FIFRA SCIENTIFIC ADVISORY PANEL (SAP)

OPEN MEETING

THE POTENTIAL FOR ATRAZINE TO AFFECT AMPHIBIAN GONADAL DEVELOPMENT

U.S. ENVIRONMENTAL PROTECTION AGENCY

CONFERENCE CENTER- LOBBY LEVEL

One Potomac Yard (South Building)

2777 S. Crystal Drive

Arlington, Virginia 22202

October 10, 2007

8:35 A.M.

Page 2

U.S. ENVIRONMENTAL PROTECTION AGENCY
FIFRA SCIENTIFIC ADVISORY PANEL
OPEN MEETING
OCTOBER 10, 2007

DR. BAILEY: I think we're going to get going here. Welcome back to the second day of the FIFRA Scientific Advisory Panel Meeting. And this meeting is looking at the potential for Atrazine to effect amphibian gonads development. I want to thank Dr. Heeringa for being here as the chair. For anyone who wasn't here yesterday, I'll just run over a few things.

The FIFRA SAP is an advisory committee. It only provides advice and recommendations. The agency is responsible for making all decisions and implementation of those decisions.

We have asked the panel to complete standard government financial disclosure forms; and I, along with the deputy ethics officer, have reviewed these forms to ensure that all ethics requirements are met. We have established a public docket for this meaning, and all the materials that were handed out yesterday should be in the docket bind today. And they haven't gotten them here, yet, but they should be available very shortly.

Page 3

We will prepare meeting minutes at the, after the closing meeting. They should be available in, approximately, ninety days after the close of the meeting. And one final note, we are recording the meeting again today.

So, if you have comments to make, please give your name before you make those comments so that we can clearly hear on the recording who's making the comments. I think that's it, and at this point, I'll turn it over to Dr. Heeringa, the chair for today's session.

DR. HEERINGA: Good morning everyone, and welcome back to the second day of our FIFRA Science Advisory Panel meeting on the topic of a potential for Atrazine to affect amphibian gonadal development.

As Joe mentioned, I'm Steven Heeringa. I am from the University of Michigan, where I am a statistician specializing, primarily, in population-based research studies. I'd like the other members of the panel to also introduce themselves, and maybe a little background, and I'll turn it over to Ken Portier, and we'll move around the table.

DR. PORTIER: Good morning. I'm Ken Portier, Director of Statistics at the American Cancer Society National Home Office in Atlanta. And I'm a

Page 4

member of the permanent panel, and my expertise is in general applied statistics and probabilistic risk.

DR. CHAMBERS: I'm Jan Chambers with the College of Veterinary Medicine at Mississippi State University. My area is pesticide toxicology with emphasis on neurotoxicology and metabolism.

DR. SCHLENK: My name is Dan Schlenk. I'm in the Department of Environmental Sciences at University of California Riverside, and my, I'm a member of the permanent panel. And my expertise is in aquatic toxicology.

DR. BUCHER: I'm John Bucher. I'm the Associate Director of the National Toxicology Program at NIHS. I'm a member of the permanent panel. And I have a background in carcinogenesis and general toxicology issues.

DR. ISOM: Good morning. I'm Gary Isom from Purdue University and Professor of Toxicology. My area of interest is neuromechanisms of, or molecular mechanisms of neurodegeneration, and I'm a permanent member of the panel.

DR. GREEN: My name is Sherril Green, and I'm from Stanford University. And my area of expertise is in Xenopus laevis amphibian biology and disease. And I'm an SAP member, and had some discussions with

Page 5

United yesterday about a sixty-eight person standby list, so I finally made it today. Thanks.

DR. PAULY: My name's Bruce Pauly. I'm with Environment Canada in Ottawa, Ontario. I'm a wildlife biologist with a particular interest in the effects of pesticides on wildlife. And I've been doing studies on pesticide effects on amphibians for the last while.

DR. SKELLY: My name is David Skelly. I'm a professor of ecology at Yale University. And my interests include the study of natural populations of amphibians, including developmental deformities.

DR. DENVER: I'm Robert Denver from the University of Michigan. I'm Professor of Molecular Cellular Developmental Biology. And I'm a neuroendocrinologist. And I study hormone action in amphibian development.

DR. FURLOW: My name is David Furlow. I'm a Professor of Neurobiology Physiology and Behavior at the University of California Davis. I am an endocrinologist, specializing in thyroid hormone and steroid hormone action, and including in amphibians.

DR. YEATER: My name is Kathy Yeater, and I'm an area statistician with the U.S. Department of Agriculture, Agricultural Research Service, where I

Page 6

specialize in applied statistics of biological and agricultural statistics.

DR. BAILEY: Tim Bailey, Professor at Iowa State University in Department of Statistics.

DR. DELORME: Peter Delorme. I'm a Senior Science Advisor in the Environmental Assessment Division of the Pest Management Regulatory Agency of Health Canada.

DR. LEBLANC: Gerry LeBlanc. I'm a Professor of Toxicology at North Carolina State University with research interest in endocrine toxicology.

DR. MILLER: Debra Miller. I'm from the University of Georgia College of Veterinary Medicine, and I'm a veterinarian pathologist.

DR. PETINO: I'm Reynoldo Petino. I'm with the U.S. Geological Survey Texas Cooperative Fish and Wildlife Research Unit in Lubbock, Texas. And I'm a comparative and reproductive endocrinologist and physiologist.

DR. HEERINGA: Thank you, again, members of the panel for your willingness to participate in this multi-day meeting. We, again, we recognize very busy schedules at work, academia for many of you, and appreciate the fact that you're able to be here and

Page 7

participate. So, at this point, before we turn back to the proceedings, John Bucher mentioned that he had one clarification from yesterday that he wanted to get in the record.

DR. BUCHER: Thanks. John Bucher. Yesterday, I was asking a question about the relative doses of estrodyle used in the Syngenta studies versus the Hayes studies. And I misread my notes and indicated that the Hayes positive control list, actually, concentration of half of what Syngenta was using, and in fact, that's incorrect. And the dose that he was using was much higher than that. So, I withdraw that question.

DR. HEERINGA: Thank you very much for that clarification. At this point, just to review where we've been, I asked the panel. We have had our first of initial overview presentations from the EPA scientific staff.

We have, then, had a period of extensive public comments, including a description of the data call-in studies, represented by Syngenta crop protection. We've had additional public comments from representative of the Agriculture and the Natural Resources Defense Council. And, so, we've had a fairly good introduction to the topic.

Page 8

I understand that, while I was in College Park teaching, that you've managed to finish up the presentations on the studies and the statistical analysis of those studies, and that, at this point, we are ready to approach the conclusions, I think, that would be presented by Dr. Steeger. And, also, we have one additional presentation on the power analysis that was done for the Syngenta DCI study.

So, we'll turn to those in order, but we'll complete our discussion and presentation with EPA scientific staff. And then we'll go to the power analysis presentation. And then we'll have a general set of wrap-up questions for clarification from the panel on any remaining items that you'd like to have discussed before we move to the charge question.

So, at this point in time, I'd like to turn to Dr. Thomas Steeger of the Office of Pesticide Programs at the EPA, and ask him to, maybe, introduce the staff that's with him this morning, scientific staff, and then, also, to proceed with the conclusions and the EPA's reviews.

DR. STEEGER: Okay, good morning, and thank you for the opportunity, again, to address the FIFRA Scientific Advisory Panel. Sitting with me at the table today is Mary Frankenberry, a senior

Page 9

statistician with the Environmental Fate and Effects Division; Dr. Sig Degitz, who is a Research Biologist and order at the Office of Research and Development, The Mid-Continent Ecology Division.

Next to him is Arthur, Arty Williams, who is the, I'm sorry, Acting Director of the Environmental Fate and Effects Division; Dr. Stephanie Irene, who is a Senior Advisor at the Environmental Fate and Effects Division; and Anita Pease, who is a Senior Biologist in the Environmental Fate and Effects Division. All of us are co-authors on the 2007 white paper, and many of us were co-authors on the 2003 white paper.

Yesterday, we spent most of the day discussing the data that the agency considered relative to the effects of Atrazine on amphibian gonad development. While focus has been on the registrants submitted studies that are in response to the agency's data call-in for tiers, tier one studies, the agency has continued to look at open literature.

Over the past four years, since the last FIFRA SAP on this same issue, laboratory and field studies have continued to examine potential effects of Atrazine on amphibians.

The Agency has determined that none of these studies have taken into account the recommendations

Page 10

made by the Agency and the 2003 FIFRA SAP to reduce potential sources of variability that could effect the endpoints being measured in the studies. Taken at face value, though, both the laboratory and field studies have failed to demonstrate a consistent, dose-dependent effect due to Atrazine exposure.

Some authors, that claim to show a consistent effect in both the laboratory and field studies in previous studies, now demonstrate no effect on Atrazine, of Atrazine on timed metamorphosis or gonadal development, and claim that the lack of response is due to biological variability.

The Agency is obliged to follow specific processes in its risk assessments; and yesterday, we discussed the risk assessment proc-, the paradigm used by the Agency. Ecological risk assessment process is further described in the document entitled, "The Overview of the Ecological Risk Assessment Process" in the Office of Pesticides Program document.

In considering open literature, the risk assessor adheres to guidance to assure this consistent approach is used to evaluate all studies.

The criteria used to evaluate open literature were discussed yesterday and include, experimental design; study protocols and quality assurance

Page 11

measurements; the strength and shape and cause, of the cause and dose, cause and effect relationship; whether there was a dose response; whether observed effects have a plausible mechanism of action consistent with the known; what is known about the chemical; and finally, whether the measured effects are ecologically relevant. These same criteria are applied to guideline studies, as well as non-guideline studies.

Contrary to what was suggested yesterday, the EPA cannot dictate to non-registrants how to conduct studies. Nor does the agency, typically, require researchers, other than the registrants, to provide raw data. The agency has to rely on the journal peer review process to serve as the primary reviewers to open literature. For the previous SAP, the agency did try to work with researchers outside those that are regulated by the agency.

That process proved to be resource intensive and completely unproductive. As well as-, as will be discussed-, as was also discussed yesterday, the agency makes use of multiple lines of evidence to determine whether the use of a pesticide represents a threat to human health and the environment.

The prim-, the Agency primarily relies on guideline studies that are intended to provide

Page 12

information on whether the chemical affects acute mortality or chronic survival, reproduction, growth of non-targeted plants and animals. Guideline studies include both laboratory and field studies.

The Agency also relies on nine guideline studies, either submitted by the registrant, and or recorded in open literature, to determine whether additional effects, not measured by guideline studies, are associated with the use of a particular chemical. Field studies are typically used to determine whether effects observed in the laboratory are apparent in the field on their actual use conditions.

It is possible that well designed field studies can identify chemical effects prior to laboratory studies; however, it is likely that the laboratory studies would then be required to verify that the effects, to verify the effects, and to better allow the determination of the dose of response; and if necessary, establish the mechanism to which the effect occurs.

In 2003, the Agency appeared before the FIFRA SAP to discuss its assessment of data for non-guideline laboratory and field studies. Based on the open literature studies, the Agency recommended that, while there was sufficient information to conclude that the

Page 13

Atrazine causes gonadal effects in amphibians, there was insufficient information to formulate a hypothesis, or there was sufficient information to formulate a hypothesis. The agency proposed, and the FIFRA SAP concurred with, the tiered process for examining whether Atrazine exposure results in effects on amphibian gonadal development.

We are meeting here this week to discuss the results of studies designed, based on recommendations from the EPA and the FIFRA SAP, to address whether exposure to Atrazine effects amphibian gonadal development. We have, also, discussed the available open literature. Yesterday, Mary Frankenberry provided an overview of the statistical analysis and of the DCI studies.

Based on that statistical analysis, the Agency has developed the following conclusions. The agency, again, has reviewed the total of thirty-six documents, representing both interim and final reports from open literature and registrant simulated studies related to the potential effects of Atrazine on gonadal development.

In doing so, the agency used experimental designs, study protocols, extent of quality assurance and control, strength of the shape, strength and shape

Page 14

of the cause-effect, and or dose-response relationship, both classical and U-shaped dose response curves, mechanistic plausibility, and ecological relevance as evaluation criteria.

Based on the available data, the Agency concludes that Atrazine does not produce a consistent reproducible effect across the range of exposure concentrations and amphibian species tested. The lines of evidence do not support that hypothesis, that Atrazine exposure causes effects on amphibian gonadal development.

In 2003, in the white paper, and as previously described by Dr. Degitz, the Agency proposed a tiered approach to determine the effects of Atrazine on gonadal differentiation in anuran amphibians. At that time, the FIFRA SAP concurred that the approach was reasonable.

In response to the recommendations made by the SAP, consistent with what was proposed in the 2003 white paper, the agency required the technical registrants of Atrazine to conduct tier one studies to test for apical gonadal effects. During this SAP, the agency has provided its analysis of the registrants submitted studies that were responsive to the data call in.

Page 15

The agency has also provided its analysis of the open literature studies and its effects, and the effects of Atrazine alone on amphibian gonadal development. Based on its analysis, the agency has concluded that Atrazine is not affecting amphibian gonadal development. The agency is also concluding, based on the tiered study approach proposed in 2003, that since the tier one studies reveal no effect on apical endpoints in Xenopus laevis, no additional testing of amphibians for amphibian gonadal effects is warranted.

The 2003-, in 2003, the SAP was asked whether there was anything that would preclude the use of the African clawed frogs as surrogates for amphibians, and whether there were any important differences to conclude that any developmental processes of Xenopus laevis would not, also, occur in ranids. In response to these questions, the SAP could not identify any differences. Therefore, the agency concludes that additional tier one testing with other amphibians is not warranted.

Consistent with the Agency's iterative process for evaluating ecological risk, the Agency will continue to evaluate data as it become available; and where necessary, the Agency will develop appropriate

Page 16

measures to mitigate risk to ensure human and environmental health.

Similar to what was done in 2003, the Agency is seeking input from the FIFRA SAP on the Agency's evaluation of the available literature. In the next presentation, Dr. Stephanie Irene, Senior Advisor in Environmental Fate and Effects Division, and co-author of the 2003 white paper; and Ms. Anita Pease, Senior Scientist in the Environmental Fate and Effects Division, will read the charges to the panel.

The SAP is being asked to comment on the Agency's evaluations and conclusions regarding the effects of Atrazine alone on amphibian gonadal developmental data. Additionally, the SAP is being asked to comment on the Agency's conclusion that higher tier testing is not warranted. The SAP is also being asked to comment on the Agency's conclusion that no additional testing on other amphibian species is warranted. Thank you.

DR. HEERINGA: Thank you very much, Dr. Steeger. Members of the panel, any questions for Dr. Steeger at this point? Not seeing any additional questions, at this point, I'll have one more opportunity before we move to the charge questions. I would like to propose that we move to the-, you're

Page 17

finished with the presentations for the formal portion at this time?

DR. STEEGER: That's correct.

DR. HEERINGA: If that's correct, we move to the final item, which was left for me yesterday afternoon, which was requested by the panel, and that is a discussion or presentation on the power analysis that was conducted for the Syngenta study, and then permitting that is a, I think, an important contribution to the statistical understanding of the study data analysis. So, I think, at this point, Mr. Hosmer, Bob Sielken, will be doing that presentation. Panel members, you should have, I think, a handout with the slides in front of you.

DR. SIELKEN: Thank you, Mr. Chairman.

DR. HEERINGA: He wants you to mike, Robert.

DR. SIELKEN: My name is Robert Sielken. I'm a statistician with Sielken and Associates, and a consultant to Syngenta Crop Protection. Thank you, Mr. Chairman, and thank you to the rest of the panel for this opportunity. Dr. Portier raised the question yesterday about power, and I do have a presentation on that, on the power studies that we did for the Atrazine study. I have in the slides, the post hoc power

Page 18

calculations that we did, and I'll discuss those.
I'd like to, also, indicate that prior to the conduct of these DCI studies, and in fact, prior to the conduct of the preliminary estrodyle studies that were used to kind of iron the issues of conducting these Atrazine studies, we did do a power analysis for those estrodyle studies. We had been involved with the Karr, et al, and some of the earlier studies sponsored by Syngenta, and we had familiarity with the type of experiments that were being done.
Drawing on that experience, for example, that we were seeing a background rate of what went on, a .04 percent in some, or .4, point four percent, in some of the back-, or 4 percent in the background for some effect, like newt six, in some of the earlier studies, we used that as a nominal background rate, and asked some power questions, relative to that background rate as a planning step for the estrodyle studies. And we looked at a range of tanks from four tanks per treatment, up to about forty tanks; and then, also looked at a range of animals from two per tank to eight per tank.
So, we did an analysis early on to see what would be reasonable. We concluded that using eight tanks in the controls and eight tanks in the treatment

Page 19

groups would be a reasonable, give us reasonable power relative to a four percent background.
Then that would give us the opportunity to detect anywhere between ten and twenty percent of, say, something around the mid-six, if it were to occur against a four percent background. And we would be able to detect ten to twenty percent with an eighty percent, we'd have eighty percent chance of detecting such and end effect.
In the estrodyle study, this was the first study that EPL and Dr. Wolf were going to be asked to look for abnormal findings in the gonadal development area. And in order to accommodate, give him a background in what was normal, we included, in that early estrodyle study, a reference group. I don't want to call it a control, because it wasn't used for treatment control comparisons.
But it was a reference group that was untreated that would, allowed him to look at those animals to see, to give him an index of what was kind of normal. So, the, we did end up with sixteen tanks being untreated, that eight of those were really controls, and eight of them were a reference population.
When we went to the design of the Atrazine

Page 20

study, it was logistically just as easy to go with sixteen controls, as it was to go with eight, given the dimensions of the room and what was going to have to be done to actually accomplish that study.
So, they went, in the planning stages, with sixteen controls to give them, sort of, a cushion, a security, if you will, that if something should happen, that the whole study wouldn't just disappear. So, that's the reason why there was sixteen and not eight in the planning stages for the Atrazine study.
Now, let me talk about, in terms of the slides that I have, what the power was, looking back, now, at the eventual design and eventual fate of the Atrazine study. And as most of you know, here, power is the probability of getting data that yields a significant result. It's a property of the design of the study.
It does not equate to what was the least significant difference that was observed in the study. It's, the power is, rather, the probability of getting a significant result if something is true. And as the curve indicates up there with the horizontal axis being the size of the effect that might be there, you're likelihood of detecting, your power of probability of detecting it increases as the effects configure.

Page 21

And, so, to talk about power, you're really talking about a power curve. However, in order to facilitate conversations, it's pretty convention to talk about the effect size that gives you an eighty percent power, or a probability of .8 to detect something, and to use that as an index of this whole curve, but bear in mind that you do have a greater probability of detecting effects that are larger, and correspondingly smaller power to detect smaller effects.
I'm going to talk about power in two contexts. One is for the measurement employs, which are the continuous variables like age in metamorphosis, body weights, snout vent length, gonadal image area, those types of endpoints. And, initially, I'm going to be talking about effect size or some people would call this a standardized effect size. And it's really what mean difference from control can you detect with the eighty percent power in looking at that mean de-, mean difference as a ratio, with the underlying between tanks standard deviation.
And so, in the studies, we've got power numbers for both IGB, which ended up with sixteen control tanks, and the Wildlife International studies, which ended up with eight control tanks. And starting

Page 22

first with the IGB number of 1.5, that's saying that in, for a measurement endpoint, if the true mean difference was one and a half standard deviations, we would have eighty percent power of detecting that. And that's a, really a result, it's a simulation of the whole process.

And by the whole process, I mean that for most of these endpoints, you go through a protective F test, and if that's significant, you follow with a peer wise comparison at five percent level.

And, so, we've got that sequential testing, and if you go through that whole system, the probability of detecting the difference is eighty percent, if the, if you're talking about a difference that's one and a half standard deviations. On the, for the Wildlife International study, that number moves up to 1.6.

So, you have a little bit less power when you have eight control tanks. It's not a dramatic shift in the power as you from sixteen to eight. In fact, I can hardly see if there's much shift at all.

This is, although, this might seem a little bit surprising at first, there is two things going on here. There's sort of a point of diminishing return, standard deviations act like the square root of the

Page 23

sample size, but also, here, we're still talking about having, at both IGB and Wildlife, eight tanks per treatment.

And the next slide is also in response to something that is, more than one of you mentioned, but I remember Dr. Portier mentioned specifically, was what was the observed variation in these two studies. And could I give something that kind of gave a feel for what was the variability in these studies.

And this slide shows the coefficient of variation, which is the ratio of the standard deviation to the mean, so, giving an idea of the relative amount of variation. And that is different, of course, for the different endpoints, because there's underlying different standard deviations for age, body weight, snout vent length, and gonadal image area.

