DR. HOLMES: You know my general attitude of being in favor of it.

Let me just give one example. If you take the common seizure meds like phenytoin, phenobarb, carbamazepine, they're probably a reasonable example of something that a registry would address. They are associated usually in most studies with roughly a twofold increase in the frequency of major malformations.

Well, you've heard people already today citing a figure, an historical figure. Well, the historical figure we have from our work with newborns at Brigham and Women's is that the baseline rate is 2 to 2.5 percent. Others use a figure if 3, 3.5, 4. So, just picking that one example, if you're trying to show 4 percent is a high number and you pick a baseline of 3 or 3.5, you'll never show it. Whereas, if you've generated your controls from a system that would truly show it is 2 percent, then you've got a chance to show a doubling in the usual kind of sample size you need for that. But you won't know that it's really a low number unless you have concurrent controls to prove it.

This 3 percent comes out of the ceiling tiles. There's no study of newborn infants that's going to show it's really 3 percent I don't think. Maybe there's a reference somebody knows. I'd love to hear it.

DR. GREENE: I'm sorry. What that 3 percent number is --

DR. HOLMES: I don't think, Mike, there's a study of newborn infants that shows a prevalence rate of 3 percent with the kinds of definitions we're using in our registry and I would bet most people use in their registries. I think the numbers are generally going to be lower than 3 percent.

DR. GREENE: Yes, I agree with you. Certainly the Collaborative Perinatal Project, which was a very large project obviously with a huge number of infants, had a lower number than that.

DR. MILLS: That's obviously a very difficult question, and I think there are several areas that need to be considered. One is, if you select controls, are you really getting a good match? Are you getting people that do what a control should do in terms of being similar to the cases in -- are they exposed in the ways that you need them to be similar. You always have to ask yourself in that context are there major biases in the people who agree to participate in these things. As we've discussed, there are certainly some studies where participation demands a lot of commitment where the only people likely to do it are the people who have their own concerns or their own problems. So, as a general rule, you may find that the

easier it is to be a control, the better chance you have of getting people who actually represent the population you want to represent.

The second part of that question is, obviously, are there alternatives? One of the ways that this can be addressed, sometimes more satisfactorily than others, is to look in a birth defects situation at the people that are exposed during organogenesis who therefore are at risk versus the people that you identify who come in exposed after organogenesis. This has the attraction that you're getting both your populations from the same general source and they're people with the same general thoughts about participating in the studies. Whereas, in the first group, you know that they have the potential exposure and in the second group you know that they do not, so that they, in effect, do form a control group, so that that in some instances can be a satisfactory response to that problem.

DR. GREENE: The question I would further ask then is, is the bar that we want the registries to clear eliminating a twofold increase in all congenital malformations or eliminating a 10-fold increase in a specific malformation like valproic acid neural tube defects?

DR. MITCHELL: Well, at the risk of giving away my talk, that's really the issue. I think the way I like

to think about it is -- and to put sort of names to your numbers -- is the first task, which I think everyone would probably agree, to be sure they're not thalidomides and Accutanes. That has its own set of questions in terms of sample size. It has its own set of questions in terms of the need for controls. I think most of us would agree that a sample of 50 exposed moms for thalidomide or Accutane is sufficient to identify the problem and you probably don't need controls other than historical controls because the risks are so high.

When you get into valproic acid, as Lew points out, there are an awful lot of other issues you need to take into account, and valproic acid is a good example in terms of a high rate for a specific kind of outcome.

As you get further out to the right side of the curve that I'll show -- and it's probably analogous to Elizabeth's curve as well -- then the issues of confounding and bias become paramount. If you're talking about twofold increased risks for rare defects for infrequently used drugs, apart from the issue of whether it's even feasible to try to get that information, then the issues of controls become absolutely critical.

I think Jim's suggestion is a very useful one, but it's what we all struggle with. Thalidomide and Accutane is a no-brainer in terms of the need for controls

and the concerns about bias, by and large. On the other hand, if someone were to come up with an allegation that an anticonvulsant was associated with a doubling in the risk of neural tube defects, we'd be very concerned about what they were using for controls. So, I don't think there's a single answer.

DR. WISNER: I think whenever there is a discussion of controls, the first question I always ask myself is, control for what? Because, for example, in some studies that I'm doing where we're looking at antidepressants in pregnancy, one control group might be a normal population who's unexposed to just look at malformation risks. But in fact in a study we're doing, we're comparing the antidepressant treated to depressed non-antidepressant treated patients and a normal control group because for drugs you almost always have an indication. So, you have the underlying disorder that could be an additional exposure. So, the question about contemporary controls for me really means what's the question specific to the study and what kind of control best answers the specific question at hand.

DR. GREENE: Yes, please.

DR. ANDREWS: I'd like to make a couple of comments on controls. I think one of the key questions to ask is are you comparing data across methods or databases

where the ascertainment method has been the same. That's really the point that Lew was making earlier.

There was a very useful study that was performed by the CDC that was in the package in which the ascertainment from registries was compared against the Metropolitan Atlanta Birth Defects Program. They categorized birth defects as external birth defects easily identifiable at birth and internal defects. And we compared some of the registries that we've participated in against those external defects identified at birth. There was very good concordance. The methods were fairly similar, although we certainly have under-ascertainment of those categorized as internal that were less obvious.

A couple of strategies that we've looked at, in terms of comparison groups using similar methods, is in the antiretroviral registry to compare exposures that were in the first trimester to exposures that began in the second or third trimester. So, that gets to the point that was raised earlier. It's not always easy to do because not many drugs are used fairly chronically.

Another thing that we've begun to do is to look across our different registries, and we've been amazed at how similar our results are across drugs that we've studied using the same method.

DR. GREENE: Yes.

DR. FRIEDMAN: I like the point that Allen made that it depends on what you're looking for. It seems to me that there's another group, though, that we haven't talked about. At one extreme there's the Accutanes and the thalidomides, where if you look at 50 or 100 kids, you're going to see it and you probably don't need controls. At the other extreme there may be a lithium where you look at 1,000 kids and you still may not see it because the frequency is so rare. The issue of valproic acid and neural tube defects sort of is in between, and clearly controls are very important there.

The other place that controls are very important is if you're looking for a syndrome. Could the acyclovir registry have picked up a fetal alcohol syndrome? I think not. And the reason is that you need to have a control for minor anomalies and for subtle patterns of anomalies. The expectation is probably zero in unexposed that you'll see them, but the identification is subjective and you need a control to deal with the subjective identification not just the frequency of the defect. The frequency may be high enough so that you wouldn't have any trouble seeing it if you had a sensitive enough assay.

DR. WEISS: I think another thing we were talking about left and right truncation earlier, and if the defect is identified during pregnancy and there's a

therapeutic abortion, you'll never see it. And if there's an increase in therapeutic abortions in the group that are exposed, you'll never know that without a comparison group. That is very specific to populations and areas of the country. You can't just take a number from the literature.

The same with other outcomes that we're interested in besides birth defects such as spontaneous fetal loss, which is an important outcome I think we should think about.

So, I don't see why there would be a question that scientifically or methodologically do you really need a control group to answer the kind of questions that we would like to see answered.

I think the big issue is with the method that's been done in the past with the typical surveillance based registry that it may not be doable in an industry setting because of the constraints on them. I don't know if there's an argument there that it should be done.

DR. GREENE: Other thoughts about this issue. Please.

MS. CHAMBERS: I think it's also important to think about the issue of lost to follow-up. Whatever the rate of lost to follow-up is, you're interested in how those people might differ from the people who completed follow-up. And it's important in registries I think, or

any other type of prospective study, to have a control group that generates a lost to follow-up as well, and if you use population resources, you don't have that.

DR. GREENE: Further discussion of this issue before I move on to my next question? Yes, please.

DR. MONTELLA: I think, as you listen, this is just screaming for collaboration between industry and clinical centers because what's going on is we're trying to figure out a reason why you don't need controls because they're hard to get, and that's not a reason not to use controls. So, if we establish that the best way to do this scientifically is to have controls, then the real issue at hand is how you can do that and how you can collaborate to do that. I think what you need is collaboration between industry and clinical centers to do that.

DR. GREENE: The next issue that I'd like to raise dovetails with this in a way. What we've spoken about so far is how big an increase in risk is it reasonable to need to detect.

The other question that logically flows is how big an increase in risk of what. Much of what we've addressed is risk of major malformations which are generally determinable within a short period of birth.

The next question, of course, then comes how about other things that are more subtle that don't show up

until somewhat later. Short-term follow-up would be 6 or 7 years when you can do a valid IQ study, but there are other issues, obviously, that don't show up until even later than that.

One of the great sayings is that the advantage of experience is that it helps you to recognize your mistakes when you make them again. It may be that we need to acknowledge that there is just no way to detect the next DES, for example, in which case let's just say it up front and acknowledge it and move on.

So, beyond control groups, the next question is, how long is a reasonable follow-up period and how subtle an abnormality do we want to -- again, where's the bar?

DR. MITCHELL: I'm always happy to speak to that. Again, I would translate your "what's reasonable" to what's feasible. Again, to give away -- I can just not give the talk.

(Laughter.)

DR. MITCHELL: But we cannot delude ourselves into thinking that we're going to resolve uncertainty in this process. There will always be uncertainty, and the uncertainty is going to be directly proportional to the statistical power that we have to resolve it. So, a frequently used drug has public health consequences beyond

clinical consequences. For that, we may very well choose as a society to demand more information of that drug, and we might go to the point of 7-year follow-up and IQ, and in certain circumstances we might even go further.

For other drugs that are less commonly used, you can do the math. There are about 4 million pregnancies a year, and I'm going to do some of that. And there's a finite resource in terms of exposed pregnancies, which also has public health implications. If there aren't very many pregnancies, it would seem to be a lesser order of concern than the commonly used drugs.

So, I think ultimately the task will be to create a hierarchy of priorities, and they may differ for different drugs. So, a drug that's used by 15 percent of the pregnant population might demand -- FDA and others might demand of it much more information about risk. We might demand IQ information. We might demand to know what the risks are of relatively rare malformations. Whereas, a drug that's used by perhaps a half of 1 percent of the population, we may say, well, let's at least assure ourselves it's not a thalidomide or Accutane. And if we can do a little better, let's do a little better. But I can't imagine a situation where we can reasonably have the same demands for all drugs.

DR. MONTELLA: I would be really cautious about

that in terms of being consistent about those demands, however, because you can influence the outcome of that just by saying a drug we're not using as much we don't need to look at as hard because it's not being used as much.

Well, alpha methyl dopa is a perfect example of this. It's the most frequently used drug for the treatment of hypertension in pregnancy. It's probably not been used in internal medicine circles for hypertension in 15 years. Yet, it's the only drug -- and it's quoted over and over again -- for which there's data followed out to 6 or 7 years with no neural developmental delay in the offspring and no IQ lessening in the offspring. So, that's the reason the drug is used.

So, if we take that approach, then we're going to study alpha methyl dopa 15 more times for 20 years and we still won't know could we have used a different drug with comparable data. So, I'm worried about that. I'm worried about the effect that's going to have on how we look at it and would argue for some consistency there so that we don't end up making bad choices.

I personally don't think it's a great drug, in case you didn't notice.

(Laughter.)

DR. MITCHELL: But I need to qualify what I was saying off the cuff. It's still a clinical question that

you're raising, and I think that if there's a clinical basis for concern, then that ramps up the priority. And that clinical basis may be that we need to consider drug B as an alternative to drug A and therefore it becomes important for us to collect a lot of information on it. So, I don't mean to suggest that it's only driven by the current clinical use or the projected clinical use. It may be driven by the projected reasonable clinical use.

DR. MONTELLA: And in that same vein, then anytime you choose a category, maybe we need to address that our priority is to choose two or three drugs in a category, as opposed to a single agent for a category.

