

Question 5: Future Trials for Coxibs

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
Discussion Text: Question No. 5	3

Highlights

- **GET EXPERT GROUP TO DESIGN STUDY:** Dr. Farrar suggested that a group of experts get together to develop a “really good design” and have an “ongoing process with a group of academic advisers”.
- **TWO TYPES OF STUDY – VS PLACEBO AND VS ACTIVE COMPARATORS:** Dr. Wood said that there could be two groups of studies – against placebo and against active comparators. One would choose the comparator based on the indication, and also on “information on what the comparator looks like on its own”. He suggests that it would be good to know how naproxen+PPI versus placebo affects CV risk.
- **COMPARABILITY TO IBUPROFEN MAY NOT BE ADEQUATE:** Dr. Temple questioned how acceptable an ibuprofen comparison would be.
- **WILL PLACEBO-CONTROLLED TRIALS BE FEASIBLE?** Dr. Temple asked how feasible it would be to do additional placebo-controlled trials. If you have to wait until you have a definitive naproxen-placebo trial “we are talking almost never...”.
- **APPARENT COX-2 RISK REQUIRES CAUTION IN TRIALS:** Dr. Wood said a concern is that the COX-2 drugs may be risky, so that “you would be cautious” about doing a trial that would reproduce the toxicity already seen.
- **CELECOXIB/NAPROXEN STUDY SUGGESTED:** Dr. Temple suggested a 1-3 year comparison of naproxen and celecoxib.
- **DON’T USE ANOTHER COX-2 INHIBITOR AS YOUR CONTROL:** Dr. Wood would be “unimpressed” with a “study against another selective COX-2 inhibitor. I think that is likely to be negative”.
- **USE NAPROXEN CONTROL, NOT PLACEBO OR ACETAMINOPHEN:** Dr. Nissen said you could not do a placebo-controlled trial in arthritis, and he did not think an acetaminophen-codeine control group would be “practical”. It will be difficult to interpret the data if every sponsor uses a different

comparator, and he favors using naproxen.

- **CELECOXIB 200 MG/ NAPROXEN 500 MG BID/ DICLOFENAC COMPARISON:** Dr. Nissen suggested a study of celecoxib 200 mg/day versus naproxen 500 mg bid – and it would “make a lot of sense” to add diclofenac as a third arm.
- **ADD IBUPROFEN TO CELECOXIB/NAPROXEN/DICLOFENAC:** Dr. Dworkin largely agreed with Dr. Nissen but suggested a fourth arm using ibuprofen.
- **MULTIPLE COMPARATORS NEED FUNDING BY LARGER GROUP:** Dr. Temple said that a single company might reasonably say that a single comparator such as naproxen should be adequate, so that studies with multiple comparators might have to be funded by “a larger group”.
- **NON-INFERIORITY TRIAL OF COX-2 INHIBITOR:** Dr. Fleming suggested a 2-3 year, OA/RA, non-inferiority comparison of celecoxib and either naproxen or aspirin+PPI using 10,000 patients/arm to provide 90% power of picking up a 50% increase (i.e. an excess risk of 17%) with only a 2.5% false positive error rate. If results were favorable that should be enough to remove the celecoxib black-box. For valdecoxib and rofecoxib where the safety signal is higher, a comparable study would be needed within an acceptable time frame “if they are going to be on the market”. In the case of rofecoxib which may have been studied at excessive dosage, the dose should be lower.
- **NEED TO SHOW SAFETY OF SHORT-ACTING NSAIDs:** Dr.

Hennekens said that “the short-acting NSAIDs” appear “at least as hazardous as the coxibs” and are OTC drugs with DTC advertising. Accordingly, they should not be ignored when planning future studies.

