

# Committee Questions to Dr. Platt and Dr. Graham

## JOINT MEETING OF THE ARTHRITIS ADVISORY COMMITTEE AND THE DRUG SAFETY AND RISK MANAGEMENT ADVISORY COMMITTEE

February 16-18, 2005, Hilton Gaithersburg, 620 Perry Parkway, Gaithersburg, Maryland.

Highlights.....	1
Presentation Text.....	6

### Highlights

- **MAKING INDIVIDUAL DRUG RECOMMENDATIONS WHEN DATA ON SOME DRUGS ARE RANDOMIZED AND ON OTHERS IS OBSERVATIONAL:** Dr. Shafer said that coming up with “some sort of common warning as a class” would communicate “no relevant information”. However, to give “individual drug-specific recommendations” would have to be on the basis of controlled trials for the COX-2 drugs and on the basis of the less convincing observational studies for the older drugs. Dr. Graham gave a detailed answer that implied that one has to make decisions on the data that are available, even if these data are from observational studies. One should “weed the garden of the bad actors” (he mentioned indomethacin as a bad actor) and “shift the market” towards those that appear to have less “risk in the totality of their evidence”. Later in the discussion, Dr. Abramson said that it is not necessarily bad to “call attention to this class effect” since it may stimulate doctors to do “the simple thing of checking blood pressures”.
- **POSSIBLE BIAS IN INTERPRETING DIFFERENT OBSERVATIONAL STUDIES:** Dr. Friedman commented that because of the known limitations of observational studies, “it is easy after the fact to critique away those whose results you don't much care for”. Dr. Elashoff later commented that “just because you put a lot of variables in some model doesn't necessarily mean that you have adequately removed the confounding effects even of those variables”. She also pointed to the omission of the Ingenix data in slide 13 on “excess population risk” that makes it “a biased presentation”. Dr. Graham said “OK, fair enough”. Later in the discussion, Dr. D’Agostino later expressed concern that Dr. Graham criticized some studies as bad because of confounding but not others that had the same potential confounding, saying “Why do you throw out a result you don't like and keep all the results you like?”
- **PATIENT YEARS ARE DIFFERENT FROM ACTUAL DURATIONS OF THERAPY:** Dr. Friedman also that using “patient years of exposure” can obscure the

fact that most of the patients may only have a few weeks of therapy. So his preference is for randomized studies that do have long exposure in individual patients. During another detailed answer, Dr. Graham said ‘I think the epidemiologic data, in my mind at least, answers the question about when the effect begins’ but he did not elaborate on this position.

- **RANKING ROFECOXIB RISK VERSUS OTHER CV RISK FACTORS:** Dr. Bathon asked where the cardiovascular risk of the high dose of rofecoxib ranked among the known cardiovascular risk factors. Dr. Graham said his “ballpark” would be “probably more significant than smoking or diabetes or hypertension, maybe more important than the combination of several of those factors in a patient.” For the lower dose he estimated “probably more than hypertension, a little less than diabetes, and a little less than smoking”. He conceded that Dr. Bathon and Hennekens knew the risk factors “much better than I do.”
- **NASAIDS APPEAR AS RISKY AS COXIBS FROM OBSERVATIONAL DATA:** Dr. Abramson challenged Dr. Graham on his assertion that “the coxibs were more risky” since Dr. Graham also said that both indomethacin and meloxicam have “a risk” (Dr. Abramson also mentioned the McDonald & Way paper that ibuprofen had a 2-fold higher mortality). The message he got from Dr. Graham’s presentation was that cardiovascular risk is dose-related for both coxibs and non-selective NSAIDs. Dr. Graham said ‘I think your observation is correct’ and

“you do need to look at it drug by drug”.

- **RECALL BIAS AND REVERSE RECALL BIAS:** Dr. Day commented about “recall bias” (“flashbulb memory” can be “notoriously false” and there are problems with “eyewitness testimony”) and “reverse recall bias”. She said “it is not trivial how you ask people questions”. The actual questions asked are often not included in the publications.
- **ADJUSTMENT FOR INDICATION AND DOSE-INDICATION INTERACTION:** Dr. Gibovsky asked if Dr. Graham stratified by indication (rheumatoid arthritis has increased cardiovascular risk) and factored in dose-indication relationships (since higher coxib doses are used in rheumatoid arthritis). Dr. Graham said that in most of the studies there were very few patients with rheumatoid arthritis so that it would not “materially affect things”. The California Medicaid study was “limited to patients who had diagnoses of osteoarthritis or rheumatoid arthritis” and rheumatoid arthritis did not “seem to affect things”. He did not respond to the question on dose-indication relationships.
- **SURVIVOR BIAS:** Dr. Gibovsky asked Dr. Platt about “the concept of survivor bias” in relation to differences in the amount of previous NSAID exposure and the recent study “suggesting that discontinuation of an NSAID may itself be a risk factor for a thrombotic event”. Dr. Platt agreed with the point saying that “if we start the clock after a person has already been

exposed to a drug or to one that has the same effect, then, it is very much less likely that those individuals will have a problem.”

- **TIME FROM NEW USE VS. TIME ZERO:** Dr. O’Neill (FDA) asked Dr. Graham if the observational studies were able to “identify new initial use, and then track continued use for that individual” so as to allow comparison of the hazard in randomized trials with the hazard in the observational studies. Dr. Graham said that “time on drug” was available for some cohort and nested control studies, and the Wayne Ray cohort study included “prevalent and incident users” allowing a “new user” subanalysis. However, none of the studies presented data as “a survival analysis” which he thought “is what Dr. O’Neil would like to see”. Dr. O’Neill said “my question is not so much in survival” but he does not think that these studies were designed to “define any time from new use, which is essentially critical to when those risks start”. Dr. Graham said that “time zero for rofecoxib was identified” and that looking only at the data in the 30 days since time zero, the increased risk with rofecoxib was still seen. One could also know that an increased risk ~~æen~~ occurred before 18 months if no one in a study had 18 months of therapy.
- **RISK MIGHT BE BASED ON RISK FROM DISCONTINUING PRIOR NSAIDS:** Dr. O’Neill said that if “discontinuation from an NSAID alone raises risk”, the increased risk following initiation of coxib therapy could merely reflect the discontinuation from the NSAID.

This would not be a problem in randomized trials where there is a comparator group that has similar NSAID discontinuation. Dr. Graham said that “even in the clinical trials, study 090 was 6 weeks long, 12.5 mg, and it had a cardiovascular effect”.

- **SIGNIFICANCE OF LEVEL OF RISK:** Dr. Farrar asked Dr. Graham “what level of risk is acceptable” bearing in mind that “treating pain or not treating pain and not treating the disability of arthritis also has very serious risks, even of death”. Dr. Graham said that a reduced GI risk would contribute to such a decision but that the data are conflicting – although only rofecoxib has the GI protective claim, in two “large” published studies celecoxib was associated with a “lower rate of hospitalization for GI than rofecoxib”. We “actually know very little about the actual population benefit of any of these products” but “The case fatality rate for myocardial infarction in the United States approaches 40 percent. The case fatality rate for hospitalized GI bleeding is probably somewhere around 5 or 10...”. Dr. Nissen said that a “hazard ratio of 1.5 to 2” is “equivalent to raising a cholesterol from 200 to 260, or taking up smoking”. The drugs that are most effective in reducing cardiovascular morbidity and mortality are the statins, and they “reduce risk about 35 percent. So, a hazard ratio of 1.5 to 2 is really a very, very big effect when you are talking about the most common cause of mortality”. Dr. Boulware later suggested that even with a risk approaching 2.0, there might be some patients for whom

that risk is acceptable, for example patients for whom physical impairment as a result of a lack of pain drug therapy could increase risk. He mentioned data showing that mortality increases with increasing functional impairment in rheumatoid arthritis patients.