But looking, for example, at Wildlife International, the standard, the coefficient of variation is about 2.1 for females, 3.4 for males, and then 3.4 and 3.9 at IGB. And so, you can look at these numbers and, yes, they're not identical, of course, but they're fairly close. They're fairly comparable. Which is saying something about the comparability of these two studies. And there are some numbers here on this slide for you to look at. I apologize that I

Page 24

added this after Dr. Portier's request. This one slide is not in what I distributed, and I will reproduce it and get it back to the panel this afternoon.

SPEAKER: Thank you very much.

DR. SIELKEN: All the other slides are in there and in order. The earlier slide where I showed that in will, if you had power, eight percent power at 1.6 standard deviations for will, and 1.5 for IGB. Well, what does that translate into, in terms of percent differences from the control.

And, for exam-, and these are the numbers that, kind of translate standard deviations, if you will, into percent differences from control, if that makes it kind of easier to understand. And that's really just a product of that effect size expressed in terms of a ratio difference in standard deviation, times, or factoring in the coefficients of variation, gives you these numbers.

And these would be the numbers if, in fact, the observed coefficients of variation are then true underlying coefficients of variation, and the, render then powers of theoretical calculation.

Okay, and you can see that these numbers, like age, that you're looking at in the will study, being able to decipher the percent of power of three to

Page 25

five percent change in age, depending on whether you're talking about females or males, and in IGB, it's five to six percent, right. Now, in case, I mean, and this gives you the idea that these are really fairly comparable, I mean, you've got 11 and 9.8 for body weight, and two elevens for IGB.

So, these are fairly comparable. They're, this calculation shows that the age in sixteen didn't really make too much difference in terms of power. It, also, shows that, you know, with a little bit more variability at IGB, that little bit of variability, coupled with a few more control tanks, really made these numbers quite comparable.

So, it's a coupling with both of those factors in there. In terms of a power curve, not just looking at the eighty percent point, you can, actually, draw a power curve, the power increasing on the vertical axis, and the true difference, in terms of difference divided by standard deviation, on the horizontal axis.

The solid line is the power curve for sixteen control tanks. And the dotted line is eight control tanks. This particular picture is for a measurement endpoint. So, the point of this picture is, you know, to remind you that these are power curves. And also,

Page 26

that the difference between sixteen and eight is not a dramatic difference, and to quantify that.

When it comes to looking at incidence, that is, the things that we evaluated at percent, presence or absence, and what percentage was present, what percentage did you have in the treatment, what percentage did you have in the control, and here, effect size is not relative to a standard deviation, but just difference in those two percents. So, that first and, your power would differ, of course, depending upon whether you're using something that involves all frogs, a one-sided test or a two-sided test, or whether you're using males and females separately. All of these cases were, you know, in one analysis or another. We've got lots of analyses, and they're one of these types.

Here, the power actually depends upon the background rate. At a low background rate, which I've, I couldn't do exactly zero. I did a nominal, something close to zero, .1 percent, and asked, what difference from .1 percent, how big would the percentage have to be in order for me to have eighty percent chance of detecting it. And it will, it's 3.5 percent or 2.7 percent there in that first line. And you can see that there IGB has slightly more power than will, and you

Page 27

can quantify the magnitude at those differences. And, kind of, if you go all the way down the page, they're comparable. They're not identical.

This calculation, also, involves the, since we're dealing with tanks as the unit of analysis, it involves, this calculation requires to specify a correlation within the tanks. We've picked the number here of .02, which is about, is the upper confidence limit of the correlation that we estimated.

So, this is a fairly conservative value for the correlation. And then, of course the great the correlation, the less the power, because the correlation effectively reduces the sample size. And these, the power for incidence endpoints would change the picture as you move to a higher percent backgrounds.

From these first few slides, you can see that, even though the number of control tanks in the will study did end up being reduced from sixteen to eight, four are lost to a microbial blurb, and four lost to trace Atrazine contamination. The power of the will study was not significantly compromised.

And that the power of the will study is still, approximately, equal to, or maybe marginally less than, the power of the IGB study. So, it's still

Page 28

reasonable to compare the results in the two studies from the perspective of power.

The other types of comparisons that are being made at different times, at least in the white paper in Syngenta's reports, is to compare what happened for the positive controls, the estrodyle or E2, and what happened for the Atrazine. And for a lot of endpoints, there was an effect for E2, but you did not see that same effect for Atrazine.

So, the positive controls were responding. The Atrazine, there was not that same adverse effect. And, in order to illustr-, in order to, for that comparison to be fair, it has to be true that the power for the E2 statements and the power underlying the Atrazine statements are fairly comparable. I mean, it wouldn't be fair to say, I found differences in one place if I had a lot of power, and I didn't find them over here when I didn't have power. That wouldn't be fair. So, it's important to show from the power perspective that the powers were comparable, and that it was a fair comparison to say, we sought the positive controls. We did not see it in Atrazine.

And, there are a whole bunch of cases that I could have enumerated, and of course, in the finite time available, I chose to just give a few examples.

Page 29

And, really, I've got examples of where Atrazine compares then to its slightly greater power than the E2.

Some where they were, approximately, equal, and some where Atrazine had slightly less. In almost all of these comparisons, you'll see very little practical difference in the powers, but let me just illustrate that for you. Here is a couple of the endpoints where Atrazine had comparisons at slightly more power than the E2 comparisons. And, what you're looking at is, really, in the bottom is those pairs. I can point to them one place, but of course, I can't point to them everywhere. And from here, apparently, I can't point to anything.

So, that's all right. Since I can't point to three screens, and I am not real sure how to use the mouse here, I'll just say, look at that line for will. And in that line for will, that age for metamorphosis, it says that in order to have an eighty per-, I mean, 80 percent. You have 80 percent power, thank you. It's working, but it just doesn't go to that far out. A little more powerful. Yeah, there we go. Thank you for all.

The 6.4 says that if you had the, if you have 80 percent power to detect, and age at metamorphosis

Page 30

that's increased by 6.4 percent over the controls, when you're talking about until you, the power for, you would only need to be 5.4 percent increase in detecting Atrazine. And I'm not, that's really not much difference. I wouldn't make much of an issue of it. And for IGB, you're looking at 6.3 versus 5.9, and you can see the numbers on the right-hand side as well.

The point here is not the exact numbers, but the comparability of what you get for E2 and for Atrazine. And the same, I mean, there are some that are really almost exactly the same numbers for both E2 and Atrazine. And there are also some where Atrazine is slightly more powerful. That first column is a little confusing, because if you're talking about percent male, you have 80 percent power.

If the percentage had dropped from controls down to 29, you could detect it at 80 percent. And from 48 down to 27, it would have to be how much it dropped before you pick it up with Atrazine. So, those two numbers are fairly comparable.

For a member of the course that, when you're talking about percentages, the points in the middle, around 50 percent, are the most difficult to detect changes. So, that's why those numbers would go from, like, 48 or around 50 down to around 30, because it's

Page 31

very difficult to detect changes in 50 percent. If you go from, say, 1 percent to 5 percent, that's a lot of, it's a fivefold increase. That's a lot easier to detect, than, say, to go from 50 percent to 55, which is a relatively small change, and given that things are quite variable at 50 percent.

So, the whole point of these slides was to say that, if you're looking at the Atrazine comparisons, and you don't find it in Atrazine, and you do find it with estrodyle, it's not because they were different in the power. It was different in the way estrodyle works and Atrazine works.

In the DCI studies and in EPA's follow up to those studies, we not only looked at a protective test followed by a pair wise comparison. We, also, did trend tests on everything. And we did those trend tests regardless of the protective F test or the protective Kruskal-Wallace test.

Trend tests were done across the board. And they were done to increase the likelihood of finding something, if something was really there. And, I'll just illustrate that.

Here, I really haven't assumed a linear effect with dose. I've just said, if you've got linearity, in terms of dose order, something like that,

Page 32

an increase with the doses. And what does that do, if you had an underlying trend like that, what would that do to your power, or how was the power to, how was power affected by having that sort of trend actually existing in the data.

And what it does is, if there is a trend, the trend test has more power to find a difference. Here, you can see in a comparison, the first column there, where it says comparison to control, if that's the analysis, a variance of a Kruskal-Wallace comparison, those tests had an effect size of around 1.5 standard deviations. You had 80 percent power to pick that up. Trend tests will pick it up at 1.1 or 1.2. So, if there is a trend, if underlying trend, the trend test will give you a little bit of power, more power to pick that up.

Now, this slide is to restore your faith in, kind of, what's going on with both types of tests. On the left hand side, there's a scenario where you only have a high dose effect. Okay, and I've chosen this maximum high here, and then the point begins 80 percent power. That's 1.4 standard deviation in this scenario. And then, I took the same maximum difference over here, but I had a different sort of trend.

The protective Kruskal-Wallace or F test

Page 33

followed by T contrast for the analysis and variance of the Wilcoxson, and when they, for the Kruskal-Wallace, it has 80 percent power at this departure in this scenario.

It has almost the same power, even in this dose order situation. It's slightly less. And that's because an F test looks for differences among treatments. It doesn't look strictly at treatments versus controls, and do an F test, and you make it a difference in treatments, because two of the dose, treatment doses are different, and not, necessarily, different but the controls.

On the other hand, if there is a trend, although the trend test doesn't pick up much over here, where there isn't a trend, if there is a trend, the power of finding a difference jumps up to 94 percent in this case. So, the point here was that, by including both non-trend tests and trend tests, you've got an increased chance of finding a difference if there was one there.

Since I mentioned protective tests, it's important to say that we used a reasonable protective test methodology. And as I've indicated, we did a, for example, the measurement endpoints when we do a analysis of variance.

Page 34

If that F test was significant, we would then do peer-wise comparison in the 5 percent level for each of those. If you do that, the, you really have a little bit higher false positive error rate using that procedure than something like a Dunnett's test, or a Bonferroni correction.

So, we've kind of used a little bit of a protective test that was conservative in the sense of increasing the false positive error rate, and by consequence, also increasing the power. So the screening method we did choose, with some added on power, to choose something that would give us a little bit more power than say Dunnett's or a Bonferroni type correction to deal with multiple comparisons.

Lastly, we were concerned about tank effects, because we'd seen tank effects in the earlier studies. And what this slide shows is that, as the within tank correlation increases, what does that do. That can have a dramatic impact in your false positive error rate. So, the tank effects, if they exist, they, it's important to incorporate them into the analysis, and the tank effects may have an effect, even though you, they're harder to detect.

Our solution for the tank means, and since there are four statisticians on the panel, I'm sure

Page 35

that they, any one of them could come up with some more complicated approaches, nested models, mixed models, to address this issue of tank effects.

And so, it's probably important for me to say, why didn't we do that. And the reason why we didn't do that is, by going to the, kind of, simpler approach of going just to the tank means, and using that as our unit of analysis, that would work for both the measurement data and the incidence data. Whereas, some of the other models, particularly the mixed and nested models, they have, they struggle when incidence rates get low.

And so, they're very, they're a problem when the incidence rate gets low. But tank mean analysis will work in both situations. Also, the tank means, because their averages tend to be a little bit more normally distributed, and we weren't forced to do one transformation for one endpoint, and a different transformation for another endpoint. So, we did consider those other approaches, but we thought that the tank means themselves made an intuitive, understandable, workable approach that applied across the board.

The other thing that's relative, perhaps, to power, a little bit more than biological veins and the

Page 36

statistical vein, is the choice of Xenopus laevis. And what this slide shows is, starting up at the top, and on the far right, these are looking at the dots, the black dots are looking at the static studies, looking at the literature, and asking, what was the smallest dose reported in the literature that gave 100 percent females. And who fell up there at the top is one where you need the largest dose to get 100 percent with females. And moving on down, that's kind of a cumulative plot, if you will, all the way over, and you can see that Xenopus laevis is down there at the end, where it requires less dose to get to 100 percent females. And those are in the static studies. So, picking Xenopus laevis was at the more responsive end of this curve, as far as 100 percent females, and that was an important consideration.

The green dot to the left of the black dot represents the study by Knotts, et al, in 1999, where they were looking for 75 percent females. And that was in a static system.

The similar number in the flow through system for the Xenopus laevis in the DCI studies was around .2. So, Xenopus laevis is at the more responsive end of that curve, as far as the species selection, and in a biological way, then, likely to have more power of

Page 37

turning up something in the doses that we've studied than some of these other species.

The last thing that I have that is the last slide, if I can figure out on how to get there. Sorry, bear with me for just a second. I prepared something that I hope would be helpful for you. It was something that I found helpful. And it's a one page overview of all of the DCI results.

It makes a lousy slide, but it makes a great reference sheet. And, I'm, I give it to you because it helps you, me to see what was going on with E2, what was going on with Atrazine, what was going on in will, the Wildlife International Study, and what was going on at IGB. And I won't belabor this. This is for your reference, but it's a cheat sheet that I found helpful, and I thought that you might as well.

And what this is made up of is, obviously, there's two columns for will, Wildlife International, and two columns for IGB, results for E2 in the column marked E2, and the results for Atrazine there.

Within a column, if there was no peer-wise or comparison that was significant, there's nothing. If, for example, under number three, the endpoints are indexed on the left, for age at completion of metamorphosis, in 42, the males were significantly

Page 38

different than the controls.
So that's why there's an M there. Similarly at E2, looking down here, gonadal image area for males was significantly different for E2. Looking down for mixed tissue types, there was no sex specificity. I mean, it was a test using all the animals. It says none, but that just means that there was no sex.
It was significant there, but there wasn't a sex to put there, so none should not be taken to mean not, as not significant, but just that it wasn't a sex specific test, so it's a no sex, but it was a significant result. Under Atrazine, for gonadal image area, the males were significant at dose of 100, and that's it. So, it's looking at all five, and it's saying only 100, is saying that there was not significant at 25 one, .1 was a .01. So, wherever it's significant, it shows on this table.
The one page I gave you allows you to look at gross on the top half of the page. Histological at the bottom half, the left side is the non-trend test comparisons. The right side of the page is trend tests. So, I'm not going to make any comments.
I just thought you might find that a handy sheet. These results are consistent with what's in the EPA report. They're consistent with what's in the DCI

Page 39

study report. There were a couple of cases where EPA used a one-sided test. I have marked those in parenthesis where those differences occur.
And they, for example, down there under gonadal segmental translucence, it shows a two-sided test is what the DCI study was, when, and EPA F in parenthesis would mean that EPA found female to be significant at one side of the test. So, this is just a handy page that you're free to use or discard at your pleasure, but it helped me, and so I give it to you, okay.
The other thing that relates to power is the sticky wicket that came up yesterday about the cluster effect. And Mr. Chairman, since that does relate to the power discussion, I have just a few words - -
DR. HEERINGA: I'll allow it, yes.
DR. SIELKEN: Thank you.
DR. HEERINGA: Dr. Sielken, please proceed.
DR. SIELKEN: Okay. As I've indicated, we did use the tank as the unit of analysis. And I'm just going to say it that, as a unit of analysis. We did, when we were considering the design of this study and the layout of this study, we were aware of the layout of the room that was going to be necessitated by

Page 40

the pumping system that led to what was identified as clusters and has been discussed as clusters.
So, we were aware of that element of the design, and we discussed with the principle investigators, well, why don't you just simplify this whole thing and put one pump for every tank, and spread all the tanks around the room. Well, we got a lot of feedback on that, and, of course, as a statistician, I thought that that was the simplest solution. But, then, we listened to the discussion about why not do that.
Well, first of all, they said, well, you know, if we put all those pumps in the room, we're going to have additional heat. We're going to have to figure out a way to deal with that. We're already rebuilding the rooms to accommodate this study, but those extra pumps are going to be a problem. And, they said, well, you've got to look at the overall error rate in the study.
Since everything had to be blinded, instead of having sixteen things to control, you'd have sixty-four things to handle all the equipment, and handle all of the preparation of what goes in each. And they, basically, convinced us that running a lot higher error rate for all of these other things that could go wrong,

Page 41

they didn't want to do the study blinded, but for scientific reasons, it had to be blinded.
But the blinding and all of those extra tanks was going to create a problem and, probably, increase our error rate. So that was one of the reasons that they argued to try and break up the number of tanks into, I mean to keep a certain amount of clustering, just from a practical point of view.
We did try and make the number of tanks per cluster as small as workable. There was, at one point, a discussion of having them all being fed by the same thing, having eight tanks instead of four in cluster. The difficulty with that was a practical consideration that, that was difficult, that if we lost something there, we lost everything.
I mean, if we, if, as it turns out in this study, the things that they were worried about, we had seen in the earlier studies of Karr, et al, back in those days, that the tanks themselves, there was a lot of problem with the tanks, and you could lose tanks. And we, you know, and here we had a microbial bloom, and that caused us to lose one cluster, and did not cause us to lose everything.
Similarly, by having the clusters, we had a little bit better control over cross-contamination,

Page 42

where, as the sixty-four tanks was increased and the blinding would have been sixty-four clusters, if you will, the breaking it up into one tank per pump, would have given a greater chance for cross-contamination, and we were trying to avoid that.

There was also some discussion, and I'll let you turn to others if you want for whether this was consistent with the aquatic tox guidelines and using the tank as a replica, it certainly was. We did, of course, randomize the tadpoles to tanks. They did come in and were randomized.

The tadpoles were randomly assigned to the tanks. Since we were aware that we did have these clusters, of course, that creates the possibility of some dependence, so we did tests for cluster effects. We, the tests were not greatly powerful, but we did test for them. And, we tested all endpoints, so when you're looking for cluster effects, you know, we didn't just look at one. We looked at every endpoint tested.

As I mentioned yesterday, we had one out of 176 of those tests that is significant at the 5 percent level. And I really looked, not just at that, but this 5 percent is just a bright line in the sand, this would be sort of meaningless.

If you look at the bulk of the test, were

Page 43

there any others that were close to 5 percent. The next closest was 8 percent. There's one of those. There's one in ten. And then there was 18 and up, it's only three of them out of 176, were under 10 percent. And most of them were 18 percent and above, so, a P values of .18 or above. So, there really wasn't much close to being significant.

So, what we concluded was, was that, although the tanks are not technically independent, they were a reasonable approximation to independence. And we treated them as the iterative analysis, with that understanding that where they were as close as we could get to independence of the situation and have a viable study.

We did notice that we did get largely comparable and parallel results with the E2 positive controls, with this method of analysis, in treating the tanks as the unit of analysis. We also asked ourselves, well, what would be the effect of a cluster, the fact that it was there.

And, as I mentioned yesterday, the effect would be, with the cluster effect, you would be, kind of, overestimating the number of real tanks that you had. So, you would be pretending that you had a bigger sample size than you actually would have had. And that

Page 44

would increase the false positive error rate. And of course, when you increase the chance of rejecting a null hypothesis, that's true, that where there was no differences, you're actually increasing the power, the likelihood of detecting something as different. So, we felt that it was a conservative approach in that, if it did occur, we were only going to be increasing the false positive error rate. We weren't going to be hiding anything.

It was a robust design. We thought it was a robust design and gave us a certain amount of protection against the things that the experimenters themselves were worried about.

Turns out, it was a darn good thing we did, because they did have some of the problems that were expected. And because it was a well designed study, they had the opportunity to actually detect these problems and survive. I mean, they detected the plume. They detected the trace contamination. They were able to detect those things and still have a viable, strong experiment when they were done. So, that's it, thank you.

DR. HEERINGA: Thank you very much, Dr. Sielken. What I would like to do is to give opportunity for a few questions of clarification.

Page 45

Remember, we're going to discuss the statistical analysis for your charge question 9B, but if there's anything about the information that Dr. Sielken has just presented, if there are any questions on that, now, but again, let's try to keep our comments or, kind of, dialogue on this for the actual response. Yes, Dr. Yeater.

DR. YEATER: Kathy Yeater. Could you clarify on the paralyzed comparisons, 'cause that was kind of confusing yesterday. Did you use any adjustment correction or are these strictly all paralyzed comparisons. And then, you looked, individually, at the T tests then for your comparison and estrodyle to the control and the Atrazine to the control, or did you use a Dunnett's or a Bonferroni or any, or any type of that thing?

DR. SIELKEN: Okay, thank you for the question. This is Dr. Sielken. We did, for the measurement endpoints, do a protective F test. For the incidence endpoints, there was a protective Kruskal-Wallace test.