- **NEED TO FOLLOW-UP DROPOUTS:** Dr. D’Agostino said that follow-up is important for patients dropping out of the study “because their blood pressure is building up, they are getting hypertensive, or because of G.I. problems”.
- **OA BETTER POPULATION THAN RA:** Dr. Hoffman suggested that the proposed study should be done on OA(since in RA patients are sicker, are at greater CV risk, and an “analgesic arm” would not be feasible). He suggests that the groups consist of: 1) acetaminophen (with addition of codeine if necessary), 2) an NSAID (such as naproxen or ibuprofen) plus a PPI, 3) the COX-2 drug.
- **PRACTICALITIES OF ALLHAT-TYPE DESIGN:** Dr. Cush supported a 2-year trial using Dr. Temple’s proposed ALLHAT approach. However, he expressed concern about the “impracticalities” of the study suggested by Dr. Temple and Dr. Fleming. He does not think patients would stay on high dose aspirin for the duration of the study and “to provide some modicum of protection by putting a PPI on top of that is not going to be practical”.
- **SAMPLE SIZE OF 10,000/GROUP ANTICIPATES POINT ESTIMATE OF RISK NOT MORE THAN 15% TO PROVIDE CONFIDENCE THAT**

TRUE VALUE IS LESS THAN 50%: Dr. D'Agostino expressed concern that the suggested sample size would only be sufficient to pick up a 50% increase in risk. Dr. Fleming responded that to increase the ability to detect smaller effects would increase the sample size from the 10,000/group he had suggested to

20,000 for a 33% increase and 60,000 for a 20% increase. However, he pointed out that the point estimate of risk would have to be within about 15% of zero effect in order to have the upper confidence limit be below the 50% increase in risk that must be excluded.

Discussion Text: Question No. 5

DR. WOOD: What additional clinical trials or observational studies, if any, do you recommend as essential to further evaluate the potential cardiovascular risks of celecoxib, rofecoxib and valdecoxib. What additional clinical trials or observational studies, if any, to you recommend as essential to further evaluate the potential benefits--reduced G.I. risk--of celecoxib, rofecoxib and valdecoxib. Please be specific with regard to which COX-2 selective agent to study, trial design, patient population, control groups, endpoints, duration, sample size, et cetera. And it is five to 4:00. There is a three-day task right there, it seems to me. Do you really want that before we leave? Bob?

DR. TEMPLE: I guess I was struck by the fact that several of you, but not everybody, said that celecoxib or valdecoxib has to do something to get rid of a certain nasty thing in the labeling, get rid of the box. So it raises immediately the question what would they have to do to do that; comparison with some other drug, not be worse than naproxen? What do they have to do? That is why this deserves some attention. Nobody expects you to design the whole trial perfectly or fully in five minutes.

DR. WOOD: Four-and-a-half, now. Comments on that? How are we going to design a trial? Can we break it out easily? What would we need to evaluate the potential cardiovascular risk if we think there is a cardiovascular risk of celecoxib, rofecoxib and valdecoxib. Comments on that? Yes, Dr. Farrar.

DR. FARRAR: I think this actually begs an issue that we ought to address which is that we cannot possibly, in the half an hour or forty-five minutes that is left do all of the issues that are being requested here. But it does suggest--

DR. WOOD: Did you think you were going home at 5:00?

DR. FARRAR: My mother is down the road. It's fine.

DR. WOOD: She'll be glad to see you tonight. We're all coming. (Laughter.)

DR. FARRAR: But it suggests, in fact, that there needs to be a process that, perhaps, even expands beyond this particular class of drugs to really examine the issue of how the safety of drugs needs to be considered with regards to the patient populations that in whom the drugs are likely to be used and

with regards to the potential uses for a particular drug. I actually would strongly recommend that, for those of us who--I was one of those who recommended that there should be some trials or some studies done to try and remove some of the black-box labeling that, at least, I was in favor of. Rather than trying to design all of that now, what, really, I would suggest is that a group of academic folks made up of some of the people here but, clearly, including people with pharmacoepidemiology, statistical, epidemiological skills as well as the particular specialty skills of arthritis, pain or whatever is necessary, be put together to formulate a really good design based on the type and the discussions here and that the recommendation of this group ought to be that, not just the folks at the FDA, but that there should be an ongoing process with a group of academic advisors to really formulate an appropriate study.