- **IS RISK RATIO A FUNCTION OF DURATION OF THERAPY?**

Dr. Nissen asked if the lower apparent risk in the observational studies might represent a somewhat increased hazard with short term therapy, but that the larger degree of risk in the long term randomized studies might represent an increasing hazard ratio with longer duration of therapy. Dr. Graham thought it was “more likely” that the “hazard is the same” but there are insufficient events early to show the effect. He also suggested that the lower apparent risk in the observational studies might reflect the higher “misclassification of exposure” (with more intermittent therapy) and “misclassification of outcome” than in randomized trials.

- **STRONG SIGNAL IN RANDOMIZED STUDIES REDUCES VALUE OF OBSERVATIONAL STUDIES:**

Dr. D’Agostino said he had “spent a good part of my career in the Framingham Heart Study” which is a cohort observational study but he was concerned that, since we already had “very strong” randomized studies (the APPROVe study and the APC study), the observational studies did not add much. Dr. Graham said that the observational studies were the only ones with adequate power to demonstrate the increased risk during early therapy.

- **RISK AS A FUNCTION OF AGE:** Dr. Boulware asked if there was evidence that the level of risk was age-dependent. Dr. Graham said that “Nobody in any of the studies where they have looked at it” suggested “that the level of risk differs at different ages”.

- **IMPACT OF MISSED DOSES ON RISK:** Dr. Cryer asked if there was a way in observational studies to take into account missed doses, since a missed dose results in antiplatelet and other effects being “immediately reversed”. Dr. Graham said “No, there isn’t” and the impact of missed doses would be to bring the risk ratio closer to 1.

- **NSAID COMPARISONS VALID BECAUSE PATIENTS NEED TREATMENT:** Dr. Temple made two points: The first was that people were going to get “one drug or another” for their chronic pain so that “comparisons with other NSAIDs seems like as good a comparison as we should make”.

- **OVER-INTERPRETATION OF SMALL RISK RATIOS:** Dr. Temple’s second point was that differences of less than 2 (or even less than 3 according to “Jerry Cornfield, who sort of invented all this stuff”) are generally agreed to be hard to interpret, but “we are talking about differences here that are 0.1 differences, not that they wouldn’t be hugely important if they were true”. Yet “just as an example, there is a very great consistency that you cite that celecoxib looks sort of okay, but you found one study where there is a little hint that maybe the higher dose is a problem, and since probably we all think dose response is likely, that looks good to you.” Dr. Platt said “a

relative risk of 3 in an epidemiologic study, as David found” (for high dose rofecoxib) “is meaningful”. Dr. Temple said “I would not dispute that at all”. Dr. Platt said that data on lower doses “gains weight by borrowing” from the clearer data at higher doses, and that replication of results “in a number of different environments” where similar biases are less likely is also important. Dr. Graham said that we are reaching the limits of “the available tools we have to define the levels of risk that we are talking about”. He agreed with Dr. Platt that “consistency across different studies” was important but that “some light in a storm is probably better than no light in a storm”. Dr. Temple asked if Dr. Graham was saying that “very low hazards need at least multiple support before they are credible”. Dr. Graham agreed.

- **INCREASED RISK RATIO WITH LONGER DURATION REQUIRES LONGER STUDIES:** Dr. Temple said that new data alters one’s thinking. He gave the analogy of the antiarrhythmic drugs which after the CAST study had to show “that you don’t alter survival unfavorably” with the result that “there are hardly any being developed”. So, if you form a new hypothesis that the risk ratio increases with long term therapy, you have to study drugs for longer durations. Dr. Graham said that larger short term studies could provide the same precision in evaluating risk. Dr. Temple said that this would not be the case if the increased risk involved something like “a small, long-term increase in

blood pressure” for which the effects would take longer to develop.

- **CONCERN ABOUT UNPUBLISHED NON-PEER-REVIEWED STUDIES:** Dr. Stemhagen expressed concern about interpreting unpublished studies without peer review.
- **DEFINITION OF INCEPTION COHORT:** Dr. Stemhagen also asked about Dr. Graham’s definition of “inception cohort”. Dr. Graham said that “Inception cohorts are where people enter the cohort with their first-time use of a specific agent, so it’s basically like an incident cohort, it’s new users. That is to be distinguished from a prevalence cohort where starting January 1st, everybody who was on an NSAID is in our cohort.”
- **MOST PATIENTS ARE NOT DE NOVO NSAID USERS:** Dr. Stemhagen said that most patients “are really switching” from previous drug therapy rather than first time users of NSAIDs.
- **OBSERVATIONAL STUDIES DID NOT ADDRESS IMPACT OF UNDERLYING DISEASE:** Dr. Stemhagen also pointed out that the observational studies did not “tease out” the impact of the underlying disease, such as colon polyps or Alzheimer’s in which increased risk versus placebo had been shown in the controlled trials.
- **EVALUATION OF ABSOLUTE RISK DIFFICULT IN OBSERVATIONAL STUDY:** Dr. Fleming said that it is difficult to conclude that there is increased hazard of a drug (and he specifically mentioned naproxen) versus no therapy in an observational study since the “non-use people” may have

been “intrinsically better”. Dr. Graham agreed that this was an issue and “you adjust for all the confounders you are able to measure” but “there still could be effects that you cannot remove”.

- **CCB SAGA ARGUES AGAINST PREMATURE RESPONSE TO APPARENT RISKS IN OBSERVATIONAL STUDIES:** Dr. Hennekens pointed out that 10 years ago “a large body of basic science, clinical studies, case-control, and prospective cohort studies consistently showed that patients with hypertension prescribed calcium blockers had 1.5 to 2-fold increased risk of MI” and he

interpreted Dr. Graham’s position as being that “you would have asked the agency to withdraw the drugs” (even though later randomized studies showed no increased risk). “Protecting the public from harm” is “simple and straightforward” but may not have the effect of “doing the most good for the most people”. Dr. Graham responded that “when you are faced with a large risk that affects large numbers of people, and has a large consequence, that you don’t have the luxury of time to wait 10 years to get clarification on the issue, and you have to use what data you have available at the time.”

## Presentation Text

DR. WOOD: From the committee, we have questions. Let’s start with Dr. Shafer.

DR. SHAFER: Dr. Graham, tomorrow we are going to be asked, as a committee, to consider the question about a class effect for the selective COX-2 antagonists and for the non-selective NSAIDs. One of the things that I am finding, that I am having trouble putting together here, is we have a lot of conflicting data, and for the COX-2 antagonists we have a lot of data from randomized controlled trials. Certainly for the NSAIDs, we are going to have to go with a lot of these observational studies because we don’t have a lot of data on the topic at hand from randomized controlled trials. As I look at this, if we come up with some sort of common warning as a class, and it

applies to everything, we have, in fact, communicated no relevant information. On the other hand, if we are going to come up with individual drug-specific recommendations, we are going to have to have very different evidentiary standards in some ways, because for some of these, we have very little information, as you pointed out, and yet your data, particularly the unpublished data from the Medi-Cal trial, and I appreciate that there is all the issues of not being previewed and stuff, but we are all familiar with that process and know how it works. What can you tell us to guide us? Should we try to go drug by drug specific? How do we set our evidentiary standards when we talk about class effects where in some cases, we are just not going to have a lot of data here?