If those tests indicated at the 5 percent level if it was two-sided, or 10 percent if it was one-sided, if they indicated a significant difference in looking at the group of treatments or the group of E2

Page 46

versus control, that second group.
If, at a group-wise level, there was a significance somewhere, it rejected the null hypothesis of equality, then we followed with a peer-wise T contrast, or a peer-wise Wilcoxson in with the test, at the 5 percent level, there was not any protection beyond the initial test. So, everything beyond that initial test was a 5 percent significance for that one test.
DR. HEERINGA: Dr. Bailey?
DR. BAILEY: You said that you did examine for cluster effects, and could you tell me how you made that test, what, how you actually carried that out. Was it a T test or an F test or, and what did you use for an error turn?
DR. SIELKEN: Okay, good question. This is Dr. Sielken in response. When we did that test for treatments, there would be two clusters. And we would look at the, do it, do a comparison of those two clusters at a particular treatment level, control or treatment. We'd look at those two and ask the question, whether those two were the same or different.
That's only a sample of size two. We did that for every dose, got a significance level, combined those significance level producing Tippett's minimum P-

Page 47

value as a method of combining those, and then evaluated whether that Tippett's minimum P-value was less than 5 percent or not. So, it was a, looking at the pair of clusters, followed by, then, looking across all the treatments and controls, and then combining those P values, using Tippett's minimum P-value method.
DR. HEERINGA: Dr. Sielken and Dr. Bailey, my experience in these panels on an item like this is, I would encourage you, maybe, just during the break, to sit down with a pad and paper, and so that you're very clear as to exactly how this was done. Don't you agree? Or are you comfortable with it. If you would be willing to do that, I, and we'll come back, and we'll report specifically on that, but I think, I don't want to, avoid any misunderstanding later on on that, so. It's a critical question. I know you've raised it earlier yesterday, too, so. Dr. Portier?
DR. PORTIER: Bob, thank you for the presentation. It was really good and answered a lot of questions. Kind of, the only, kind of, remaining global question would be combining the data from the two studies into one overall analysis, and looking at the data that we could see, we don't see a lot of differences in variability.

Page 48

A lot of effort was put into making the two sides as comparable as possible with similar protocols, similar sources of the frog larvae. So, I wondered if you, actually, went off on the side and did that analysis, like a one way anova for all of those things, both sides, kind of, combined together?
DR. SIELKEN: This is Dr. Sielken in response. That's a good question, Dr. Portier. Of course, it is possible to create an analysis that would combine the two laboratories. We did not do that. We discussed.
We discussed some of the complexities involved in doing that. And we, also, discussed what could be gained by doing that. And the feeling was that what we wanted to strive for was two comparable studies. And then, as you've indicated, they are quite comparable, and they were designed to be that way. They were designed more to be reproducible results. And intended to be portrayed as reproducible results, rather than, rather than combining them.
We tried to design the studies to be powerful enough, each on their own, to stand alone, so that the usual argument for combining, namely to increase power, would not be as relevant, given that they were strong enough on their own. And, so, that was the intent of

Page 49

leaving them as separate reproducible studies, and not combining them, but we did not do that analysis.
DR. HEERINGA: Additional questions for Dr. Sielken? Well, I would like to thank you very much for this presentation and clarification. I think there are, have helped considerably. And I think, if Dr. Bailey and Dr. Sielken would like to meet briefly, just to discuss the test for the cluster effect, so that the panel has a clear understanding, and we'll rely on Dr. Bailey for that, and it's my best position to do that.
DR. SIELKEN: Thank you. I will have that sidebar discussion.
DR. HEERINGA: At this point, I'd like to call back to the table, the representatives of the EPA's Scientific Assessment team. And we'll move on to any final questions for them from the panel. And, then turn to the, and actually, we'll probably take a short break and go on.
But, at this point in time, are there any remaining questions of clarification or points from the panel that you would like to address to Dr. Steeger before we move the charge questions? Obviously, if something comes up during the charge questions, we will certainly permit points of clarification there, too, but I'd prefer to get most of that out of the way at

Page 50

this point, if there are no objections. Yes, Dr. Delorme?

DR. DELORME: Dr. Steeger, I was just wondering if you could clarify something for me. When you look at field studies, I think you're talking more about experimental field studies, where somebody's going out and trying to do the experiment in the field. Given that, would you use the result of these field studies, if they are done appropriately, to determine in effects endpoint for use in risk assessment?

DR. STEEGER: We would use well- defined, well-described, properly conducted field studies in our assessments. But it is necessary when we're able to determine the relationship between the measurement endpoint and the, or the Agency's assessment endpoints. Our assessment endpoints are for acute studies, that would be mortality; and for chronic studies, whether reproduction, survival, or growth are impaired.

One of the problems that's been present with the current battery of studies that we've had on the effects of, potential effects of Atrazine on gonadal development has been trying to link this phenomena of intersex or hermaphroditism to our assessment endpoints of reproduction and growth and survival.

The, none of the studies that we're aware of,

Page 51

although, intuitively, you might think that hermaphroditism could be problematic for animals, it would impair their reproductive capacities, we don't have that many studies that, actually, demonstrate that.

And, in fact, in many of the studies where field collections were made, apparently, the researchers did not have a problem collecting animals where hermaphroditic animals were relatively common. So, it's unclear how hermaphroditism would impact the reproductive capacity of the animal.

DR. DELORME: And just another question, in some of the field studies that you reviewed, you indicate that the Atrazine concentrations weren't properly characterized to allow an analysis. Can you just share with us a little bit what would be a properly characterized in a field study, what do you consider properly characterized determinations of whatever the stressor is?

DR. STEEGER: Well, ideally, we'd like to know what the concentration of the chemical in question is at each of the sites. And where you have reference sites, ideally, we'd like to see that the chemical is not present at the same concentration that it is at the treatment sites.

Page 52

It would also be important that you demonstrate that other pesticides or chemicals that are capable of impacting the measurement endpoint are characterized, and ideally not present at your reference sites, so that, if you're going to make a, draw a conclusion that the levels of a particular chemical are correlated with a particular measurement endpoint, there has to be data from all the sites that you've collected those levels from, to demonstrate that from where you've had a zero concentration, to where you have high concentrations, you're not having overlap in the measurement endpoints themselves.

DR. DELORME: And what about temporal aspects?

DR. STEEGER: Well, temporal aspects is studies that was problematic for a lot of the field studies that we reviewed from the open literature and registrants submitted datas, where field collections were made of animals in different, that would clearly have been in different stages of their reproductive cycles. And the animals were combined and an effort was made to draw conclusions regarding plasma steroid levels, the condition of the gonads, when in fact, it would have been very difficult to say that all the animals were at the same stage of development.

Page 53

Because, clearly, they were collected at different periods, and they were of different sizes, different ages, and so their reproductive capacity would have been different, just by age structure alone.

DR. DELORME: Okay, thank you.

DR. HEERINGA: Dr. Miller.

DR. MILLER: Debra Miller, EGA. I'm not sure if this is an appropriate time to ask this question, because it's probably really to Dr. Wall. I know that some extra histopath was examined. And I was just curious if things like, specifically, the brain and the adrenal glands, were they looked at?

DR. STEEGER: To our knowledge, the only, the, during the gross morphology stage of the study, a number of organs were looked at. The histopathology only focused, to my knowledge, on the gonad and the kidney was sectioned.

DR. HEERINGA: Dr. Denver.

DR. DENVER: Bob Denver. I'm curious if the EPA is aware of any studies that have been done, either in the public literature submitted by the registrant, on Atrazine metabolites or degradates on amphibian development?

DR. STEEGER: I'm not aware of any studies that have been conducted on the three major

Page 54

degradates, relative to gonadal development.
DR. FURLOW: Have any analysis, analyses been done of the accumulation of such degradates or metabolites on either static renewal or flow-thru systems of significant or significant to you?
DR. STEEGER: There is one study that was discussed yesterday that Dr. Solomon was presenting on where they did autoradiography suggested that metabolites were not accumulating to any significant effect to any significant extent, at least based on radio labeling, in animals in a bioconcentration study.
We have not reviewed that study to great extent, though. So, in answer to your question, no, I'm not aware of, to the extent to which those degradates are accumulating in static or flow-thru studies looking at amphibians.
DR. HEERINGA: Dr. Skelly.
DR. SKELLY: David Skelly. Are there, it's my, well, let me just you this question. Are there any new field studies that EPA has reviewed since the 2003 meeting, and if there are not, is the EPA's assessment of field study evidence different than it was in 2003?
DR. STEEGER: All the field studies at EPA reviewed since the 2003 SAP represented the final

Page 55

versions of internal reports that have been reviewed in 2003. So, there have been no new studies submitted for our review. Our conclusions have not changed regarding those studies.
DR. HEERINGA: Dr. Petino.
DR. PETINO: Reynoldo Petino. We had a brief discussion yesterday about this, the effect for solution, I guess, I was told that's the theme, the term to use of, when it comes to some of the categorical variables that were measured, especially things that was the presence or absence, and that given that the analysis was done bisex, that the sample size, there was no variable between ten and fifteen individuals for those analysis.
And, therefore, the effect for solution can be as low as 10 percent. And, so, I'm just curious, not being a statistician, what is, is there an acceptable level or a guideline for, that the EPA has to, you know, how, what, you know, about the effect for solution, how low can it be before there's a problem with the analysis?
DR. STEEGER: None of the studies that we were reviewing, or guideline studies, that's one of the difficulties with looking at these endpoints. However, our guideline of studies in effect is considered a

Page 56

significant at the .05 level.
So, that's, regardless of the frequency, if it's statistically significant, that's where we would flag it as meaningful. But, the risk assessor is then expected to try and determine whether that statistically significant effect is biologically significant.
And that's what we're hoping to have input from this panel on, as to whether many of the histological endpoints that are reported in the studies, many of which were statistically significant, they're occurring in very low frequencies, are they biologically significant. And even those that were not statistically significant, was it just because, as you say, the frequency at which the animals were, either male or female, in the tank, were such that it would have been difficult to detect a significant effect. I'm not a histologist.
These are new endpoints to me. It's been very difficult to try and link them to what, do these represent some primordial stage that an animal, it might be in, as it's moving towards becoming an hermaphrodite or a mixed sex animal. I don't know, but that's why we're asking the panel. What's your opinion as to whether these low frequency events, whether

Page 57

they're significant or not, should be meaningful in our assessment of these data.
DR. HEERINGA: Good questions, both statistical and biological. Unfortunately, it's hard to separate out. Any other questions at this point from the panel? Seeing none at this stage, what I'd like to do is, I'd like to thank everybody, the EPA scientific staff that's prepared this white paper, all the commenters and the representatives from Syngenta who presented their data yesterday for your contributions to the initial information session of this panel. I'd like to call a break at this point for, say, twenty minutes. We'll reconvene at 10:15, and at that point in time, we'll turn to the charge questions that have been posed to the panel.
(WHEREUPON, there was a break.)
DR. HEERINGA: I'd like to welcome everyone back to the late morning session of second day of our FIFRA Scientific Advisory Panel meeting on the topic of the potential for Atrazine to affect amphibian gonadal development. At this point, we have had the presentations from the EPA scientific staff, and the period of public comment. And, I believe, that we are ready to proceed with the first of the charge questions to the panel. And I think I'll leave it to Dr.

Page 58

Steeger. I believe that my near-sightedness, I'll have to - -

DR. IRENE: This is Stephanie Irene.

DR. HEERINGA: Yeah, Stephanie Irene is going to be reading - -

DR. IRENE: Stephanie Irene with EPA.

DR. HEERINGA: Dr. Irene is going to be reading the charge questions into the record. And also, Dr. Pease, I believe. And we will respond in turn. We will have a lead discussant who will begin the initial response of the panel.

Followed by associate discussants that have been identified. And then we will open it up to the panel at large for their comments. After we've successfully completed each response, we'll move on to the next charge question, obviously. So, Dr. Irene.

DR. IRENE: Thank you. The Agency now requests that the panel discuss the first charge question, which is, in reviewing the available laboratory and field studies, the Agency used a number of criteria to evaluate individual investigations. Criteria such as experimental design, test protocols, and quality assurance information were used to evaluate the reliability of the study's ability to adequately assess the hypothesis, that Atrazine elicits

Page 59

developmental effects in amphibians, and if so, the nature and strength of the associated dose-response relationships.

A. Please provide comments and recommendations regarding the EPA's approach, and criteria used to evaluate the studies; and

B. Given the evaluation criteria employed by the Agency, please comment on the EPA's overall application of these criteria for the currently available studies.

DR. HEERINGA: Dr. Green is the lead discussant, but since she was delayed on her flight yesterday, I asked Bruce Pauly to step in, but I have not determined quite where we came down on that.

DR. PAULY: And I've passed it on to Dr. Schlenk to it.

DR. HEERINGA: That's appropriate committee work. Everybody will have their say, but let's begin with Dr. Schlenk.

DR. SCHLENK: Yeah, Dave Schlenk, UCI. Just to fill everybody else in, Dr. Pauly and I met this morning, and we met Sherril this morning as well. So, hopefully, we'll give it a first crack here, and then, please feel free to add in as needed. So, I'm just going to, what I did is, I sort of combined A and

Page 60

B together as addressing both of them, sort of intermittently.

So, with regard to experimental design, EPA noted several issues that prevented their utilization of previous reports in open literature regarding the risk assessment of Atrazine on gonadal development in amphibians.

To reduce uncertainties, the 2003 SAP recommended that a tiered approach be utilized to determine causality between exposure of Atrazine and adverse affects on the model amphibian. Standard aquatic toxicology methods, with endpoints associated with apical effects were recommended.

In most cases, the experimental design implementing these recommendations and the criteria were sound. A tiered approach is a logical step in determining causality. However, it should be noted that laboratory studies may not always exclude causality in field effects. And, I think the 2003 panel addressed that in one of their papers.

Ten being a great example of where you see 100 percent of the animals showing endocrine destruction in the field, but very difficult to replicate that in a lab. A lot of metalloid types of chemistry, selenium, it's very difficult to replicate

Page 61

those studies in a lab. But, of course given the situation here, and due to the lack of a proved field of studies and cost considerations, the laboratory approach seemed like a sound alternative in this particular case, but it shouldn't rule out completely field studies, I guess.

As previous studies focused exclusively on static exposure systems, often in containers of questionable and confounding material constructs, the use of a flow-through system in class is valid, especially since it appears that the life history organism does not seem to be impaired under these conditions.

One of the things I thought was very good about this study was the replication of the experiment in two concurrent laboratories. I thought was an excellent approach. And, actually, in contrast to what Dr. Portier mentioned about combining the two, I think the strength is actually in the separation of the two, as far as the different effects. And I'll give an example of that here in a minute.

The test organism with a well-characterized genetics and life history was utilized. This allowed efficient exposure of the toxicant during a previously determined sensitive window of exposure, which again

Page 62

maximized efficiency in the laboratory test.
Since adverse effects were previously observed at the .1 and 25 microgram per liter concentrations, exposure concentrations bracketing those were utilized in this experiment, which was a sound approach, I think.
While, however, while one laboratory was successful in conducting exposures at the critical window of sexual differentiation, bracketing the 0.1 microgram per liter, which again was identified in these previous studies, the other laboratory was less successful in reaching this nominal concentration. And we'll deal more with that when we get to question four.
So, as an example of the benefit of this inter laboratory comparison, it was, basically, you could see that the effects for particular, the gonadal hypoplasia that was observed in the males of the .1 concentration, which was observed in the WLI, but not the IGB, were determined, were measured concentrations for consistently below the .1, because you only saw 50 percent of the nominal, one wonders whether this prevented replicated response in the IGB study. However, this effect was not dose dependent at WLI, where the exposures at the .01 microgram per liter were adequate in the sensitive window of development, and

Page 63

the effects were not considered significant. In addition, this effect was not confirmed by other measures, such as gonad area or histological abnormalities.
Thus, utilizing multiple endpoints to confirm adverse effects was a strength. And clearly, the quality assurance protocols that were implemented allowed the observation and measured concentrations during critical windows of development were successful, at least in one laboratory.
Talking about the positive control, having a positive control is an essential component of the DCI study, and estrodyle is an effective positive control for feminization, and the responses of estrodyle on this organism have been well characterized, which is, again, another strength of using that particular organism.
Given that the tested hypothesis was based upon up regulation of aromatase, even though, at least to my knowledge, increases in estrodyle have never been observed, which would lead to an increase in estrodyle during critical windows of development. That's the hypothesis. However, given the knowledge base for this compound and this species, it was probably the most cost effective choice for a positive control.

Page 64

So, overall, the evaluation criteria seemed to be extremely focused on gonadal development, which was a plus, but I think, may have also been a minus, and maybe some of the other panel members can chime in on that. And the studies did not focus upon other endpoints, which may also be important with regard to the toxicity of Atrazine. However, with the exceptions of the caveats mentioned above, EPA appropriately applied their criteria in carrying out these particular DCI studies. Bruce.
MR. PAULY: Bruce Pauly, Environment Canada. Yeah, I'm just picking on one of the final points that Dr. Schlenk made. One of the things we discussed this morning was the fact that the charge question off the bat reads that the studies were evaluated to determine whether or not Atrazine elicits developmental effects in amphibians. And, I think we all realize that it's more narrow than that.
The open literature, which was reviewed for the white paper, included studies which focused on Atrazine alone and on gonadal developmental effects alone. So, the fact that we're only asked to, or the review only included studies on Atrazine alone and gonadal developmental effects alone, narrows that charge question, in my opinion, to a certain extent.

Page 65

So, I think the charge question is the one that was used in 2003. I think this year's charge question is na-, it should be narrowed. Because we are only considering gonadal effects, not developmental effects.
That said, and as Dr. Schlenk mentioned, the studies were evaluated to determine the effects on gonadal development. Every study was looked at for experimental design protocols to update equality and then, eventually, if there was, they had to go further for dose-response relationships and ecological relevancy. I think we would probably agree with the 2003 panel response that the reviews were thorough. The approaches and criteria were appropriate.
What we discussed to a certain extent, too, was that when these reviews were done on all of the open literature studies, at least, these criteria that were established were very strictly applied. And, for instance, where a recommendation was made for a particular test method or protocol, if any individual study did not employ this design element, those were grounds, that was sufficient enough for the study to be removed from further consideration. And one of the things that we specifically discussed there was flow-through methods.
All of the studies, except for the DCI

Page 66

studies, were static renewal, so they were immediately removed from future consideration. So, the open literature studies were reduced quite substa-, well, completely, basically, by those considerations on their design elements. There were other ones that we, I've already heard about, other considerations, in terms of the open literature studies, which effectively removed them from consideration. High loading was one.

Few Atrazine concentrations, so a dose response wasn't possible to determine. And poor responses to positive control was another criteria for evaluating the studies. High mortality, and I think the final one that I had written was inadequate study design to overcome high variability in the endpoints, was another reason for taking a lab study out of consideration.

And we already heard, again, from Dr. Steeger's presentation, yesterday, presentation three, there are studies, or criticisms of the field studies as well, which removed all of them from future consideration.

So, what happened there was that, applying, strictly applying the criteria, and I think we're going to talk about this in the next question as well, strictly applying those criteria resulted in an

Page 67

effective study of one, a study that we're, basically, considering in this process. And, while we say that the evaluation criteria are appropriate and were appropriately applied, I think we'd like to recognize that, as a result of that, we're, basically, considering only one study for this process. Thank you.

DR. HEERINGA: Thank you Mr. Pauly. Dr. Green, who we recognized was delayed due to airline problems, but what's your response at this point in time?

DR. GREEN: I concur with the comments made by the other discussants so far, and I'd like to bring to the attention of the panel and the people in the room today that one of the criteria listed on the to-do list from the previous session in 2003 was that studies be performed such that bio-residues of Atrazine in the specimen studied are looked at.

Tissue levels, plasma levels, if possible, to document that these animals did, indeed, absorb and had the substance in their body at the time that the gonadal abnormalities were observed histologically. There was a dearth of studies submitted where that kind of dose response testing and toxicological analysis was performed. I believe there was just one that reported

Page 68

a half-life that was very short for Atrazine in the animals that were under study.

So, in addition, grow out studies were recommended in the past, I believe, in 2003 by the panel. And there were very few studies that reported that.

So, the interpretation of that criteria would be difficult, and despite the fact that some of these studies have shown no effect, or an effect, of Atrazine on gonadal development, we cannot, at this time, use a criteria that allows us to report on functionality as a result of Atrazine exposure. In other words, would these animals go on and lay eggs that can be fertilized, and then produce more healthy animals. So, other than that, I have no further comments.

DR. HEERINGA: Comments from other members of the panel on this first charge question. Yes, Dr. Delorme.