DR. WISNER: I would like to push the point about exposure and the operational definition of exposure a little further here because I think it's related to a technical question. And that is, right now, when we report outcome information, exposures seem to be linked or seem to be kind of lumped to somebody who took one dose and discontinued it, somebody who perhaps took a moderate dose or the typical dose chronically, or perhaps people that took a high dose. And all those folks are put together as being exposed to the meds and their outcome is here. Depending upon how that population is constructed, if there are a lot of one-dose and discontinued patients, we may actually be deflating the risk for reproductive toxic

outcomes that we're looking out at here.

The technical point is that if we could divide those exposures into strata and then begin to look at them more carefully, we would know, first of all, which populations the reproductive outcome data may be generalizable to.

But the other issue is, in terms of thinking about a hierarchy of how we might spend our resources, I would argue that I would like to know, in that population of exposed patients, the outcomes longer term for those patients with high risk exposure than perhaps a lower risk exposure if we're thinking about hierarchies of expenditures.

DR. GREENE: Ken, you had your hand up a minute ago. Did you want to make a comment?

DR. JONES: I did but my comment has been taken by others. I would like to say I agree with Allen Mitchell completely on this issue.

DR. GREENE: Jim?

DR. LEMONS: This issue of neural developmental outcome and subtle changes I think is really an important one. My reaction is similar to Jan's; that is, I can't imagine that a registry in any fashion could accurately discern subtle effects on neural development no matter when you test them because of the confounders, the lack of

adequate controls. Most people who have done follow-up studies of newborns I think recognize how incredibly difficult it is and the arguments concerning appropriate controls, whether it's siblings, whether it's maternal IQ, which would have to be included, which varies tremendously in populations around the country, socioeconomic factors, whether they breast fed or didn't breast feed, if you believe those data.

So, unless I guess the question of subtle outcome can be tied to clear dysmorphism even with fetal alcohol -- I mean, now that fetal alcohol association has been touted to be linked to learning disability and hyperactivity, that's a very dangerous association to draw. It has a lot of social implications and stigmata attached to it.

My sense is that as difficult as that is and looking at a minimalistic database for this and the incredible lost to follow-up issues and control issues, I would be afraid of even going into that arena with a registry. Plus the bias from an unmasked trial. We know there's tremendous bias, both in the investigators and if you ask subjects to report. There are accurate ways to do that, but there is such inherent bias built into a mother responding, for example, to her child's development versus a blinded or masked expert who's assessing.

But those are very focused, very expensive studies to do and would have to be designed very, very carefully prospectively I would think. So, I can't imagine this type of methodology could be applied effectively to address those issues.

DR. MATTISON: Some of the comments and the draft document talked about two types of outcomes for registry studies. One was hypothesis generating and the other was hypothesis testing. I think thinking about both the need for controls, the characteristics used to define exposure, as well as the kind of risk that the studies would potentially describe are going to be different for those two kinds of approaches. In a sense, you go into a hypothesis testing study with a much more potentially precise set of guidelines. But if we're concerned about identifying maybe less structural but more functional endpoints, then we're probably going to find ourselves in the context of hypothesis generating studies. And the issues of controls are going to be very critical.

Coming back to the issue of biological markers, are we going to require the actual expression of adverse health consequences in these studies, or can we identify steps in the progression of the endpoints that we're looking at or the disease states, whether they're either functional or structural, and ask if we will allow

biomarkers to be surrogates for the actual expression of disease? And then if we do, how does that change the precision with which these kinds of studies can be conducted?

I think actually we ought to spend a little bit more time thinking about not actually requiring the development of disease, but some antecedents and the use of those antecedents in potentially enhancing the sensitivity of studies.

DR. GREENE: Do some of the epidemiologists want to comment on use of proxies for endpoints?

DR. MITCHELL: How can I not take the bait?

I guess I'd have to ask you, Don, what you're thinking about. I have enough difficulty identifying endpoints.

(Laughter.)

DR. MITCHELL: And I think that the discussion here reflects the frustration that we all have in being able to quantify, in a reproducible and structured way and systematic way, endpoints. I mean, good gosh, cleft palate. That ought to be easy, and yet I'm sure there are clinicians here who would differ in the way they would design a study to do that.

So, I guess the question would be, what's the endpoint? Is it removing us one step further from the

outcome of interest? And while it would be attractive, if it existed, I can't think of any.

DR. MATTISON: Yes. I think you're right in that context because it does require a mechanistic understanding, and the issue of how many of the individual endpoints that we might be interested in looking at can we quantify mechanisms and antecedent steps. But for some of them we may be able to.

Just what I'd like to do is raise the possibility that as we begin to look more in a hypothesis generating mode, we may end up having to look at alterations in gene expression or alterations in levels of protein produced in a particular set of tissues and use that as the surrogate for disease. But it does require that we have a fairly good understanding of the disease process itself.

DR. GREENE: Lew?

DR. HOLMES: Just to follow up on the point about how long do you continue the follow-up, if you think of what Don is talking about with the biomarkers, if you think of the question of IQ as an outcome, to me all of these are spinoffs from a registry. If a woman has enrolled in a registry through an informed consent process and you include in that permission to contact her later and she agrees to that, then if you develop a hypothesis about

a biomarker, you could contact folks, and those who said yes could then be enrolled in the study where you would then explore among that subset whatever you wanted to explore.

The same would be true for IQ. If you developed concern that cognitive function was impaired, then folks in some geographic area could contact everybody that enrolled in the registry and ask would you be willing to participate and so forth. So to do those as spinoffs rather than trying to prolong the follow-up of the registry as a whole because the personnel involved in that would be enormous and I think you'd want to focus.

DR. GREENE: Jim?

DR. LEMONS: Just a quick question, Lew. How would you become concerned about the risk of decreased IQ? Are you thinking that some type of animal data or unusual reporting? Because that is a very subtle --

DR. HOLMES: Let me just use my neighbor here as my example.

(Laughter.)

DR. HOLMES: It's not his IQ that I'm worried about.

If Ken did a clinical study in San Diego, of a group of children with a specific exposure and said I don't have a sample size that is big enough to resolve this, but

as an experienced clinician, I see cognitive issues, major anomalies, minor anomalies, whatever pattern he saw in a focused, case-driven study that came along after the registry was established, then you could say, okay, we could address this with less bias in a group of folks who enrolled in a registry prospectively in pregnancy. And Ken would be representing clinicians all over who are going to continue to develop these hypotheses from their own personal experience. So, it would just be that.

DR. GREENE: Jan?

DR. FRIEDMAN: It seems to me that the issue of how far you go, whether you go to these spinoffs, whether you look at IQ, for example, or other behavioral or developmental endpoints, is clearly not something that you do for every single drug that's out there. In the decision of which ones to include, we're sort of suggesting that that group of drugs would generate itself.

It seems to me there might be a more useful way of doing it. If there were sort of two levels of information or two levels not exactly of approval but two levels of drugs with respect to pregnancy -- in other words, if there were a pregnancy formulary where the drugs on that formulary were required to have a higher level of knowledge, those are drugs that you, like the anti-hypertensives, that you might feel comfortable using in

pregnancy. If you know you're going to be using them, if you know you're going to be using them in the third trimester and the second trimester largely, that's the kind of information you'd gather. And if you concentrate on just a few drugs, you can gather enough information for it to be meaningful.

In order for there to be that sort of thing, there has to be some sort of a carrot. There has to be some benefit to gathering those additional kinds of a data and getting into that kind of information. That carrot could be some sort of legal protection for physicians and companies that have drugs that meet that higher level of standard perhaps. I don't know.

But I think that thinking about having two levels, having a pregnancy formulary where we have some drugs that we are more comfortable prescribing during pregnancy and others that have to meet a minimal standard -- we can't allow them at all -- is worth doing.

The other point that I would make about this is these others studies go beyond the registry, so they can't be seen just as part of the registry. But they still need to have a tie back to the label. One of the frustrations I have is that there's often lots of information out there about certain drugs and it never seems to see the label. So, if there are good data, if there are good studies, if

there is good information that suggests there's something in humans, it seems to me that should supplant the animal data that is sort of the stuff that we usually see on the label.

DR. GREENE: This gets naturally into one of the other questions that I wanted to address and that is among the questions that we've been asked to address by the agency. So, before we get into the issue of this information feeding back onto the label that Jan has sort of opened up, are there any other comments on the original question that I asked, which is how high and how far do the registries have to go?

DR. MITCHELL: I just want to offer sort of a general comment on whether there really is a duality in the registries for both hypothesis generation and testing. I think without question they serve a very useful purpose in generating hypotheses. I'm not sure if registries ought to be viewed as the appropriate construct for testing hypotheses for a number of reasons, which we can talk about, but not the least of which is that once a hypothesis is out there, biases become even worse than we had feared in the sort of naive state. I think there are other approaches, case-control studies and perhaps databases, that may be able to respond to hypotheses that are out there.

I think Lew's point is a very valid one, that the existence of a registry can be very helpful in responding to hypotheses that come along the pike. But to set up a registry specifically with the notion in mind that we're doing it to test this particular hypothesis is probably not the most efficient way to go.

DR. GREENE: Let's address then the issue of labeling and information getting back on to labels. That again was a theme that came through in the industry responses to the draft guidelines. Specifically with respect to reassuring information, negative data, how should this get back onto the labels, and should it or shouldn't it? Comments about that.

Then it was also brought up this morning with respect to the acyclovir registry in that the data did come back and change the label and change the grade according to the classification.

Thoughts about that issue. Jan, it seems to me that your group regularly reviews information to rewrite your database.

DR. FRIEDMAN: Well, I think the answer at the first cut is simple. The labels should reflect all of the information that's out there and not just the animal studies. Clearly when both animal studies and good human epidemiological data are available, I would leave the

animal studies off and put the human epidemiological study on there because it's more relevant to the clinicians.

DR. GREENE: Other thoughts. Yes, Jim.

DR. MILLS: This is another question that comes up all the time and you'll be asked, I'm sure, or have been asked by the press or by lawyers, how do you know that this drug doesn't cause birth defects or this birth defect? Of course, I'm just stating things that we probably all know but need to go on the record, and that is that it's difficult to rule out a drug being a cause of birth defects globally with the sample sizes that we often have available, and it's extremely difficult and often impossible to rule out a specific birth defect, which of course is the real issue in most of these cases.

I think that in many instances, in terms of the label being helpful, we have to put down what we do know. In other words, this has been the subject of a number of studies, the rates in the exposed were in the same general range as we expect to see in the general population.

I also think sometimes it can be helpful to note that if a study is large enough, that the spectrum of birth defects that are seen are the garden variety birth defects in the garden variety distribution because sometimes that's also helpful to say that you did not see 7 cases of a particular birth defect, but you saw the heart

defect and the neural tube defect and the cleft and all the things that you would expect to see in a general population.

DR. GREENE: Yes, please.

MS. CONOVER: Just to follow up on what Dr. Mills said, one of the things, sort of the S word, for teratogen information services is that we never use the word "safe." When I talk to clinicians, that's the very word they want to hear, and they want to know is this agent safe. Give me a list of safe agents. So, it's extremely difficult to handle and complicated, but we don't use it. We use the words "low risk" or something of that sort.

But I think that one of the sad things in being a teratogen information service is that we don't use the labels and we don't use them because they aren't very helpful for the kinds of questions we're being asked in terms of risk assessment or choice of agents during pregnancy. So, it is telling you that the information that is out there -- and we certainly have lots of other pieces of information we use in counseling, providers who are prescribing or counseling patients. The information is there. It doesn't make it onto the label.

So, it's sort of like we have our little secret in the teratogen information services. You know, we know the secret answer about this information, but it's not

being disseminated to the greater group because it doesn't show up on the label. So, the question is how to mechanize feeding that information back in so that it does show up there.

DR. GREENE: Yes, please.

