Workshop on What Constitutes an Optimal Biomarker in Cancer Trials

Introduction

The process of conducting clinical cancer trials, particularly those dealing with nutritional chemoprevention, is an extremely difficult one from a logistic perspective. The ideal endpoint of such a study is the appearance of cancer. Nevertheless, using such an endpoint is rather difficult and often impossible for several reasons. First, since one begins with a population which does not already have existing cancer, the absolute rate at which cancers appear is rather low, necessitating the use of very large numbers of subjects. Second, the evolution of cancer in individuals who do not have an existing neoplasm is, in most instances, an extremely slow process that not uncommonly occurs over the course of a decade or more. This necessitates the conduct of trials that last a very long time.

There are several strategies that attempt to deal with these logistic difficulties. A common one is to select subjects who are known at the time of enrollment to have an exceptionally high risk of developing the cancer. Two examples would be selecting subjects for colon cancer trials who have had large adenomas removed in the past, or selecting women for a cervical cancer trial who have existing dysplasia in their cervix. Another strategy is to select an alternative endpoint that serves as a surrogate marker for the subsequent development of cancer, a so-called ‘intermediary biomarker of cancer’ or ‘surrogate endpoint biomarker’. Ideally, such a biomarker: (1) precedes, by a long span of time, the occurrence of cancer; (2) predicts, in an accurate fashion, the subsequent occurrence of cancer; (3) is modulated by environmental influences in a manner that parallels the subsequent risk of cancer, and (4) is a practical and accessible parameter to assess. In nearly all instances, a suitable intermediary marker is one that is integrally involved in the mechanism(s) by which the cancer evolves but this is not an absolute requirement.

Suitable biomarkers tend to be specific for each tissue but common ones include: histologic evidence of dysplasia, immunohistochemical evidence of an increased proliferation rate in the tissue, biochemical alterations indicative of increased cellular proliferation, as well as a variety of genomic anomalies such as mutations and other critical alterations in DNA. Note should be made that what this chapter pertains to is intermediary biomarkers of Cancer, not Diet. The latter are biomarkers which reflect the exposure of an individual to a particular nutrient or the status of an individual in regard to that particular nutrient. Such biomarkers are the topic of a different discussion.

Although the concept of using intermediary biomarkers of cancer is an extremely attractive one, and a strategy that is heavily pursued nowadays, it is an infantile science and a rather imperfect one. There are essentially no intermediary biomarkers which have a perfect
Workshop

concordance with the subsequent occurrence of cancer and much work needs to be done to thoroughly validate biomarkers that are in use today. It is critical for the clinical scientist to be cognizant of the limitations of each biomarker if one is to correctly interpret the results of contemporary trials in nutritional chemoprevention.

Discussion

**Dr. Meier:** When you do studies on intermediary markers, do you stop the study when you have resolved the intermediary markers, or do you go on to the end until real tumors develop so that you are able to validate your marker?

**Dr. Mason:** That’s an important point. My personal feeling is that to really validate an intermediary marker, at some point along the way you’ve got to do the full trial, because you’ve got to find the concordance between this marker that you’ve decided to use and the cancer itself. But that defeats the whole purpose of using an intermediary marker, because you end up having to do the whole trial. I wonder how many people would agree that to truly validate an intermediary marker at some point you’ve got to do the full trial.

**Participant:** Let me take a different viewpoint. I’m an editor-in-chief of a journal and I’m being confronted with this problem all the time. In talking about a marker we’re considering doing *in vitro* studies, where you culture tissues and look at a biochemical event – there are a lot of biochemical molecular markers you can use on the Petri dish. But that’s a very different marker when applied to the whole human being because you then have to take into account the intermediary metabolism of that marker. So if I talk about carcinoembryonic antigen (CEA) or α-fetoprotein I have to take into consideration how the body as a whole handles the substance. Indeed, α-fetoprotein may not be a specific tumor marker; it’s a normal antigen produced by our bodies, so is CEA; it’s still a tumor marker, but it is not a cancer marker. So it all depends upon what stage of the disease you’re looking at, whether in the human, or in a transgenic animal model, or in a transplanted tumor, or if you’re dealing with a biochemical mechanistic event, because with all those markers you may end up looking at different packages. So I can’t look at one ideal marker to answer all your questions; I have to look at the stage you are talking about. So the kind of marker depends on the question you are asking. Once that is settled, and once you’ve defined your model, then I want the marker to be validated in the Petri dish, on the nude mouse, and in the human if necessary. The validation depends on where you want to use it.