DR. DELORME: Just following along on what doctor, or Bruce Pauly said, and what Dr. Schlenk said, with respect to the field studies, I think it might be helpful for EPA to develop some general guidance on what's expected in field studies to help the people who are doing them. And, essentially, to help the overall success rate of field studies.

Page 69

I think we've seen from this that there hasn't been much success. A lot of resources expended, but I'm not sure that they resulted in anything. And to improve the utility of the studies. So, I think some general guidance from EPA to, whether it's registrants or universities researchers, on what's expected and what's needed, what the criteria are for evaluation might help. And that would be a recommendation.

DR. HEERINGA: Dr. Chambers.

DR. CHAMBERS: Just to respond to an earlier comment, I'm really not too bothered by the fact that there weren't any residues in the animals. If they were exposed in the water column, such as an animal in the environment would be, as they were in that experiment, then, you know, they were subject to the exposure. And I think, I think it's reasonable not to have that data set as a real concern.

DR. HEERINGA: Dr. Green.

DR. GREEN: If I can make a comment to that.

DR. HEERINGA: Sure, absolutely.

DR. GREEN: I think, in certain studies where the exposure was chronic, and we do or don't see effects, those are well-designed studies. But if there

Page 70

was a brief exposure for a short duration at a short time point in an animal's life, that may or may not have an effect.

If the half-life truly is as short as it is, that wouldn't allow for accumulation or chronic exposure over the time, and for us to see the results, which I think would probably be more likely to parallel what you would see in a field study, where animals are growing up in a pond, and spend their whole life exposed through periods of rain and runoff, and repeated applications of the chemical.

DR. HEERINGA: Dr. Chambers.

DR. CHAMBERS: Jan Chambers. The, this was continuously renewed in the tanks, though, so they were continuously exposed and, really, probably, more continuously exposed than they would have in rain, runoff events, I would think. And this is the critical time for the gonad development, which is the subject of this particular question, this particular study.

So, again, I'm really not concerned that the analytical chemistry on the animals was not performed. The analytical chemistry on the water column was performed, and the compound was there at, pretty close to the nominal amount, so I'm not bothered by the lack of analytical chemistry data in the animal.

Page 71

DR. HEERINGA: Other comments from the panel. Yes, Dr. Furlow.

DR. FURLOW: David Furlow. So, one thing that came out of discussions with Dr. Steeger was, I think I said I was disheartened by the fact that he found the investigators in the academic lab or his laboratory said that those discussions either that they were non-responsive, or it was unproductive. And, I guess, I'd like to go on record as saying that I wish that could be improved in some way.

If we could come out, come up with some means for the EPA or other regulatory agencies to improve their interactions with academic scientists. I know that in my world, not in terms of regulatory agencies, but, say, if there's some sort of database or repository, I know that in microwave data, there's a, you have to submit microwave data to, the raw data, say, okay, here it is.

Here's our evaluation of the microwave data, but the raw data is available for everybody to look at and make their own conclusions. And here are the criteria that we use to do the experiment. And, maybe some of my toxicology colleagues can comment on that, so if there is some sort of, maybe I'm not aware of it. Some sort of open toxicology database for quality

Page 72

assurance that studies can be submitted to, that is in addition to the peer review, where it's only one or two peer reviewers, right, the times with the journal.

So, I think it can be a strong, that it should be a strong level of control at the peer review level, but sometimes it is not. And I wonder if there is some sort of database or if that's something we can discuss to help the interaction between the EPA and academic institutions.

I guess, just the one last comment I had about that is that, if we are only restricted to considering the studies that are in DCI, really, when it comes down to it. If that's what we're left to really consider as what is acceptable to the EPA, in my heart of hearts, I don't see any conflict of interest. I haven't seen any evidence of that at all.

But the close interaction between the EPA and those studies might have the appearance of conflict of interest. And I don't that's something that we want to have, even the appearance of it. And, so, if there was some closer contact with, and trust and interaction between academic scientists and the EPA outside of the industry to study, from the studies, would be, I think, beneficial to all of us, just as a general comment.

DR. HEERINGA: Yes, Dr. Steeger, please,

Page 73

I think it's important to respond.

DR. STEEGER: Yeah, I'd like to make a couple comments. In the 2003 SAP, prior to it, we did make an extraordinary effort to contact the researchers that were generating some of the data that were showing adverse effects. We requested, we don't have the authority to require researchers to provide data, nor do we have the authority to tell them how to conduct their studies.

But, we were, we, actually, went to the lab, and to the study site, and we requested, on three occasions, to have access to those data. There were four staff scientists that attempted to review the data, and the researchers are provided the data in the way that they, typically, would record data. It's not the way that we would require data to be presented to us.

So, we're at something of a disadvantage from the get-go, but, at no time, were the data supporting the researchers conclusions. Each time we received a dat-, a different data set, each time the data could not be related to the study or its conclusions. By the third time, the Agency had determined that it had spent enough resources attempting to make sense of this, and we ended up evaluating the study at face value.

Page 74

The other researcher with whom we communicated with had noted very dramatic changes in histology after very short exper-, exposure periods, and eventually, some of the more provocative results that were recorded, in follow up discussions with that researcher, it turned out that they had been artifacts of histology.

So, again, rather than get into contradictions between EPA's analysis of an open literature study versus the author's interpretation of it, we just let it go and said that we would just evaluate the studies at face value.

The, so we do make an effort, we did make an effort to evaluate, or to work with the study researchers, but it did not prove to be productive. Again, we have no authority to tell people how to do their studies.

And with the Syngenta studies, as we indicated in our presentations, we had extensive Q-A going on throughout the study. EPA wanted to assure that the protocols were properly developed, that they were consistent with what EPA had required, and that the SAP had recommended in 2003.

I have never participated in a st-, in a Q-A process that has been more extensive than the one

Page 75

that's been applied to these particular, these two particular studies. It is unusual for the agency to have access to a study as it's actually being conducted to know exactly what is going on at any moment of the study.

When the Atrazine contamination took place, we knew about it within a day. And, so, it's our very clear understanding of what was expected, what actually occurred, and the results of those studies, that we've put such an incredible amount of weight on what appears at face value to be a very small number of studies, relative to what is available in the open literature.

And to follow up on a previous commenters statements, the criteria that were applied, a single criteria was not implied to discount the utility of the open literature. We attempted to, did they follow the recommendations that were made by the SAP, that they did or they did not.

That would not have discounted them from further considerations. It was the concordance of information from each of the studies that would have allowed us to either use it or not use it in determining whether Atrazine was having an effect on gonadal development. And most of it had to do with, you were, the, you were either not seeing dose res-,

Page 76

not, in addition to not just seeing, following the recommendations that were made, it was a matter of not getting a dose response, that the response has changed, and were inconsistent. So, it was a concordance of information, not a single factor that deemed the acceptability of the studies.

DR. HEERINGA: Dr. Furlow, please.

DR. FURLOW: So, thank you, Dr. Steeger. I just wanted to clarify that, 'cause earlier you said other, I think I had it down as non-responsive versus unproductive, and I just wanted to clarify what the interaction was, for the record, with the academic researchers, so we do, more about the nature of that interaction, so I think what you said was helpful to me in that regard. So, I just want to improve the process, too. I mean, that's something we can discuss later.

DR. HEERINGA: Thank you, Dr. Furlow. I think that's a very important point. Steve Heeringa, just to make a general comment that I think the study that, the studies that are in the general literature review and the studies that are being reviewed here are a little bit of an exception in the normal process that I have observed, in terms of the involvement of academic research, and the collection of peer review

Page 77

literature, it's input to processing.

I think a lot of that was, as a member of the 2003 SAP, and I'll let others contribute here, it was very much a result of the deliberations in that meeting where we systematically went through a review of the scientific quality and the weight of evidence from all of these studies, and concluded that what was really needed was a consistent, certainly to begin with, a consistent and well-conducted with the criteria laid out, laboratory-based study.

Not to the exclusion of field studies or other studies to supplement that, but we really needed to get back to a fundamentally direct and accurate laboratory-based study before the decision making on this could proceed on it.

That's my personal view, and other people from 2003 SAP could, but generally, the process, I think, that, that has gone through with the SAP, and with the EPA sciences, is broadly inclusive at academic based research. And so the criteria here, that sort of sweep away a lot of the previous researches. It's a little bit atypical, but it was in large part reviewed and endorsed by the 2003 SAP. Yes, Bruce Pauly.

MR. PAULY: Bruce Pauly. Just, real

Page 78

quick, clarification, if you had a study that had good water quality, low mortality, low loading, and if it was static renewal, a static renewal studies are, would be acceptable?

DR. STEEGER: Static renewal studies would be acceptable, provided they measured the concentration of the test material, and that the water quality parameters were well reported.

DR. HEERINGA: Dr. Delorme.

DR. DELORME: Just following up on what Dr. Heeringa said. I think it's important to note that the criteria that the EPA used are based in science.

I think another thing that's important to note is that studies on amphibians are not the norm in pesticide regulations. And, you know, the protocol work that was done in developing this, I think, is going to help us in the long run, not only for Atrazine, but for other compounds in the future.

So, when they're looking at, you want porta-, pseudo-protocols during the study for the first time. Maybe they're going to be a little bit harder on it, but you know, I think that the criteria that they've used is appropriate.

DR. HEERINGA: Comments from other panel members. Turn then to Dr. Steeger to see whether you

Page 79

have any, require any additional clarification. Do you feel the panel has addressed these two points of question number one?

DR. STEEGER: I agree that the panel addressed the two questions, and I, also, agree that the scope of the question is intended to focus on amphibian gonadal development.

DR. HEERINGA: Thank you for putting that on the record, too. Implied, certainly, and that correction. Okay, at this point in time, I'd like to turn to either Dr. Irene and Ms. Pease to read the second question, the tag team this year, so Doctor or Ms. Pease.

MS. PEASE: Yes, Anita Pease here. The second question to the panel deals with questions concerning open literature studies.

The Agency has concluded that the information contained in the open literature published in 2003 SAP does not provide any additional information that could be used to refute or confirm the hypothesis that exposure to Atrazine alone causes adverse developmental effects in the amphibian gonads. This is a three part question.

A. Please comment on the comprehensiveness of the Agency's literature reviews relative to the

Page 80

potential effects of Atrazine alone on amphibian gonadal development.

B. Please comment on the Agency's evaluation of the open literature studies, and the Agency's conclusion that the data derived from laboratory studies, both individually and collectively, are not sufficient to refute or confirm the hypothesis that Atrazine exposure causes developmental effects in amphibian gonads.

And C. The Agency concluded that that the field studies are not adequate for assessing the hypothesis at hand. Please comment on the Agency's conclusion. If the SAP concludes one or more of the field studies do provide the means to assess the hypothesis, the Atrazine exposure results in effects on amphibian gonadal developmental. Please suggest interpretive and statistical methods that should be employed to evaluate the data.

DR. HEERINGA: An important, Dr. Skelly will be our lead discussant. I'll leave it up to you, David, to, whether to do three parts together or individually.

DR. SKELLY: Okay, this is David Skelly. I think what I'll do is I'll run through A, B, and C; and then ask my co-discussants to add in their

Page 81

comments.

On the first question on the comprehensiveness of the Agency's literature reviews, I am not aware and the other people that I've spoken to are not aware of additional studies that have examined the potential effects of Atrazine alone on amphibian gonadal development.

And, in fact, if you examine the review that EPA has done, they consider studies that went beyond this specific charge as well. So, in that sense, the review was thorough. There are certainly other studies that have been conducted on Atrazine and amphibians since then, but not, specifically, on Atrazine alone or amphibian gonadal development.

For part B. to comment on the Agency's evaluation of open literature studies derived from a laboratory, in general, I agree with the review. There are certainly many concerns with the open literature laboratory studies that have already been gone over.

We don't need to go over them again in detail. I will just reiterate one point, so, many of these studies involve a lack of information. So, tadpole food was not tested for the presence of Atrazine in many of, if not most of the open literature lab studies. And this is viewed as a deficiency.

Page 82

Many of them used static renewal techniques that Bruce Pauly talked about. And that is also viewed as a deficiency. So, it seems like there's a, there is an ongoing gap between these expectations and what is being done in the open literature. I was reassured by Dr. Steeger's response that these alone would not have refuted, or did, completely discounted the studies; nevertheless, all of the studies in their estimation were discounted.

And there was a comment yesterday referring to GLP standard studies. I guess that means good laboratory practice studies. Dr. Steeger said that the open literature cannot hope to compete with that standard. And I, actually, hope that that's not true. I think that it's actually very important for these standards to be developed in a way so that scientists that are not doing things in close collaboration with EPA can participate in this process.

I think it's, that's actually very important. If that's statement is taken at face value, it's going to be very difficult for open literature studies to be used as part of the basis for decision making in context like these, and by default, that means the GLP studies are going to predominate in a situation where weight of evidence is important. And so, I hope that

Page 83

we can do some more thinking about that, and it's a part of our, the SAP's recommendations, maybe, come up with some advice that we can offer to EPA to try to increase the participation of scientists working in the open literature.

One comment I wanted to make about the laboratory studies that goes back to something that Dr. Leblanc said yesterday, was this idea of reviewing the open literature studies in the context of what was seen in the data call-in study. So, without getting in, too much, into the, we're going to have plenty of time to discuss the data call-in study. One of the patterns that was observed in a number of open literature laboratory studies was the influence of Atrazine on growth and development of larval Xenopus and other species.

So, this result was confirmed in the portion of the DCI study, and it was confirmed in the laboratory that would have had the best ability to detect such an effect. And I appreciated the presentation on power today, but what I would really like to know is, what the power of detection of that effect was. Because it might have been less than 1.6 or 1.7 standard deviations of an effect size. So, but we know that at least the trend was that if we have the

Page 84

best ability to see that in the Berlin lab, and it was there, and it's been seen in other places.

And the only other comment I would add to that is, is that, that effect was dismissed as an anomaly in the DCI study, and yet, it's something that we've seen in these other studies. And so, I don't think it's fair to characterize it as an anomaly, and the other argument that was made is that, this isn't biologically significant. And, in fact, if you go to the ecological literature, and try to estimate the influence of a 6 percent decrease in body weight and metamorphosis, I think you'll find that studies have shown that post-metamorphic survival and the size at maturity, and so on, can be influenced by an effect of that size.

So, while the growth and development result is not consistent among the open literature studies, and it's not even consistent within the DCI study, it has emerged frequently enough that I think we can't treat that as an anomaly.

As for part C. field studies, earlier this morning I asked Dr. Steeger the question, have there been any new studies done since 2003. And, essentially, the answer is no. So, the answer to the overarching question that we're supposed to be

Page 85

entertaining here is, the Agency has concluded that the information contained in the open literature published since 2003 SAP does not provide any additional information.

So, in fact, there is no additional information on the field side of things. And, I view that with some concern, because while I certainly agree that there are significant issues with each of the field studies as they have been published, the white paper fails to acknowledge something that I think is an important observation that was also raised in the 2003 SAP report.

And that is, that several field observational studies, or surveillance studies, conducted by independent research groups, have detected gonadal abnormalities in wild populations of North American amphibians.

And in at least some of those studies, those patterns were heterogeneous across landscapes where either the expected or confirmed application rate of pesticides, including Atrazine, also differ. So, it does, these kinds of studies, I think, are probably incapable of giving us that gold standard type evidence that you can get in a study design like the DCI study.

But, what they do show is that these kinds of

Page 86

gonadal abnormalities are not simply an artifact of a lab practice of some kind. They, they seem to be happening out in the field. We don't have a good understanding of background rates, but what we do know is that it varies in space. And it varies in space in a way that, at least, raises the possibility that was expressed in the 2003 deliberations, that this hypothesis is worth entertaining. And rather than viewing the field part of all of this as subsidiary to the development of a laboratory study, at least some members of the 2003 SAP viewed those as complimentary efforts.

In other words, these observations of gonadal abnormalities in the field stand on their own. And so, in my estimation, they constitute a lot of evidence that's worth further evaluation. That may or may not lead us to a conclusion that Atrazine's even involved.

As Dr. Steeger and others have pointed out, it's very difficult to tell, in most of these studies, what the driver may be, because of various problems in the way those studies have been carried out. Nevertheless, the abnormalities are there. They're in the field. And, there, we're not here to talk about limb deformities, but there are also studies of limb deformities that show similar kinds of patterns.

Page 87

So, none of what we're talking about here today, with the DCI study, takes away from the fact that there's something going on with frogs out in nature. And we need to pursue that. And with that, I will turn it over to my co-discussants.

DR. HEERINGA: Dr. Denver is the first co-discussant.

DR. DENVER: Well, first I have to say that I concur with the points that Dave Skelly made. I don't know of any other literature that is available, open literature that deals with the effects of Atrazine alone on amphibian gonadal development. However, I want to point out that there is a body of literature that deals with Atrazine effects on other aspects of amphibian life history development and survival that's not being considered. And as an environmental risk assessment, gonadal development is only one poss-, but not the most important potential endpoint to be analyzed.

Regarding part B., the data that are recorded in the open literature are derived from studies that vary in their experimental design, and the questions being asked, the endpoints analyzed, et cetera. And there clearly are flaws in these published studies. But, I, my opinion is that this does, in and of itself,

Page 88

disqualify all of the open literature studies for applications in the testing of the central hypothesis.

Several lab studies, in fact, studies that we considered in the 2003 SAP, suggested that there may be effects of Atrazine on amphibian gonadal development, of course that prompted the subsequent DCI. A major difference between the published studies and the DCI studies are static renewal versus a flow-through system. And I think that, I think that, and I agree that the flow-through system provides much greater quality control, but I would like to have seen a comparison of the static renewal versus the flow-through. Because those data that are in the published literature are still out there, and still need to be explained. I don't think they can be completely discounted.

Now, studies in which the conclusions were negative are not informative in that there are significant flaws in the experimental design. There are many problems with the field studies. Unpublished studies, I think, are sufficient to refute the hypothesis. However, some of the published studies do provide support for the hypothesis, despite flaws in their design. That is, these flaws may or may not be fatal flaws.

Page 89

Whether such flaws are sufficient to remove a study from consideration is arguably a subjective determination. The possibility that components of these studies may provide some relevant data appear to not be considered. Limitations and flaws may render some data suspect, but I find it remarkable that, virtually, all of the published data are being discounted. I think that it is misleading to suggest that all conclusions of all of the studies reviewed for the 2003 SAP are suspect.

And finally, part C., I agree that the field studies are not adequate for testing the central hypothesis; but I also agree with David's assessment that it requires further analysis before concluding that there is no risk of not showing it in populations.

DR. HEERINGA: Okay, Dr. Denver. And the final discussant is Bruce Pauly.

MR. PAULY: Bruce Pauly, Environment Canada. I agree on question Dave, with Dr. Skelly and Dr. Denver. I'm not aware of any studies that have been published that would help to address this question published in the open literature.

I also agree with Dr. Skully and Dr. Denver on question B. I guess, I also have some concerns, and I guess maybe that was why I was trying to determine

Page 90

how, which criteria would trigger no further consideration of a study. I think there is some, there is something going on, and it's been seen, effects have been seen in so many studies, that I think that it, it would behoove us to try to see, maybe, doing a consideration of published studies versus the DCI.

In one study in particular, I'm not sure if we'd want to go into it in any detail, I'm, the one study that seems to have disappeared, to a certain extent, from 2003 to now was the Karr study, where it was static renewal.

There was some water quality issues, but there seemed to be, to a certain extent, acceptable mortality. There was effects seen, not a dose response, but I think, I'd, just maybe, like to reiterate that, when these studies are done, things are seen. And, we have to try and reconcile that with a, you know, an evaluation of all of the existing information that we have at hand.

Finally, on C., again, I agree with Dr. Skully and Dr. Denver. Field studies are problematic, and I think we need to do them. I think we need to do them for ecological relevance, and I think we need to do them to figure, try to figure out what's going on in the real world. And I think I'd agree with Dr. Delorme

Page 91

from the previous questions that we need a certain amount of guidance and instruction on how we might conduct a field study that would be acceptable for a risk assessment paradigm that requires rigid considerations of the protocols, and the data quality.

DR. HEERINGA: Thank you much, Bruce. Comments from other members of the panel on this particular question, charge question. Yes, Dr. Leblanc.

DR. LEBLANC: Being on both the 2003 and the current SAP, as several members here, I've participated in both. I struggle with the fact that with the 2003 SAP, we had data, we evaluated data, and we tried to reconcile the data when we struggled. And we felt that there was something going on.

We just didn't know what it was. And the best we could do was to agree that the data was sufficient first to hypothesize that perhaps Atrazine elicited effects on gonadal development in the anurans. And the recommendation was that we now proceed and conduct well-designed GLP studies to reconcile these ambiguities, these uncertainties that we were struggling with, and that's been done. And we have that data in front of us now.