DR. WEISS: Thanks. I think one of the problems as kind of a traditionalist here that I have with putting, you know, we saw a 3 or 5 or 6 percent rate of adverse events and that compares with the population, so therefore there's no problem, or 100 people had pregnancies and there were 2 adverse outcomes, I think it leads to a false sense of security and also false sense of knowledge and tells people that we think we know more than we do. I think it's interpreted wrong.

I think Dr. Kweder had a good point when she put her arms out and said margins of safety. I think maybe we need to train clinicians more to look at confidence intervals and maybe think about putting those on the label where we say we can never rule out a risk. However, given the data, it's no higher than 3 or 4 times, and while we don't want it to be 3 or 4 times and we're pretty sure it isn't that high, that's as good as we can do with the data that we have. With better and more data, you could continually narrow that window. I'd like to see us go towards something like that.

DR. GREENE: Some years ago, Abby Lipman Hand wrote an editorial in JAMA, the title of which was something like "If Nothing Went Wrong, Is Everything All Right?" It deals with the issue of how do you deal with a zero numerator. If you have whatever the number of observations is, but the numerator is zero, how do you convey to people what an estimate of risk that really means?

I think that one of the senses of the committee that we could certainly convey to Dr. Kweder and her colleagues is that we would like some statement to become routine in labeling as to what does this observation mean in terms of the maximum risk estimate, not simply we saw 3. But what does that mean in terms of the maximum level of risk given the denominator, given the number of observations made? Presumably if that became a standard part of the labeling, eventually clinicians would become educated to its usefulness and meaning.

Other comments? Allen?

DR. MITCHELL: Yes, but that's the easy part. That's just math.

How do you compare the validity of two different studies that may find similar or different things? While of course the studies we do are perfect, everyone else's are quite poor of course. It's a real

conundrum.

I think a reasonable example of that is diazepam in oral clefts. We've done studies that have shown no meaningful elevation in risk. Others who have done very nice studies have found perhaps some evidence of concern. How is that reflected in a quantitative way?

What I would argue is that somewhere along the line, the agency, or some group anointed by the agency, needs to be able to make a qualitative as well as a quantitative statement. It may be something that says, well, for this drug there's conflicting data, but given the level of conflicts, the upper magnitude of risk, if there were a risk, may be this high. So, it's a modification of what you're proposing but one that somehow takes into account a validity assessment, which is to me the most contentious difficulty because it's very hard to reach agreement even among so-called experts.

DR. GREENE: Dr. Kweder.

DR. KWEDER: I'd just like to comment that we have taken the approach that Dr. Mitchell just described. We're simply putting in labels what's out there and acknowledging that the data do conflict in other parts of the labeling, not specifically for pregnancy instances that I can think of, although there may be some, but for other adverse events. I've seen it recently for several drugs,

particularly neuropsychiatric where there are studies that are done, surveillance type studies, that show different findings, and we're not sure what that means and put the information out there and try and characterize where there's controversy as best we can so that at least their prescriber has the information and can made a judgment about how much risk or uncertainty they're willing to accept.

DR. GREENE: Jan, you have some experience with an expert panel reviewing data and sort of adjudicating differences of observations. Do you care to share that with the FDA?

DR. FRIEDMAN: Well, I think the bottom line is that there isn't an easy way to do this, but you have to face it. In other words, you're never going to be able to do it with a computer. You're not going to be able to do it with a policy that says every case has to be handled like this and these are the rules and this is how it works. It's going to take some expertise. It's going to take people who know what they're doing and have experience in the area, and you're going to have to treat these decisions on a case-by-case basis.

But you've got to have the information. You have to take into account all the information that's out there, and someone has to take the responsibility of

looking for it and getting it and putting it together and actually reading it and thinking about it and making these decisions. So, there isn't going to be an easy way or an automatic way to do this.

DR. GREENE: Jim.

DR. MILLS: I was just thinking about sort of a case example of this, and that is intrasitus cytoplasmic sperm injection where there was a controversy recently about a study in Belgium where they followed a large group of couples undergoing this procedure. They reported that there was not a significant increase in birth defect rates in the offspring produced by the procedure. This was published.

Then a group from western Australia reclassified the birth defects the Belgians had reported, compared the rates to their own registry rates, in western Australia and came to exactly the opposite conclusion, that there was, in fact, an increased risk of birth defects.

Then the Belgians wrote a response and said, well, if you take out cases such as atrial septal defects that were diagnosed ultrasonographically, basically en passant, and had no clinical importance whatsoever, then you would discover that we were right in the first place and that there wasn't any increase in the birth defects rate.

I think this is useful for illustrative purposes because, first of all, it shows that just a rate is not terribly useful, even with confidence intervals, that you have to have some idea of what's going in to determining that rate.

Secondly, it shows that even if you're going to use controls, that you better have concurrently evaluated controls, evaluated using the same procedures, because had it not been for those ultrasounds -- and I never did figure out why they did the ultrasounds in those particular instances -- they would never have found those atrial septal defects that seemed to have caused a good deal of the confusion in those studies.

So, it can be very, very tricky, even if you have a control group and you can calculate a rate, to know what's going on unless you have it very, very clear in your mind what your malformations are and how you're going to look for them and you look for them consistently throughout the entire study population, exposed and controls.

DR. GREENE: I suppose the cynic would view that experience and ask if there's anything quite as suspect as the advice of experts.

(Laughter.)

DR. GREENE: Yes, please.

DR. ANDREWS: A couple of other comments. I

think this is a very tricky issue and I don't think there's any way of getting around judgment. There's no cookbook way of presenting the information.

One of the things that's been the most useful in all of the registries that we've been involved with is pulling together an advisory panel that has helped advise, steer, evaluate the methodology, and more importantly evaluate the data, the specific cases, the bulk of the data, along with everything else we know about the drug in pregnancy, to help come up with what we call our committee consensus of what the data mean. That has been an enormously helpful process.

So, when we did get some information in the acyclovir label, it was couched very carefully, and I think it probably is appropriate to have some description from some basic registries that talks about the scope of the study, the size of the study, and puts the observed rate of birth defects with the confidence interval with a notion of what the baseline comparison rate was in a population with similar monitoring.

I think where the tricky issue comes in is trying to present the upper bounds of detectability of specific birth defects. I mentioned in my talk that with our acyclovir registry, we felt that we had the statistical power to detect a 7-fold increase in the risk of a birth

defect that has a 1 in 1,000 baseline risk. That's a lot of information to convey to clinicians and there are all kinds of permutations of that that aren't covered in that statement. It's not an assuring statement. It raises a lot of alarms. It takes a lot of education for people to understand what that means. So, I think I would be very cautious about what kind of standard we move toward in trying to put that in a standard way on labels.

DR. GREENE: Other comments?

(No response.)

DR. GREENE: I'd like to ask the next question or make a statement maybe, and that is, as an obstetrician I was struck by the desire to circumscribe really these registries to the notion of fetal and neonatal effects. There was no specific address of potential maternal toxicity. We know that the maternal liver, and possibly to a lesser extent kidney, is uniquely sensitive to toxins during pregnancy. We don't know exactly why that is, but we learned that lesson the hard way certainly in the 1960s with high doses of tetracyclines when they were one of the few acts in town available to treat severe pyelonephritis.

Without looking specifically for evidence of maternal toxicity, are we going to not notice the next pregnancy troglitazone, if you will? I worry about maternal toxicity, and that wasn't really addressed in any

of the registry information. Thoughts about that. There aren't any other obstetricians here.

DR. MONTELLA: I'm not claiming that I'm an obstetrician, but I'll claim that I'm a clinician.

I completely agree with you, that we really have to look at maternal toxicity as well. It's very, very important. When you see it most importantly is when the mother is used as a vehicle for treatment of a fetal arrhythmia, for example, and what level of toxicity are you going to tolerate in a mother to use her as a vehicle in that way. And yet, you have to treat those arrhythmias in fetuses. So, there's a whole host of that.

I think it comes up most with us with probably INH and liver toxicity, if I had to pick the time it came up the most. But it's there and we do need to address it, and it would be a pity to gather all that data and not find out outcomes in mothers.

DR. GREENE: One of the jobs this afternoon or the opportunities this afternoon is to include a public discussion. In raising the next couple of questions, I would like to invite members of industry to participate in the discussion. Specifically what I'd like to do next is to really address the gauntlet, if you will, that was thrown down earlier today about which drugs need registries. Is it only drugs where there's a suggestion in

animal data that there could be a potential problem for humans? Or, as was pointed out, most human teratogens that have been recognized have been recognized without a priori thoughts that they would be teratogenic in humans.

So, I'd like to invite industry in particular to respond to that issue. What are the criteria for suggesting or requiring that a registry be established for a new drug or of an old drug, for that matter? Please.

DR. ANDREWS: Well, my slide number 3 was points to consider that we find particularly useful because it addresses the issue of is there some background reason because of the class of drugs, the underlying disease, the animal studies, but also the intended population and the extent of the exposure that have to be considered in terms of understanding the possible public health risk as well as understanding feasibility of addressing the question.

DR. GREENE: Well, certainly target population and numbers of individuals that might be expected to take the drug is one set of criteria. That's very different. That's fairly straightforward. It would be, for example, probably not very useful to establish a registry for a new drug that was to be used to treat Alzheimer's disease. But that's fairly straightforward.

That still doesn't get to the central question which is if it is a drug that is to be used in young

people, particularly women in their childbearing years in relatively significant numbers, whatever the drug is, should that be the only criterion or ought there be other criteria to set up a registry? Dr. Teter, would you care to comment?

DR. TETER: I think Dr. Sharrar had his hand up.

DR. GREENE: Oh, okay. Please.

DR. SHARRAR: I'm Bob Sharrar from Merck & Company.

I can tell you the registries that we've established at Merck and why we've established them.

The first registry that we established a number of years ago was for our new Varivax or our varicella vaccine. It is a live attenuated viral vaccine.

We know that natural chickenpox infection that occurs during pregnancy can lead to a syndrome called congenital varicella syndrome. So, the question we asked ourselves is can our vaccine do likewise. Clearly anything that can lead to an infection in the newborn isn't one that we should do.

The other ones that we've established registries so far would be Singulair, Vioxx, Maxalt, and Crixivan. Singulair, Maxalt, and Vioxx are new chemical entities. Consequently, we have a lot to learn about what

impact they have. This isn't just on the newborn. This is also on the mother during pregnancy. So, we do collect information on the mother as well. The Crixivan registry we do in conjunction with the PharmaResearch thing done with Elizabeth Andrews and that group. So, these are new entities.

We have not established any registries for drugs that have been on the market for a long time, and I'm not sure if we ever are going to do that either, to be honest with you. They're difficult to establish and I guess we still have to give more thought to it. But clearly drugs that can affect the newborn, new chemical entities, these are the ones we're interested in evaluating.

DR. GREENE: But at least in the case of Vioxx, it's related closely enough to other medications which are known to be associated with adverse effects if used during pregnancy, it's a small leap.

Other comments.

MS. CONOVER: I can tell you again in the teratogen information service we see really remarkable exposures. I think I'd be more interested in hearing someone from industry describe which ones that are medications that are going on the market now that you would choose not to study really again with the exception of

things that might be topical where you wouldn't get any fetal exposure. For me, the medications are a question mark until shown otherwise, and it is remarkable how women in childbearing age get exposures to agents you wouldn't automatically think would be first-line drugs for them.

So, I guess I'm kind of interested. Have there been new agents that you've chosen not to set up a registry for and why?

DR. SHARRAR: There hasn't been a new drug that we've put on the market at Merck that we have not developed a registry for in the last four or five years. But I'm not saying that that's going to go on forever, but that's currently what we've done.

We have had some ophthalmic products that have come on that we haven't done birth registries for, but I'm not aware of any new product that we really haven't set up a registry for.

DR. GREENE: Other comments. Please.

