms in children and adults, I think it can provide supportive evidence. But I think the lack of similarity does not necessarily mean that the mechanisms are different, because they could also reflect different levels of sensitivity report. For example, people might be more likely to report symptoms in children than in adults. Just higher sensitivity.

Also, different routes of exposure

might occur in children versus adults. Kids put everything in their mouth. Adults are a little more selective.

And different sizes of the population exposed so that if you really have small numbers that are exposed, this could be resultant in less certainty in the distribution of symptoms.

CHAIR HEERINGA: Thank you, Dr. Gold.

Dr. Pope. Carey.

DR. POPE: Well, just similar to Dr. Chambers, I haven't really looked at this kind of information much in the past. I thought that it appeared to be that all these reporting systems were pretty effective at detecting changes over time that made sense with the changes in use of this particular one.

And, therefore, use of this information looks pretty good for looking at effects of risk mitigation and management processes.

In general the reporting systems, I thought, picked up relatively similar frequency, some signs and symptoms. There were some differences in categories like miscellaneous signs that may fit in with other categories from other reporting systems.

So, there were differences in terminology. And, also, it was noted that severity of signs or the symptoms could be different between these different reporting mechanisms.

So, it seems like maybe some more useful information could be gathered if you make more common recording instruments, if that's possible, between these different procedures and systems.

I think with regard to knowledge of repeated exposures, it seems this was not going to be nearly as relevant as incident reports for single exposure incidents.

Regarding the subpopulations with

the caveats that Dr. Gold just said, I think there is potential to gain something out of differential subpopulations and their responses from this kind of data.

The limitations regarding the knowledge of specifics of exposure, I think, is for me the biggest problem with this data as far as what a person or particular incident, what kind of exposure the person had.

It would seem to me that this may not even be for a - in this case, it may not be a diazinon exposure with certainty. So, I think that's a serious weakness.

I think a good strength of this kind of information is unanticipated responses. And I think someone mentioned earlier the idea of vigilance. I think this fits very well in here.

Vigilance for this kind of information resource can pick up unanticipated responses that might suggest you look at alternative mechanisms of toxicity that haven't come - haven't shown up in the risk assessment processes before.

So, for me the bottom line is that I think any information on pesticides can be used in the risk assessment process, but I think, for instance, data probably have relatively minimal influence on the quantitative aspects of the risk assessment process.

CHAIR HEERINGA: Thank you, Dr. Pope.

Dr. LeBlanc.

DR. LeBLANC: I don't have much to add here. I certainly concur with my colleagues with respect to trends of incidents over time that this is a good meausre of success in the risk management process.

And I think diazinon sort of represents a gold standard exemplifying the utility of evaluating trends and incidents. And it may very well be that a few other chemicals will meet that standard, but nonetheless I think it's an important measure. One that should be used in the analysis of incident.

Frequency of reported symptoms, I think that this is important and can be used to help establish plausibility.

For example, we've talked a lot about biological plausibility in our discussions at this meeting. But independent of biological plausibility if a hundred farmers report an implausibility toxicity following exposure to a pesticide, I think the agency needs to take notice of that and act accordingly.

Assessment of children versus adult symptom profiles I think is warranted. I think we all recognize the possibility, the potential for susceptibility differences. Certainly differences in exposure patterns that may exist between children and adults.

And just something I wonder, and

that is whether there are similar incident databases for non-human populations such as pets. And if there are, whether they are considered in looking at incidents, because it seems to me that pet incidents might serve as a good surrogate to exposure as it relates to children.

CHAIR HEERINGA: Thank you, Dr. LeBlanc.

Dr. Manibusan.

DR. MANIBUSAN: Sure. I just want to respond to the question of whether the agency considers other incident data beyond human data.

We do consider pet incident data as well as ecological incident information. And rather timely that you asked this question since we've just completed a really comprehensive review of the spot-on pet incident data for fleas and ticks, controls of fleas and ticks.

And we actually look to using

these same kind of opportunities to scrutinize this information in terms of from a surveillance standpoint, as well as understanding consistency, reproducibility, understanding where there's opportunities where we can do a better job in labeling, because we're seeing a lot of misuse of the product. For example, there might be a lot of situations where dog products are used on cats.

Where can we do a better job in risk reduction is very important for the agency.