What bothers me is that we're still

Page 92

struggling with people from literature data that we struggled with back in 2003, with all the uncertainties that are associated with it. I agree that we can't ignore it.

We didn't ignore it then. We made recommendations. We've moved forward, and we have new data now to evaluate. And, as I said yesterday, I think there would be value added if we revisited the open literature data now, and compared it to the GLP studies, but I truly feel that our mission, at this SAP, is to look at these GLP studies and look at the spread and what they have to say.

DR. HEERINGA: Yes, Dr. Petino.

DR. PETINO: Reynoldo Petino. I just want to say that I agree that the field studies, despite the problems that they may have, that they should not be just discounted.

That, you know, they either conform or don't conform to a hypothesis may not serve to prove it or disprove it, but they may provide some information to, in support of the laboratory studies. So, they should not be just discounted, I mean. Then, on the other thing, and I don't know if this, you know, if I read correctly the report, the white paper, that there may be, may have been an error in the white paper in the

Page 93

review of the Hayes pesticides mixture or study.

But, the analysis of the white paper states that this paper found no significant effects of Atrazine on the size of metamorphosis. And, by, if I read the paper correctly, they had a figure three that had that met, Atrazine actually did affect the size of metamorphosis, to a degree that was very comparable to one of the laboratory results of the DCI studies.

So, I don't know if we may want to look at that, and maybe correct the white paper if necessary, but if I read it correctly, you know, that on page, at the top of page 65 of the report, that statement is made that the Hayes, et al, paper did not find an effect on size, when, in fact, they did do a report of an effect of Atrazine alone on size. So, I just wanted to point that out.

DR. HEERINGA: Additional comments, yes, Dr. Williams.

MS. WILLIAMS: Thank you. I'm not a doctor, but thank you.

DR. HEERINGA: Just on tv.

MS. WILLIAMS: Not even on tv. I appreciate all of the comments and the struggle with literature studies. And I think one of the things that was pointed out a minute ago is, kind of, where we're

Page 94

coming from on these, this point.
For this particular question, the open literature studies that were available were not discounted. They, in fact, did form the basis for the hypothesis that had led us to where we are today. So, I just don't want to leave without saying, that we're not ignoring the information. It was not adequate to form a conclusion, we didn't believe, and the 2003 SAP agreed with that.
I also wanted, though, to point out, because we're focused so discreetly on this one issue, in talking about open literature studies, the non-GLP studies, we're talking about it in the context of this issue. There are other issues in which we have used well-conducted field studies, and in fact, some of those issues related to this particular chemical, where there was a body of mesocosm studies that were very well conducted that we were able to use in a regulatory context.
So, I appreciate the time to be able to just articulate that, that when we're talking about these things, we're talking about them specifically in the context of the issues before us today. And I hope you don't take it more broadly to imply that there are no field studies that are of value, and that nothing that

Page 95

isn't done in a laboratory under GLP standards is of use to us, 'cause that's just not the case. Thank you.
DR. HEERINGA: Thank you very much. Yes, Dr. Bucher.
DR. BUCHER: It's John Bucher. I think, though, that the issue that's been brought up about evaluating the new studies in light of the older studies is, still has value to me. Because, for example, I think that some of the static renewal studies that might be reevaluated in light of the fact that nothing was seen in a flow-through study.
So, you might want to begin to look at metabolites and things of that nature. You know, the issues of buildup of potential metabolites as being responsible for some of those other effects seen. So, I think there is, certainly, certainly value in going back and looking at the old literature and trying to understand it, in light of this new literature. And I would agree with those statements.
DR. HEERINGA: Yes, Dr. Bailey.
DR. BAILEY: Open, the field studies, I think, have a very important role. They serve a different role than lab studies, but for people who are examining this kind of work, having that kind of data, the field studies is very important. Otherwise, you'd

Page 96

just have lab work, and you try to talk to other people about policy and so forth, they say, well, that was just lab work.
DR. HEERINGA: Thank you. Dr. Delorme.
DR. DELORME: I just wanted to ask Dr. Denver for a bit of a clarification perhaps, or a bit more information. You indicated in your comments that you thought some of the lab studies would be, or could be useful. And I was just, I'm sitting here looking at the study deficiencies that EPA has identified. There is numerous ones for some of those, and looking at this from a risk assessment process, as a risk assessor, you know. There's enough concern expressed in those that I would have a hard time using them to set an effect endpoint to do a risk assessment.
DR. DENVER: Right, so I'm looking at it from the perspective of a scientist trying to test an hypothesis. And what I'm considering is whether there are any data in the public studies that might be considered in providing support for or against the hypothesis.
DR. DELORME: Okay.
DR. DENVER: Not in a risk assessment. And so, that's what I'm referring to. So, for example, Professor Petino just mentioned data in the paper that

Page 97

were, actually, supportive or perhaps parallel to the DCI study. So, the question becomes, then, what do you do with those data.
Do you cherry pick, or do you consider data, you know, in one study and not another. Or, you know, are there some data in some studies that you might consider reliable, or at least might consider as a basis for, either formulating a hypothesis, or a test for a hypothesis, or design the experiments to further test the hypothesis, so that's what I meant.
DR. DELORME: Okay. Thank you.
DR. HEERINGA: Ms. Pease.
MS. PEASE: Yeah, I just want to clarify the Hayes study that you're referring to that did find treatment related effects at Atrazine at the .1 PTD level for growth in the snout vent lengths. And the problem with that study is that they compared that treatment to an ethanol control, and there was no negative control tested along with it. So, it kind of confounds our ability to discriminate between potential treatment related effects and solvent impacts. I just wanted to clarify that. Thanks.
DR. HEERINGA: Thank you for that clarification. Dr. Steeger.
DR. STEEGER: I'd like to make a couple

Page 98

of clarifying statements. On the rejection of studies because of static renewal conditions, many of the studies that employed this methodology were using the static renewal after three days, and only 50 percent of solution changes at that time.

The studies where we had data on water quality for following similar methodologies, and Karr was one of those studies, indicated that the quality of the water, given the loading rates that, the high loading rates that were used, and the infrequent solution changes, resulted in a compromise of the animal's ability to develop. And that was apparent through the lack of metamorphosis; the animals weren't undergoing metamorphosis, in many cases; decreased weight with increased time to metamorphosis; high mortality rates; and poor response to positive controls.

And I think that the final effect that I just that I mentioned was poor response to positive controls, where you weren't seeing any developmental effects in the animals suggested that the animals were simply trying to survive, and that their development was seriously compromised. And that was associated with this static renewal phenomena, where you have three days of flow, or three days of static, and only a

Page 99

50 percent change. And, many of the open literature studies were following that protocol.

The, I want to clarify something. The field studies, when I was asked are there any new field studies that we were aware of. Several of the studies that were reviewed in 2003, were still in the process of being completed, and were two to three year studies, and only one year of the study had been completed. But the problems that had been identified with the study were methodological problems, and those methods had not changed in the way the studies had been conducted, even into their second and third year.

And so the issues still remain. I will say that, in the growth study, he did, eventually, look at Atrazine in the control sites. That was one of the issues that had been identified in the interim report. That, as I indicated yesterday, there are a number of other confounding factors that, really, limited our ability to use those studies, in addition to what had been identified in 2003 SAP.

On the comment regarding GLP and inability of universities or academia to compete with that standard, in '75, when I was working as a bench biologist at Stanford, I had the misfortune of doing experiments for EPA.

Page 100

And, at that time, the GLP was just in its infancy. And I told them that it really wasn't worth the money that you could get from EPA to impose the GLP standard on any study. It, at that time, in '75, it added $10,000 dollars to the cost of the study. And it was just a nightmare to try to satisfy the EPA. And, ironically, now I'm imposing it on other people.

DR. STEEGER: It's, to expect that standard. I'm not saying that universities, academia cannot meet that standard, but the monies that are involved in accomplishing it. Here you have right in front of you contract labs that regularly function under GLP, and even trying to meet that standard is difficult for a contract lab.

It's very difficult to imagine university labs from the get-go being able to meet that standard, and at a cost that would be competitive with contract labs. I'm not saying it can't, but it's a difficult standard to meet. A study does not have to GLP for us to consider it, though. A well-conducted study would be given heavy weight in any type of analysis that EPA has done.

And, okay, that's, I do have one question to the panel, though, in follow up to your comments regarding the laboratory studies. Can the panel

Page 101

suggest methods of statis-, or statistical methods to consider these studies, given the limitations that have been identified in the methods that were used to collect the data?

DR. HEERINGA: A question posed to the panel, this is snap quiz, possibly. We can do one of two things. We can try to address it now, or we can certainly return to it after a little thought has been given to it. I think Dr. Denver has raised this issue, and I think about the possibility of looking back at some of these now in the light of the DCI studies, but I'm not sure whether any of the statisticians or other panelists present here would have a suggestion.

DR. BAILEY: Ted Bailey. I'd like to answer that, in our current discussion about university research. Is it possible to consult with those researchers, and maybe you did do that for the DCI studies, is it possible to consult with experts about the design and the type of analysis, as well as the treatments in the conduct of the experiment? Thank you.

DR. STEEGER: I'll give you my personal perspective. I am more than happy to talk to researchers that could potentially undertaking studies to address this effect. I attended an amphibian

Page 102

conference earlier this year. And a message that I brought to those researchers was that, you know what the Agency's standards are that are being applied to these studies.

If you hope to generate information that is going to be used in a regulatory context, understand the process that we use. Don't be surprised that your study doesn't meet that, those standards.

When we're applying the same criteria to evaluate your studies as we apply to registrant studies. It's a level playing field. I, personally, have 27 chemicals on my plate that are active. Atrazine is only one of them. I have a 2300 page report that I have to analyze in a narrow period of time, and that's only one of my chemicals. To, we're willing to work with researchers to develop, help them understand how we evaluate studies, but as I said before, we cannot tell them how to conduct their studies.

DR. HEERINGA: Dr. Portier.

DR. PORTIER: I would like to attempt to address your question about statistical method design. And I'm addressing it from a background of twenty plus years of consulting with researchers, both in agriculture and health and ecology on study design.

Page 103

And the problem is, when you have studies that have confounded factors that are important, and they don't do anything in the study design to take that into account, by either measuring co-factors, so we can adjust for them in the analysis, or blocking, or controlling for them at the design phase, you can't do anything statistically. You have the data, the confounding is a fact of life, and it's what leads to all of this discussion that we're having.

Every one of these criteria you look at is a confounding issue that can only be fixed by another study, the next study. You know, in an academic setting, the studies are small, so we tell the graduate student who usually makes this mistake, right, to go do it again and get it right, or take additional measurements. When you're doing $20 million dollar field study, and you have confounding, which often happens because you don't have the control in field studies that you have in lab studies.

You really got to think these things through very carefully before you step in the field. And I don't, and once it's started, and you haven't accounted for it, the data's going to be only marginally useful. It'll be useful for hypothesis generation and raising a lot of questions, but it's not going to answer any

Page 104

questions. And that's, that's my take on it. It's a fact of life.

DR. HEERINGA: Dr. Delorme.

DR. DELORME: Just to echo a few things that have been said already here, as I sit here and look at the study deficiencies that have been identified, in, it's one of the tables in the white paper, there's some things that can be recovered from. Unfortunately, at the time many of these studies were done, there was no protocol for looking at developmental effects on Xenopus. And we know have, at least the beginnings, of a protocol.

And our conclusion in 2003, and as a member of that panel, we decided that there was enough evidence that there could be something, and that's how we came up with the hypothesis, to go forward and try to clarify it.

To retroactively go back and look at these, you might be able to find some support for the conclusions that we've made, but beyond that, I'm not sure what you're going to get. I mean, many of the issues that EPA have identified were issues that were identified by the panel as well as being confounding factors, as Dr. Portier has pointed out. I don't know what you could do to actually make the data more

Page 105

useful.

Things like high loading rates, high mortality controls, contamination, I mean, I would contamination controls a fatal error, and that would take any study out. If you have nothing to compare it to, how can you use that data.

So, and I think you have to give a little bit of consideration that, you know, people at EPA have looked at these and as Thomas has pointed out, Dr. Steeger's pointed out, it's not usually one thing, but it's a series of things that are going to knock a study out. It's unfortunate. We look at this from our perspective now in 2007. You know, having seen a study that was very well conducted with the protocol.

But try and put that in the framework of the studies that were done back in five, four or five years ago, they didn't have that. We've used the knowledge that we gained. We've looked at it. We've recognized what those mistakes were and moved forward.

DR. HEERINGA: Dr. Denver.

DR. DENVER: I just want to respond to that. I want to make it clear that I don't disagree by and large with the evaluation of the studies that are in the white paper by the EPA.

Because, in fact, many of those studies, you

Page 106

know, and many of the same issues we raised and discussed at the SAP in 2003. I think it's a matter of respect of risk assessment, as you point out, and the process of doing science, which is often dirty, you know.

And, you know, we take a result and we use that to formulate hypotheses. So, that's where I'm coming at it from, that, in fact, there are results in the literature that suggest that there could be an effect.

And so, as a scientist, I'm curious why those results were obtained. They weren't obtained simply because of the flaws in the experimental design, right. So, there are flaws in the experimental design that make us wonder whether the results are reliable or repeatable or dose response, that sort of thing. Okay, so that's it.

DR. HEERINGA: Very good discussion, and I appreciate that, encourage that, I guess, as we move through these other questions. Yes, Dr. Williams.

DR. WILLIAMS: Can I please ask a clarifying question of Dr. Denver?

DR. HEERINGA: You sure can.

DR. WILLIAMS: What I heard is the following, and you can tell me if I heard it correctly

Page 107

or not. For effects other than the one that we're looking at, there's literature that shows those effects, and the study showed the potential for those effects. Are you suggesting that it's those that we go back and look at as different effects? That's what I heard, rather than going back and comparing the DCI study to the literature for the effects that we're focused on right now.

DR. DENVER: Now, so, I think you're referring to my reference to the point that was made by Professor Petino, and the, is that what you mean?

DR. WILLIAMS: No, I'm referring to the comment that you most recently made. I guess I just am not, I guess I'm not understanding - -

DR. DENVER: My point of view. So, my, as I said, my point of view is of a skeptic, as a skeptical scientist wondering if there is literature, or if there's data, that points to some type of response that needs to be entertained.

So, I'm not making any comments about how the EPA would use the literature for regulatory decisions. I'm just simply approaching it from the perspective of a bench scientist, who would look at results and decide whether there's something there or not to - -

DR. WILLIAMS: I appreciate that. I

Page 108

guess the question that I'm having in my mind is that I think that this effect that the EPA is focused on right now, I think we did do what you're talking about, because that very literature helped to form the basis of the hypothesis that we brought to this group. So, I guess, what I'm wondering is, are there other effects, that you've seen in the literature, that you're suggesting are more points of curiosity that we ought to be exploring?

DR. DENVER: Well, yeah, I mean there are lots of, there's a lot of literature, and, that goes beyond amphibian and points them out, that suggests that there may be effects. So, but, as far as gonadal development, which we're focused on?

DR. WILLIAMS: Yes.

DR. DENVER: I don't know of any other literature that deals with that, maybe other than the literature that's been reviewed.

DR. WILLIAMS: Okay, thank you.

DR. HEERINGA: Any additional comments for the panel? Dr. Steeger.

DR. STEEGER: I'd like to follow up on that, Dr. Denver. I would, as a scientist, wonder about the fact that you still have people who are generating effects data showing Atrazine and gonadal

Page 109

development.

Right now, it's my understanding, based on my review of the literature that Dr. Hayes has demonstrated in the past, that exposure to Atrazine, at various concentrations, results in various levels of an effect. In his most recent effort, though, to study even the same species, ranid, rana pipiens, there is no effect. And when I try to understand how does, how does he get a response like that, since it seems to be so clear in his original work; and yet, when he reproduces it, he doesn't get it. And when we have a study that attempts to control for all the sources of variability, we don't seem to see the effect.

Yet, there his study is, the original one, where there is an effect. And I really am at a loss as to how to explain that. And I believe that, that might be what you're getting at, that you do have some investigator showing that there's an effect there, and it's, it stands. I don't know. I can't explain it.

DR. DENVER: I can't explain it either. And, I, you know, and I'm actually going to mention this in a moment, but there, there are effects that appear, and in fact, in the Karr setting, there's also an apparent effect that is similar to what Hayes found. I don't know why he, he has not repeated it in other

Page 110

species.
But, and I'm not trying to defend his work, or anyone's work. I'm simply coming at that this as a skeptic trying to look at the data. I think that, you know, I'm still puzzled by results that are reported. And I wonder why, you know, why did they, we find these results.
Is it something that is related to the experimental design. Now, the experimental design in the DCI studies are arguably impeccable. I mean, there's, you know, they're highly controlled, following the STM standards, but nature is not like that.
In fact, the static, you know, the static renewal studies, which had been done, I wonder whether there's some aspect of those studies that led to the generation of an effect in some labs and not others. And that's a curiosity that I have. And I don't know the answer to that.
DR. STEEGER: I share your curiosity, and I struggle with what kind of conditions could have accounted for that. And I, I also struggle for, if a researcher has, or had, if we've identified what could be potential sources of variability in the study, how much you eliminate them, why someone would proceed to conduct a study knowing that those are potential

Page 111

sources of variability, and not make any effort to correct for it.
The use of plastic mouse cages, and for those of us that have worked in labs with mammals, with rodents, know that as you work with plastic mouse cages, they have a limited life span in the lab, because they become brittle with washing, and eventually, they become useless. They'll just crack. And the reason for that is they're leaching out, they're plasticizers.
EPA very strongly recommends the use, or discourages the use of plastics in aquatic studies, because of the potential for plasticizers to leach out of the container. There are published literature that indicate that phthalate ester, it's a common plasticizer, can produce gonadal deformities, and to continue, knowing that, and to continue to do a study that could potentially be confounded in that way, is amazing to me.
Yet, even in his, like I said, if, that aside, if you look at, as the researcher continued to produce the effect that he claimed occurs under thousands of times of experimentation, thousands of times, tens of thousands, the answer is no. And it's attributed to biological variability. And I say var-,

Page 112

biological variability might account for a decrease in effect, but not account for an elimination of an effect.
DR. HEERINGA: I'd like to thank everybody, yes, this has been a very healthy discussion, and certainly one that we'll, I think will be reflected in the report of the panel. And I appreciate the contributions from everybody on this topic, because I think, clearly, it's critical to the consideration, not only of the current experimental results, but the general issue of scientific evidence here. Okay.
What I'd like to do at this point, is to move on to charge question. Let me just look at the panel. Let's tackle charge question number three, and we may, we'll take the appropriate time, but, obviously, if we get to 12:30 and we haven't finished, I may call for a lunch break and resume. But, Dr. Irene, if you could read charge question three into the record, please.
DR. IRENE: Stephanie Irene, question three. The following questions will concerning the DCI study. Please comment on the Agency's evaluation of the final study design.
For example, the Agency concluded that the minor changes in the experimental design, such as,

Page 113

omitting Atrazine degradated analysis for DACT, DEA, and DIA, and not conducting differential cell counts for ovarian and testicular histology, did not compromise the means to assess the hypothesis that Atrazine exposure can affect amphibian gonadal development.
If the SAP concludes that the alterations in this study design preclude, or significantly compromise, the ability to assess the hypothesis, please discuss the extent possible, to the extent possible, how the specific design modifications could impact the means to assess the hypothesis. Please provide comments on other aspects of the Agency's evaluation as well.
DR. HEERINGA: We'll return to Dr. Denver, who is the lead discussant of this question.
DR. DENVER: Okay. So, omitting the Atrazine degradated analysis, that is for Deethyl Atrazine DEA, Deisopropyl Atrazine DIA, and Diaminochlorotriazine DAC, of the water, in my opinion, in itself, did not compromise the means to assess the hypothesis that exposure to Atrazine alone, and I emphasize alone, can affect amphibian gonadal development. The limited analysis that was done suggested that the metabolites are at or below the

Page 114

limit of detection, and subsequent water analysis were not conducted.

We know that Atrazine is rapidly metabolized by frogs. It's unclear whether the tests have been conducted with tadpoles, but I know that frogs have been analyzed. And there's little bio-accumulation in fish or frogs. So, it's like that if metabolites or degradates were generated to any significant degree that they would be rapidly cleared in the flow-through system.

Thus, these studies were able to test whether Atrazine, in its native form, affects amphibian growth and development, without potential confounding effects of Atrazine metabolites or degradates. And I just want to read for a moment from a document that was submitted from the eco risk panel that is the effect of Atrazine on aquatic wildlife, a critical review.