DR. TETER: I would just mention two small points. One is that we still have in place spontaneous reporting, which means that physicians can call the pharmaceutical companies to report any cases of what they think is an adverse outcome related to pregnancy. That may in some cases be the kind of initial signal that is being detected that would indicate, whether it's to industry, to

sponsors themselves, or also to the FDA who are eventually getting these reports as part of the normal reporting process, that there is a risk and that a particular agent should be looked at.

But we still also feel that our animal data has been done and studies have been done for many years, and they're done to look at that very possibility of could this drug cause a risk in humans. It may not. It may strictly be an animal effect. But that might be the drugs that you would start with and that you would want potentially to set up a registry for, either initial signal in humans when the drug is first marketed or because of the animal findings.

DR. GREENE: I think my colleagues at the table would say we're not worried about the drugs that cause a problem in animals that then are okay in people. What we're worried about is the studies in animals that are negative and yet turn out to be a problem in people.

DR. TETER: Were you thinking of any specific drugs besides thalidomide?

(Laughter.)

DR. GREENE: Well, that's a good start.

DR. TETER: I guess we were testing and maybe Patrick might want to comment on that because we only tested in one species in those days years ago. So, it wasn't picked up.

DR. GREENE: Yes, but in fairness, even after it was tested in multiple species, humans turned out to be uniquely sensitive.

DR. TETER: No, not really.

DR. WIER: I don't think really you brought this up to debate the merits of the preclinical testing, so let's just put that aside.

But one comment I did want to make is that the chairman brought this up as if the question would be answered in absolute terms, and there's really no need to approach the question in that fashion. It's not a matter of absolutely we would not do a registry on a given compound. Unfortunately, it's always a matter of practicality in terms of resource availability.

Where I think we could probably have more fruitful discussion is in terms of what are the characteristics of a compound that would suggest it to have a high priority for a registry and that that's where the resources should be applied, rather than to pretend that other than drugs that are absolutely not absorbed or that would only be used in an age population or the third one I can think of is some questions about ethics of doing a registry in my mind, especially a registry defined to be proactive and prospective if the drug is outright contraindicated in pregnancy because ostensibly some

exposures would be identified at a time when the exposure could potentially continue if you did not contraindicate, which is sort of the antithesis of the purpose of the registry in a way.

So, I think there are those three sort of exclusionary causes that we could probably all agree to. I think beyond that it's a matter of priority setting, and there are so many factors that go into that it's difficult to codify them, but I think the group could begin to prepare a list of the considerations that make them particularly suitable to a registry approach.

DR. ANDREWS: Just another couple of comments to that. I'd echo the comment about feasibility. I think that clearly it would be desirable to know this kind of information for every single medicine that's taken by pregnant women. I think we would all agree we would really love to have that information.

If we're talking about very labor intensive physician or patient intensive follow-up studies, then what's the feasibility of actually doing that kind of study for every single product? So, I think we have to be very selective.

While it's perhaps hard to come up with a list of those drugs you wouldn't study, because we do want the information, I think case examples are instructive. Just

looking through our experiences, the antiretroviral registry was clearly important because we knew that there was a real possibility that retrovir and perhaps other antivirals would be taken and recommended in pregnancy to reduce maternal-fetal transmission.

The antiepileptic drugs -- obviously, we really need to understand the risks of these drugs because some are known to be associated with an increased risk, as is epilepsy. Women have to keep taking medicines for control of their disease, and we really need comparative information. So, it needs a different kind of study and a more carefully designed comparative study.

We're also looking at bupropion which is used for treatment of depression and smoking cessation. I'd have to say we've had not very good success in recruiting exposures when it's given for depression. Smoking cessation I think is a very big issue because it's suggested that women stop smoking when they intend to become pregnant. So, we thought that the likelihood of exposure to this drug in pregnancy would be higher.

With our migraine products, we've had good success with one of the registries and very poor enrollment in another, but these products are vasoactive. We tried to study sumatriptan using a prospective follow-up study enrolling patients through physicians, using consent in a

typical clinical program. That wasn't successful. A registry has proven to be much more successful.

DR. MATTISON: I think the focus on practicality is important, and from that perspective simply exposure to a large number of reproductive age women needs I think some refinement. For example, what are the structural domains that are present in the drug and what do we know about those structural domains? If they do represent classes, for example, of structures that we have no information on in terms of development, then I think it's critical. But if the domains are all those for which we have good experience from other drugs, and more than just class, but I think a real critical structural analysis, it's probably less important.

I think the highest priority ought to be given to those drugs with large exposures which represent new chemical domains for which we can't draw any inference at all.

DR. MONTELLA: One of the more clinically useful ways to look at it may just be categories of use. The AED registry is a good example of that. Because what a clinician or a patient wants to know isn't is this new drug that's put out by Glaxo going to be okay in my patients. What they want to know is if I have a patient who has asthma or a seizure disorder or hypertension, what's going

to be the drug that has the most data on it, what is going to be the drug that feels the safest to me and my patient. So, maybe looking at things in categories of drugs.

Again, that makes us have to cross the board. We have to have collaboration between industry and the rest of us. We have to have collaboration inter-industry in order to accomplish that. But that kind of registry by category seems more clinically useful to me.

DR. FRIEDMAN: Just to follow up on that point, an issue that I know we're going to talk about tomorrow is having a central registry that might collect data on all or a large group of drugs.

I think from a practical point of view, new agents are easier to start with than all agents.

But the issue of choosing ones that are or are not used, if you had a central agency where all were in and somehow had the cost reflect how many cases were actually ascertained of each type, it might be that drugs that turn out not to be used wouldn't cost that much to gather some information on, whereas the ones that turn out to be used, whether you're anticipating it or not, might be a little more expensive to investigate.

There are two other advantages of a central registry. One is that you can compare drugs within a class or within an indication group sort of side by side, and the

other is you can use one as a control of sorts for the others and you deal with this issue of controls.

DR. GREENE: Yes, please.

DR. WEISS: I think this question is really unfair. I'd like to answer the question when should we do a study, when should we gather information, as opposed to when should we do a registry. I think it's very difficult to answer this question. If you're saying this type of study design is one question, if you're saying where do we want information, it's another.

I'd like to go back to what Franz Rosa had said to WHO on where we should collect data on pregnancy exposure, and he said all the things that have been coming up: when there's a condition that's either chronic so that there will be a chance of exposure during pregnancy or if a lot of women are going to be exposed. I think we all agree that we want comparative data, that we want to know which antihypertensive, which AE, antiepileptic drug, which antidepressant is the best. That's a different question. If we have a drug that is a new molecular entity and we don't know anything, we want to do maybe a registry study and follow that specific drug. If it's a drug where there's a suspected risk in animals, there's maybe a different design. If it's a drug that we have some signals in humans, we might want something more intense than a

registry design. I think we have to look at not just the registry but all types of data collection and studies, including even clinical trials, depending on what the question is, what's feasible to do, and the fit between the question and the design.

DR. GREENE: One other question that came up in the responses to the draft guideline was the understanding, it seemed, that it would not be necessary to have any sort of a registry for a known teratogen. I guess the question that I would ask is, is that true and necessary?

Obviously, on the one hand, you don't want to have a registry implying that people ought to continue taking a known teratogen. On the other hand, if there is some method in place for disseminating information about this drug that it shouldn't be taken in pregnancy, should there not be some way of knowing for sure that people aren't taking it during pregnancy?

Allen, you have some experience with that.

DR. MITCHELL: Yes. It's an interesting question because I think I could imagine a situation where a drug is a known teratogen and a clinician and female patient may choose that drug as the best treatment for their condition. Multiple myeloma is certainly one for thalidomide these days, but I could imagine others.

I think Pat raised a provocative question which

is if the contraindication is somehow absolute -- and I suppose someone could always come up with something that isn't -- then there's an inherent ethical problem. Again, I think it's a no-brainer. I don't think society could justify that.

It would in my mind come down to an individual judgment. Is it useful to collect more information on the frequency and distribution of birth defects associated with thalidomide or Accutane? I don't think so. I would rather put our resources into things we know. Whether those drugs have a 60 percent penetrance or 30 percent penetrance isn't really going to change the clinical judgment, isn't going to change the public health equation very much.

DR. GREENE: I'm sorry. I didn't mean do we really need to define whether there's a 30 percent or a 50 percent incidence of malformations, but rather how effective a job are we doing in avoiding those exposures.

DR. MITCHELL: Oh, yes, I wouldn't think that normal registries would be able to do that. That I think is a specialized activity.

Frankly, I think from a public acceptance standpoint, to mix known teratogens with the registries that are trying to eliminate ignorance may have a deleterious effect on the perception of the public of what registries are all about. So, I'd just raise that as a

concern.

DR. GREENE: Lew?

DR. HOLMES: If we go back to the anticonvulsant model, if you look at monotherapy and do your power calculations to identify a twofold increase for phenytoin, phenobarb, carbamazepine, and you take a sample of 300, 350 infants, so you could say, well, all of those are on my list of teratogens, and yet none has been studied in sample sizes that big. So, you could say, well, yes, they're drugs we're concerned about but they haven't been studied well enough.

Then the other point that you made earlier was most people think in terms of all malformations and of course most teratogens don't produce an increase in all malformations. So, even if you had drawn a circle around a drug and said, well, I'm worried about that, the registry could not only establish it with greater statistical certainty, but begin to address the important issue of the increased frequency of specific disorders, which takes a lot longer time.

DR. JONES: Furthermore, to go back to the Accutane issue, as far as this is concerned, a registry permits you or at least ascertaining patients prospectively permits you to delineate the total spectrum of abnormalities, be they functional or structural, that are

associated with prenatal exposure. Accutane is a perfect example of that where data is coming out now indicating that even children who lacked the structural abnormalities associated with prenatal exposure to that drug do, in fact, have problems with neurobehavioral development. So, I think I would say that despite the fact that we know of drugs that are known human teratogens, we should be subjecting them at least to follow-up and through some kind of a registry methodology.

DR. MONTELLA: I think the purposes are just different. If you have a known teratogen, what we really want to know is really what Dr. Greene is saying. What we really want to know is are people still using it anyhow, and if they are, who is, where are they, how can we disseminate that information differently. I think that's a different question.

But there's plenty of models for having very good information available and having it not disseminated properly or used properly. Glucose is a good example of a teratogen that we've known about for a long time, and yet many, many patients and physicians don't do pre-pregnancy counseling or pre-conception control of their diabetics. So, there's something about the way we disseminate that information that's at issue there, and I think that may be a separate issue than registries of drugs we don't have

information about.

DR. GREENE: One question that Jan raised a few minutes ago and I'd like to pursue for just a minute is the idea that there may be some advantage to having more detailed knowledge of safety to some medications and that, if you will, a carrot might be a pregnancy formulary, if you will, that drugs for which there was extensive experience and demonstrated safety could be listed as such.

The question that came to my mind, as I listened to that, was given the current realities of finances, of how many dollars worth of drugs you could expect to sell to pregnant ladies, the magnitude of the potential liabilities involved if there were a problem later on down the road, could that ever possibly really be an attractive idea to a manufacturer?

DR. FRIEDMAN: I'd just mention there are other possible carrots that might have to do with prolonged patent protection, for example, across the spectrum of the whole drug if it was known it was safe. At least the idea ought to be explored that we would like to have drugs that we have more confidence about the safety. You'd never have complete safety, but about which you have more information so that we're not shooting in the dark so often.

DR. GREENE: Other thoughts.

(No response.)

DR. GREENE: Well, I'd like to then proceed to address the questions that we haven't already addressed in our discussion so far this afternoon. I think that question number 1, specifically, under what circumstances are registries most useful, I think we've discussed that reasonably thoroughly. Sandra, you are okay with that.

We haven't really addressed number 2, the most important data elements that should be routinely collected in a registry. Thoughts about that?

DR. ANDREWS: The epidemiologist's answer: it depends.

DR. GREENE: Lew?

DR. HOLMES: I think the good quality information on the phenotype of the infant alleged to have a birth defect.

DR. GREENE: Other comments?