DR. WOOD: I guess I was the person that suggested the sort of "get of out jail free" card if they came out with the studies. It seems to me the studies break into two different broad groups. There are studies that would be potentially against placebo that would establish whether the drug had an absolute risk and there are studies against some other comparator that would establish whether the drugs were superior or inferior or the same as the other comparator. It seems to me the choice of the comparator would depend, first, on the indication, clearly, and one would like at least to be able to get some information on what the comparator looks like on its own. So I am sort of going back to the question that Tom Fleming raised yesterday, or whatever day it was, not it all merging. But, with so many of these studies, we

are in the position of trying to impute what we would expect to see with a placebo or what we would expect to see--yes; what we would expect to see with the placebo--in the absence of the placebo, or even we are trying to impute what this drug would do versus that drug based on another study. So it would be important to know, for example, unequivocally, whether naproxen plus, I would think, a PPI inhibitor does something good or bad in terms of cardiovascular risk. If we knew that, we might be in better shape to make judgments about how to design the trials. So I am not sure I would jump in immediately to all these comparisons. We are going to get to some of that, I guess, in the next series of questions that look at the other non-steroidals. But I think it is a complicated issue that would need to be addressed for both the placebo-controlled group and the active control group and would need, actually, a third comparison which is a research program that looks at the active comparator so that we establish what it is we are actually looking at there because a lot of that we have imputed. Bob?

DR. TEMPLE: Suppose you knew--there are a bunch of non-steroidal anti-inflammatory drugs out there. Everybody agrees somebody is going to--you are going to treat pain with something. Several of you said that staying on the market with this box places you under some--that, ideally, at least, there would be some further studies designed to show something. So, just to pose a couple of questions, suppose an adequately sized study of adequate duration showed that this drug was no worse than ibuprofen, a standard treatment, would that be reassuring up to

a point even though you have never had a placebo-controlled trial of ibuprofen and you probably never will. Or would you have to use naproxen which people sort of have an inclination to think is a little better. Or do you have to try to dream up another placebo-controlled trial which is not easy to think how you do these days unless sort of polyp reduction raised its head again, maybe at a lower dose. It would be helpful not to design the whole trial but to think a little bit about some of those things and what is possible. If you can't do anything until you have the definitive naproxen versus placebo study, we are talking almost never because we don't have any of that.

DR. WOOD: Of course, the other issue that is on the table is that some of us believe these drugs were risky. And so inherent in that assumption is that you would be cautious about recommending a trial to be done because the likelihood is it would revalidate or replicate what has already been shown. So there is some hazard in suggesting that, I think. But I actually think there is a value in trying to demonstrate an effect against naproxen. I don't see a problem with that. Naproxen may be beneficial. We ought to know that, though, and we ought to be able to find that out fairly quickly, I would have thought. And let's get that. After all, it is not that we are trying to define the origin of life or something here. This is not some fundamental discovery we are trying to make. We are trying to divine what the optimal therapy is for something. If we can evaluate naproxen plus a PPI and work out how that stacks up against placebo, and then move on from there, we could get a lot of information fairly quickly, I think, that would be very valuable.

DR. TEMPLE: So at least the initial study, perhaps you would add other groups like one of the other selective ones that isn't so named, but the first study would be a study of reasonable duration. You also--I hope you will say something about just how long it needs to be, too. I mean, is it a one-year or a three-year study? The initial comparison might be against naproxen and some dose of celecoxib. They already have a study against ibuprofen so that wouldn't be too helpful to do again, I guess. Is that the sort of thing you are thinking of?

DR. WOOD: Yes. I would be unimpressed with a study against another selective COX-2 inhibitor. I think that is likely to be negative.

DR. TEMPLE: And it has been done.

DR. WOOD: And it is being done right now. I am not sure of what that will teach me. At the end of that study, if you gave two doses of the same drug, you would expect to see the same effect in both groups. If you give two drugs that are very similar in their pharmaceutical effect, you are unlikely to see a difference between them and I am not sure what that would tell me.

DR. TEMPLE: So, so far, at least, your thinking naproxen, if I hear you.

DR. WOOD: Right. Dr. Cush?