**Dr. Mason:** I would agree with that.

**Participant:** I would like to add an epidemiological viewpoint. In general, biomarkers can be used to identify exposure and also disease, and in the evaluation of technical error and of the error resulting from biological variability. Let’s say you want to measure a biomarker in the blood. You have to make sure that your technical error – in other words the error that you made during the act of measurement – is low. The other thing that you want to be sure about is that the biomarker is stable. In other words, it has a high predictability and doesn’t change from one month to another, is not influenced by diet or the menstrual cycle, or by any other potential factors. When considering biomarkers for the preclinical phase of a disease, you have to evaluate the specificity of the biomarker, trying to validate it in relation to the development of actual disease, as mentioned before. So in general as a criterion to reject or accept a paper in relation to the suitability of a biomarker, you should go through a simple list of sources of variability. If the biomarker is characterized by a large technical error then it is not suitable; if the biomarker is characterized by a high level of predictability and is not influenced by diet or other exogenous factors, then it is indeed suitable. These are simple guidelines that can help in identifying the best biomarker for your purposes.
Dr. Mason: Let me just summarize some of what has been said in the last few minutes. First of all, I agree that you have to make this distinction in your mind about whether you’re talking about biomarkers of exposure or biomarkers of disease. Now, just for the sake of focus, I would suggest that we concentrate on biomarkers of disease, though you’re right, there is a whole field of biomarkers of exposure as well. I believe it is absolutely correct that you have to modulate your biomarker concept depending upon whether you are doing a clinical cancer trial versus a laboratory study, a cell culture, or whatever. But the theoretical construct is the same. If you are doing a cell culture study and your ultimate gold standard endpoint is p53 mutation, you might look at preceding events, such as p53 strand breaks or something like that, which might precede mutations, but ultimately the theoretical construct is the same.

Participant: I’m sorry, I think that’s irrelevant. In tissue culture you can take only 3 or 4 weeks or 6 months, never 20 years. And the same is true for animal studies. So what really we are looking at today is to be able to save years of clinical studies, with all due respect.

Dr. Mason: Does anyone want to respond to that?

Participant: Yes I will. I want to take this argument back to 1970 when the National Cancer Institute came out with the biomarker concept. Although we have evolved over the past 30 years, any construct you try to set up is still basically the same. We need to remember that, considering the multiple stages of malignancy, markers will change over time, and many of the markers that could be in current use have not really been validated – that is, if you want to use biomarkers clinically, you’ve got to assess the specificity, sensitivity, predictive value, and clinical trial validation in relation to the particular stage of the disease. This has not been done for many of the markers used right now. The second thing is that there are a lot of markers available in basic science that cannot be applied clinically. I can use angiogenesis markers or proliferation markers in basic research on nutrient-gene interactions because I have a suitable controlled laboratory environment, but I can’t do that in the human. So when I review a paper, I always put it in the proper frame. If it’s a clinical paper, then I agree with the last speaker that a biomarker has to be validated in terms of whether it is to be used for prognosis, for diagnosis, or for prevention. But when you talk about tissue culture, about the nude mice model, about transgenic mice, and so on, my frame of reference is very different, because you have better access to what you want to look at. So when I accept or reject a manuscript, these are the points I’m looking at.

Dr. Mason: I would argue that sometimes the development of an intermediary marker in cell culture or animal studies is highly utilitarian. I sat on the thesis committee for a graduate student who was looking at modulation of an animal model of colon cancer. One of the things she was looking at was aberrant crypt foci as an intermediary marker, because it takes 35 weeks to develop tumors in these animals, and the diets are expensive, so we can save tens of thousands of dollars by looking at aberrant crypts as opposed to waiting until all these animals have developed tumors. So I think there are instances, albeit less impressive than in clinical cancer trials, where intermediary markers probably are valid.

So there’s a general agreement that at some point you have to find concordance with the gold standard endpoint, which in recent clinical cancer trials is the development of cancer itself. Now let’s talk a little bit about sensitivity and specificity. Obviously, in a perfect world you’d want something with perfect sensitivity and perfect specificity. Where do you draw the line? How sensitive does it have to be? How specific does it have to be? For example, ornithine decarboxylase is an intermediary marker for colon cancer, a rate-limiting enzyme in the polyamine pathway. There are many things that up- and downregulate ornithine decarboxylase other than carcinogenic processes, and yet even to this day it’s used by a fair number of people as an intermediary marker, because among the things that cause it to go up and down is expansion of a neoplastic cell clone. So I can think of instances where you have less than perfect specificity and yet you have a marker which people at least perceive to be valid. What do you use as a minimum threshold for specificity?
Workshop

Participant: It depends on your study design. For instance, if you are doing a clinical trial, I imagine that you would prefer an instrument that gives you high sensitivity. You want to detect the lesion. You want to have a low number of false negatives. So in this case you decide that your task has to be highly sensitive. If you are running a screening program in a general population, you will want to have only a small number of false positives, because these will cost a lot; you are involving a huge number of people and false positives are very expensive, from several points of view – from the human point of view and from the economic point of view. So in that case you are looking for a task characterized by high specificity. So it depends on your study design.