DR. GRAHAM: Right. What you are going to be getting now, of course, is my opinion, not FDA's opinion. Probably if you were to talk to Bob Temple or John Jenkins, or anybody else, everybody is going to have a slightly different answer. What we are talking about now I think to some extent is philosophy, so what that preamble, first, I believe based on the evidence that there is a COX-2 effect and that that COX-2 effect is dose dependent, and that we see evidence of that with rofecoxib, with celecoxib, and with valdecoxib. The difference between rofecoxib and the other two coxibs on the market is that a safe dose for rofecoxib wasn't identified, the dose wasn't low enough. That raises a question in my mind about what is an appropriate therapeutic index for a drug.

I am giving you my opinion now, but when I listened to Dr. Cryer's presentation yesterday, the bottom line conclusion I came to at the end of that was there really doesn't appear to be a need for COX-2 selective NSAIDs based on what I heard yesterday. There is probably other information out there why I am wrong, but that was the conclusion I came from. So, in any event, that is answer one. I believe there is an effect and it's dose related, and with celecoxib and valdecoxib, I think we have evidence. You said before we have a good evidentiary base based on clinical trials for the COX-2s. I would challenge that in the sense of the survival curves and the things that I talked about there, that we have a very weak evidentiary base for things like protective, you know, is there a grace period for use, and also on the dose issue, we really don't have a great evidentiary base. But that being said, you understand me.

Now, for the non-coxib NSAIDs, my own view is that as an epidemiologist first, I try to report the phenomenon I observe and leave it to brighter minds to figure out why what I observed happens. You are asking me sort of: What do I think is happening underneath it all? I am attracted to the COX-2 hypothesis personally. Dr. Gurkiepal Singh, my colleague and co-author in Medi-Cal, he has a different view on that, but I think that we can have these in vitro tests that say, oh, this is the COX-2 selectivity of this NSAID, you know, in a test tube. What happens in the human body could end up being surprisingly different. We saw yesterday that the dynamic response of these differences, that the platelet effect is very quick, the thromboxane effect is a very quick effect, the prostacyclin effect seems to be a more gradual effect, that this creates very complex interactions that ibuprofen, that any of these drugs could, in the end, end up with a deficit, a prostacyclin deficit that results. I think Dr. FitzGerald showed that slide yesterday with the normal distribution of the time area under the curve and then this little sliver where they are not protected, and that may be the reason why, for these different drugs, that we end up with these different relative risks and these different odds ratios.

In the end, for the non-selective NSAIDs, my own advice would be let's look to see are there somewhere in studies--it is going to be observational studies--in observational studies that we believe have been reasonably well done. By "well done," here, they have to be large. The literature is full of really small studies. I mean I could have presented meloxicam studies, 5 patients, no risk. Well, Duh, you know, you have got a

confidence interval that goes from zero to infinity. They need to be large. Look in a systematic way to identify what the body of evidence is. Can we identify bad actors? I believe indomethacin, for example, is clearly a bad actor, and if people looking at the data concluded that, take appropriate action, weed the garden of the bad actors. Try to identify drugs that based on the evidence we have, appear to be less risk in the totality of their evidence, looking for consistency study to study to study, and then, in a rational way, suggest these are the drugs we think that the public should use, and these other drugs, well, then you have to decide do you want them on the market or not. I am not really going to comment on that, but I think that is the approach I would take. I would be trying to sort of identify right off the bat the bad actors and let's get rid of them. Things that look like they may actually be safe, and when I say "safe" now, I mean that they don't appear to have cardiovascular risk, identify them and shift the market towards that, and then deal with the others.

DR. WOOD: Dr. Friedman.

DR. FRIEDMAN: Thank you. Several comments. First, as both Dr. Graham and Dr. Platt have mentioned, observational studies are essential, but they have a number of limitations, and because of those limitations, it is easy after the fact to critique away those whose results you don't much care for as we have seen. But a couple of other points. One, can these particular drugs, their primary use, we are dealing with chronic conditions, conditions that last years, sometimes many years, and so the drugs are intended for use over those many years potentially. Yet, most of the

clinical trials we heard reported yesterday are 12, 18 weeks, a few of them go longer. You mentioned that one of the reasons we didn't see the problems early on may be numbers, and I agree that is potentially it, but the fact is we didn't see problems arise in the studies until 14, 18 months. We often see analyses by patient years of exposure. In this particular setting, I don't know whether patient years are always equal to patient years, and therefore, I guess I would say why aren't we doing more bigger, longer randomized clinical trials for these chronic conditions?

DR. GRAHAM: I am not speaking for the agency now.

DR. WOOD: We got that. Don't say it each time.

DR. GRAHAM: Okay. I think they are incredibly expensive and companies don't want to do them. There is not an incentive for them to do them, and you would have to talk to the people from the new drug side of the house, but the fact is that they are not requiring them. So, that is a very legitimate question. You know, working as an epidemiologist, we try to make do with what is, and so we use the observational data. You are going to get better quality data if you are able to do this, but just to give you a sense of the size of the studies that I think you would need to do, I mean you talked about before that you have the APPROVe study and we see no effect until 18 months, but there was study 090 that was talked about briefly by Dr. Villalba yesterday. It was a 6-week study at 12.5 mg, and it showed a difference, the suggestion of a cardiovascular risk within the 6-week

study at the lowest dose. Now, it's a small study, as well.

But I am just saying that to say that I think the epidemiologic data, in my mind at least, answers the question about when the effect begins. The question is if you want to have--this is the philosophy--how much certainty do you need to make a decision. Right now, when it comes to efficacy, the effect, does the drug work, you are looking at the lower bound of the confidence interval, and you want to see is that different than 1, because if it is, then, I will conclude with 95 percent certainty or greater that the drug actually has an effect. When it comes to safety, you are doing the same thing. You are looking at that lower bound. You want this 95 percent certainty that the drug is harmful. You are presuming that the drug is safe rather than let's presume we want to do no harm to patients. Let's start off at the beginning assuming that the drug isn't safe, and we want to have a certain level of confidence about how bad this drug could be, and that is still tolerable to us. We want to cap the risk. It will be a completely different way of looking at studies for a safety perspective, one that actually gives a priority to safety and it maximally protective of patient safety, just as that high standard for efficacy is maximally protective of patient safety, because by keeping drugs off the market that don't work, I am protecting patients from unsafe drugs, and if I have pneumonia and I am given a drug that doesn't work, well, I get a harm from that. But that's philosophy, and I think it's an outcropping, it's a development, a natural extension of the development of clinical trials in the United States where the focus has always been on efficacy.

DR. WOOD: Let's try and keep both the questions and the answers reasonably short, otherwise, we will be here until after midnight.

DR. GRAHAM: I apologize.

DR. WOOD: That's okay. Let's go on to Dr. Elashoff.

DR. ELASHOFF: First, I have one comment and then one question. In terms of confounding, just because you put a lot of variables in some model doesn't necessarily mean that you have adequately removed the confounding effects even of those variables. The second has to do with Dr. Graham's slide 13, the excess population risk. I note that the Ingenix data has been left out of the bottom category.

DR. GRAHAM: That's right, because for the high dose.

DR. ELASHOFF: Yes, but the negative sign needs to be on the slide, otherwise, it's a biased presentation.

DR. GRAHAM: Well enough. I take that correction. Okay, fair enough.

DR. WOOD: Dr. Bathon.

DR. BATHON: Yes. As we weigh the risk-benefit ratio of these drugs, one consideration is that there are subgroups of patients in which the benefit might outweigh the risk possibly. With that in mind, it would be helpful for us who are not cardiologists or epidemiologists to be able to put the relative risks that we have been seeing over the past day or two in context with all the cardiovascular risk factors that exist. So, for example, if you were take the

presumed relative risk of rofecoxib of 1.5 to 2.0, at least at the higher dose, and put it into some context for us of the 20 to 40 cardiovascular risk factors that exist in a sort of rank order, where would you put the COX-2 drugs?