DR. CUSH: I think, by going through this data in the last few days, that we have acknowledged that there are a number of signals that exist--that are worrisome that exist for the nonselective non-steroidals especially for ibuprofen and diclofenac which have been often

comparators in these trials. I think we also have been impressed by the performance of the naproxen. Hence, I would say that, really, the whole class, all non-steroidals, should have a warning that would include some lesser version of what may be in the black-box warning about cardiovascular risk and that everyone should basically carry that forward, maybe with the only pass being provided to naproxen which becomes a comparator drug for future trials. I think that, to get off the list, to get that warning removed, you basically have to, as a sponsor, do a trial against naproxen or, in some other manner, show that you do not show a significant cardiovascular hypertensive risk to your patients. I would also favor the performance of an NIH and/or FDA-funded--ALLHAT trial has been proposed--and such a trial should be two years duration.

DR. WOOD: Dr. Nissen?

DR. NISSEN: Let me see if I can get very specific here. For each of the marketed COX-2s which I assume, at least for the moment, will be--who knows? Actually, I am not sure what we decided. But let me say that, in arthritis, it is very clear you can't do a placebo-controlled trial. While it might seem appealing to do your acetaminophen codeine control group, I just don't think it is a practical approach. I think we need to have some clarity and some consistency in comparators because, if we don't, if every sponsor compares to a different active comparator, we will have no clarity at all. So I happen to think that the evidence is pretty good that naproxen is no worse than neutral. So I would like to see celecoxib 200 milligrams, a dose that has not, at this point, been shown to have excess cardiovascular risk, against

naproxen, 500 BID, with adequate size, and Tom has mentioned some numbers--we are talking about around 100 events, at least, maybe a little bit more--to get the upper limit of the hazard ratio to be at a level that would provide some comfort. If you are going to do that trial, then it makes a lot sense to add a third arm to the trial which includes a conventional and non-naproxen NSAID. I happen to like diclofenac because it is an agent that looked, in CLASS, an awful lot like celecoxib. So now you have clarity in a single trial of acceptable size on how a low dose of celecoxib, the most commonly prescribed dose, compares to an agent that you believe is, at worst, neutral and to an agent that has some potential suspicion to be worse than neutral. When you are done with that trial, you will know a lot more. Now, Merck has already set up a diclofenac comparator with their agent and that is helpful. The problem is--

DR. TEMPLE: Not naproxen.

DR. NISSEN: Not naproxen. You know, obviously, is it very costly to redesign that trial but there is a problem for me. If Garret FitzGerald is right, that diclofenac is similar to celecoxib in selectivity and, therefore, in cardiovascular risk, then the comparator to etoricoxib could be a comparator that is not neutral. It is not a naproxen comparator. So we may not have clarity, the clarity that we would need. So I think that, in the absence of being able to do placebo-controlled trials, you have got to pick an agent that you think is probably no worse than neutral and try to show whether new drugs that are proposed and existing drugs are not worse than that agent on cardiovascular risk. So that is one guy's opinion. But I

am not an epidemiologist. I am just a knuckle-dragging cardiologist.

DR. TEMPLE: That sounds good. Actually, that is getting close to the ALLHAT study that we are hoping for.

DR. WOOD: Dr. Dworkin.

DR. DWORKIN: I was going to say much of what Dr. Nissen said except that what I would prefer is a third comparator, ibuprofen, because I think we have this large class of traditional NSAIDs and, if we had a series of studies with these COX-2-selective drugs, whether it is celecoxib, rofecoxib, valdecoxib, and each of those studies had a comparator of naproxen and ibuprofen, while we wouldn't have a placebo baseline at the end of the day, we would have a lot of information about naproxen, which we would all like to know a lot more about, and, with the ibuprofen arms across all these studies, we would have a lot of information about the traditional NSAIDs that we know very little about at this point. Of course, we would also then know a lot about our coxibs as the third arm.

DR. TEMPLE: Not to state the obvious, there is a difference between what you can reasonably ask a company to do and what you could ask a larger group to do. We all want to know about diclofenac but it is not clear that some company wants to know about diclofenac. So it could be slightly different, but this is a very helpful discussion.