I think, Dr. Chambers, you brought up a very good point. This information is very, very useful to look at trends and patterns and clusters in that way. And in particular, looking at product level clustering.

Because that gives us a sense going back, at trying to target whether it's the active ingredient or the inert or the combination that really is causing the effects reported.

CHAIR HEERINGA: Dr. Levine.

DR. LEVINE: Tina Levine. I just want to point out in a way it's kind of a separate kind of analysis.

One thing, the kind of data that we get on pets is aggregated. We don't get the individual case reports that we get for humans. They just send aggregate data.

And, also, we're not - we generally don't look at pet incidents in trying to evaluate what might be happening in humans. We're looking at the pet incidents in terms of what's happening in the pets.

CHAIR HEERINGA: Dr. Reed.

DR. REED: I think what I wanted to say is probably all said early on already with the Question 1.2. And I think what I heard is pretty much what I have in mind. But since my list is very short, I'll just repeat that.

But a philosophical difference, I

think, with being a risk assessor and looking at data is that when we talk about this particular type of data seldom has sufficient exposure data, I, in my brain, I translate it in terms of so it's not none, it's just seldom.

And so if we say that it has limited usefulness, in my mind I'm saying okay, so indeed we didn't say that it was never useful.

So, in risk assessment, we do take into account all the data that we have. Sort of how we do it at least within our group, is that we will still look at all the data that we have, including incidents data, and you can quickly scan through what might be useful, what might not be, and then go from there.

And as I said, there's two ways of looking at these types of data. One is looking at holistically as a group, and the other way is to pick out which ones might be useful.

And I think early on we mentioned something about the aldicarb in watermelon. And if that is considered as incidents data, it turned out to be it was very useful. We even came up with dose-response out of what's in people's refrigerator when the next day you go and collect these there.

So, I would not want to discourage the agency by saying it's not very useful, but I think it's important to look at these type of data with all the caveats that we have mentioned today, but not to dismiss it.

CHAIR HEERINGA: Thank you, Dr. Reed.

I think, Sarah, we're willing to go on to Question 3.2.

MS. WINFIELD: OPP plans to conduct analyses of human incident data like those described in Case study C for other pesticides undergoing registration review. In light of the scientific issues discussed in Question 4.1, OPP requests input from the panel on factors to consider when evaluating the reliability of human incident data and determining the relative weight that should be placed on such data in risk assessment/risk characterization or in problem formulation.

CHAIR HEERINGA: Our lead discussant is Dr. Chambers.

DR. CHAMBERS: Thank you. In my opinion, very little light should be placed on incident data for all the reasons that have been discussed in the last couple of questions.

Now, notwithstanding Ruby's comments a few minutes ago that can identify certainly something like aldicarb in watermelons that was unanticipated, and you should be alert to situations like that where you've got a cluster of something that is unanticipated.

But my sense of most of this incident data is that they have a very diverse nature with regard to estimated dose levels, product characteristics, the ability of the observer to accurately assess symptoms, and just kind of a scatter of information.

So, for the most part I don't think they're going to be very, very useful in any kind of quantitative risk assessment since the numbers of incident reports are great, such as what Ruby pointed out for Temik in watermelons.

If the exposures are well estimated and the symptoms are highly consistent, then perhaps the incident data would be useful, and you've done this in that particular case.

In cases of abuse or suicide, the data would not be very helpful for overall risk management because these exposure levels would be well beyond label recommendations and wouldn't be accidental exposures.

If reports are mainly described like flu-like symptoms or symptoms that could arise from physiological stress from the fear of poisoning or something like that, then I think that could be general symptoms for a variety of illnesses or conditions and it may be impossible to distinguish the pesticide effects from other confounding conditions such as infectious diseases or just general stress.

Certainly fear of poisoning could lead to symptoms that are just sympathetic nervous system reactions. And so I think you need to be alert to that.

The incidents are mostly going to be short-term type things. In many cases you're interested in long-term effects. And they won't give you much information along those lines.

The diazinon situation is probably uniquely suited for such an analysis. We've mentioned that before because of the types of clear symptoms that can be experienced acutely from an anti-cholinesterase and because of the risk mitigation measures of removing diazinon from residential uses and the consequent reduction in incidents.

Most other pesticides that you're probably going to try this for would not probably be, in my opinion, as adaptable to such a clear presentation.

And I'm not saying don't do it, but I don't think you're going to get something that's as clear-cut as this since it's an organophosphate and you have the same sort of risk mitigation type thing.