One of the issues again that was raised in some of the responses from the industry to the draft guidelines was the whole notion of trying to accumulate lots of other data about potential confounders, including illicit drug exposure, smoking, lifestyle and behavioral issues, that while they might be desirable to have in terms of eliminating confounders and dealing with confounders, they might have a negative on your ability to garner the core information that you're really interested in gathering.

How much potential confounding information is necessary to collect in a registry?

DR. MATTISON: I'd actually like to go back to the previous question, but maybe comment on this a little bit.

The issue of data elements. I think as much as we need to characterize the outcome of the pregnancy, we need to characterize the exposure. I think simply looking at frequency and amount of dosing is probably inadequate for a good characterization of exposure. That's why I think it was important that in some of the earlier presentations the discussion of the use of biological markers as providing a better characterization of exposure will be important.

With respect to the confounders, again in the context that they can potentially modify other kinetics or dynamics probably should be given consideration in terms of how that data needs to be collected with respect to exposure. And similarly, with respect to outcome, what is the likelihood that these exposures may also be produce or be associated with these types of pregnancy outcomes.

DR. MITCHELL: Again, I think it's, you know, "you pays your money and you takes your chance." It seems very clear and almost intuitive that the less information you collect per patient the more patients you'll collect.

At the same time, the less information you collect per patient, the more likely the result is to be confounded in ways that you can't handle, you can't understand. That would be quite an unfortunate circumstance.

So, I think again it has to do with the nature of risk that the registry is being designed to identify. My own view is that we should go for the big ticket items. Let's get the first circle of certainty. Let's get the Accutanes and thalidomides, and if we're lucky, the valproates. But even a valproate is subject to much confounding.

In the example that Elizabeth cited about a medication that might be used for sinus conditions and found to be associated with an abdominal wall defect, one really has to know about the indication and what other drugs might be taken for that indication before one blithely indicts the drug.

So, it's not an easy question to answer, and I think that there needs to be a sense of priority. Again, I'll come back to, yes, we all want everything, but I think we really need to be sure we're not letting a major teratogen loose upon the land, and then I think the second tier, if you will, needs to be explored in conjunction with other approaches, whether they're computerized databases or case-control surveillance or case-control studies. I just

don't think one size fits all. It's a theme that I'm echoing from others.

DR. GREENE: Yes, please.

DR. SHARRAR: If you're looking at post-marketing surveillance data, you have to realize the limitations of the data from which you're working. We try to restrict our questions to demographic data and to pregnancy history and to actual drugs they've taken. So, we think we get a better response rate that way.

I think really what we're trying to do is to generate a signal here, and if in fact we identify a signal, then we have to design a more formal epidemiologic study to evaluate that. Then at that time you might want to ask those additional questions.

Furthermore, I don't think it's possible to get accurate information about drug abuse, alcohol abuse, or tobacco use because people really don't want to admit all the different things that they do. So, we keep our questionnaire focused on those things we're concerned about.

DR. JONES: Yes, but you're not talking to the mother, are you? You're talking to the obstetrician.

DR. SHARRAR: Yes, we're talking to the obstetrician. That's true.

DR. JONES: Right. And so, you're not going to

be able to get that information in that way. And if you're talking to the mother in a situation like this, there's much greater likelihood of being able to get all that information.

I would really just like to echo again the need to be very individual, as far as these particular issues and the way you want to design these studies. We're doing a study right now on this new drug for rheumatoid arthritis, Arava, or leflunamide. One clearly needs to know information about the severity of the rheumatoid arthritis and a variety of other issues related to that in terms of being able to evaluate whether the drug or the disease or whatever is leading to the outcome. So, one really has to design one's study based upon the drug that one is studying.

DR. GREENE: Jim?

DR. LEMONS: I know this is obvious, but in discussing these kinds of data, usually the fewer data that are collected, the clearer the definitions can be made, but it's important to have very consistent, clear definitions, depending on whom you're asking the questions of.

Secondly, some control over the quality of the individual submitting that data which may or may not be the obstetrician or the nurse or whatever. I know with the vital statistics form evaluation, that has been a major

concern as to who's really entering the data and from what source.

DR. MILLS: It's been said already several times that it depends on what you're going after, but I think in the case of looking at developmental outcomes, it's particularly difficult. For example, at a minimum you'd want to be thinking about education of both parents, family income, what the native language of the parents is, what the IQ particularly of the mother is, who the primary caretaker is, what the family constellation is in the home, and the last items, of course, over time since you're going to have to wait 4 years, 5 years, 7 years to get the data. I think it illustrates it's extremely difficult sometimes to get even the minimum amount of information you need for these things.

DR. GREENE: Thank you.

DR. MITCHELL: I hope that most people would agree that no data are better than poor data. I don't know that that's universally held, but I think that when it comes to some of these issues about confounding variables, one of the points that Jim made, that rather than say, well, we can't be assured that the confounders or potential biases that you're collecting are going to be collected rigorously, but we'd still like you to collect them. I think we know up front that's a recipe for disaster and

ought to be avoided. Even though it's a tough decision, I would hope the committee would feel comfortable with that.

DR. GREENE: Let's move on to the group of bulleted items here under number 3, several aspects with respect to really sort of concrete details of recommendations in terms of follow-up here. In general, what is the minimum length of follow-up required to assess pregnancy outcome?

That's obviously going to depend on what outcomes you want to look at. Again, I'm not sure that a one-size-fits-all response is ever going to make it here. Dr. Kweder, you're not looking for a one-size-fits-all I trust.

DR. KWEDER: No.

DR. GREENE: Okay.

Under what circumstances is it most helpful or appropriate to end data collection and follow-up? Yes, Sandra.

DR. KWEDER: Let me just add to that. Let's just take the big picture scenario. If what you're looking for is what would generally be considered major congenital malformations, is it enough to end follow-up at delivery? Because that's an issue we frequently confront.

DR. GREENE: Does anyone care to respond? Lew?

DR. HOLMES: Sandy, I would argue that budgets

are always limited and you'll do a lot better enrolling a lot more people in stopping at birth than you would spending an enormous amount of time to follow them in that period after birth, which is a very labor intensive process. I'm not sure you will get as much for the personnel time as you would enrolling more people.

DR. WIER: Mike, I want to make a general comment about both these questions that deal with registry design. I think it's worth reminding ourselves of the obvious, and that is not only is not one-size-fits-all, it's not the only time we may go to buy a suit, if you will.

That is to say, we have to remind ourselves that research is largely an iterative process and this includes both the preclinical studies as well as the clinical studies. It's possible that the first registry generates some questions that require further preclinical assessment, and that in turn suggests other clinical assessments that should be done. I think that allows you to take a more sanguine view of the registry. Maybe the first registry can be designed with more practical considerations in mind, but we have to keep our eyes open to subsequent clinical studies. They may not all be registries, of course; they could be other types of epidemiology studies.

Maybe it helps to illustrate this with a specific example. We can talk about the case of a drug where preclinical studies in mice showed the possibility of a neural tube defect and then suppose the registry is designed looking for this and finds, in fact, there is significantly increased incidence of neural tube defects. It still could only be in 1 to 2 percent of the exposed pregnancies. That's a big increase above background, but you're left wondering, well, what is the exact susceptibility determinant here. And you go back to the preclinical setting where you can ask those types of questions and perhaps identify susceptibility biomarkers now that could be then assessed in a clinical evaluation.

I think if we take this view of an iterative process, it also helps to understand how both the preclinical studies and the clinical studies work together, that neither one is ascendant, but their both important and they both work back and forth.

DR. GREENE: Other comments on this issue? Yes, please, Jim.

DR. MILLS: Lew gave, I thought, a good practical answer to the question because it's certainly true that by decreasing the time, you can increase the number of people you can see, but I'll give more of a theoretical, scientific answer and that is that if you

follow longer you can perhaps double the number of even major malformations you'll be able to diagnose. And that's a consideration in the sense that you will be finding different types of malformations with a longer follow-up so that if you have a teratogen that causes the type that isn't that obvious at birth -- the classic thing being fetal alcohol syndrome where even very sophisticated observers may not be able to spot it at birth -- you may be able to find things that you would totally miss.

There is also some payoff in the sense that by being able to find malformations, you would decrease the sample size required in some cases to find an effect.

DR. HOLMES: To follow up on that, Mike. If you think about it though, Jim, the things that are picked up postnatally in that first year or second year, one of the problems will be the chance nature of some of the discoveries. One of the things that haunts you when you do studies is when you have a group of people who by chance had studies other people didn't have. That's one of the reasons why I exclude anatomic variance picked up prenatally. It's worthless in this process because you can't have every fetus have the same exam by the same skilled sonographer and so forth.

Likewise if you have children who by chance have an ultrasound of the kidney or by chance have a this

or a that, and yes, you get another item in your numerator, I think it's a dangerous addition to your numerator because everyone hasn't had the same evaluation.

DR. LEMONS: Plus, I think the fetal alcohol syndrome was pursued because there was the typical clear phenotype in the newborn that was recognized initially and then it was backtracked to look at the longer term and broader -- is that not true, Ken?

DR. JONES: No. To be honest with you, Jim, that's not true. The initial children that were picked up with a fetal alcohol syndrome were 5 and 6 years of age, and then it was backtracked to the newborn period.

On the other hand, I'm not sure to follow a child up to 6 months to a year is any better in terms of picking up the fetal alcohol syndrome in the newborn period. You have to wait till probably 4 years of age to feel very confident about the fetal alcohol syndrome.

I would, however, take exception to Lew on this issue. I think that there are a number of things that you're going to miss, central nervous system abnormalities, renal defects that are picked up by virtue of urinary tract infections and the like, not just due to a chance ultrasound, and a variety of other even cardiac defects that you're going to miss in the newborn period if you're not following these kids, I would say, at least to 6 months

of age. So, I would go beyond the newborn period at least to 6 months of age to follow up these kids.

DR. GREENE: Other comments on this issue?

Just let me ask one question of Ken and Lew. Would you be better off doing a more extensive set of studies on everybody in the immediate newborn period in the first 48 hours of life and study them from guggle to zatch rather than leaving it to chance 6 months later?

DR. HOLMES: Are you talking about the registry model?

DR. GREENE: Yes.

DR. HOLMES: I think in the registry model you're totally dependent on the routine pediatric exam, and for me that's your gold standard you've got to work with. Given a lot of children are now in the hospital very short periods of time, it's true some things can be missed, but I think that's got to be your gold standard. It's only when you do these offshoot studies that you're going to be able to do the guggle to zatch exam by someone like Ken.

DR. JONES: And I would agree with that. Furthermore, one would have to do guggle to zatch in your control group, which Lew just pointed out, which I think would be absolutely unmanageable.

DR. GREENE: Sandy?

DR. KWEDER: I think that Lew has clarified, to

some extent, the question that I had for him because I think that, Lew, you're talking about the model of actually using the pediatrician's exam, which is quite different than getting information from the word of the obstetrician.

DR. HOLMES: Absolutely. I think there's a lot of reason to believe that relying on the mother's self-reporting, the obstetrician's best effort versus the pediatrician's exam -- those are your three usual alternatives. Or I guess the fourth would be reading the medical record. Clearly the pediatrician's input would be the best.

DR. KWEDER: And you would have information by the pediatrician, generally likely to have a little bit more information than the obstetrician whose contact with the infant ends in the delivery room.

DR. HOLMES: Right.

DR. GREENE: One issue that we really haven't addressed yet, which is the next bullet, is what strategies might registries consider to enhance patient recruitment and retention, as well as facilitate follow-up. Any thoughts about that?

DR. MITCHELL: I would certainly defer to Lew on this. But it seems to me that one of the major incentives -- this is from our own somewhat different experience -- is that if the physician promotes the

activity to the patient, it is an amazingly strong incentive. So, in much the way that folks at NICHD are working to try to reduce sudden infant death syndrome by getting to the grandmothers, I think that getting to the physician may have much more bang for the buck than any sort of general advertising. To the extent that community of practitioners can be encouraged to see this as an asset to them in their management of patients rather than a risk or a pain in the rear end, I think it would enhance greatly. I can't prove that and I'd defer to others.