DR. HEERINGA: This document, by the way, is in the docket for this particular panel meeting, and panel members have been provided a copy as of yesterday morning.

DR. DENVER: Well, just a couple of sentences. Measurements of a persistence of Atrazine transformation products are limited, however, aqueous half-lives of Atrazine degradates and metabolites are

Page 115

not available.

Bioconcentration data for Atrazine transformation products are also limited, as studies reporting biological concentration focus on metabolite production in biota upon exposure to Atrazine, rather than uptake a few degradates from the environmental media.

Based on the results of the studies in mammals and other organisms, these metabolites should be considered in assessing potential risks. So, therein lies the rub. One wonders whether the flow through paradigm that was used in the DCI studies, versus the static real paradigm that they used in all the open, in the open literature studies, which we've discussed already, and have significant flaws.

And prior registrant sponsor studies could account for the differences in the results among the studies. And that's what, that's what I wonder about. For example, one cannot rule out the hypothesis that a metabolite or degradate of Atrazine is responsible for the xenobiotic activity of the compound that's been observed in some other studies, some, but not all studies.

Now, these effects could be minimized in the flow-through system, since such compounds would be

Page 116

rapidly cleared. And to my knowledge, the effects of Atrazine metabolites have not been tested for effects on amphibians. Correct me if I'm wrong. Nor have these metabolites been routinely analyzed in exposure studies. These metabolites have been measured in sites where amphibians live in South Africa and other areas.

So, they do occur, and they are present in the environment. They have been shown to have effects in other species. For example, they have been shown to, perhaps, delay puberty in rats. This is work of some of the EPA lab in North Carolina. So, there are other precedents for considering the activities of metabolites, degradates. One of the better known being the breakdown product of DDT. It is DDU, which has demonstrated activity, endocrine destructive activity.

The second part of the other question was the omission of cell counting for ovarian and testicular histology could have provided a means to backup the hypoplastic scores of the gross morphological level.

However, as was pointed by the participating histopathologist, on stage 66, it's very difficult to differentiate primordial germ cells and primaries for that value. He suggested that it was impractical and of minimal value to do differential sub-counting of those omissions.

Page 117

DR. HEERINGA: Our second discussant would be Dr. Furlow.

DR. FURLOW: So, first off, I have to say that overall, the study design was, in fact, impressive to me, in terms of the controls that were included, and the rigor to which the water quality control and health of the animals was maintained. And that is certainly something that has confounded amphibian research in the lab for years and years and years.

And, I think that the setup that was used in the DCI experiments was justified and was, well, you can see by the consistency between the two labs, one in the States and one in Germany, that the amazing, to me, having worked with amphibians for a while, in laevis Xenopus, in particular, the amazing degree of overlap between the two labs is really a tribute to the careful conduction of the study.

And I think, Peter Delorme's earlier statement that we really need some sort of standardized testing setup for an amphibian, I think this is an excellent starting point and framework, with which to continue to test potential endocrine disrupting or other kinds of chemicals that compromise amphibian, all kinds of aspects of amphibian physiology.

It's my understanding that, at least part of

Page 118

the original design of the flow-through system, maybe, actually, Dr. Davis can mention something about this, was to develop a system that would give you extremely reproducible time to metamorphosis. And as someone who studies that particular aspect, I appreciate that, and I'm jealous at the tight error bars that have, and the times metamorphosis, and the, actually revealing potential effects on body weight in the snout vent length of Atrazine. Small effects, but nonetheless, I think it's physically significant ones.

So, all of that being said, I think, I definitely agree with what Dr. Denver said earlier, in the terms of does the study design allow us to test the hypothesis of whether or not Atrazine alone effects amphibian, or at least, let me say, Xenopus laevis gonadal development. I think so. And I think the data produced list was remarkably good. However, I do agree with Dr. Denver that, in trying to reconcile some of the earlier studies, that metabolites should have been considered.

In addition to the studies that Dr. Denver mentioned, it's also, I think, reasonably well established, that DDT metabolites to a safe DDE can actually be a potent anti-androgen. Whereas, DDT is a weak estrogen. And so, I think these things certainly

Page 119

need to be accounted for.

In addition, I was actually impressed by the mitochondrial DNA testing that was mentioned in the earlier presentation, and actually getting at actual potential subspecies differences in Xenopus laevis, and differences in where animals were collected from South Africa, and where people actually get their animals to do these studies. That, actually, was very impressive to me, and very important work. It was, in fact, curious that the animals that were, sort of in the maize growing regions, and north, east South Africa, in fact, were sensitive.

Or, I should say, showed more of these testicular oracites than the animals were actually, most of them were collected down near the Cape for, to supply the animals for, for virtually all of us. But it does raise the question of where the animals come from, all of those earlier studies.

Now, the large majority of them are known to come from Nasco, is that Xenopus One, or Xenopus Express, but to my understanding, the Hayes studies were using his own colony, and it's unclear where those animals came from. Were those animals, in some way, more sensitive for whatever reason. So, this is something that, at least, ought to be considered.

Page 120

The other thing is just a general comment, and that is, in the design and, this is something that goes back before Hayes presentations, and back to the original suggestions made by the 2003 SAP, I think. And something that was brought back to my mind by something Dr. Steeger said. And that is, mixtures are not being considered, and I think that for, I can understand and appreciate the EPA's stance on this, or at least, approach to the Atrazine question, that it, you start to open, I would say, a Pandora's box, when you start thinking about, well, maybe it's not just Atrazine.

It's Atrazine in combination. We touched on this, in terms of the metabolites, but it is also formerly possible that Atrazine is, essentially, sensitizing the system to the effects of other compounds, including possibly estrogens.

If there are phthalates potentially coming from the containers, perhaps Atrazine is sensitizing the animals to the effects of phthalates. This is possible, but I don't think it can be ruled out. I think that it's becoming more of an important problem that we're going to have to address sooner rather than later about the effects of mixtures.

All of that being said, I think that the

Page 121

study design was, in fact, in my mind, remarkably robust. It touched the specific hypothesis about Atrazine, and pretty much, Atrazine alone on Xenopus gonadal development. And I'll stop there.

DR. HEERINGA: Dr. Miller.

DR. MILLER: Debra Miller, UGA. I agree with the other discussants, and I do, also, want to say that your design was very good. But I, also, would like to say that I, it would be a great advantage, I think, to look at parallel studies of flow versus stagnant or, a stagnant type of system. Simply because, in other situations, we are starting to design studies that way and are noticing a difference. And most of those are with dizzy studies, but it may be similar here just because of the degradate component.

And, as far as the differential cell counts, I understand, and I totally agree with omitting them, and I don't really have a problem with that. The problem I do have is how you dealt with the scoring, because for, as a pathologist looking at the results, it's really hard for me to look at, okay, so you have significant differences here, but not here, but what does that really mean. Because we can see such a wide range of variation in changes.

So, if you have hypoplasia of the tubule, for

Page 122

example, looking at it histologically, did you just have, like, less than 1 percent, and you might see that in all of them, and that is going to be a standard. And so, then, that becomes a presence versus absence. But how many did you see were most of the tubules were affected, or 50 percent of them.

So, that becomes a real issue when you're looking at where there was no variation with increasing dosages. Because you don't really, you've got to be able to break that apart. And I understand the variation, and I am so amazed with the numerous slides that you went through to look at these and grade them, that it's really quite a feat in and of itself.

But, I totally, also, understand that if you review a slide more than one time, yes, you can easily see differences. And then, you have to go back and say, well, what do we need to do to compensate for that, rather than just omit it. Does that mean that your pathologist has to then go through the slides two more times and take the average. Which would be, I'm sure Dr. Wolfe would have absolutely no hair left by that time.

Or, do you need to have a panel of pathologists, then, to, you know, to say, well, we'll have two others. And Dr. Wolfe is the lead, and so

Page 123

then, we can compare those. So, for me, omission of that was more of a factor than omitting the differential cell count.

DR. HEERINGA: Dr. Petino.

DR. PETINO: Reynoldo Petino. I, generally, agree with what has been said. I think that the lack of the information of the degradates of Atrazine, even if one takes a, that the effect of Atrazine was the main interest, does not effect the ability of the study to determine whether the Atrazine is having an effect or not.

I am, I also, share, though, the concern that the change from static to a flow-through, and what the level of metabolites, and you know, that accumulated in those two systems, and whether that might be a factor. And, you know, and perhaps, you know, the effects being different now than, you know, some of the studies in the past that reported effects. But, I think we're providing information.

The answer does show that the level of metabolism was quite low. So, I don't know that I'm terribly concerned about that, but I agree, it's a question, it's a question mark, and you know, when you change from a static to a flow-through, or that would definitely affect in relation to metabolites.

Page 124

The lack of the differential cell counts, I'm not bothered too much about that either, especially because there was some other measurements made, you know, gonadal image area, for example, where there seems to have been some effect where they were small, and if I remember correctly, they were different of the, there was a trend towards negative values in one live, and an effect of the high concentrations won't effect the other live, so if the effects were inconsistent, and were small.

So, I mean, if it had been otherwise, I would have said that maybe the differential cell counts would have been important to consider. But, it's not, so I'm not too concerned. I would like to make a point, though, that perhaps, if the concerned one is the, you know, the, the effects of Atrazine or any compound on gonadal genesis, and the ability, you know, the productive potential of the individuals as they grow as adults. That probably, other endpoints would have been other than, I guess, I wasn't here in 2003, or I wasn't a member of this panel.

But, if there would have been other endpoints probably, it would have been more appropriate to measure. If the concern was the effects gonadal genesis future potential, and if they're commenting on

Page 125

that, for example, size and age to maturity are, you know, important parameters, they have also an ecological relevance. But, actually, those are common measures that people in the fisheries field make. So, you know, is there an impact on those, too.

I think, the, that we were presented with a study yesterday, a grow up study that can, attempted to address that concern, and I think it's, that study has a lot of value in determining, for example, the effects of the, even, lifetime of exposures on reproductive potential.

But, the, if the impact is, you know, if the question is fecundity, you know, germ cell, numbers of germ cells, I think that the way that study was conducted, probably, was not appropriate to address that question, because the frogs, after the exposure was conducted, and they became mature, the frogs were selected because they weren't mature, even, they were mature.

They were, in other words, it was a biased sampling for selection of the rootstock in order to determine the F2 generation effects. So, I think points such as age, size of reproduction are important, would, could provide some information in view of the this differential sub-colonies. But, that would be

Page 126

useful, that information would be useful there, too.
DR. HEERINGA: Dr. Green.
DR. GREEN: Sherril Green, and I just wanted to make a comment. I concur that I'm not too bothered, particularly, if the histological data was not recorded as, or collected as recommended, but I do think omission of the metabolites is probably something that should be considered as important.
One of the reasons, as I recall the 2003 recommended the flow-through tank testing system to start was because, at the time, the papers and the data that we had had largely been collected from animals that were in static tanks, and as Dr. Steeger pointed out, the water quality analysis was, either, not done or what was reported, reported some parameter changes in waste materials that would have prohibited normal growth in that tank anyway, or compromised the animal's health.
And we did have one paper, at that time, that showed, in the presence of nitrate, Atrazine did have some effect on those animals. So, before we could begin to sort things out, that recommendation was made that we try to begin basic studies in flow-through tanks. And I think that data has been done, and it's pretty good data.

Page 127

And, again, in a static pond in an environment, though, in the field, animals might not get the benefit of a flow-through, wash away effect of Atrazine or metabolites if they're toxic, or the metabolic waste that you would find normally, or any other chemical in conjunction with Atrazine. So, I would make the recommendation that those studies be reconsidered as important. And the studies I'm referring to are the ones where the metabolites are evaluated as well.
DR. HEERINGA: Dr. Chambers.
DR. CHAMBERS: I'm a little confused about the metabolites, and I guess we don't know for sure that the tadpoles can create metabolites at their level of development, but any in vivo study, of course, studies not only the parent compound, but any metabolites that the organism forms while the parent compound is being metabolized in them.
So, if metabolism was possible by these tadpoles, then obviously, that the, the studies studied not only the parent compound, but any metabolites that the tadpoles formed. So, that was part of the study design. Now, the degradates that, environmental degradates, that's another question altogether.
DR. GREEN: I should correct and say

Page 128

degradates, instead of necessarily, metabolites.
DR. HEERINGA: Dr. Leblanc or Dr. Delorme.
DR. LEBLANC: Jerry Leblanc. In regards to metabolites, I think there certainly would be value in, if we had information on metabolites from these studies, with respect to, perhaps, explaining ambiguities between studies, between the flow-through and the static renewal studies. Unless, I'm significantly less concerned about the relevance of metabolite data with respect to trying to assess ecological hazard of the material.
If, and I was a little bit confused as well. If we're talking about biotransformation products, metabolites, I think it's addressed in these studies. The animals were taking up Atrazine in a flow-through situation, and they're constantly being exposed to Atrazine. And they're little machines. They're producing these biotransformation products, and so they're getting a good dose of them.
If we're talking about environmental degradates, then certainly in a static renewal situation, these things could be accumulating. But, we're talking about a little aquarium, as opposed to a big pond, and I think that the pollution, a,

Page 129

considerations there would be again, would certainly reduce my concern.
The other thing is that, and I could be wrong here, but I, but it's my understanding that the degradation products of Atrazine are no mo-, have no more of propensity to bio-accumulate than does Atrazine. And Atrazine doesn't bio-accumulate. So, it's not like these animals were being cumulated. They may be exposed, but they won't be accumulating significant amounts of these materials.
And I, I just want to add a voice of support to something that Dr. Furlow had stated with respect to variables. The ability of Atrazine or the toxicity of Atrazine to, perhaps, be confounded by other components in the exposure environment, perhaps, metabolites, perhaps, estrogens, perhaps, nitrates, whatever.
Certainly, this is an issue, and again, it may be, in time, it may turn out to be the issue that explains a lot of the uncertainty that we're struggling with. But this isn't unique to Atrazine. It's something that EPA has to deal with with respect to every chemical that they're charged with regulating in the environment, and it's something that they need to get a hold of and address, but I think it's warranted.

Page 130

DR. HEERINGA: Dr. Delorme.
DR. DELORME: And with respect to metabolites, or environmental transformation products, whatever you want to, metabolites I usually associate with the organism, itself. And transformation product is what's in the open environment. EPA might want to consider looking at the existing monitoring data to see to what extent larval amphibians or eggs might even be exposed to determine whether or not the concentrations in the real environment are high enough to warrant that. If you consistently find high concentration, it might warrant some further investigation.
Obviously, when we, when the panel in '03 looked at and recommended a flow-through design, it was the water quality parameters that were uppermost in our mind, trying to reduce the number of confounding factors so we could get some clear answers to some of the questions that we had posed.
But I think that this is a case where maybe there is some field data available that could help you make a decision as to whether or not you need to move forward on that.
And it doesn't, necessarily, mean you have to repeat this kind of, or the details spending with those as well. There may be other methods that can tell you

Page 131

whether or not there's potential for finding to estrogen receptors or whatnot and help you out in that to begin with.
So, you might want to consider looking at the existing monitoring data for the metabolites. I know it exists out there, to see, you know, what the state of the environment is, and what role or what exposure there is.
DR. HEERINGA: At this point, I guess I'll turn to Dr. Steeger for any questions of clarification or comments.
DR. STEEGER: I just have one comment on the scoring issue. Out of fairness to Dr. Wolfe, the veterinary pathologist, when he was asked to repeat those scorings for the benefit of the new panel member, during an inspection, EPA inquired that the pathologist re-read at least six to ten of his slides. I had his original results in front of me, and I just tracked on how well he was able to duplicate his results. And for the actual lesions, he was able to duplicate them very well.
But the scoring, in terms of the severity of the lesion, was where he deviated. Those deviations were not striking. If it got a score of a one versus a two, that type of difference might have occurred. If

Page 132

it was a high number like a four, he did not score it down to a one.
So, it was not a striking difference in his scoring abilities. But, at that time, I viewed as somewhat as a subjective interpretation, and because of that, I recommended that the scoring would be collapsed. In doing so, it made it more likely to be able to detect effects, but the registrant themselves actually conducted their statistical analyses, based on the score data, as opposed to the collapsed data.
DR. HEERINGA: Dr. Frankenberry.
DR. FRANKENBERRY: This is Mary Frankenberry. I think we collapsed the upper levels into a more severe score and then a less severe score, or, and I think it was, actually, all effects and then the most severe. And with regard to Dr. Miller's point about it's contributing to the variability, I think we did, we probably eyeballed that only before making the decision to combine them. And, my recollection is that the numbers were so infrequent that it, we didn't think it would affect them. On the other hand, if they were extreme enough, I suppose that could have happened, and we should go back and we'll look at that.
DR. HEERINGA: Dr. Miller.
DR. MILLER: Just to clarify, then, so

Page 133

you did or did not include them in the statistical analysis?
DR. FRANKENBERRY: We combined everything into the effect or no effect. The company looked at severe effect or any effect.
DR. HEERINGA: At this point, it's 12:05, and we all deserve lunch. So, I'm going to recommend that we break now until, let's make it until 1:30, and we'll reconvene at 1:30 and pick up with charge question number four. But we're making good progress with the agenda, and I think we're having a fairly thorough discussion on each of these items, and so, I'm very pleased. Thank you. See everyone at 1:30.
(WHEREUPON, the morning session was concluded.)
DR. HEERINGA: This is the meeting of the FIFRA Science Advisory Panel on the topic of the potential for Atrazine to affect amphibian gonadal development. At this point in the process, we have begun to consider the charge questions that have been presented to the panel, and we have completed discussion of charge questions one through three, and are about to move on to question four.
But before we do, Dr. Steeger and Joe Bailey and I had a discussion. Dr. Steeger indicated that one of the studies that was mentioned in the discussion of

Page 134

the charge question this morning, and relevant to the literature on this topic, the study published by Dr. Carr, that there's been some re-analysis. And Dr. Steeger has asked permission for Dr. Carr to present that re-analysis at this point in time, and we've granted that, I've granted that at this point.

DR. STEEGER: The original work done by Dr. Carr on the analysis of whether Atrazine affected amphibian gonadal development was based on gross gonadal morphology. Subsequent to the publication of his article, Dr. Carr worked with experimental pathology labs with Dr. Wolfe and, actually, conducted histology on those same animals. And I've asked Dr. Carr if he would be willing, along with Dr. Wolfe, to provide a very brief overview of what the results of the histological analysis of those study samples were.

DR. CARR: Jim Carr, Texas Tech University.

DR. WOLF: Jeff Wolf, EPL, Incorporated.

DR. CARR: There was some discussion this morning about some of the previous studies.

And one of the things that came up was the question of intersex gonads in the Carr, et al, report from 2003. And what we have done, subsequent to that study, is taken slides that were read for that study of

Page 135

animals that had ambiguous gonads, that we could not tell were male or female, based on gross morphology, and sent those to EPL for them, and this was done blinded, for them to do their analysis, in an attempt to harmonize terminology, make sure that we're calling things the same when we see them at the microscopic level, and EPL prepared a report, based on the re-analysis of those slides. And they were also sent slides from the Cardell study from Michigan State, which was also reviewed as an interim report for the 2003 SAP. And, I think we can make a copy of that report available to the SAP.

In the 2003 Cardell paper, we used the term intrasex at the gross morphology level to identify animals that we couldn't sex as male or female at the gross morphology level. That was the initial protocol. And upon subsequent histological analysis, as we reported in our paper, we could tell at the histology level that they were males and females. This was about twelve out of three hundred animals. It was about 4 percent of the 25 PPV study group.

But we could tell that they were males and females. And, so, subsequent to that study, with EPL developing these more rigorous criteria for evaluation of intersex, as well as a lot of other things, we sent

Page 136

those slides to Dr. Wolf. And, do you want to mention what you found?

DR. WOLF: So, we received the slides, and we also received, although I was blinded to it at the time, the histologic assessments that had been done previously. And, essentially, I think except in one case, I agreed with their histologic assessments, which, actually, instead of saying ambiguity, they used terms like male-like, or female-like. And I confirmed, well, that is a male or that is a female. So, there, really, isn't that ambiguity there.

DR. CARR: All right and we discussed that in the overview document that's on the docket that was submitted by the eco risk Atrazine panel, and that's on page 38, and we cite the EPL report. And I think the EPL draft report can also be made available to the panel, so that they can see this subsequent histological analysis.

DR. HEERINGA: Thank you very much, Dr. Carr and Dr. Wolf. Panel members, any follow up questions on this. It seems as though the data is available to us in the eco risk report that was provided to us yesterday morning. And if there's interest we can apparently see the draft report of Dr. Wolf's analysis, also.