DR. GRAHAM: For the high dose it would be probably more significant than smoking or diabetes or hypertension, maybe more important than the combination of several of those factors in a patient. For the lower dose, it is probably more than hypertension, a little less than diabetes, and a little less than smoking. I know, David, you know the cardiovascular risk factors much better than I do, and so does Dr. Hennekens, but that would be my ballpark on that.

DR. WOOD: Dr. Abramson.

DR. ABRAMSON: Yes. I want to go back to the question Dr. Shafer asked about if these classes of drugs or this group of drugs could be if there was a hierarchy of risk, and you first answered that you thought the coxibs were more risky, but I would challenge you a bit simply on your own presentation. I would like you to discuss your data, because you then went on to talk about how indomethacin has a risk. Meloxicam has a risk. Based on your data, the message that came through is that there was a dose response risk for cardiovascular outcomes, that we saw it within the coxibs, but we also saw it where the data were available in the non-selective NSAIDs. There are data that we have seen that ibuprofen might increase risk. We didn't talk about the McDonald and Way paper that in cardiovascular discharge patients, people given ibuprofen had a higher mortality 2-fold. So, as the smoke clears, I am not

sure that the simple answer that the coxibs were different was actually supported by your data, nor your ultimate explanation. Can you defend that?

DR. GRAHAM: I think you are accurate. What I was saying was I was referring, I think, to the underlying COX-2 hypothesis and that it is clearer, I believe, and, well, maybe it's an overgeneralization, because we have the n that we are viewing is so small, that looking at rofecoxib as sort of the example where we can see very clearly the dose response at all the levels and its progression, and understanding its mechanism of action, and then seeing similar things with celecoxib and valdecoxib. I think what you are saying is fair. Maybe a better thing to say is, in the end, that you do need to look at it drug by drug. What I was saying, though, in that answer that I gave to Dr. Shafer, I was really talking more about sort of the COX-2 mechanism and the coxibs as being, in quotes, "COX-2 selective," but I think your observation is correct.

DR. ABRAMSON: Add to that, that although there is a hazard that we don't accomplish a lot by simply saying the class of NSAIDs may have risk, I think we have under-appreciated that over the last 10 years. It is not that different from the mid-nineties recognizing that there was a class GI effect of these drugs, and that compared to placebo, whether it's hypertension or long-term potential adverse outcomes, this is something that doctors have to be aware of, even the simple thing of checking blood pressures when you put people on any nonsteroidal drug. So, I don't know that it is necessarily a bad outcome to call

attention to this class effect until we get better information on each of these individual drugs.

DR. WOOD: Dr. Day.

DR. DAY: I have a comment about recall bias and reverse recall bias. There is a huge research literature on how memory works both in the laboratory and in the every-day world, and there are two phenomena that have been very heavily studied that I think might be relevant here. One is called flashbulb memory, and the idea is when an emotional spectacular event happens, such as when you first learn that JFK had been shot, or the Challenger blew up, or the World Trade Center had been hit, it is as if the old-time flashbulb from an old-time flash camera went off and captured all the details, and you remember all of those details forever afterwards associated with the event that you might otherwise have just not even noticed or forgotten. So, there is a lot of research on flashbulb memory that shows many of those details are indeed correct, but some are notoriously false. For example, there are accounts of people who remember a certain even with great emotional aspects to it, and they remember listening the world series when so-and-so is pitching and it was the bottom of the 9th, da-da-da, all these details, and when you go back and check the evidence of what was going on, on that day and time, that particular game was not on. So, that phenomenon number one, flashbulb memory, and the second is eyewitness testimony. How you ask a person a question will affect what answers you get. So, if you have in the courtroom, someone who has witnessed a car accident, if the lawyer asks this witness, "Did you see the

broken glass," then, the witness is more likely to say yes than if you ask, "Did you see any broken glass," because the broken glass presumes that there was some, and so forth. So, I take your points seriously about potential recall bias and reverse recall bias, but we would have to look at both, whether there is an emotional component or not. Those who have had an MI, for example, would have that most likely, but also how the questions are asked in these surveys, and it is not trivial how you ask people questions about were you taking any medications or were you taking medication X, and for how long, and what was the dosage, and so on. So, I don't think that these details are always published with the studies, and I would like to encourage people who ask people about their experiences with drugs, take a look at the memory literature for some of these points.

DR. WOOD: Dr. Gibofsky.

DR. GIBOFSKY: Dr. Graham, I am wondering if you separated out your populations based on the indication for which they were taking the drug. I ask that because we heard yesterday, and it's well known, that rheumatoid arthritis is itself a risk factor for cardiovascular disease, and higher doses of coxibs, in particular celecoxib, are usually given to patients with rheumatoid arthritis as opposed to osteoarthritis. So, I am wondering if you look at that in your breakdown.

DR. GRAHAM: Several of the studies that I reviewed have looked at the indication, but in automated claims data, it is very difficult to be sort of be sure does the patient have rheumatoid arthritis, and there are different

algorithms one could use, but in general, what has been found in the studies where they have looked at that, that the prevalence of rheumatoid arthritis in the study populations has been low, very low, and that its impact on the results when they adjusted for it didn't materially affect things. Now, in the California Medicaid study, one difference in that study was that our base population was limited to patients who had diagnoses of osteoarthritis or rheumatoid arthritis. Now, these are diagnoses, and so does that mean that they really had osteoarthritis or rheumatoid arthritis, I don't know, but when we did try to eliminate in that study at least were the people who might be using an NSAIDs for a muscle injury, a short-term complaint as opposed to a chronic illness. In none of those does the presence of rheumatoid arthritis seem to affect things, but again I think the prevalence is pretty low in all of these studies.

DR. GIBOFSKY: One quick question for Dr. Platt, if I might. I need to understand the concept of survivor bias somewhat in that I think there is a difference between a patient who is drug-naive, then put on a drug, and then an event happens versus a patient who may have seen a drug, perhaps seen another drug after that, 3 or 4 agents of the class, and is then switched to another agent and something happens. I think we have talked about remote versus current, but there is also this issue of sequential effect, and I am wondering how you deal with that as a survivor, particularly because of the paper we saw a few weeks ago in the Archives suggesting that discontinuation of an NSAID may itself be a risk factor for a thrombotic event.

DR. PLATT: Your point is exactly right. I think that the concern about survivor bias is that if we think that some people are particularly susceptible, which is almost certainly the case, then, if we start the clock after a person has already been exposed to a drug or to one that has the same effect, then, it is very much less likely that those individuals will have a problem. That may be the explanation, for instance, for the reason that the literature was so badly wrong about postmenopausal estrogens and heart disease, that most of the epi studies started with prevalent users. I think the majority of the studies that we were reviewing here, these were individuals who are known to have had at least a year of prior experience without exposure to the non-steroidals. Your study in Kaiser I know was an exception cohort at least with regard to a year of prior history, but I am not aware that any studies have a longer drug-free prior interval than that.

DR. WOOD: Dr. O'Neil, do you want to comment particularly on this?

DR. O'NEIL: Yes, this is an important point and a lot of things have been covered in Richard's and David's presentation, but one thing I think that is relevant that Richard did not cover, that is, the value of a randomized trial, is the ascertainment and follow-up, and knowing the status of individuals in the sense of who goes off therapy and how long they stay on therapy. That is very critical relative to the time dependency of the risk. It was mentioned, for example, the use in the observational sense of recent and remote and current use. Those are all terms that are nice, but they don't get at the issue that we are

trying to get at with regard to the clinical trials, and that is essentially when does time zero start for you. So, I think the appropriate question to ask is what is the duration of exposure since your initial exposure to the drug, because I think that is very relevant to the interpretation of the three clinical trials that we have, two of which are in placebo-control populations. There is a rofecoxib-naproxen control trial for one year, there is a placebo-control trial in polyp prevention for three years, and there is a placebo-control trial in Alzheimer's disease for four years, and the time dependency from time zero matters as you have seen in the plots. It is relevant to the excess risk calculation. So, I would ask the committee, as well as I would ask David, of the observational studies that you have reported, how many of them are cohort studies, and how many of them are able to identify new initial use, and then track continued use for that individual, so that one could look at the relationship between the hazard rates and the hazard ratios that we are identifying in the randomized trials and match that to the odds ratios that are being reported in the observational studies.