DR. GREENE: How do you do that?

DR. MITCHELL: It's not easy. But I think physicians tend to operate, as most people do, in their own self-interest. I think the very problem that obstetricians face in not knowing what drugs to use has its roots in the fact that there isn't information. I think that both using a little bit of guilt-tripping and a little bit of carrot to say that here's an approach that will help you -- and it has to be structured in a way that doesn't suggest that the physician prescriber is going to be nailed by this. This isn't a way of identifying blame. It's tricky but I would think that would work well.

Lew, I would just invite you to comment on it.

DR. HOLMES: The model we have, as you know, is the one where the woman herself calls. We are convinced

that that has made a difference in retaining her. It has not been as easy to get her in because she has to decide to call. We do a second call at 7 months gestation. We call her, and we find about 10 percent of the people have changed address and/or phone number. So, if we wait until the postnatal period, we'd never find them. Our lost-to-follow-up rate after about 1,700 to 1,800 enrollees is about 2 or 3 percent with a system where she is the person you work through. So, that's an argument for doing it that way.

But it may be you could argue if all doctors reported their patients, it would clearly generate a lot more enrollments, but I don't think all doctors would.

DR. GREENE: Yes, Dr. Wisner.

DR. WISNER: Just thinking about this from a clinical perspective, I have a point similar to yours to make, and that is when you do that risk-benefit decision making with the patient and you look at what are the possibilities in different reproductive toxicity domains, you have this kind of partnership where the patient begins to express what they want or what they value as far as components of that decision making.

But what always happens is I'm always saying there are certain things that we know with limited amounts of certainty and that there are a lot of unknowns. I would

see joining a registry as kind of making a partnership with a patient that we need more information and this is a way that we can work together to help collect the information, almost like a responsibility to contribute to the information for our daughters.

The second issue has to do with in all the material I looked at, I didn't see much about engaging the pharmacist. The reason I thought of that is, although I don't usually have a lot difficulty with pharmacists, I had a recent experience where the pharmacist, in screening my patient for giving her the medication, was distressed that the medication was one that he was uncomfortable with. So, I thought how could we switch that around. Well, we certainly could use pharmacists in terms of educating about registry materials.

DR. MONTELLA: I think when you're dealing with physician recruitment, you have to make it incredibly easy, and somehow you have to do that. You have to put a post card attached to the prenatal record. You have to get to office managers. You have to do something that makes it very, very easy, and then be very careful to offer to share the information that comes out of it with them, to do a mailing of 6 months' worth of data, 5 years' worth of data. Whatever comes has to get back there so that it was worth it.

DR. GREENE: The next question, the last one under this group of questions, is: Should an additional mechanism be put in place to recontact subjects after an extended period of time, should the need arise? If so, considering the practical aspects of conducting registries, under what circumstances should it be done?

Lew, you said that you recontact your subjects.

DR. HOLMES: Our system is three interviews. So, initially she calls. The informed consent document is the next step. Then she has her initial interview. Then we call her on the 7 month roughly and then call her between 4 and 8 weeks after expected date of delivery. So, we initiate that. We do not do a year later or 2 years later.

DR. GREENE: Yes, please.

MS. CHAMBERS: Through the TIS collaborative studies, we contact the women three times during pregnancy, sometimes twice depending on how late they enroll, and then we contact them either by mail or by phone every 6 months up to 5 or 6 years of age.

DR. GREENE: What additional information do you glean from those contacts out to 5 and 6 years of age, and how do you use it?

MS. CHAMBERS: The primary purpose of the contacts, after about the first 6 months of age is the one

that Lew brings up because otherwise we would have a huge lost-to-follow-up when it came time to offer them neural developmental follow-up. So, we try to keep in touch with them basically to control for the moving issue.

But the other issue does come up, the one about getting information on birth defects or developmental problems that wouldn't have been identified or were not identified in the first few weeks after birth. So, we do collect information on that in the long term, but it's treated differently, obviously, because of the issue that we don't have that length of follow-up on everyone.

DR. GREENE: Other comments about this issue, relatively long-term follow-up?

(No response.)

DR. GREENE: The fourth question is the criteria that should be used to determine when a registry should be closed. You closed a registry on acyclovir. Would you like to speak to that?

DR. ANDREWS: You've got to start somewhere. When we started that registry, it was our first effort, and we struggled in defining what our overall objectives were. So, we went back and forth and said, well, we're shooting for a target of 300 first trimester pregnancies followed completely, but we didn't have as clearly defined an endpoint as we would have liked.

We found that our need for information changed over time. So, when we got to the 300 patient mark, we were interested in pursuing over-the-counter status for this medication for the treatment of genital herpes, and we felt that in order to do that, we wanted significantly more information. So, we kept it open beyond our sort of predetermined target. But I would say that the optimal thing to do is to establish a target number in the beginning and go for that.

The other thing that happened along the way is that information about safety was out there. So, we were receiving fewer calls because so many different organizations were generating and distributing information. So, we found that enrollment was, in fact, decreasing and we weren't learning that much more by the additional information. So, at that point we decided to shut it down.

DR. GREENE: Dr. Sharrar, I believe Merck has closed the Varivax registry. Haven't they?

DR. SHARRAR: No. The Varivax registry is still very much in operation.

DR. GREENE: Oh, I'm sorry.

DR. SHARRAR: It has been in operation about 5 years. We have about 400 and some people registered in there. The day may come when we feel that we have a sufficient sample size to close it, but that has not

happened yet.

DR. GREENE: Do you have a number in mind right now, a notion of when that day will come? What's your target? What's your goal?

DR. SHARRAR: We'll keep it going until we decide. It is a collaborative registry with the Centers for Disease Control and we have an advisory board. I think once we all come to an agreement that we've done it long enough, then we'll stop it. But we're certainly not at that point yet, and I don't expect to be at that point for a number of years.

DR. GREENE: Certainly even the CDC eventually closed the rubella vaccine registry.

DR. SHARRAR: Yes. They closed the rubella registry I think after about 10 years, and I think they had something like 700 reports in there. So, it will vary. And we may in fact reach that point.

But they were also looking for a specific syndrome, the congenital rubella syndrome, and that's very similar to the Varivax registry. We're looking for a specific defect, which is the congenital varicella syndrome. So, I think you do reach a point where you could say enough is enough.

DR. GREENE: But at least you do have some idea of what the incidence of the congenital varicella syndrome

is in people who are exposed at the right time in pregnancy to natural varicella.

DR. SHARRAR: There are a lot of advantages to the varicella syndrome, which I'll tell you about tomorrow.

DR. GREENE: Okay, great.

Other comments about this issue of closing down a registry?

(No response.)

DR. GREENE: Then the last question. I think we talked about this really in terms of labeling. Dr. Kweder, have we not given you adequate guidance in any particular area?

DR. KWEDER: No. For us, question number 5 really is critical because when we speak with companies about conducting registries, this is understandably one of the issues that's on the table. What are we going to do with this information? How will we get information from registries out to clinicians? Then we get into lots of discussions about because registries don't offer us sometimes the level of certainty we would like, is the information still worth putting in labels. So, if anyone else has any further comments on that, please feel free to make them.

I guess the only other question that I have was one regarding a comment that Elizabeth made earlier in the

day and it's something that we struggled with ourselves. It may be because we're not always all talking about the same thing when we say pregnancy registry. Maybe "registry" isn't the right term and maybe we ought to discipline ourselves to be calling some of these things something else. I would love to hear what folks' suggestions might be.

DR. GREENE: One of the issues that came out of the June meeting was the fact that labeling drugs the first time around is going to be a herculean task. The idea of revisiting it at a regular interval is daunting. Has the agency given any further thought to that issue? Because that certainly would impact upon this.

DR. KWEDER: This will necessarily be revisited with the new regulation that requires safety update reports, where if there are additional data to bring to bear, it will naturally be revisited.

DR. GREENE: Jan?

DR. FRIEDMAN: With respect to incorporation of pregnancy registry or follow-up study data into the label, I guess I don't fundamentally understand why this is different from a study that happens to be done in Czechoslovakia or a study that's done from the CDC or from an OTIS study.

It seems to me that all of these studies have

limitations. All of these studies have certain strengths, and all of them provide a certain quantity of information that has to be evaluated on a case-by-case basis. I would think that whether or not and to what extent the data gets put into the label has to do with its overall quality and significance in the context of the whole body of data that's out there and that there should be no bias toward or against putting pregnancy registry data in because it happens to be under FDA auspices or is funded by the sponsor or whatever.

DR. MITCHELL: I'd absolutely second that.

DR. GREENE: Any other comments before we take our break?

(No response.)

DR. GREENE: Okay, then we are adjourned for 15 minutes. Thank you.

(Recess.)

DR. GREENE: We'd like to reconvene please and see if we can keep the meeting moving roughly on time. First I'll give Dr. Kweder to make a few remarks.

DR. KWEDER: I'll only make a few. I want to thank you for your discussion. I think that I speak for all the FDA people who aren't back at the table yet and say that your discussion has been very, very helpful.

While the questions that we asked you to

discuss are not specifically related to the draft guidance document on registries, your discussion does help us in identifying areas where we need to be more clear in that document, where we need to leave room for flexibility, and how we can develop some of those areas further.

My other comments are simply to remind you that at this point we feel that you've addressed most of the questions for the first-day questions 1 through 5 adequately, although I would not be surprised if some of these issues don't come up again further on in the meeting. You shouldn't feel that you can't go back to those things if you've changed your mind or you have additional things that you want to say about them.

As we move into the next section of the meeting, when I gave the introduction this morning, I suggested that one of our goals is to get people to think beyond what they've always thought about as a pregnancy surveillance study or a pregnancy registry, to think creatively about additional ways of collecting data, to think creatively about how FDA and others at this table can build partnerships to facilitate better and more data collection of a type that would be useful to clinicians. This is not something that we've heard discussed at an advisory committee meeting before. We've had lots of internal discussions trying to brainstorm about this, but

we'd like to hear what some of your thoughts are. So, that's where we take it from here.

DR. GREENE: Without further ado then, I'll introduce Dr. Allen Mitchell, my fellow Bostonian.

DR. MITCHELL: Thank you. I know you're not supposed to begin with an apology, but I apologize for two things. One is that my voice is terrible, worse than usual, and the second is that many of the points that you've heard have already been made by me, as well as probably others. But bear with me because now they're in context.

I was asked to give a talk called "state of the art" and was given no more direction than that, which could be dangerous. I'm going to try to give what amounts to a bit of a personal perspective, I hope not one that's completely out in left field, but to try to address some questions that I think are relevant to the considerations of this subcommittee. What I will do is provide you with a perspective which I would like to argue is complementary to the work that the committee has focused on.

I guess I ought to start with the first slide. Really, the issue here is identifying teratogens. I think for starters, we ought to focus on the critical questions. From my point of view, the clinician and the patient alike really want two questions answered. I'm focusing on

structural birth defects, but I would like to believe that the same outcomes could be applied with some modification, of course, to other considerations.

First of all, is this drug another thalidomide or isotretinoin? It just seems quite obvious that that is the first question that needs to be answered. But a second and important question are, are there other less frequent teratogenic risks associated with a given drug? Those are the questions that the clinician and patient need to answer. But in addition, I think the regulatory agency and the public also need to know what is the teratogenic impact on the public health.

While those seemingly may be the same question over again, it's not, and let me just give you two extremes. These are both hypothetical for the purpose of discussion.

For carbamazepine, let's assume that .6 percent of women use the drug in pregnancy. Lew, I took that from yours. I hope it's close. That would amount, if you believe that there are 4 million pregnancies a year, plus or minus, to 24,000 exposed pregnancies a year. Let's assume that there's a 4-fold increase associated with that drug in oral clefts. Remember, it's 4-fold, not 30-fold, but not 2-fold either. And the baseline for oral clefts is roughly 1 in 1,000. Well, if you do the math, the drug

would cause an extra 72 cases of oral clefts a year in this country. Well, it's certainly not trivial.