DR. DWORKIN: Could I say one thing, Bob, about that. Isn't it the case that, in Europe, the European regulatory authorities really require a comparator arm. So you would not be doing much

more than is done in Europe by saying that we would like to see at least one trial with an ibuprofen arm and also a naproxen arm, in addition to your drug. I don't think that is unreasonable.

DR. TEMPLE: That is fair. But if they were to come back and tell us, if we are as good as naproxen, aren't we okay? It would be hard to say the answer to that is no.

DR. WOOD: Dr. Fleming.

DR. FLEMING: I will defer discussion about ibuprofen and diclofenac until we get to the later questions. My sense is there is a trial that I believe should be done with celecoxib although it is optional, although I would tie it to the black boxes, as you have previously. I believe that there are, however, studies that should be viewed as essential for valdecoxib and rofecoxib. Relative to celecoxib, what I had written down parallels what Steve Nissen had said with a few extra specifics. The design that Bob Temple had put forward, to me, makes a lot of sense. It would seem logical as one approach here that for celecoxib could lead to the kind of evidence that would remove the black box, is to do a trial. I would urge that the comparator be naproxen or aspirin plus PPI, agents for which there is a considerable sense that the effect on cardiovascular excess risk is minimal, and it be a noninferiority design, essentially ruling out the magnitude of effect sizes that we are seeing overall which is actually going an achievable task; that is, ruling out a 50 percent increase. Basically, if truth is no increase, you can rule out a 50 percent increase with 90 percent power with only a 2-and-a-half percent false-positive

error rate with 250 events which, essentially, is a trial that would have about 10,000 people per arm. That would be the basic target that I would put forward. That trial is positive if your observed excess risk is in the neighborhood of 17 percent. So anything that is not worse than about a 17 percent increase in the trial of that size would rule out a 50 percent increase. In my view, if that type of evidence were available, and I would be inclined to think it would be the OA or RA setting and I would like to see it for two, to two-to-three years follow up. You had staggered entry and then additional follow up so we are looking at at least a couple of years of follow up. We are looking at duration of outcome. That is the kind of evidence that, from my perspective, would provide a considerable reassurance. I don't consider it mandatory, but I would link it to the black-box issue. On the other hand, for valdecoxib and rofecoxib, linked to the fact that I voted no, to my way of thinking, if these product are going to be on the market, it should be essential that we get additional evidence. I am very troubled that, that, for valdecoxib, we have 3,000 patients. We have minimal evidence here upon which to base a clear sense of whether or not there is excess risk. I believe the FDA should consider it essential, within an acceptable time frame, to perform a study that allows us to get a clear sense of whether there is an excess risk. The dose should be chosen according to what the sponsor believes would be an appropriate marketable dose that we would want to be able to establish safety. For rofecoxib, my sense is that, what we are hearing is that Vioxx may have gone forward with an improper dose. I think, if we are, in fact, going to get it back

onto the market, there should be studies done at a dose that is, in fact, going to be marketed that needs to be established to be safe. Similarly, as for the celecoxib, if these studies are done, and I believe they should be considered essential, they should be done in a manner to allow us to rule out a 50 percent increase using a proper control and that control would depend on the indication, but either a placebo control, and aspirin plus PPI or a naproxen control would seem acceptable.

DR. WOOD: Okay. Dr. Hennekens.

DR. HENNEKENS: The randomized comparisons of the short-acting NSAIDs suggest to me that they are at least as hazardous as the coxibs. These are over-the-counter drugs that have direct-to-consumer advertising. I think there is a signal here that we should not ignore, so I would not limit the comparisons to naproxen.

DR. WOOD: We are getting to that, though, in a second. Dr. D'Agostino?