Page 137

But, not seeing any additional questions at this point, we appreciate that clarification and additional information on this point, and we'll certainly take that into consideration in deliberations here. Thank you Dr. Carr and Dr. Wolf. And I think, in our order of question entry, it's Ms. Pease is up to bat, here.

MS. PEASE: Okay, question number four. The Agency has described exposure profiles for studies conducted in response to the DCI and has stated that mean measure concentrations in the studies were lower than the target nominal concentrations. However, the Agency concluded that the frequent analytical measurements provide a sufficiently comprehensive understanding of the exposure profile over the course of the studies.

Please comment on the Agency's conclusion that the Atrazine exposure concentration profile is reasonably characterized and sufficient for documenting the potential effects of Atrazine over a broad range of exposure concentrations. In addition, provide comments on whether the actual concentrations were consistent and sufficiently stable to establish the means to analyze exposure concentration response relationships.

DR. HEERINGA: Our lead discussant on

Page 138

this is Dr. Schlenk.
DR. SCHLENK: Dan Schlenk, UC Riverside. Again, in conversations with Peter and Bruce, we, this is sort of a synopsis, but I'm sure they'll have their comments as well after this, but. As mentioned earlier, EPA attempted to bracket concentrations of previously demonstrated adverse effects in test species. The lowest of these was .01, or .1 micrograms per liter.
In order to demonstrate a dose response effect, a 0.01 microgram per liter concentration was utilized. This was also the LOQ for the toxic end. It is very difficult to carry out an exposure at the LOQ, and WLI should be commended in this effort, since they were able to, actually, maintain appropriate concentrations during the window of exposure.
Since IGB was unable to maintain the exposure concentrations consistently, multiple evaluations of water chemistry were essential in both laboratories in determining accurate exposure concentrations. Given the LOQ, it probably would have been more appropriate to use something like .05 micrograms per liters of bracketing concentration, as the IGB exposure failed to approach these values.
In particular, the, and this is in the IGB

Page 139

study, the 0.01 microgram per liter was under the limit of detection, which is 0.005 micrograms per liter for tank seven on day thirteen, tank four, six, and eight on day twenty, and all of these were within the critical window of exposure.
All tanks at the .1 microgram per liter concentration during the critical window from day thirteen and day twenty were below .1 microgram and were approximately .05, which is, would have been an appropriate value perhaps.
But, of course, this is all sort of moot, since the adverse apical effects, again, in morphological and histological effects, were not observed consistently at the .1 microgram per liter in either lab. And the WLI laboratory was able to meet the exposure requirements during the critical window of differentiation for both concentrations.
The Atrazine exposure concentration was adequately characterized when considering the studies together. And, again, this goes back to my earlier point. It was nice to, actually, have both studies done in both laboratories because, if the WLI hadn't been able to attain that, I think we would have had some issues with regard to the, that low dose, the lowest dose, or lowest concentration.

Page 140

So, and again, if the WLI had not been successful, a concentration of .05 might have been a better choice than, rather than the concentration equivalent to the LOQ. Again, one may wonder if this may be the reason why secondary effects were not observed in the IGB lab for the .1 microgram per liter nominal concentration, but just a thought. Overall, using a multiple chemical concentration analyses adequate to a allowed appropriate exposure over the range recommended by the 2003 SAP.
DR. HEERINGA: Thank you very much, Dr. Schlenk. Dr. Delorme.
DR. DELORME: And I'm just going to echo several things that Dr. Schlenk said. I'll just read what I've written.
The frequency of analysis is sufficient to allow a good understanding of the exposure profile, and we have weekly samples over an extended period of time, gives us a good idea of what the exposures were in the different nominal concentrations each week. And it's well characterized over a broad range of concentrations, which encompass concentrations, which, in the past, have been associated with effects on gonadal development in Xenopus laevis.
So, they did bracket properly the

Page 141

concentrations that before were of concern.
I had some minor concerns with respect to the stability of the two lower exposure concentrations at IGB. These were considerably below nominal concentrations as noted by the EPA. And I had some concern that the measure of concentrations at 1 microgram per liter exposure concentration at IGB was a little low. It was reported at 72, a mean of 72 throughout, with a 74 percent in critical periods. I believe, normally, we would look for around an 80 percent of nominal to have it acceptable.
So, that being said, for those three concentrations, you may want to consider in your report, reporting against a mean average concentration or time weighted averages of some respect, rather than the nominals, because they are so different. I think that would, you know, provide a little bit extra information.
And as Dan noted, the establishment of an exposure level at the LOQ, I don't think is a really good thing to do, and would encourage or recommend that in the future, EPA consider of establishing a guideline for a minimum distance between lower doses and LOQ, just to avoid problems in the future.
We both noted, I went through the raw data

Page 142

that was in the Syngenta report that we received from Syngenta. There are a number of instances that Dan noted where there were no detects at the lowest dose level. And had there been effects at that dose level, it could have seriously compromised your ability to interpret the results.

So, there's a number of tanks where, during the critical period, there is apparently no Atrazine, based on the data that's in that document. So, I'd be careful about that. Other than that, everything else is.

DR. HEERINGA: Bruce Pauly.

MR. PAULY: I would concur with both Dr. Schlenk and Dr. Delorme. I think that the frequent analytical measurements provide a sufficiently comprehensive understanding of the exposure profile, but again, I might come back to the general question, as I did this morning.

We are, we have got data on the parent compound, and that maybe the question should be that the exposure to parent Atrazine was adequately assessed over the course of the experiments, and I won't belabor the discussion we had this morning about the possibility that there are degradates, metabolites in there. That said, I think that I would concur that

Page 143

there are sufficient data to assess whether or not these, the exposure concentrations were well characterized.

DR. HEERINGA: I just want comments from members of the panel. Dr. Steeger, I believe what I heard that we had a fairly clear and consistent response from all of the primary discussants of this. Are there any questions that you have, or do you feel that we've addressed this question satisfactorily?

DR. STEEGER: I think the question has been satisfactorily addressed. I would comment, though, that the registrant notified us before the start of the study that the lowest test concentration would be at the level of quantification.

They had originally proposed testing to .1 micrograms per liter, which has, in previous studies, been identified with effects on gonadal development. The Agency insisted that they test at a lower concentration that suitably bracketed the effects concentrations in previous studies, and would account for a potential U-shaped curve, if one did, indeed, exist.

My recommendation to the principal investigator was that he attempt to sample larger volumes of water, possibly using salt phase extraction,

Page 144

so that they could concentrate the sample and better count for their level of quantification. My understanding was that they were limited in the volume of material that could be transported back to the analysis lab, which was at Syngenta, with, using tandem aspect.

But, we were aware going into the study that it was, that the lowest test concentration was close or at their level of quantification, but we had hoped that there would still be reasonable recovery. And at Wildlife International, they were, indeed, able to be fairly close to the nominal concentrations at 0.1 micrograms per liter.

DR. HEERINGA: Thank you very much. Dr. Schlenk has a follow up comment.

DR. SCHLENK: Yeah, just, again, let me reiterate. I think you didn't have to be stuck on the .01 concentration. You could've gone up, you know, fivefold on that, and still been, and got the cursive view, if you were trying to demonstrate that dose response curve.

DR. HEERINGA: I think that I'd like to move on, then, to question five at this point in time. And ask - -

DR. IRENE: Stephanie Irene, question

Page 145

five. The Agency described Atrazine contamination of negative controls in one out of two studies, and concluded that, since the experimental design had twice the number of controls, relative to other treatments, the data from these Atrazine contaminating controls could be removed from the analyses without invalidating the statistical interpretation of the results.

Please comment on the Agency's decision to omit half of the controls from the Wildlife International study in the statistical analyses, and on the conclusion that the study is still scientifically valid. If the SAP has an alternative approach to treating these control data in the statistical analyses, please provide specific recommendations.

DR. HEERINGA: Dr. Portier.

DR. PORTIER: I like easy questions. I concur with the decision to omit the contaminated controls from the WLI study, and have no alternative methods for treating or using these control data. Also, the control clusters does not invalidate the overall WLI study, nor does it have a large effect on the ultimate power of the study comparisons.

DR. HEERINGA: I think that's the briefest I've heard you. Dr. Bailey.

Page 146

DR. BAILEY: The study is valid. Not using half of the controls may have been unfortunate, but it's not a problem. Thank you.

DR. YEATER: Kathy Yeater. I also agree that there is no effect on the validity of the study by removing these tanks.

I just wanted to comment, also, on how the nature of the study led to a discussion earlier on the appropriate laboratory layout of the tank, mixing container, pump setup. And I really appreciated the discussion on the possibility of, either, the one pump per tank, which could have resulted in increased heat, and also, by affecting increased confounding in the study; or one pump per eight tank cluster, which would have caused greater restriction and more problems, due to the fact that we had to toss some tanks.

So, also regarding the comments earlier on the assessment of a possible cluster effect, I don't disagree with the use of the tank to be the unit of analysis, with the understanding that it is not a true experimental unit.

DR. HEERINGA: Thank you very much, Dr. Yeater. Other comments on this particular question from the panel? Dr. Steeger, I think the answer was fairly concrete, and do you have any comments or

Page 147

questions?

DR. STEEGER: No comments.

DR. HEERINGA: Before everybody runs to re-book their flights, we won't -- So, we're on to charge question number six, Ms. Pease.

MS. PEASE: Anita Pease. The original white paper from 2003 identified measurement endpoints that included the possible enumeration of specific histological structures, such as the number of oogonia in ovaries and the number of spermatids in testes. Such a detail analysis was not conducted in the studies that are in response to the DCI. Rather a qualitative assessment of the instance of ovarian and testicular gonad oocytes was conducted.

The Agency has concluded that the lack of these data does not limit the means to assess hypothesis that Atrazine exposure affects amphibian gonadal development. Please comment on whether the lack of these histological data limits the utility of the available information to support the hypothesis that Atrazine exposure affects amphibian gonadal development.

And B., if the SAP concludes that these data are necessary to adequately assess the hypothesis, please provide options to processing, processing and

Page 148

analyzing the data in an efficient and robust manner.

DR. HEERINGA: Dr. Miller.

DR. MILLER: Debra Miller, UGA. First of all, it's excellent to have multiple endpoints and generally good selection, because they're doable. As discussed in, back in number three, on the mission of the number of oogonia in ovaries and the number of spermatids in testes is acceptable, because, basically, they were not doable. And, qualitative assessment is acceptable, given that scoring was performed.

With the clarification of the scoring variation, I don't have an issue with that. But I would recommend that you revisit the statistical analysis and do so with Dr. Wolf's input. Dr. Steeger is correct in part that pathology is very subjective. Pathology is often described to be more of an art than a science, but in research, we have to combine the two, and mold art into science, and scoring is one way that we do that.

Combining the scores may be acceptable, but I would do so with Dr. Wolf's guidance. Example, yesterday, renal mineralization was brought up, and yes, no one wants mineralization in their kidneys or gonads.

But, Dr. Wolf point out, renal mineralization

Page 149

is not uncommon in fish, and in my experience, it's not uncommon in amphibians or reptiles either. Thus, a renal mineralization score of 1 could easily be expected in almost all of the test subjects, because it can be a common response to various stresses on that organ.

But a score of 2 may be significant, and thus, it may not be appropriate to group scores 1 and 2, and similarly, for the various gonadal scores. So, you might wonder why to revisit the analysis with Dr. Wolf's guidance.

It's because he can tell you whether or not it's appropriate to group those scores. With regard to whether the data collected support that hypothesis that Atrazine exposure alone affects amphibian gonadal development, and here, I need to state that I'm not really certain if using the term amphibian is really correct, because, are we including caught eggs in that, or just anurans and, perhaps, more specifically, should we keep saying Xenopus laevis.

I think you are addressing morphologically and, or anatomically, if you will, development. But this does not, necessarily, equate to function. And to evaluate whether or not development was complete or successful, it's necessary to follow through to the

Page 150

adult stage and assess reproductive ability, i.e., was that individual able to reproduce, and was reproductive output changed, increased or decreased.

As Dr. Petino stated, you presented a grow out study that was impressive, but there were aspects of that study that may fall a bit short, such as, what was the age inside sexual maturity, because I believe you waited until they were two years of age, and then you chose that were sexually mature for your study, which may enter in some bias.

So, for B., as far as study design, as I indicated, I would recommend revisiting your analysis with, on the histological parameters, and, perhaps, consider incorporating size and age of sexual maturity into your broad study.

DR. HEERINGA: Thank you, Dr. Miller. Dr. Green.

DR. GREEN: Sherril Green, and I concur with Dr. Miller in all her points. The only other additional point I have to make is a very fundamental one, that I still think we should consider that we really don't know what is normal in Xenopus laevis in terms of some of these gonadal that we are, gonadal differences that we're calling abnormalities.

And would this be something that we'd see in

Page 151

a very large population of healthy Xenopus in the wild that would go on to develop without incidence and continue to reproduce normally. And that's my only other comment.

DR. HEERINGA: Dr. Petino.

DR. PETINO: Yes, I agree with, Reynoldo Petino. I agree with what has been said so far, and that the lack of the histological data does not impair the ability of the study to determine, to assess the hypothesis.

And the reason being, you know, we talked earlier about the same, the same situation, but you know, the numbers of cells and the types of cells present at the time that this anomalies were analyzed are, you know, that's a dynamic state of changes at that point in time, that also depends not only on the state of development, but also on the age of the animal. So, there's a, you know, counting cells, at this point in time, may not have a lot of value.

So, and as we discussed, and it's been repeated that there's other ways to assess the effects on gonadal genesis and, or, I guess, more to the point, fecundity, or reproductive fitness, better ways to do that than, than, you know, this, this endpoint that was included in the study. So, not having that endpoint

Page 152

does not effect the study's conclusion, overall study's conclusions.

DR. HEERINGA: Thank you, Dr. Petino. Other members of the panel have comments to contribute on this particular charge question? Dr. Portier.

DR. PORTIER: I was just going to say, we are going to address some of the analysis issues related to this in question 9. B.

DR. HEERINGA: Thanks. Dr. Steeger, I'll turn to you again, before we move on, to make sure that a clear interpretation of what was said, and if you have any questions for the panel, clarification on their comments.

DR. STEEGER: My interpretation of what the panel said is that the endpoints that were measured were suitable. And that if additional endpoints, or additional information was needed, a further examination of the grow out study that the registrant conducted may be appropriate. Is that correct? Thank you.

DR. HEERINGA: Dr. Green, do you concur with that? Your questions, of course, will, the final report will reflect the full discussion. Any other comments on this particular item from the panel? Let's move right along then to charge question. I'll ask Dr.

Page 153

Irene to read it.

DR. IRENE: And as you said, moving right along. Question number seven. The Agency has described a number of measurement endpoints, for example, translucent gonads, unpigmented ovaries, and pigmented testes, based on the histology results that were reported in the studies.

The Agency, however, based on its understanding of relevant scientific literature, could not conclude that these measurement endpoints are biologically relevant indicators of the effects on growth or reproductive success.

That is, the Agency did not interpret these responses as adverse effects, per se; nor was the Agency aware of any information that established these responses as precursors to the apical endpoints of primary interest; time to size at metamorphosis, sex ratio, and the presence of mixed and or intersex animals. Please comment on the biological relevancy of these endpoints, and the extent to which they may reflect reliable measures in developmental abnormalities.

DR. DENVER: Well, I'm not aware of any literature that links translucent gonads, unpigmented ovaries, or pigmented testes to effects on growth or

Page 154

reproductive success. That's not to say that these phenotypes are not associated with fitness consequences, but to my knowledge, as I said, such relationships have not been described.

The lack of pigmentation could reflect abnormal migration of cells from the neuro crest, but presumably, that's not the case in these studies, because the exposure was initiated well after that developmental stage.

Alternatively, it could be the failure to express melanin in the melanocytes. But, I don't know what the significance of internal melanocytes are on internal organs. And I'd like to ask anyone in the room if they do know. I think it's a fascinating question, but I have no idea what the significance is.

DR. HEERINGA: Dr. Leblanc.

DR. LEBLANC: The value to get these, what I view as secondary endpoints, are certainly to provide support when effects are observed on more apical endpoints. That is, endpoints that we, as biologists, can ascribe with a little bit more confidence, some ecological relevance, some effect on fecundity.

So, for example, if we observe that male frogs have feminized testes, we're concerned about

Page 155

that, and that may go into the formula in describing ecological relevance; whereas, hypoplastic testes may not. However, in conjunction, it may provide additional support. It gives us confidence in our decision making. However, in the absence of effects on these more apical endpoints, I don't see much value to these secondary endpoints. So, I guess I concur.

The, I can't we can't lose sight of the fact that there is always the possibility that there's some biological significance associated with these secondary endpoints directly, or that they are a precursor to some biologically significant effect. And I think EPA has acknowledged that. But at this point in time, that's only speculation. I don-, I'm not aware of any information in the literature to support that. That's all. Thank you.

DR. HEERINGA: Dr. Petino.

DR. PETINO: I agree, Reynoldo Petino. I agree, I don't think I have much new to contradict, so I'll just agree with what's been said. Thank you.

DR. HEERINGA: Dr. Miller.

DR. MILLER: Debra Miller. I agree with what's been said, too, and I will just comment. Pigmentation is one of those things that we don't really have a good grasp on why it occurs or doesn't

Page 156

occur. We don't know the biological significance of it in many cases. There is some thought that it might help with UV exposure and things like that. But we do see, however, and it's a good point, that you always have to combine these things and put together the whole picture.

Because, if you see some sort of irritant to a specific organ, pigmentation in amphibians and fish and reptiles, it's one of the things that tends to increase in many cases, or it can decrease in the skin, depends on the organ. And so, it can be an indicator that that organ is distressed in some way. But again, you always have to look in a big picture. In this case, you know, I agree with what's been said, in that, we don't really understand the biological significance of it.

DR. HEERINGA: Dr. Denver.

DR. DENVER: I guess we don't know, but correct me if I'm wrong, whether there were any other pigmentary changes in these animals, other than the gonads. I don't know if that was recorded.

DR. HEERINGA: You mean internally, Dr. Denver?

DR. DENVER: Internally or externally.

DR. HEERINGA: Dr. Steeger.

Page 157

DR. STEEGER: If there were changes in pigmentation, it would have been noted on a gross morphological basis for any of the organs that were evaluated, and for the external appearance of the animal. None of which is, was specific to either one of those.

DR. HEERINGA: Comments from other members of the panel on this particular question. Dr. Delorme.

DR. DELORME: I forget who it was yesterday, but somebody made the comment that without a grow out, there's always going to be some uncertainty as to the importance of these kinds of effects.

So, just for the record, I mean, I just went back and looked at the original panel report from '03, and the panel did suggest that some grow out studies be started immediately as well. I understand the logistics behind it are horrific, but again, it's going to remain an uncertainty, I think. It may be nothing, but we won't know.

DR. HEERINGA: Dr. Steeger, I think, get your reaction and any questions of clarification or follow up points on this charge question.

DR. STEEGER: My interpretation is that the panel concurs that it's an uncertainty.

Page 158

DR. HEERINGA: Moving so fast here, I don't even keep up with the numbering. I only have to go even odd, 'cause I know that Dr. Irene did the last one, so Doctor, Ms. Pease, number eight, please.

MS. PEASE: Question eight. The Agency's analysis of potential developmental effects and study's responses to the DCI has focused on histological data, as opposed to gross morphological data. Histological data from these studies were based on the analyses of a single certified pathologist. While this approach eliminates the potential variability associated with having multiple pathologists analyze histological slides, it may introduce an avidity bias.

Please comment on whether a single pathologist is sufficient for interpreting the histology data. If the SAP believes that a single pathologist is not sufficient, please comment on the potential value of convening a pathology review board to evaluate findings of the DCI study.

Please also comment on the potential value of having a pathology review board evaluate materials, for example, the digital images of gross morphology and microscope slides containing histological sections from studies published in the open literature. These data can be submitted voluntarily by the authors and could

Page 159

include slides to evaluate similarities or differences in identifying or describing histological features, and or describing and quantifying histological responses.

DR. HEERINGA: Dr. Miller.

DR. MILLER: Debra Miller. First of all, I do not expect to get through this question today. First of all, as far as, I believe it was a good choice to choose Dr. Wolf. There are very few pathologists that have a lot of experience with amphibians, and experience with fish, reptiles is always a good addition. Internal correction, and I mentioned this this morning, of a bias is possible if you have that one pathologist read the slides like three times.

But if that, the best way to function, I'm not really sure that that is, and in many experiences, especially when you're talking about a study that's designed for regulatory purposes, a panel can be very advantageous, and we tend to use panels a lot when there are large scale studies.