But let's consider another example. Let's consider ibuprofen and let's assume for the moment that 15 percent of women use the drug in pregnancy, and I would argue that's a conservative estimate. That would amount to 600,000 exposed women per year. And now let's assume a 4-fold increase in the risk of something like TE fistula, which has a baseline rate of roughly 2 per 10,000. If I've done my math right, that drug would then cause 360 cases a year. We're talking about cause. So, you have to subtract baseline. Well, that's roughly 4-fold or more cases attributable to that drug exposure. Of course, ibuprofen is largely a nonprescription drug.

I think the point here is that we need to consider the public health impact as much as we need to consider the clinical and patient side of the equation.

This is nothing new to anyone here. The post-marketing approaches for identifying teratogens rely on case reports, experimental studies, which are almost never informative for reasons that have been discussed, and epidemiologic studies which largely fall into the cohort and case-control design.

Just to very quickly summarize the kinds of study options that we have available as a cohort study, we

can have broad-based studies which might look at a wide range of exposures. I think the best American example of that is the U.S. Collaborative Perinatal Project that recruited 58,000 women in about seven centers throughout the U.S., followed them through their pregnancies. They were not selected based on their drug exposures. These were all comers.

There are also focused cohort studies which look at specific exposures, and of course, registries, whether they're run by manufacturers or TISs, are clearly in that category. And computerized databases might also be seen as a focused kind of cohort study.

Case-control studies, on the other hand, are broad-based as one form, and this would include case-control surveillance, which is what we happen to call our design, risk factor surveillance, which is what CDC happens to call its design. They're both the same in design where a wide range of defects are investigated. I gave the two examples, ours which began in the mid-1970s and the CDC's large effort, with which a number of us here are involved, began more recently, in the last 3 years.

There are also focused case-control studies which examine specific defects rather than a broad range of defects. The examples are simply too numerous to count. For those of you with clinical background, that's all I

remember.

One can spend a lot of time talking about the problems and the strengths of focused case-control studies, but I won't. What I will do is try to suggest to you that I can play both sides of the street because while we have certainly focused our attention since the mid-1970s on our birth defects study, which has enrolled over 19,000 malformed infants in a number of centers, which is a case-control design, we really did our intellectual teething on the Collaborative Perinatal Project data, which was a rather large cohort, as I've mentioned, and since then, have had about 11 or 12 years' experience with the Accutane survey, which is a form of registry, where we've enrolled a little over 500,000 women, and in more recent years, 1998 and since, have been involved with the thalidomide survey which has enrolled somewhere under 10,000 women.

That gives you a little bit of sort of my background in terms of how we come to this, but one message that I think I will carry to the grave is that prospective does not necessarily mean it's good and retrospective does not necessarily mean it's bad. There have been people who have tried to make that argument. I would argue that there are bad prospective studies and good retrospective studies. I will also admit that there have been a lot of bad case-control studies which have given this study design a bad

name. But there's nothing inherent in the designs that make one better than another. They have different strengths and weaknesses. You'll be surprised to know that I'm going to touch on some of those.

Well, keep in mind the epidemiologic nitty-gritty. We need to talk about exposure. We need to talk about outcome, covariates, bias and confounding, and estimating risk. I'm only going to touch on some of these. This is a sort of stream of consciousness presentation. But we have to keep in mind that the approaches we're talking about are by definition epidemiologic studies, and like it or not, you can't escape epidemiologists in this sort of undertaking.

When we talk about exposures, we're talking about medications on the one hand. Prescription medications are the focus here, but I want to remind everyone that OTC medications are an important source of both exposure and potential confounding. The issue of diet supplements I guess has to wait for another day, but maybe not a different place.

Again, to emphasize the OTC concern, they are the most prevalent medications taken in pregnancy. Their exposure is independent of a health care provider. It's direct consumer advertising if there ever was any. And they are generally perceived to be safe and not just by the

public, but I would argue that they're generally perceived to be safe by the physician community.

So, here's one example, a snapshot in time, if you will, from our own data for LMP years 1992-1993, of the drugs most commonly used in pregnancy based on 686 interviews in Boston and Philadelphia. We left off the Toronto interviews because Canadians have somewhat distributions. But you can see that roughly two-thirds of women reported exposure to acetaminophen, 17 percent to ibuprofen in those years, 14 percent to pseudoephedrine. You can see that there are relatively few prescription products on this list. The most common exposures are, indeed, nonprescription items.

But it's not enough to, I think, look simply at a snapshot in time. It's a dynamic. This gives an example, first trimester exposure to selected benzodiazepines among, this is now, 15,000-plus women over roughly a 20-year period from our data. The red line is diazepam and the yellow line is chlordiazepoxide.

So, while we and others have had a lot of concern about diazepam, which at one point was used by over 3 percent of pregnant women, at least according to our data, the data in the mid and late 1990s would suggest that this is not a drug of common use, and that has to have an effect on how we conceive -- you'll pardon the term -- of a

registry's purpose.

The other side of the equation, of course, is what about drugs that may be increasing their use over time? This is a variety of cough, cold, and allergy medications, some of which have been increasing, some of which have been relatively stable, but one that strikes you is pseudoephedrine, which has taken off rather appreciably since it was made over-the-counter and now included in a wide variety of cough/cold medication. So, if one were asking me what drugs would I be concerned about, that would clearly be a drug of concern where I would say that I have an intellectual affinity for the issue of diazepam, but as a clinical and public health problem, it's taken somewhat of a back seat.

I don't consider this an absolute statement, but it's something I call the fallacy of class action teratogenesis. There's a presumption that members of a give class of drugs have the same teratogenic or non-teratogenic activity. What I would argue the fallacy is is that's not necessarily the case.

What is problematic in the area of teratogenesis is we don't know what causes malformations. We don't know that it's the pharmacologic effect of the drug. We don't know that it's not some methyl group hanging off the end of one drug compared to another. Until

we know that -- I think the argument could be made even for retinoids -- a drug class doesn't necessarily behave as one.

The example I'll give is these two chemical entities which are both comprised of a glutarimide ring, and this is the extent of my biochemistry. I apologize for it. If one were to go on that basis, one would say, well, maybe we should assume that they have similar teratogenic potential. In reality, the one on the left is thalidomide and the one on the right is glutethimide, which used to be sold in the 1960s and the 1950s under the brand name of Dordin, a common sleeping pill. Actually there are data in the Collaborative Perinatal Project which indicate that it isn't a thalidomide.

So, I think it's very dangerous to make the assumption that, gee, this drug is safe, therefore we don't have to worry about this, and conversely, this drug is dangerous, therefore we need to worry equally about this. I'm not suggesting that this being the case, one shouldn't worry about glutethimide, but the opposite is not necessarily the case.

So, moving on to outcomes, this has been a theme of this morning and afternoon, and so I won't spend a lot of time on it. But we have to make judgments about what defects are of interest and importance and recognize

that the prevalence of birth defects -- this is from the Collaborative Perinatal Project -- varies considerably depending on whether you're talking about any major defect, which might occur in 3 percent of the population, or any selected, specific defects, which can be as common as 1 percent for inguinal hernia; oral clefts, about 1 in 1,000; and hemimelia, phocomelia, 1 in 10,000. TE fistula, as I pointed out, was about 1 in 5,000. So, the specific birth defects are extremely rare events in the normal setting.

I believe that back when we analyzed the Collaborative Perinatal Project data, we made a fundamental mistake in collapsing categories of outcomes based on cardiovascular defects or urogenital defects. I think we've learned a lot since then, and I think it would be a mistake -- and the same would be true of oral clefts and other groups of malformations. So, as we learn more, we raise our level of anxiety about the specificity of teratogens.

On that theme -- and this is an area where I think Ken Jones and I will agree to disagree, that most drugs that we know cause marker or signal birth defects or a cluster of birth defects. There certainly may be teratogens that cause syndromes. I wouldn't challenge that. But by and large, the teratogens that we know of are associated with a peak increase in one or a few particular

defects. That makes for some difficulties in study design.

I'm probably the only one here who's using this antique form of audiovisual presentation called the slide projector, but I didn't call them lantern slides. So, that's probably good.

(Laughter.)

DR. MITCHELL: This is really going back. This was a chart that we photographed that Dennis Slone, for whom our group is named, presented to Congress in the late 1960s to a congressional subcommittee, trying to educate them about the forms of study design. What's remarkable is it hasn't changed. This is here for clarity but also for sentimental reasons, and you'll see two of these slides for cohort and case-control.

The point is that in a cohort design of 100 pregnant women -- this is not exposed to any drug in particular necessarily, but perhaps in most cases, even when exposed to a drug -- 97 of those women will deliver a normal infant from the perspective of malformations. From a human reproduction standpoint, that's a good thing that it's no more than 3 percent malformed, but from a study design standpoint, it's a terribly inefficient way to collect information because for every 100 women you follow, there are 3 malformed, more or less, and then you take some controls, but you're still essentially not using 90 percent

of the data.

But there are definite strengths in selected cohorts, and by selected cohorts, I'm referring to registry designs. First of all, this is where prospective is good. The registries allow the opportunity to identify exposed pregnancies before the outcome is known, and that in and of itself is a major contribution. They are also able to assemble a study in a relatively short period of time.

Here's an example. This is Ed Lammer's paper in the New England Journal 15 years ago. Oh, good grief. It's 15 years. I'd like you to ignore the spontaneous case reports, those 23. That's really not the point of the slide.

The point of the slide is that with a cohort of 36 pregnancies, in the bottom half, there was a cohort of 36 exposed pregnancies identified before the outcome was known, and out of those 36, there were 5 malformed infants. That in itself was a trigger for concern. But what really sort of put the icing on the cake -- and I think Jim or someone had mentioned -- was a distribution of malformations. Now, they had collapsed the spontaneous case reports with the cohorts and there was no way from the paper to separate them, so let's leave them collapsed.

If you use the expected rates from the Collaborative Perinatal Project, what they found was that

for the marker malformations that we now associate so clearly with isotretinoin, there was a 237 times increase of microtia and anotia over the expected, and for congenital heart disease, it was 3 times; for face and skull defects, and particularly micrognathia, it was 32-fold. So, a small cohort can really identify very quickly and effectively an Accutane or a thalidomide.

But not all small cohorts have that same capacity, and this is just one paper focused on calcium channel blockers where the concern going into the study was that animal studies had suggested digital and limb defects. The cohort included 78 women with first trimester exposure to five different calcium channel blockers. There were 66 liveborn infants, 2 with major malformations, a 3 percent overall malformation rate. There was no evidence of risk, and they were able to rule out a 5-fold increase for the overall rate of malformations. But remember, in this case the concern was limb reduction defects, for which there was simply no power in a cohort of even 78. And what about other specific defects? So, there are certainly situations where one is extremely useful and another may not be quite so useful.

The limitations of cohorts in general terms have the same limitations that any epidemiologic study has. I'm not picking on them particularly. Bias and confounding

are the major concerns in terms of the cohort designs we're talking about. We're worried about selection and referral bias, confounding by smoking, health behaviors, and other factors, and as many people have mentioned, what is called confounding by indication. Is the disease state itself accountable for the increased risk that might be observed in association with the drug exposure and how does one separate that? One even has to consider the severity of the disease state. Ken was talking about in the Arava study looking not just at the disease but how severe the disease was, so that it becomes a real issue.

But the particular concern that I want to focus on today is the specific concern of a cohort study which is statistical power. Let me just give you a couple examples.