DR. GRAHAM: On one of my initial slides, you can see what the cohort studies were, and in some of the nested case control studies, you are also able to get the time on drug. Actually, in Wayne Ray's cohort study, most of these cohort studies include prevalent and incident users, so they will do what is called a "new user" subanalysis, which is to try to get to this issue of when does time zero begin. We addressed that problem in our study here by the inception cohort design in our base population, so that we can identify what time zero was for the

cases. Now, none of those studies presented data in the form of a survival analysis, which I think in the end, that is what Dr. O'Neil would like to see.

DR. O'NEIL: No, my question is not so much in survival. I don't believe, and again that is why I am asking you, I don't think any of those studies were designed or able to capture the question I am asking. In fact, if I am not mistaken, in the Wayne Ray study, he defined new use, but he did not define any time from new use, which is essentially critical to when those risks start.

DR. GRAHAM: That study isn't cited as one of the studies where we are able to derive that information. This slide was a slide that I presented to show that from the epidemiologic literature, those studies where the investigators had identified when time zero began for rofecoxib use, and they didn't present the data as a survival analysis, but they identified when time zero began and then, in various ways, showed you either what the distribution of the cases were, so that you can see that it was impossible for the risk to have been delayed for 18 months, because nobody in the study used the drug for 18 months, or they parsed time out and looked at the first 30 days of use from time zero, and found the risks that they found down here. But you are right, those studies aren't designed that way, and we haven't had time in our Medicaid study to do these analyses yet, but we have the data to now do the cohort study and time to event, so we will have an opportunity actually within the data to actually compare and look to see exactly the question you are driving at. But I would say that from the published data, in each of these studies, time zero for rofecoxib

was identified and in some way or another, information that I think could be useful to the committee in establishing when does risk begin was contained in those studies.

DR. O'NEIL: Well, the other point here, which is the value of clinical trials, and it was the question that was discussed yesterday with regard to the intent-to-treat analysis, and that is to say to analyze all outcomes once randomized to the trial regardless of whether you want to track the individual to 14 days post-exposure. You can't really maybe get access to this information in the observational studies. That is a conjecture, but it's one or the other biases, and it was interesting to the comment, whether one would believe this or not, that discontinuation, discontinuation from an NSAID alone raises risk. If that were to be the case, that is a different analysis altogether.

DR. GRAHAM: In that actual paper, it could be that people were discontinuing the NSAIDs because they were having chest pain and it was being interpreted as dyspepsia or something, and then they go to have their infarct. I mean you are right about that, but this is the nature of how epidemiology is done, and I can't change it. I didn't make the rules, I am only following them. Nobody is arguing that clinical trials, if they could be large enough, that they would give all of us answers that we would have greater comfort trusting what they are saying. What I am proposing is that we don't have that kind of data in the clinical trials. As large as the clinical trials are, for the questions that this committee is facing, you don't have the data you need, and what I presented is the epidemiologic data, and it is imperfect

and it has its warts, and that is why I would emphasize looking at consistency and trying to sort of derive from that a general sense. I mean does it make pharmacologic sense that you would have an 18-month delay? I mean I guess I suppose it depends on what you think the mechanism of action is for the underlying disease, but even in the clinical trials, study 090 was 6 weeks long, 12.5 mg, and it had a cardiovascular effect.

DR. WOOD: I am happy to facilitate a discussion among the FDA, but I think we would rather hear from the committee right now. Dr. Farrar, you are next.

DR. FARRAR: I think that the recommendations of the committee tomorrow are going to depend on the assessment of the overall risk and the overall benefit of this class of drugs. As a researcher and after all the data that has been presented, I am more than happy to accept the fact that there are serious risks even of death from taking NSAIDs. In fact, though, there are serious risks in taking any medication at all. For some of the NSAIDs, it is cardiovascular risks, for some of them it is clearly GI bleeding. As a doctor, though, who takes care of patients, I know that treating pain or not treating pain and not treating the disability of arthritis also has very serious risks, even of death. Given the extensive work that you have done, on the risk of both the cardiovascular and the GI bleed, I wonder what level of risk is acceptable you, and remembering that the only other drugs that are really available is analgesics or narcotics, and the only other drugs that are really available in terms of limiting inflammation are

biologics or immunosuppressants, I wonder what drug is safe enough that you would recommend that I actually would be able to use it in patients to prevent some of their suffering.

DR. GRAHAM: Well, I am not going to give a product endorsement. A couple of things, though.

DR. WOOD: Try and make it brief.

DR. GRAHAM: One, the benefits of the treatment for the traditional NSAIDs compared to the COX-2 selective NSAIDs with GI bleed, we have clinical trial evidence that suggest that there may be a difference, but here, to me, is an anomaly. Rofecoxib got the indication for being GI-protective, celecoxib didn't based on the clinical trials data you guys looked at yesterday. There are two published studies in the literature looking at what I would say is actual benefit. There, they were looking at hospitalization for GI bleed--they didn't look at death from GI bleed, but I wish they had--but hospitalization for GI bleed, and what they found was, in both of these studies, that celecoxib was actually more beneficial, you know, lower rate of hospitalization for GI than rofecoxib. So, that is the population, two large studies. You have got your clinical trials that would have said it should be the reverse. So, I throw that out as one sort of conundrum. The second is that I don't think that the actual benefits of these drugs are understood well enough to sort of try to weigh these very well. The case fatality rate for myocardial infarction in the United States approaches 40 percent. The case fatality rate for hospitalized GI bleeding is probably somewhere around 5 or 10, it is a much lower case fatality rate. Nobody

that I have seen anywhere has sort of worked this out very well, so I would submit to you and to the committee that you actually know very little about the actual population benefit of any of these products.

DR. WOOD: I don't think we are going to get an answer to that question, so let's move on. Dr. Nissen.

DR. NISSEN: Let me briefly answer the earlier question about what does the hazard ratio of 1.5 to 2 mean. Before I came to the meeting, I made a point to look this up, because I thought it would be very relevant. It is equivalent to raising a cholesterol from 200 to 260, or taking up smoking. Another way for the committee, I mean as a cardiologist I have to deal with this all the time, the most effective drugs we have for prevention of morbidity and mortality are statins, and they reduce risk about 35 percent. So, a hazard ratio of 1.5 to 2 is really a very, very big effect when you are talking about the most common cause of mortality, and that is why this discussion is so important.

Now, my question is this. We are going to be asked to balance risk and benefit, and so the magnitude of the hazard ratio is very important to all of us, and I am trying to reconcile what we see in the randomized control trials with, let's take rofecoxib for a moment, where it looks like the hazard ratio in the randomized trials is in the range of 2, 3, 4, maybe even higher, and in the observational data it is significantly lower. I would like to propose a hypothesis to you and just ask you if you think this is right. In your observational data, you are looking at mostly short-term exposure, so you are looking at less than 12 months typically

of exposure. It may well be that the hazard increases over time, so that by the time you get to 18 months, you can actually see it in a much smaller randomized trial, and so it doesn't rule out the possibility that, in fact, both observations are right, that, in fact, there is an early hazard, but that early hazard has a smaller hazard ratio than the hazard at 18 months or 24 months or even 36 months, and if we ever were to look out 5 years, it might still be increasing. Do you think that is a reasonable hypothesis?