So, although it may be adequate or appropriate for a single pathologist, there's that old adage that if you ask five pathologists, you get five opinions. And that the panel of three or so pathologists can be best to, for studies designed for regulatory purposes.

Page 160

And in this case, one pathologist is usually designated as the lead, for example, Dr. Wolf, and then all the pathologists look at the slides separately, and write their findings, and the lead does this as well. And then distributes the findings to the others. And they all convene in a multi-headed scope to discuss discrepancies. And then, what generally happens after that is, the lead pathologist submits the group's consensus.

And this is how histopath is generally done, like, with marine mammal species, which is one of the areas that I've worked, and this is how we've done it. And it's worked quite nicely. Or another way that you can do it is have the pathologists separately review the slides and submit them separately. And then you add in that into your statistical analysis. This may, you know, I'm not a statistician, but this may, actually, be more, add more soundness to the statistics.

The second part, I'm not really sure exactly what you're suggesting here. As written, I'm not sure it would be necessarily appropriate, but, yeah, it would be great if you could researchers to send in their slides for a panel pathologist to review. However, the reason that I'm not sure it would be

Page 161

appropriate or all that useful is because collection is a huge factor.

You have to make sure that everything was collected in the same way. That the serial sections were done in the same way. The stainings the same, it all makes a difference. Just, you know, one slide from a testes that's done at one facility can be totally different than what's done at another facility. And so, yeah, you could do it, but you would have to give a lot of weight to how you interpret those results.

DR. HEERINGA: Dr. Green.

DR. GREEN: On point A, I think ideally you would like to have more than one pathologist review those slides. There would be a, if it were me doing 100,000 or 83,000 slides over a period of time, there would be some, I believe, the statisticians call it, inter-observer error after a while.

So, to repeat the test, as Dr. Miller suggested, three times and taking an average or do the actual statistics to see how good that pathologist is in calling and recalling the same variations over and over, would be very useful.

In terms of having a pathology review board look at the slides from a particular study, I think Dr. Wolf has done a fine job. It's, his interpretations I

Page 162

agree with. I think the slides are good quality. It seems relatively straightforward to detect gonadocytes and gross morphology differences as well.

So, if you did have a panel of people reviewing that particular set of slides, then, again, as, although I'm not a statistician, I should think that it would add strength to the study statistically if inter-observer agreements were reported, so that we can get a standardized, this set of pathologists all looked at it and they're pretty close in agreement that this is what it is.

But again, I think, Dr. Wolf has done a great job with that, and I feel fairly comfortable looking at most of those slides, and can see what he is pointing out. So, I think it is a straightforward evaluation.

And I concur with Dr. Miller about the utility of having a panel of pathologists look at sets of slides from the open literature, because of the variations you're going to get in processing and serial sectioning and thickness and all that. It might not be so helpful. So, other than that, I have nothing else to comment on.

DR. HEERINGA: Dr. Petino.

DR. PETINO: Reynoldo Petino. I agree with what's been said. I think, I agree with the fact

Page 163

that normally you would want to have, when you, especially when you're ranking, providing grading ranks on categorical variables, it would be ideal to have more than one individual do the gradings. And then having, if there's any disagreements, have the individuals, I mean, this is standard practice, for example, reading otoliths for aging fish.

If you have two individuals reading the otoliths and then, if there's a disagreement, there's a discussion, and if they cannot agree, well, there's procedures on how handle that data. But, so, ideally, then, it would have been nice to have more than one pathologist.

But I also agree with the conclusions of the other two discussants, that, you know, the parameters of it were being looked at and all the catalog of slides provided for us to understand what the, what was being looked at was very clear here. So, I don't know what the value at this point in time would be of having those slides re-reviewed by a panel. But, that's not really my area of expertise, so I'm just saying that I rely on the bonafide pathologists here to give you that advice, but, that's what I think.

DR. HEERINGA: Thank you, Dr. Petino. Dr. Portier.

Page 164

DR. PORTIER: I can't even contemplate looking at 100,000 slides, but one thing I can think about is that if you were going to do this, I would definitely recommend some kind of structure, stratified sampling approach, where they don't, you know, your panel doesn't have to recreate every slide that Dr. Wolf has looked at.

It's clear that he's looking at certain things on certain slides, and so you get representatives of each of those things. And you just see how consistent your panels are. And that would give you your inter and intra observer variability, and that should give you a feeling for how reliable the overall study is.

So, my recommendation would be to not recreate the whole thing, but actually, be a little smart and design a study that gets at resources of variation to figure out where you would have, expect variation among your experts, and where you would expect them to be pretty consistent.

DR. HEERINGA: Dr. Yeater.

DR. YEATER: Kathy Yeater. Just a further comment about getting a panel to read and a stratification setup, also, that would allow for more randomization to occur in the reading of the slides.

Page 165

As was noted in the Syngenta report, Dr. Wolf was unable to truly randomize the reading of the slides by treatment group, because he was getting them as they were coming at him out of the lab, and he just had to go through and start working through reading all those slides. So, I would think also that having a panel and some sort of stratification design would help in that randomization, so that you wouldn't have any sort of trend problems.

DR. HEERINGA: Additional comments from the panel. Maybe for Dr. Yeater and Dr. Portier, a question comes to my mind. In recommending a replication or, essentially, an inter rater reliability test, even on a sample basis, the objective here would be, sort of, quality assurance, quality measurement, quality of assessment.

Not something that would, ultimately, then, proceed to some sort of measurement error or adjustment in the data. It would just be an additional step in quality assurance.

I heard a lot of confidence in Dr. Wolf's reading and quality of the preparation, but we had one assessment, so, no statistician could sit here and say we have est-, I mean, could be highly accurate and zero variance, or it could be zero variance and not so

Page 166

accurate.
We don't know, but I think the general consensus is that it was probably done well, but in terms of external validity, at this point, there would be benefit in, from a Q-A perspective, and possibly, at least a sample base. I don't want to overstate things, but is that what you're suggesting? Additional comments, Ken.
DR. PORTIER: I was going to say, this is an, this is different than what Dr. Miller was saying about averaging the results from multiple experts. I mean, if the goal is really to try to get the, get to the truth for each slide, then certainly, you're going to have to get more than one expert, look at each slide, and average the results to get the right, closer to the true value of the scale, right.
But I'm, I think the feeling is, what you probably need is some kind of assessment of the overall quality of the pathology work that's done. Something that would support your vague feeling that he did a really good job. You might have something that would back that up that would be short of redoing his study all over again.
I mean, it's got to be months of work to re-read all those slides, and a lot more work for somebody

Page 167

to go in and put them all together, and average them all, and I don't get the feeling that the cost of that's going to match the benefit of doing that study. But a smaller study that I'm talking about could be done pretty quick by even a larger panel.
That way, you could have, although it doesn't sound like there's a lot of experts out there you could draw on your panels. You can get all the experts in the world to look at a small set of slides.
DR. HEERINGA: Dr. Miller.
DR. MILLER: Yeah, because I, you know, I agree that Dr. Wolf's assessment is, you know, fantastic. And, what you're suggesting is to take a subset, just to, just kind of do a generalization, a quality assurance type of thing, and that would definitely be doable.
DR. HEERINGA: Additional comments on this particular question from the panel. Yes.
MR. PAULY: Bruce Pauly, Environment Canada. I just wonder, just to clarify, would we follow the approach of Dr. Yeater, maybe, or would it be recommended that an approach is followed where, by sent, slides were selected by some mechanism that would, then, be the subset that would be given to all the world experts in amphibian gonadal histology for

Page 168

assessment of the quality?
DR. PORTIER: I think we were kind of saying the same thing. I mean, the hard part would be stratifying the slides into typical classes that you could, then, randomly select.
And the thing is, you could randomize the presentation of these slides to each of the panel members so that they're not seeing a sequence that might lead you into getting into a pattern of looking for certain things and not being challenged the right way to get, and that was, I think, Dr. Yeater's idea is that, we're going, we can do more randomization within a broader stratification. The stratification makes sure you get coverage of all the conditions that he saw, right.
So, the panel is to challenge, with all of the different kinds of things, but they don't have to be challenged a hundred times on each of those things. And we may only need three or four to ascertain that they can rely, they pick up that item.
DR. HEERINGA: Ask a question of Dr. Miller and Dr. Green. SAP's had a habit of, sort of, recommending additional work. I think this clearly has a benefit to the ultimate integrity and evaluation of the study, in terms of quality assessment. My sense,

Page 169

naively, is that we don't want to just sample slides. We want to sample organisms, and then have the review on all slides for that organism. Because the ultimate determination is whether the organism, a frog in this case, has a particular, in other words, we're not analyzing at the slide level. So, I assume, all of these different sections that are taking, that there would be multiple for a frog, or am I wrong here?
DR. MILLER: If I'm understanding you correctly, no, basically for this part, because it's based on the histology. And so, we are looking at the slides. Now, to be able to review, you know, the sticky part is, slow as it would comes in, is that, in order to get an assessment, a true assessment, you've got to do it on that same, you've got to pick certain slides, and everybody has to look at that same slide. Does that answer your question?
DR. HEERINGA: I believe so, and you're probably right. I'm confusing everybody in good shape.
DR. GREEN: I think, though, you would look at the same set of slides from a group of randomly selected animals on the study.
DR. HEERINGA: Exactly, so you look at, for a randomly selected set of animals, you'd look at all the sections. That's what I'm getting at. Not

Page 170

just a random selection of slides across animals, because - -

DR. GREEN: A statistician could choose, but you know, in the studies I've participated in, when we've done subsets, so that we didn't have to repeat work on hundreds and hundreds of animals, it may be as few as ten animals. But they might have had twenty slides, but we would each review them independently.

DR. HEERINGA: And you would review all slides for each - -

DR. GREEN: Three times.

DR. HEERINGA: Thank you. Additional comments on this particular question. Let's, I see Dr. Bailey smiling. And I know that we're moving ahead a little faster on these charge questions, and I think even I anticipated, and certainly some members of the panel, but what I'd like to offer here is that, I'd like to continue, because I think we're making progress.

But I'll also mention to the panel members, and to the EPA staff, that we will revisit these later on, potentially, tomorrow morning, so that there's an opportunity to go back to any question in which the presenters either haven't had a full time to prepare, or would like to, sort of, extend their comments on a

Page 171

particular question. So, we'll proceed, but also, allow the possibility of revisiting those questions just in case. We don't want to race through them. We want to cover them thoroughly, so, but let's move ahead at this point, Dr. Irene, with question number nine.

DR. IRENE: After an evaluation of the laboratory based studies submitted in response to the DCI, the Agency has concluded that these studies do not provide sufficient evidence to support the hypothesis that Atrazine causes adverse gonadal developmental effects in amphibians. In light of the responses to questions three through eight, please comment on whether the results from the study in response to the DCI are sufficiently robust to address the hypothesis that Atrazine exposure causes gonadal abnormalities in Xenopus laevis. If the SAP concludes that these results are not sufficiently robust, what recommendations can the SAP provide to efficiently and reasonably address remaining uncertainties. For example, if the SAP does not believe that the DCI study is sufficiently robust to assess the hypothesis, does the SAP believe either of the two experiments, or a specific component of the two experiments should be re-analyzed or repeated. Please provide the rationale for recommending any additional analyses and or

Page 172

experiments. Question B.

DR. HEERINGA: Dr. Irene, I wonder, let's stop at point.

DR. IRENE: Stop with that, okay.

DR. HEERINGA: I really think question nine is two questions.

DR. IRENE: Okay.

DR. HEERINGA: So, let's take up 9.A. first, and Dr. Delorme is the lead discussant on 9.A.

DR. DELORME: Okay, first off, I found the question a little bit confusing, because the first part talks about supporting a hypothesis that Atrazine causes adverse gonadal developmental effects in amphibians, which is general; but question A. and B. specify the DCI study. I believe that I do have some concerns about the general question, but I think I'll wait and address those in question number twelve later on. And just concentrate on the specific hypothesis that Atrazine exposure causes gonadal abnormalities in Xenopus laevis, which is identified in sub-question A.

And with respect to that particular hypothesis, I think what we've heard this afternoon is that we have a well-characterized exposure. There was some minor concern about whether or not transformation products should be included or not, but I think there

Page 173

are other avenues to pursue on that.

We seem to have reasonably well-characterized effects, on the apical endpoints are particularly well-characterized. There is some question, I believe, with respect to the histology, and maybe in my answer or in our answer, we can refer back to the questions, the results of question eight and what comes of that. And the statistical analysis seems to have been very good. The study was well designed, and you know, I think that, in general, we could say that, yes, you know, we can address the hypothesis in a robust manner with the data that's been presented from the DCI.

Again, there was a number of minor concerns that I had whether or not the specific strain used is appropriate, in light of the data presented by Syngenta related to relative sensitivity for Xenopus laevis. I think that's, that's the question. But it's going to have uncertainty. But there's nothing you can do about that with respect to re-analyzing.

And, I noticed that this question targets the gonadal development. There was, there did seem to be some effects on growth, but whether or not that was biological or be relevant or not, I'm not sure. But I think that needs to be looked at as well.

And, since, again, I wasn't really expecting

Page 174

to get to question nine today, I'm going to pass this on to associates.

DR. HEERINGA: Dr. Denver.

DR. DENVER: So, I first want to say that I concur, and that these studies were very well designed and controlled and robust. And they, they certainly addressed the hypothesis. If the question is whether these studies fully tested the hypothesis sufficient to allow one to reject it, that is, that one can now accept the null hypothesis, I think that the answer is no from the perspective of the scientific method.

These studies provide mostly negative results. And, as I mentioned earlier in the day, and was discussed earlier in the day, the flow-through design allows one much greater control over dosing, water quality, and maintains the health of animals much better than the static renewal. But, it does not follow that the flow-through design better mimics the characteristics of exposure that's encountered in nature. And I'm not saying that the static renewal, necessarily, better reflects that, but strictly speaking, it addresses the hypothesis, but I, personally, I do not think it fully tests the hypothesis such that it can be rejected.

Page 175

DR. HEERINGA: Dr. Leblanc.

DR. LEBLANC: Jerry Leblanc. I recognize that it's not the place of an SAP member to modify the charge questions, but I really have to modify the question before I answer it, so that, so my question, my answers are relevant. The question posed to us were, are results from the DCI sufficiently robust to address the hypothesis that Atrazine exposure causes gonadal abnormalities in Xenopus. And I simply have to qualify my answers in stating that, at exposure concentrations as high as 100 micrograms per liter, that I, or I don't think anyone else here, can make any comments about what might be happening at higher concentrations.

That said, if we evaluate, revisit some of the major considerations in testing the hypothesis, or the conclusions, or addressing how the conclusions were reached, I feel that the studies were sufficiently robust. And what I'm referring to, specifically, is if we look at the strength of the concentration response relationships as being one criteria, there was no evidence for concentration response relationships, with regards to the major apical endpoints, sex ratios, intersex gonads, et cetera.

There were, there was evidence for

Page 176

concentration response relationships, however, for some of the secondary endpoints, as we discussed previously. Another criteria, the strength of the cause effect relationship. To me, there are two things that go into answering that assessment. One is the strength of the observations themselves, and in this case, the significance, the statistical significance that was observed, but the effects tended to be quite modest, quite low. The other consideration here is, when effects were observed, were they reproducible between the two labs that conducted the studies, and typically, they were not. So, my conclusion of that is that strength of cause effect relationship, based upon these studies, was weak.

Mechanistic plausibility, I think this question's addressed in more detail in one of the subsequent charge questions, but based on the discussions that I've heard thus far, and based upon my own readings, I see no mechanistic plausibility. We've heard hypotheses that would certainly be aromatase hypothesis, I think, has been evaluated rarely carefully, and I don't think there's any plausibility for the reported actions of Atrazine. And while it's, perhaps, dangerous to speculate, I've got to assume that proponents of the aromatase theory have tried

Page 177

wholeheartedly over the past several years to use aromatase in frogs, and I've yet to see any demonstration of that if it occurs.

And lastly, ecological relevance of the effects, and as I mentioned, effects were observed, subtle effects. But again, as we've discussed in some of the previous charge questions, the ecological relevancy of these effects are unknown.

DR. HEERINGA: Dr. Furlow.

DR. FURLOW: So, I'm left with the idea that, strictly speaking, under these defined conditions with the strain of Xenopus that there does not appear to be the effect of Atrazine on gonadal development. That said, proving a negative is hard. And we have two studies here that were conducted extremely well. The animals were, obviously, quite healthy, quite robust, and when I see the data on the pigmentation changes, or some partial changes on growth that may or may not be consistent between the laboratories, it leaves one wondering if the health of the animals, when they're exposed to Atrazine, may change their sensitivity to the compound. But that's purely speculative on my point, on my part. So, other than the concerns I've voiced before, I, generally, agree with what's been said. But I do, I guess, also, side with Dr. Denver

Page 178

about the fact that these are two studies. So, I'll just leave it there, okay.

DR. HEERINGA: Very good. Thank you, Dr. Furlow. Additional comments. I want to, as Chair, I'm going to make the decision, I believe, to call proceedings for today. Because we have made very good progress and had, I think, excellent coverage of each of these topics. And discussion here has been fine, and people have done a wonderful job of responding on, sort of, short notice to the questions they anticipated later tomorrow afternoon or evening. So, what I would like to do is to call the proceedings for today, and ask everybody to return tomorrow morning at 8:30. We will have, I'll give you, Dr. Steeger, here a chance to respond, too, but I'd like to, Dr. Petino.

DR. PETINO: Reynoldo Petino. Are we going to come back to this question tomorrow, or are we - -

DR. HEERINGA: Yes, we are. I think we can revisit question nine tomorrow morning. We'll start again, and I'll return to Dr. Delorme and the others who have just responded to see if there are any additional comments, but I don't want to get out ahead of our preparation. And this is a cumulative process of learning and preparing during the course of the

Page 179

multi-day meeting. And I think, in all fairness to the responses to the question and to the panelists, I think that we're well ahead of schedule, and I think we're certainly positioned to finish our deliberations tomorrow with a good day of work. And, so, Dr. Steeger, unless you object to that, that will be my proposal.

DR. STEEGER: No, I do not object. I just have one question regarding the response to question nine, because it keeps coming. And that is, regarding the sensitivity of the strain of Xenopus that was used, the animal was collected at the Cape region, which, apparently based on genetic analysis, is less likely to exhibit testicular oocytes or mixed sex as a natural background in the population, as opposed to animals that were collected in the northeast of South Africa.

The study that was done in response to the DCI, both studies, ran a positive estrodyle control, and the animals were responsive to that, indicating that, in an exposure to a chemical that is known to induce gonadal developmental effects, it actually does respond as expected. Am I to understands of that, that test, the positive control, is not indicative that the animal, the strain being used, is a suitable surrogate

Page 180

for amphibians. You don't need to respond to that. Tomorrow would be fine.

DR. HEERINGA: If you would be willing to remind us of that question - -

DR. STEEGER: No problem.

DR. HEERINGA: - - tomorrow morning. Sounds good, just to add that to it. At this point, then, I think that I want to thank everybody for a very productive day, so a shorter day, but I think we deserve a bit of a break here, but we'll return tomorrow. And, again, my anticipation is that we would complete our deliberations on the charge questions and wrap up tomorrow. And that the panel would probably have a writing session Friday morning. But just to, again, that's based on experience. We won't shortcut the discussion of any of the questions, but I think with the remaining five questions here, that we should be able to handle those quite nicely in a full day tomorrow. So, turn to Joe Bailey, the designated Federal official, and see if he has any additional comments if he's willing to close.

DR. BAILEY: No additional comments. I just want to thank the panel and the Agency for their discussions today. Thanks very much.

DR. HEERINGA: Okay, and thank you

Page 181

everyone in the audience for your participation today, too, and all of the other commenters or presenters, the EPA scientific staff. We'll plan to reconvene tomorrow morning at 8:30 here, and we'll start again with some initial comments from the EPA scientific team on anything that came up overnight. And then, also, return again to question number 9.A., just to make sure that the discussants want to stay with their initial comments or make additional comments. And then we'll proceed on to 9.B., 10, and so on. Have a good afternoon everyone. Could the panel meet briefly next door just to discuss the progress here and writing work on the questions that we have covered.

(WHEREUPON, the session was concluded at 2:42 p.m.)

CAPTION

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

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

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

CERTIFICATE OF REPORTER

COMMONWEALTH OF VIRGINIA

AT LARGE:

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

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

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

MARK REIF, COURT REPORTER / NOTARY

SUBMITTED ON OCTOBER 10, 2007

D

F

J

N

T