If we're concerned about specific birth defects -- and I'm trying to make the case that that needs to be an area of concern, for a relative common specific defect, which might be oral cleft or neural tube defect with 1 in 1,000 births baseline, one can identify a risk of, let's say, at least 20-fold in an exposed cohort of 300 with 600 comparison pregnancies, for a total of 900. I'm not going into the various statistical givens here. Just accept that they're all the same for all the examples, so that's being held constant. This is actually really the case for valproic acid, that the estimates, as I can best determine,

are in the neighborhood of 25-30-fold risk for neural tube defects, so you need cohorts of even less than 300, and that's indeed what's demonstrated the risk for valproic acid.

If you want to identify a 5-fold increase in risk for a common, specific birth defect, you need 2,000, 6,000 total.

If you want to go down to a 3-fold risk, it's about 5,500 for a total of 16,500. Those are big numbers.

You can look in your own references or the references that were attached to the FDA document. Every sample size table is different, for reasons that I've never understood. But suffice it to say, they're close enough for government work.

(Laughter.)

DR. MITCHELL: For a rare defect, though, affecting 1 per 10,000 births, everything goes up by 10-fold. So, now to identify that 20-fold risk, if it's not neural tubes in valproic acid that we're concerned about, but let's say, tracheoesophageal fistula roughly, then we would need 3,000 exposed, and if we have a 2 to 1 comparison group, a total of 9,000 to be followed. And if you really want to get down to the 3-fold area, we're talking about gigantic numbers to be followed.

So, my sense is that while cohorts have

sufficient power and can have enough rigor to identify the thalidomides, isotretinoins, and even the valproic acids, they don't have the sufficient power and may not have the sufficient rigor to identify lesser but still important teratogen.

So, where does that leave us? It leaves us with the other poster that Dennis Slone prepared, which is the case-control design, disparagingly referred to as the retrospective or, even worse, the TROHOC design. Alvin Feinstein uses that term. It's "cohort" spelled backwards.

(Laughter.)

DR. MITCHELL: The idea here is that you begin with infants with specific malformations. So, let's say we're interested in clefts and absent limbs and heart defects and spina bifida. Then you determine, after you've identified this study group, the maternal exposures.

That's the typical case-control design, and in fact the typical design is where you have an exposure B -- let's say this is diazepam and an outcome A, or here 2, which might be clefts. And you examine the prevalence of exposure to diazepam among mothers of cleft infants. Of course, you need some comparison group. So, the typical, what might be called the semi-specific case-control study is you identify clefts and then you might look at different kinds of exposures. While you're at it -- you're obtaining

the information -- why not obtain more?

Well, the novel recognition that Dennis Slone and colleagues, Sid Shapiro from our group, actually made, which seems so obvious in retrospect, was to develop what they called at the time the non-specific case-control study, but it is really case-control surveillance. You're no longer constrained to look at one malformation only, but rather while you're at it, you identify clefts and limbs and whatever else you want, TE fistula, and then you obtain medication histories broad-based on all three. That is the nature of case-control surveillance, and it's different from a typical case-control study.

This just happens to give the selected defects in our database as of 1998. It's not even current anymore. We had over 2,100 cleft lips and palates, 1,600 VSDs, 1,200 spina bifidas, and so on. Even when you get into extremely rare malformations -- tracheoesophageal fistula, 411; hypoplastic left heart, 241 -- you can develop huge numbers. The CDC Centers for Excellence anticipates having 12,000 malformed infants in its database over a 5-year period.

So, these designs are quite powerful in being able to collect information on specific malformations. So, the strength of the case-control study is statistical power.

Let me give you the sort of flip-side example. Now the power analysis has to flip around. Now we're looking not at the prevalence of the defect, we're looking at the prevalence drug use in the controls or at baseline. So, if one wanted satisfy one's self that we need to identify a 20-fold risk for a case -- let's call it cleft palate -- we can do it with 10 case and 20 controls. If one wants to identify a 3-fold increased risk -- this is a very frequent drug exposure, bear in mind -- you can do it with 125 cases of cleft palate, let's say, and 250 controls.

Incidentally, these numbers don't take into account what you're using as a control. That's an issue of validity, not statistics. But I'm playing the numbers here, so to speak.

For a drug used by 1 percent of controls, which is now getting much closer to the anticonvulsant area and some of the other prescription drugs, a 20-fold risk could be identified with as few as 33 cases, but even a 5-fold risk could be identified with 200 cases and 400 controls. Oh, my gosh. I'm sorry about the math on the right. These are the correct numbers.

Now, there are clearly limitations to case-control surveillance. First of all, there are constraints on identifying teratogens that increase all defects across

the board. Now, I think that that's a largely theoretical concern. I think Ken and some others might argue that it's a real concern, but nonetheless, it needs to be on the table that if there are teratogens that systematically increase risks uniformly of all defects, a case-control surveillance is going to miss it. In addition, teratogens that produce high rates of specific constellations of defects may be missed. So, the clinician's job security is not compromised by this approach. The fact of the matter is that it has some limitations that are quite real.

And case-control surveillance may be relatively slow to complete, and power is by no means absolute. To give you an example of that, let's look at a drug used by .1 percent of controls, 1 in 1,000 pregnant women. If one wants to identify a risk of at least 3-fold, you would need 5,500 cleft palates, for example, in order to identify that kind of risk for such an infrequently used drug. The bottom line is there are some questions that I don't believe -- and I think most of us would agree -- that no study design can answer.

This is something that I developed in an effort to try to reflect two concepts, and I hope it does. The first is if we look here at the required sample size, I think it should be pretty clear to people now that if you want to identify a thalidomide or isotretinoin, you don't

need very many numbers. A valproic acid, you need more. A diazepam, you need more. A lithium, you need a lot more. Then you get to the point where the curve just really takes off. And in fact, those sample size requirements pretty much relate to the kind of design that might be used to deal with them.

Thalidomide and isotretinoin are the classic small cohort. Who cares about confounding design?

Valproic acid is a good example of something that's sort of on the edge of the cohort's capacity and one has some concerns about it, but it can still be identified and was identified within a cohort.

Case-control studies are really required for things like diazepam.

And I deliberately tried to put lithium on the border between case-control and unrealistic because, in fact, lithium is probably a good example of something that is so infrequently used and associated with such a rare malformation that there's really no study design, short of spontaneous reports, case reports, that are likely to make that connection.

What I wasn't able to do, or more properly I should say, what our skilled word processor was not able to do, was to color this line in an intensity that reflects the need for information on bias and confounding. What I

wanted to do was start out with this being white and this being pink and this being bright red. As you move along this line, the concerns about bias and confounding go with the takeoff in the line. It just makes life more complicated.

So, coming to a conclusion here, I would like to posit that cohorts provide a very necessary and critical first line of defense in that they are able to identify or rule out teratogens that have unusually high risks. That's not to say that's their exclusive domain, but I would argue that that is their first and foremost objective and that case-control surveillance provides a second line of defense that allows the identification of teratogens with lesser risks.

Now, the fact that I did not say rule out teratogens with lesser risks was not a typo. I think that the cohorts can rule out the isotretinoins and thalidomides, and that when those cohorts are complete, it is fair to say to the world, this drug is not one of them.

The case-control surveillance approach can help identify teratogens with lesser risks, can, to use this morning's phrase, help us as a society put our arms around some sense of confidence, but it can't rule them out.

So, in conclusion, clinical and regulatory questions on teratogenesis I believe can most realistically

be answered by taking advantage of the complementary strengths of focused cohort studies and case-control surveillance.

Thank you.

(Applause.)

DR. GREENE: Yes, we will take a couple of questions, please.

DR. WISNER: I was interested in your comment about the class action teratogenesis, that if you can't really make any kind of conclusions based upon chemical structure of a particular agent, that it might have a reproductive outcome effect because a similar agent has, say, a teratogenic effect.

My question is about grouped exposures. What we're seeing now are studies similar to the study you showed with the calcium channel blockers, where we have five different drugs grouped together as an exposure, or for example, the recent study of three serotonergic antidepressants. They are grouped together as an exposure and then outcomes are looked at down here.

I wonder if you could comment on the interpretation of these kind of grouped exposure studies and whether you would argue that each agent ought to be studied independently.

DR. MITCHELL: Hubris is a real problem in

birth defects epidemiology. It wasn't that long ago that we believed the fetus was impervious to external influences, and then we never could imagine that something like DES could happen, and on and on and on. So, I think it's very important that we maintain a perspective that isn't absolute.

I think that the approach that you've described is not necessarily good or bad. I think that one is limited in what one can do. In the calcium channel blockers, I think it's not unreasonable at all to group those exposures according to drug class, if for no other reason, if there is a teratogenic effect, it may operate through the therapeutic action of that class of drugs, and that's very important to know.

However, I don't think it's enough, and I think what ought to be done, albeit with small numbers, is to stratify on the specific drugs, recognizing that you're getting into a small number situation and chance is going to play a role here, but to try to see whether there may be, for a given drug, any suggestion of an increased risk, it's dicey because the numbers get vanishingly small, but I think that it's important. As a concept I think it's important. If I were king, it's not that I would throw away those data and say, well, because they're grouped, their useless. They're not at all useless. They're more

than what we had. It's just how we interpret them and whether we've gone as far as we can go with the data.

DR. GREENE: Yes, please.

DR. MATTISON: I'd like to take a little bit of issue with your structure activity example --

DR. MITCHELL: I figured you would.

DR. MATTISON: -- in part because I think visual recognition of structure probably isn't the best way of looking at the potential for developmental toxicity. I think just the fact that there are structural elements that are similar doesn't really say very much about how we might want to think about structure and its potential for impact.

DR. MITCHELL: But, Don, would you not agree that even if we took a much more sophisticated approach, until we know the mechanisms by which drugs cause birth defects, whatever commonality we look for among members of a drug class could be a red herring, and that in fact, we don't know exactly what aspect of the pharmacodynamics or pharmacokinetics, the structure itself are posing the risk for birth defects?

DR. MATTISON: I think that's right, and I think basically structure activity right now is in the hypothesis generating mode for the most part.

DR. MITCHELL: Which is terrifically valuable. It's not as though we don't need hypotheses.

DR. MILLS: This is just a comment. I think there's one other group that we need to just keep in the back of our minds in terms of your curve there, and that's even farther over to the left. I'm thinking in my own experience of three young people who suddenly developed Creutzfeldt-Jakob disease, which is not a disease in young people, and all of them turned out to have received pituitary growth hormone. Or children who came in to the ophthalmologist's office with cataracts and I think the mothers were smart enough to say, gee, you had rubella during pregnancy too.

The reason for mentioning it in this context is just that while you're doing the registry or while you're thinking about doing a registry, sometimes the answer is there even before you start, and the key thing is to be tuned into the very rare outcome in the wrong group or the very rare outcome with the very rare exposure.

DR. MITCHELL: I absolutely agree with you. I think that again if you look at history, the majority of teratogens, the majority of adverse drug effects are identified alert clinicians. Let's not kid ourselves. Epidemiology is really not the ultimate first line. The alert clinician is, whether it's DES or growth hormone issues. I agree with you.

DR. SHARRAR: The two examples that you used,

the cohort analysis and the case-control surveillance, require two very different mechanisms for collecting case information. The cohort study is consistent with the pregnancy registry concept, as we've been discussing it today. The case-control surveillance is a very different mechanism and is actually I think conducted in an entirely different fashion.

If you were to go back to the cohort study, could you come up with a number that you would use that once you, say, sampled or collected enough information on a certain number, you feel reasonable that at least we're not dealing with a drug like thalidomide? Is there some reasonable number that you can come up with that's reasonable to try to collect information on?

DR. MITCHELL: I hadn't thought about it. I would guess under 100. If we're using the example of thalidomide or Accutane, I think 100 gives us a fairly comfortable cushion. If I did my math, that would probably identify 10-fold risks for overall malformations quite nicely. Am I answering your question?

DR. SHARRAR: Yes. I was just trying to get some idea in terms of if we were to, quote, conduct pregnancy registries on a number of different compounds or drugs that are out there, how long do they have to go on and what's a reasonable number of information you collect