DR. GRAHAM: I think more likely it is, that in your clinical trials, early on you don't have enough power to distinguish the risk. The hazard is the same, but the lines are closer together, because we are closer to the origin. I think one other explanation for the lower risk ratios in observational studies, I would think is more likely due to misclassification of exposure and misclassification of outcome. It is likely to be non-differential, so it would tend to reduce the odds ratios and relative risks towards 1. Exposure, because people are going to take it, a lot of these people are taking it on a PRN kind of basis. In a clinical trial, you have a greater certitude that they are actually taking it every day. That introduces a lot of misclassification, so the a priori hypothesis going into an observational study, with misclassification going on, you are fighting an uphill battle to see an effect.

DR. WOOD: We have got lots of people who want to ask questions. I want to make sure that the people who are asking questions have questions they want to ask for clarification of the

speakers who have spoken rather than just general points. Dr. D'Agostino.

DR. D'AGOSTINO: I have a couple of questions along the way here. I have spent a good part of my career in the Framingham Heart Study, and it's an epidemiological study and a cohort study, and we take joy when somebody runs a controlled trial on hypotheses and then later on confirms it. The first question is I am concerned that even though you have gone through this careful analysis, your conclusions are no apparent effect, probably increased effect, probable increased risk. They really don't help us in the sense of pinning things down. We have a couple of very strong I think good studies, the APPROVe study and the APC study as placebo-controlled trials. Tell us quickly where is the weight of how we should look at these two pieces, the controlled trials we have versus what you have produced.

DR. WOOD: Really quickly.

DR. D'AGOSTINO: Really quickly, it can be done quickly.

DR. GRAHAM: My belief is that for the controlled clinical trials, for the levels of risk that we are concerned about, that they do not have the statistical power early on to show risk differences.

DR. D'AGOSTINO: I think Bob O'Neil's comment is very important here. The other two points, and again I will make them quick, I am very concerned about the high dose effect you have, and I am really concerned about the MI and the number of cases. I mean blood pressure, cholesterol, diabetes, smoking, this is what drives people to have heart attacks

and what have you, and that is completely missing on your assessment of how many new cases, so I guess it is more of a comment that I am really concerned that that sheet needs sobering interpretation.

DR. GRAHAM: But it was based on the odds ratios and relative risks where those factors were adjusted for, so as well as they are adjusted for, that is what the projection represents, the excess after adjustment.

DR. D'AGOSTINO: Yes, but I mean the comment was made by you, throwing in the analysis doesn't necessarily adjust for them. The last one, you made a very nice point about the cardio-protective effect, and you tried to show that these uses, and what have you, somehow or other all have the same risk, and your interpretation that there must be some confounding going on, why doesn't that hold for all the studies you gave, why don't that hold for the Solomon study, which you thought was a great study, yet, this one result you don't like?

DR. GRAHAM: For what, the Kimmel study?

DR. D'AGOSTINO: Wasn't it the Solomon study that had the naproxen as the cardio-protective?

DR. GRAHAM: That is because the cardio protection was present when they were on the drug and when they weren't on the drug.

DR. D'AGOSTINO: I understand what you are saying, but if that's a problem, then, it means there is some confounding going on.

DR. GRAHAM: No, it's selection bias.

DR. D'AGOSTINO: Well, it's selection bias, but why isn't it for the whole study? Why do you throw out a result you don't like and keep all the results you like?

DR. GRAHAM: No, that is not what I did. I pointed out a result where they showed the presence of the selection bias. In other studies, the Ingenix study is the only other study that looked at this. I don't have a slide of it.

DR. D'AGOSTINO: I don't know if it's a selection bias or misinterpretation of the data.

DR. GRAHAM: Well, to me it looks like selection bias.

DR. WOOD: Let's continue that conversation later. Dr. Morris.

DR. MORRIS: David, would you go to slide 14. That is the risk, the duration of use. I think one of your points was that if you look at your study, tell me if I understand this right, that with the lower dose, that the median time to an AMI is sooner than with a higher dose, did I understand that right?

DR. GRAHAM: Yes.

DR. MORRIS: A month?

DR. GRAHAM: Had more cases, a greater proportion of our cases, but the other thing is remember, down here, we are talking about 18 cases or so. The N here is small, the N here is like 58, and the N here is 10. So, I wouldn't read too much into the difference. The more important point is that at the low dose, nobody was out there beyond 18 months,

so all the action happened before 18 months, and the same for the others. I see what you are saying. I can only say that is what our data were.

DR. MORRIS: One interpretation is what you said earlier, that for this particular drug, we are talking about, as you said, no safe level. I was wondering if that is the way you interpreted it, that because we are talking about Vioxx here, and there is no safe level, that something is going to happen sooner, or is it something with the populations are different.

DR. GRAHAM: The populations could be different, but I think, you know, you would expect the higher dose to have a shorter latency to onset than the higher dose, but the numbers are so small.

DR. MORRIS: Okay, it's a small number problem.

DR. WOOD: So, the answer is too small numbers at high dose. Dr. Boulware.

DR. BOULWARE: I just want to make sure I understand something that you had proposed in your excess population risk slide, if you would put that back up. As a rheumatologist, I use these drugs in a population much greater than what you have here with a 65 to 74 where the risk of an MI is fairly high in that group. Did you want us to believe that this excess risk that you are proposing would be extrapolated to other population groups, too?

DR. GRAHAM: Well, no.

DR. BOULWARE: Do you have any numbers that may demonstrate that?

DR. GRAHAM: Well, the answer to the second is no. This was an example in conversation with people planning the talk, to try to help people connect with what it means. Cardiovascular risks go up. I mean in the next age group higher, the risks are higher. In the age groups lower, they are lower, but cardiovascular risk begins to increase in the 40s.

DR. BOULWARE: I understand, but it wouldn't be a linear type of thing.

DR. GRAHAM: No, the background risk isn't linear, the relative risks, though, are adjusted out.

DR. BOULWARE: Because one of the questions we will be faced with is are there subpopulations or groups that these may be safe in, and I just want to make sure I understand the relative risk in different age groups.

DR. GRAHAM: Nobody in any of the studies where they have looked at it have reported effect modification, which would be that the level of risk differs at different ages.

DR. BOULWARE: One more question here. I want to make sure I understand. I think I heard a comment that says when the risk approaches 2.0--maybe I just assumed that you said this--that it was an unacceptable level of risk. Is there ever a case where a drug may have a clinical benefit in which that risk is acceptable, because for the patients I see, not giving them any of these drugs will confer a great deal of risk on them, and physical impairment. And we have studies that show that the functional classification of rheumatoid arthritis patients carries with it a significant mortality as that class goes up?

DR. WOOD: I think that is a question for the committee to answer rather than Dr. Graham. Let's move on to Dr. Cryer. Do you have a question?

DR. CRYER: I do. The comment and question I have of Dr. Graham addresses an issue that I think is an important difference between the observational studies and the prospective studies, and this difference relates to assessment of drug compliance and missed doses, and I think it is critical as it relates to assessing drugs which potentially affect platelet function. A huge difference, as you know, between aspirin's effect and every other NSAID including the COX-2 inhibitors, is that with the non-aspirin NSAIDs, as soon as you remove the drugs, whatever potential effect they would have had on the platelet are immediately reversed. So, with naproxen specifically, my preconceived bias, which may be wrong, but my preconceived bias based upon everything I know about the pharmacology and the things that Dr. FitzGerald has reviewed for us, is that it should have some mild anti-platelet effects which would only be present when the drug is on board in the system. So, the specific question is, in the observational studies, recognizing that in clinical practice people miss doses of their NSAIDs, they are not taking their NSAIDs consistently, how do you account for the missed doses in the observational studies recognizing that this could potentially lead to a mitigation of whatever negative effect or positive effect that they may have?