CHAIR HEERINGA: Thank you, Dr. Chambers.

Dr. Lu.

DR. LU: One remedy for overcome the severe limitation of using incident data like reported as for diazinon, is that maybe the EPA should work out with some kind of agreement with the agency that collect those data.

So, in the future when the case is calling and reported symptoms and ask them to donate some specimen sample so in a way that you can determine the reliability of the case that it's just reported.

Say, for example, there's an OP poisoning case. And in this case it will be accidental. I agree that the suicide and so on and so forth, those data should be tossed away because it's not true data.

But if the agency were able to work out some agreement in terms of getting specimen sample from the cases as reported, not only you can test the reliability of the case as just reported, you can also see what is the level actually will trigger the symptoms that's reported and so on and so forth.

It can be labor intensive. And I understand that different agency has different mandate in terms of why they college those data. And getting a specimen sample may not be their highest priority, but I think it's really needed, essential.

Especially nowadays CDC has the

mechanism that once you submit a sample, they can analyze and turn around rather quickly. So, it's not the problem. The problem is whether those agency will be cooperative in the matter.

But without a specimen sample, that's preferable. If not, a surrogate sample will be okay. But without those hard evidence, you never know whether that's a true positive or false positive cases.

I just don't know how you can screen those data without solid evidence in terms of the level that you can measure in that person.

CHAIR HEERINGA: Thank you, Dr. Lu.

Dr. Pope.

DR. POPE: Yes, I basically agree with the comments preceding. I just hit a couple of - one point, actually. Someone mentioned earlier about the longer term effects.

And I know with OPs after acute

intoxication, you have the long-term effects that some - I'm not suggesting that it be incorporated in these incident reporting mechanisms, but all of these people gain some information whether they do have some kind of long-term effect would be useful.

And then just reiterate once again I think that this kind of information is useful, but probably in a qualitative sense for risk assessment.

CHAIR HEERINGA: Dr. Bailar.

DR. BAILAR: I agree with OPP that despite their limitations, incident data sets are an important resource of Hill surveillance information on registered pesticides.

Weak as they are, they're sometimes the only data available. And sometimes weak is good enough for OPP's purposes.

OPP uses five different sources of incident data. I assure you there is some overlap in reported cases, but how much overlap is concentrated on more severe poisoning or in certain geographic areas or in other ways.

It seems that these sources are studied independently until the last stages of analysis when results are put side-by-side to look for signals of a problem.

There seems to be a need to a research study to consider how these five sources might be used in combination at earlier stages not necessarily by matching cases, but at least by organizing the data in ways that draw on the various strengths of the sources. This might be accomplished by a focused study in one or two areas where three, four or all five reporting systems operate.

Similarly, there is a need for study of clustering of reported cases. It is not clear how multiple human problems in two or more persons exposed together are handled.

In my view, a cluster is likely to indicate a more severe problem than an isolated case, but not more severe in proportion to the number of persons reported.

This is because reporting of one is likely to stimulate reporting of others in the cluster so that the cluster is artificially inflated over what would happen if reporting of individuals were independent and it's not clear how the data sets handle multiple reports about effects on the same person at different times.

It's well-recognized in he study of accidents, generally, that some persons are accident prone, but others are report prone. That is, they're unusually likely to perceive and report a problem. Even a very minor or non-existent one. This also needs study.

It seems to me that OPP should consider somewhat different approaches for chemicals that are transient in the human body, for example, some light aromatics, and those that are cumulative, for example, heavy metals or fat-soluble chemicals.

Data needs might differ with respect to exposure parameters, history of prior adverse reactions and other things.

Of more direct importance to OPP, modes of analysis and the concerns about interpretation may differ. I'm not sure that these can be comfortably accommodated in a single analysis.

To illustrate though I'm not sure how much of this will be available, some items especially critical to the study of acute toxicants, are severity in relation to recent exposure, immediacy of response after exposure and short-term group effects.

Some items that are especially critical to study, accumulative toxicants, severity is less important, the course of development of symptoms is important, and short-term group effects can argue against toxicity of the chemicals.

Some items that are critical study in both are consistency with animal data and with findings for related chemicals, evidence of misuse and clustering effects in persons tending to be exposed together.

There's a figure here, Figure C1, that shows a high frequency of miscellaneous. This should be broken out if that is feasible.