Dr. Mason: Any other comments?

Dr. Goldbohm: I think there's something to add with regard to sensitivity, with respect, for example, to the use of something like the adenomatous polyposis coli (APC) mutation as an early biomarker. I understand that about 80–90% of colorectal tumors have APC mutations but the others don't – they go by a different pathway. Nevertheless it may well be that that is a good biomarker, without being highly sensitive because it's only covering part of the pathway.

Participant: With regard to specificity, I think you have to make a distinction between different types of markers. Some markers are very late indicators – for example, the number of tumor cells – though these are still valid because the cancer itself may not be diagnosed clinically for another 1 or 2 years. So you have markers which indicate the number of tumor cells already present, and you have markers that are upward of this process – which indicate stages that may lead to cancer. The specificity will be completely different between a marker of an established tumor and one indicating a tumor-forming situation.

Dr. Mason: I agree with that. Any other comments?

Dr. Bostick: It's hard to ask about specificity for what really amounts to a lot of different situations. This takes us back to an examination of the different kinds of biomarkers. We talked about exposure, exposure effect, and there are biomarkers of risk, and biomarkers of current presence of disease, and ones that can encompass the different stages of it. Even among biomarkers of risk, you've got genetic susceptibility and phenotypic susceptibility: the genotype and the phenotype don't always correlate well. And for biomarkers of risk, which is what I mostly deal with, the further back you go in the pathway, the less likely it is that it will be a highly specific marker for the development of disease. We know that there are a lot more people who have hyperproliferation than who get cancer, and there are a lot more people who have adenomas and never get cancer. So it depends on the purpose of the marker. If you are using a biomarker as a screen for early detection of disease or even for late premalignant conditions, you want it to be very sensitive and specific. If it's a biomarker of early risk, and particularly if it's one that you think you can modulate, then it's not going to be particularly sensitive or specific, and maybe you need an ensemble of biomarkers to profile somebody's risk. It may be that you're just looking at group differences, so a dietary intervention, for example, might be based on average values of the biomarker profile relative to a placebo group.

I think the validation process probably needs to parallel somewhat the standard Bradford Hill criteria for causality in that there needs to be biological plausibility as to why you're choosing or investigating a biomarker. This needs laboratory development and I think animals are a good place to start. Then you've got to translate it to something that you can do in a human, and you need to make sure you can measure it reliably. If you do it over and over again, do you get the same thing? And when you get past those issues, the next question is whether it is valid. In the animal, you can look and see whether the biomarker occurs if you put the animal at risk, say by giving it a carcinogen, but does it also predict the endpoint that you're interested in?

Moving into the human dimension, I think it's probably best to start with small studies that are usually cross-sectional, because if you assume that people who have a late form of
the disease are most likely also to have earlier steps in the pathway, then hopefully the biomarker is going to correlate with the presence or absence of this downstream endpoint. Then for a biomarker that you're interested in for a clinical trial there needs to be the same sort of process. Is it plausible that this intervention works via this particular biomarker in some way? Is it really relevant? If you take into account the time sequence of the development of a cancer, does the length of the intervention you're going to use match where you think that biomarker fits into the carcinogenic process? Then if you can modulate the biomarker, the next step would be to investigate the disease and make sure that you're modulating both the disease and the biomarker. Only then will you begin to feel that a particular biomarker may be useful for other interventions as well.

Dr. Mason: Let me interject here. What I'm hearing from Dr. Bostick and several other people is that you want to make the field incredibly more complicated! It sounds as though you believe there shouldn't be a standard set of criteria for an intermediary biomarker, because it has to be custom designed to the particular needs of that particular trial. So I can't write down at the end of this evening what the minimum criteria should be, because they are going to change according to the research questions of that particular study. I suppose I would agree with that, though I don't like it because it makes things more complicated. However, I think you're right that probably the criteria for an appropriate and valid biomarker change according to what the particular research questions are for that particular trial. One thing that Dr. Bostick brought up is the Bradford Hill criteria for establishing causality. Does a biomarker have to be involved in a process in a causal fashion to be valid? I would argue, no. It probably would be better if it's involved causally, but I can think of valid biomarkers that are epiphenomena and yet are still valid as biomarkers. Would you agree with that?