DR. GRAHAM: It ends up being misclassification. Generally, what that means is it will force the observed level of risk, the relative risk of the odds ratio

closer to 1. So, if we had an increased risk, it would make it lower, if we had a protective effect, it would sort of make it higher, closer to 1.

DR. CRYER: Right, we agree on that. The specific question is, is there a way to actually recognize or to account for when people do not take their doses in the observational databases?

DR. GRAHAM: No, there isn't, so when you are studying, say, an increased risk, that is why I said if you find something, you have to realize you found it despite the misclassification.

DR. WOOD: Okay. Dr. Domanski.

DR. DOMANSKI: I will save it for tomorrow.

DR. WOOD: Okay, great. Dr. Furberg.

DR. FURBERG: No.

DR. WOOD: Okay, great. Dr. Temple, who does speak for the FDA.

DR. TEMPLE: I am just asking questions. A couple. Actually, one point is it seems to me that since we expect that people are going to be getting one drug or another, comparisons with other NSAIDs seems like as good a comparison as we should make. You might want to leave out indomethacin if you are worried about it. That's one thing. I guess my main question, though, is everybody has paid appropriate lip service to the idea that very small differences are hard to interpret in epidemiology. People have said 1.5, 2. Actually, I notice in one of his editorials, Dr. Furberg cited a paper of mine where I said anything less than 2 really needs a

lot of questions. Jerry Cornfield, who sort of invented all this stuff, used to say 3. Well, we are talking about differences here that are 0.1 differences, not that they wouldn't be hugely important if they were true, that is absolutely true. So, I guess I want to know what Richard and you make of all this, because the numbers are very small, and yet, just as an example, there is a very great consistency that you cite that celecoxib looks sort of okay, but you found one study where there is a little hint that maybe the higher dose is a problem, and since probably we all think dose response is likely, that looks good to you.

DR. GRAHAM: Two studies, there were 2.

DR. TEMPLE: Okay, 2. The valdecoxib data, which shows nothing, doesn't look so good because we probably all believe that there is likely to be a class effect. What I am asking is, with numbers like this, how do you know what to do with them? That seems very fundamental for the epidemiology.

DR. WOOD: But, Bob, there are 4 randomized clinical trials here, and your comments don't apply to them, I assume.

DR. TEMPLE: No, they don't, although they are not perfectly consistent either. But, no, I am asking, what do we make of differences of this magnitude with everybody having given lip service to the idea that small differences are hard to interpret, and yet we seem to be enthusiastically endorsing them, so I just want to know what Richard and David think about that.

DR. GRAHAM: Rich, do you want to go first?

DR. PLATT: I think we have to be cautious about how we interpret it, so I would say the finding of a relative risk of 3 in an epidemiologic study, as David found, is meaningful--

DR. TEMPLE: For high dose rofecoxib.

DR. PLATT: For high dose rofecoxib.

DR. TEMPLE: I would not dispute that at all.

DR. PLATT: It seems to me that in that context, that a dose response effect, that the information about lower doses gains weight by borrowing from that. I think that is also worth keeping in mind when, in other studies that are working in that range that make us all nervous, there appears to be a dose response effect. It is the kind of consistency that makes the study, in my mind, be worth more attention. I think there is something to be said for giving more weight to relatively small excess risks if they are seen in a number of different environments when we can't have good reason to think that there is a similar kind of biases that might be contributing to it. After that, I agree with you. We are in relatively difficult terrain. I think that it is not the same as no data, though. I think we ought to distinguish between the situation in which we have no evidence from ones in which we have relatively weak evidence. We didn't talk at all, for instance, about the enormous number of spontaneous reports of myocardial infarction following exposure to non-steroidals. There are thousands and thousands of them. In my mind, they don't contribute at all to the discussion,

whereas, I think these need to be weighed in the mix when we don't have clinical trial information to depend on.

DR. GRAHAM: My answer is similar to his, but I think that what you are identifying is, is that we are hitting or at least right now the frontier is the limits of what the available tools we have to define the levels of risk that we are talking about. We are talking about small levels of risk that turn out for this particular event to be enormously important in a population level. If you are talking liver failure, we wouldn't be having this conversation. For that reason, it becomes important and what I would say is sort of emphasizing what Rich said, is I would be looking for consistency across different studies, and if I found a number of studies, say, as with Indocin, for example, to me, that is more persuasive. If I found a number of studies that pointed to a particular set of NSAIDs that seems to have low risks, I would take comfort in that in the absence of perfect information. I mean some light in a storm is probably better than no light in a storm.

DR. TEMPLE: I take it if the differences were at the level of 10 percent, 1.1 versus 1.2--

DR. GRAHAM: I am thinking more in a very qualitative sense of things that they seem to cluster around 1. I mean 1.1 for ibuprofen, it could be that, for example, may naproxen increases the risk 3 percent in the real world, we are never going to figure that out, maybe ibuprofen increases it 10 percent or 15 percent, maybe we could figure that out, I don't know, but there is going to be a place where qualitatively, if we see enough studies kind of sort of pointing

to the same place, you know, most of them, they are not all going to say the same thing, there is going to be these conflicts, just like we have in clinical trials data. But if most of the compass arrows are sort of pointing in the same direction for particular NSAIDs, I think those are the ones that at least that I sort of place on a suspect list.

DR. TEMPLE: So, very low hazards need at least multiple support before they are credible.

DR. GRAHAM: I think so, and I think that you want to try to encourage to collect that information sort of to test that out.

DR. TEMPLE: Alastair, could I take half a second to answer a question Larry raised before?

DR. WOOD: Sure, a second.

DR. TEMPLE: Well, it's a very good question, you know, if the drug is going to be used forever, why don't you study them forever. The only thing I would point out here is that what sort of started people thinking was VIGOR, and VIGOR didn't take 3 years to show anything, it showed up in 9 months. So, what you have seen is for, say, lumiracoxib, a humongous study of about the same length, but, of course, they didn't know about APPROVe, did they, and whatever you think APPROVe means, whether Bob is right that it's late, or David is right that there weren't enough cases, people were pointing toward a study that by every reasonable thought, if you think platelets are involved, ought to be long enough to show things up. But then you form a new hypothesis once you have

APPROVe, and you have to adapt it, and I think that goes on all the time. It would not be I must say for most things my first thought unless you are looking for cancer that you need a 3-year study to find it, but maybe you learned that it does. Just for what is worth as an example, you can't get an anti-arrhythmic drug approved in this country without showing that you don't alter survival unfavorably. One result is there are hardly any being developed, but, you know, we had bad experiences, we didn't like the results of CAST, so you change. I think there is no doubt that things evolve and you have to expect that, and APPROVe, depending on what you think of it, changes the nature of what you expect.

DR. GRAHAM: Bob, just one point on that. I think if the APPROVe study had been 5 or 10 times larger than it was--I am talking about retrospect now--you would be able to answer with much greater confidence what is happening month 1 to 18. I guess what I am saying is that you could also shorten the latency to identification of a problem if it turns out that the risk is early on.

DR. TEMPLE: David, I think that is entirely possible, and if it involves platelets, I would believe you, but if it involves a small, long-term increase in blood pressure, then, I am not so sure.

DR. GRAHAM: Right, but we saw yesterday--

DR. TEMPLE: We don't know.

DR. GRAHAM: We don't, but if it's prostacyclin, that effect could occur immediately.

DR. TEMPLE: Yes, but the blood pressure effect could be delayed.