DR. D'AGOSTINO: Many of the comments I was going to make have been already made, but I think that, if you shift the indication to something away from arthritis, you can get a placebo as a third arm. I think the naproxen is a good idea. I am concerned. I agree 100 percent that it should be non-inferiority. I am concerned about the 1.5 because some of the drugs that we have condemned may have something like a 1.5 or even smaller. So that may be too generous. I think that takes a lot of discussion and I don't know the answer. The other point that I visited a few times and don't want to leave is that I think the follow up is very important, that people

can leave because their blood pressure is building up, they are getting hypertensive, or they could leave because of G.I. problems. But those individuals need to be followed. They can get off the drug but they need to be followed. Should the analysis be intent-to-treat or should it be something else, one can argue that again. But I think it is very important that it is, as much as possible, a complete follow up. There is also--it goes without saying, but we need a long enough time because we don't seem to have a constant hazard over time. So we have to make sure the studies do go the two or three years and the ascertainment adjudication of these CBD events has to be a prime item in the particular studies.

DR. WOOD: Dr. Hoffman?

DR. HOFFMAN: I liked what I have heard from Steve and Tom about suggestions for a study, a long-term study, going 1.5 to three years for arthritis. But I think we fall into potential traps here when we talk generically about arthritis. I think rheumatoid arthritis, being a systemic disease, which has an increased risk of cardiovascular disease to start with, becomes a very difficult situation to deal with if one uses that cohort in a long-term study. These people are constantly having their multiple therapies tweaked to find the sweet spot which sometimes we find, sometimes we don't. However, if the study is done with a mild to moderate OA, a degree of osteoarthritis that is significant enough for which someone would take medication, then you don't run into the problems of multiple other medications and systemic illness. So I like the idea. I think with mild to moderate OA, you can have an analgesic

arm. You can start with acetaminophen. You could even increase from acetaminophen to acetaminophen with codeine, if necessary. There are no known cardiovascular risks with that. You can compare that to the NSAID group, naproxen, if you like, or ibuprofen with a PPI, and then look at your COX-2. I think that becomes a much cleaner study.

DR. WOOD: Dr. Cush?

DR. CUSH: I want to echo some of the same comments but then specifically speak to some of the impracticalities of what Dr. Temple and Dr. Fleming had suggested, very good ideas, good plans, but, again, as Gary stated, we need a team of drugs to manage these people over the long haul. They don't stay on any one drug. So, to expect someone to stay on aspirin, 4 grams a day for two or three years, is not going to happen on any drug, in fact. It is just not going to happen. Moreover, aspirin, 4 grams a day, is not used at all ever anymore by anybody who knows what they are doing. The gastroenterologists would have a field day with this. Okay? So to try to provide some modicum of protection by putting a PPI on top of that is not going to be practical and this would never work in a clinical-trial situation. As Dr. Hoffman has suggested, an analgesic class makes sense, whether that be acetaminophen, tramadol or propoxyphene, and if you want to throw in the added benefit of 81 milligrams of aspirin a day as a control, that probably would work.

DR. WOOD: Dr. Fleming.

DR. FLEMING: Under their proposal, there are, certainly options that were put

forward and an alternative to the aspirin PPI would be to use naproxen as the control arm. Just to get back to Ralph's point, he is right that it is difficult to know exactly what the margin is here. What is an unacceptable level of increased risk. I had mentioned that I would want us to rule out at 50 percent increase and that would take 10,000 per arm. If we, in fact, asked to rule out a 33 percent increase, it would be 20,000 per arm and, to rule out a 20 percent increase would be 60,000 per arm. A reassuring aspect, though, is that if we are ruling out a 50 percent increase, which is 10,000 per arm or, in essence, 250 events in the pairwise comparison, what one is doing to be successful there is getting an estimate that is far less than a 50 percent increase. It is an estimate of about 15 to 17 percent. It would have to be better than that to be a success. Thereby, what one would be getting is, for that study to be positive, a result that would indicate that the estimated excess risk is, at most, one third what we are estimating it to be in the aggregate here and, hopefully, even better. So, keep in mind that, in that trial design, it is not success if you see 1.5. It is success if you rule out 1.5 and that is going to take something that is an estimate of only about a 15 percent increase.

DR. WOOD: I am going to move us along to next that as we have already started to lose people and I think we have given them advice on this.