Also I'm concerned that here as I was for atrazine, that chemical toxicants simply do not generally have a very broad range of human health effects. Some do, but most do not, and this can be a red flag of pervasive bias.

Overall, it appears to me that this document is meant to give quite general, even somewhat vague advice, and to encourage and rely on the good judgment of informed scientists to draw appropriate conclusions which will necessarily be weak in the best of circumstances. I concur.

CHAIR HEERINGA: Thank you, Dr. Bailar.

Comments from other members of the

panel. Dr. Reif.

DR. REIF: Just briefly two thoughts about incident data and their potential usefulness.

First, validated instances of incidents regarding acute pesticide or other chemical poisonings or toxicities create the opportunity for either longitudinal follow-up studies or historical cohort studies to evaluate long-term effects of single exposures or multiple exposures.

And there are of course hypotheses partially tested about chronic exposures leading to neurobehavioral changes for some classes of pesticides. So, that also plays into the potential usefulness of those incident reports.

The other general comment about surveillance, which I think this really is, is that I think just to remember that a number of rare cancers were detected and associated with environmental exposures through occupational surveillance. Not through formal quantitative epidemiology studies, initially, but through the reporting of angiosarcoma in vinyl chloride workers or bone cancers in radium dial painters or mesotheliomas in asbestos workers.

So, that's another form of surveillance that applies the occupational setting rather than the incident reporting data here that's being discussed, but certainly historically was very, very important in identifying early human carcinogens.

CHAIR HEERINGA: Thank you, Dr. Reif.

Dr. Reed.

DR. REED: One of the kind of case study that we also cover to have through the literature search and find them and use them, is the clinical case report in the open literature, because there's usually a lot of follow-up when - by the time that they reported it. And I think for OP, that is an important source for chronic neurological sequella out of a single exposure.

And so although it's not mentioned as part of the database for incidents data, I think that's something that we used a lot or at least look into and see if there's something useful too. So, I would encourage the agency to look into that.

Maybe I'm pretty sure you already do, but not a part of this database, maybe.

CHAIR HEERINGA: Other contributions from panel members on the subject of the use of incident databases?

One obvious concern, and I think you've evaluated that, you're very close to these data sources, is that sometimes definitions change or - I think particularly in the ecological, I remember the carbofuran SAP. There was a historic, and I think it was mostly voluntary, record of incidents involving avian species.

And they lost funding and states dropped in and out. So, there are these variations in terms of the consistency of the actual tracking.

That's been brought out by others, but it's clearly part of understanding these data, too.

Other comments on the - okay. Well, we'll have an opportunity again to revisit any thoughts that you have on any of the parts of Questions 1 through 3, but it's been a long day and I'm looking around the room. I don't want to say you look tired, but you don't look as spry as you did at 8:30 this morning.

So, let me just turn to Dr. Manibuson or to Sarah. Any questions or clarifications that you'd like on -

DR. MANIBUSAN: No. I just appreciate all the comments. They're pretty much right on target.

I do want to express some

limitations from a few things that I heard in terms of recommendations for specimen collection.

I think it's really difficult to require that information from anyone to submit voluntarily. Because, again, these are probably cases are voluntarily called in to registrants or to NIOSH SENSOR or the different databases, and also to follow up on the need for understanding chronic effects through better follow-up.

Again, I know from discussions with ATSDR on the National Incident Database, that is a shortcoming of every database that we've talked about and that we've included.

The ability to follow up requires additional resources, and there's also privacy issues, people wanting to be contacted in the future.

A lot of things to consider, a lot of limitations, but we thank you for your comments.

CHAIR HEERINGA: At this point, then, I'll turn to our designated federal officials to see if there's any other administrative --

(Off the record comment.)

CHAIR HEERINGA: Just again to sort of reiterate, I think the proposed plan is that we will address Question Number 4 tomorrow and wrap up with the formal proceeding by the noon hour.

The panel will then have a writing session tomorrow afternoon. And if the panel would be willing to meet with me just for five minutes afterwards in the breakout room, we'll discuss.

With that, I'd like to bring this afternoon's session to a close, and we'll plan to see everybody again tomorrow morning at 8:30. Thanks to everyone for your participation today, and safe travels.

(Whereupon, the above-entitled matter was adjourned at 4:25 p.m.)

G

H

N

T