DR. WOOD: Right. So what, Bob, you are saying is that it is easy to be a Monday morning quarterback, but the data were not there before.

DR. TEMPLE: I would never be that rude.

DR. WOOD: I think you are right. Dr. Stemhagen.

DR. STEMHAGEN: I would like to clarify a couple things. First, I am a little concerned in terms of the unpublished data. I appreciate that we are able to get data very quickly, right at the minute that it is being generated, but none of us have had a chance to really review that, so I do have some concerns about the weight putting on this unpublished data when the rest of us haven't had a chance to look at it. I think there needs to be some clarification. There was some discussion about the recall bias, and so on. Certainly, there is a major concern about that in case- controlled studies, and we don't have the questionnaires, but there were a lot of sort of subanalysis done in the Kimmel study, about trying to look at whether recall bias is a problem, and I am not sure that you have highlighted that enough that looking at all those different things, there were really no differences found. Similarly, in the Watson study, it's a GPRD study, it is different than a lot of the large databases, the automated databases. There is a lot more personal involvement in terms of the data and the data collection and the adjudication of results, and I think it just needs to be clear that all of these studies are not the same in terms of a Medicare study where we

can't go back and validate records. A lot of them had a much more careful review, and I am just not sure that that was totally clear and if you hadn't read each of the papers. I would like to just ask a question in terms of your definition of the inception cohort, if you could just go over that again, because of your comments about the short-term use.

DR. GRAHAM: Inception cohorts are where people enter the cohort with their first-time use of a specific agent, so it's basically like an incident cohort, it's new users. That is to be distinguished from a prevalence cohort where starting January 1st, everybody who was on an NSAID is in our cohort. Some of those could be people who were on it before January 1st, and others could be people who start an NSAID after January 1st, so you are mixing people who are prevalent on the drug, who may have survived, or whatever, and people who are newly starting it. In those types of cohort studies, a new user analysis was designed to focus on those people who, during the study window, were new initiators of the particular drug under study, so that time zero could be identified for those people. That is what Alec Walker & Company did in their Ingenix study. It was a prevalence cohort, but they did a new user analysis in which they identified new users, and it was that new user analysis that showed the 1 to 30-day increased risk. Wayne Ray did the same thing in terms of new user analysis, and in our study, the nested case control, everyone was an inception user in the base population.

DR. STEMHAGEN: I guess just a comment in terms of people thinking about clinical trials where we have washout periods, is that people are really

switching. If they are RA or OA patients, they are not starting new with the drug, they have been on something for a long time, and they are switching. So, we have to think about those risks in terms of the weight we are putting onto that inception cohort, as well. I guess the last point is based on the question that Ralph had about the other studies. I just want us to keep in mind also that a lot of those studies come from very unique populations - the randomized clinical trials, the colon polyp study, and the Alzheimer's disease patients, so are very different. We can't tease out in any of these observational studies whether we have patients that meet those criteria or have those indications, as we also pointed out.

DR. WOOD: Tom.

DR. FLEMING: I think Drs. Platt's and Graham's presentations were informative, but with certainly a lot of complexities for methodologic issues that I assume tomorrow, we will give our perspectives about, so let me ask a question and then a clarification.

The question relates to the slide on the 4 positive naproxen studies, I think slide 22.

While you are getting that, just very quickly, these large linked databases certainly are very useful from the perspective of getting defined populations with numerators and denominators, but have many challenges that people have been talking about along the lines of lack of randomization, no confounder information, specificity and sensitivity.

Bob O'Neil got at a point that I think is critical, and that is the complexity of not having a time zero cohort with the ability to do what would be the analogous ITT analysis with complete follow-up or minimize loss to follow-up. You bring out in the Solomon example there, David, a very nice illustration of this very point that you recognized, which is the selection bias that can go on when you are characterizing people into these groups, and it's misleading to think that you are really seeing the causal effect of any use versus current, versus recent, versus remote, the causality could be going in the other direction. Intrinsic differences in patients could be influencing whether they are, in fact, in those four categories.

But don't you, in essence, even though your conclusion might be right, aren't you, in essence, doing the same thing at the top when you are looking at naproxen, say, when you are looking at other NSAIDs, it is protective, but you don't know whether it's, in fact, truly the harmful effect of the other NSAIDs, so you try to get in a non-use population, you are trying to simulate a placebo, but how do you know that those non-use people weren't intrinsically better? Isn't it the same issue?

DR. WOOD: I think we have had this discussion.

DR. FLEMING: But this is important, I want to get his views, because it's important for naproxen.

DR. WOOD: Okay.

DR. GRAHAM: There is no perfect reference group. It turns out that this non-use group is really they are remote

users, but it is a question and I can't answer it except to say that when you adjust for all the confounders you are able to measure, you try to remove those effects, but there still could be effects that you cannot remove. The data are what the data are, and here what I was trying to show is that based on--if these data were looked at the way most of the other studies were done, it gives a very different result. If it turns out that all of the NSAIDs increased the risk a little bit, the fact that naproxen doesn't increase it as much, could look protected, and you really don't know. The real conundrum is to get an anchor point to help you interpret everything, and there is no perfect anchor point.

DR. FLEMING: Your motivation for wanting to know what the placebo-controlled result is, is clear and justified. This analysis, though, has the same potential flaws as the Solomon analysis. So, the motivation for the question is clear, as you are just restating, but the reliability of the conclusions are suspect for this very reason that you correctly noted, due to the selectivity in the Solomon categorization.

DR. GRAHAM: You need then to sort of generalize that to all of the observational studies, because all of them, you had--

DR. WOOD: Why don't we continue this conversation later, and, Tom, you can present discussion on that later.

DR. FLEMING: Well, there is much more to say, but I will defer to tomorrow.

DR. WOOD: I am sure there is. Dr. Hennekens will be our last question

before the break. Just to encourage you, we will be back here just after 20 to, so make it fast.

DR. HENNEKENS: A question and a comment. Ten years ago, a large body of basic science, clinical studies, case-control, and prospective cohort studies consistently showed that patients with hypertension prescribed calcium blockers had 1.5 to 2-fold increased risk of MI even after controlling for a large number of available confounders. I wrote a JAMA editorial asking for randomized evidence, but I assume, based on what I heard you say, that you would have asked the agency to withdraw the drugs. So, I would ask you to consider whether protecting the public from harm is an optimal goal. It is far more simple and straightforward than trying to maximize benefit and minimize harm, which would do the most good for the most people, but doing the most good for the most people does not, strictly speaking, protect the public from harm.

DR. GRAHAM: Do you want a response to that? Okay. I think that when you are faced with a large risk that affects large numbers of people, and has a large consequence, that you don't have the luxury of time to wait 10 years to get clarification on the issue, and you have to use what data you have available at the time. I think that just as we have imperfect measures of risk, I would submit that we have even more imperfect measures of actual benefit. In the case of hypertension, I think, you know, that has been studied dramatically and we actually know that not all antihypertensives lowering blood pressure at the same amount, confer the same population benefit. I would say

that with this class of drugs, we really haven't even demonstrated--I mean yesterday, the question came up why would a company do a study on polyp prevention, had they thought about what the benefit of this was, and nobody had started to think, well, how many lives are we going to save by giving people these drugs, and I would submit that if you were to ask the agency or ask the company on this, if you don't have a good measure on benefit, so you want to make a benefit-risk assessment. We have measures of risk, they may be imperfect, but I would argue that from a population perspective, you don't really have nearly as good information as you might believe you do from the clinical trials, what the benefit in the population is, how many lives are actually saved by the COX-2s, for example.

DR. WOOD: On that note, I am told the lines are building at the men's room, so we need to be back here at exactly quarter to. (Recess.)