Dr. Bostick: Yes, definitely. For example there are genes that are commonly mutated in colon carcinogenesis, but there could also be mutation of genes related to enzymes that metabolize certain environmental chemicals.

I think we may get to a position where we can individualize risk, and determine an individual's prescription for remediying that risk, and then monitor it, just like we monitor, say, cholesterol. So I think this is probably more analogous to heart disease risk assessment than cervical cancer risk assessment. With heart disease we look at cholesterol and blood pressure, and family history, and smoking, and various other things, and come up with some individualization of risk and what people need to do about it. And then we have ways to monitor that. I think we'll be able to do that with cancers as well, in some cases relatively simply but in others the procedure will be more complex.

Dr. Hursting: So maybe Cox postulates are a better kind of criterion than causal inference rules: the marker's there when the cancer's there; when cancer's not there it's not there; when you modulate the cancer you modulate the biomarker.

Participant: The other thing is, if you want a biomarker characterized by high sensitivity in a clinical trial, you can use more than one biomarker, and maybe cluster of biomarkers. These will increase your sensitivity, but they will also increase your false negatives. That would be interesting from a biological point of view if you want to depict the natural history of the disease.

Dr. Mason: For those of us who subscribe to the philosophy that you have to check as many intermediary markers as possible because you never know what the NIH is going to want when you try to renew your grant, this brings up the issue of what happens when you start analyzing your results and you have a beneficial effect, or a salutary effect, with one of your biomarkers, but you have no effect, or the opposite effect, with one of your other biomarkers? I think many of us have probably had that experience. So you run that risk by using multiple biomarkers.

Participant: You have to take care that any additional biomarkers you want to add to your study have specific criteria that you have incorporated into your study, so the addition-
al biomarker has to be characterized by a particular level of sensitivity and specificity, and has to be valid, so in this way you can restrict the number of additional biomarkers that you apply in your study and can increase the sensitivity of your instrument.

**Dr. Mason:** One of the utilitarian aspects of looking at several intermediary markers is that if one intermediary marker is negative in your trial but another one is positive, you can try to infer a framework for the process and start to map out what those pathways actually are.

**Dr. Freudenheim:** I do think this is a situation where discordant results may tell you more about the natural history of the disease, or about multiple pathways where your intervention is only important for one particular pathway and not for another. You get information when you have discordant results.

**Participant:** I fully agree with that. That’s what science is all about. The discordance gives you an exciting avenue to pursue. In relation to specificity and sensitivity, you need to get back to exactly what you want the marker for. If you want it for clinical purposes, then is it for screening purposes? For diagnostic purposes? For monetary purposes? Because that reflects several different specificity and sensitivity criteria. And then you need to know the negative predictive value and the positive predictive value of the biomarker. But when you’re working on the bench with that tissue culture, or that nude mouse, I would like absolute specificity if I can find it, because that would define a mechanistic event that I’m trying to pursue. So you have to define what you want to use the marker for, and then you can define what an ideal tumor marker should be.

**Dr. Hursting:** Maybe in our remaining time we could focus on an example scenario. Let’s say you’ve just got funded to do a clinical study on colon cancer and dietary intervention, and we’re collaborators, and we’re going to sit down and map out what markers to use in the trial, which starts in 2 months.

**Participant:** In this situation I think you again have to draw a distinction between markers that can be obtained by biopsy, which is invasive, and those that can be obtained by blood sampling, which is less invasive. As was said earlier, depending on the objective you can be more or less invasive, and you can accept certain things under some circumstances that you won’t accept under others.

**Dr. Mason:** That’s a good point. The function of your marker is an important determinant of what the validation criteria should be. Dr. Hursting has proposed using an example at this point. That is fine with me. We’ll go up that route.

**Participant:** I would say you have three different packages: you have morphological indices, you have generic markers, and you have markers dealing with proliferation. So depending upon what question you want to ask, you either use the morphological package, which looks at DNA changes and breakage and so on, or you look at generic markers, or you look at proliferation markers. If you have a sample of the tissue, you can do all three, but if you look at blood, then you are asking different questions. If the study is on cancer of the colon, you can do a biopsy and then you can use all three packages.

**Dr. Mason:** I think it’s reasonable to make an assumption that sometimes you just don’t have the capabilities – either because of finance or personnel or whatever – to use every marker that’s available, so you have to choose one or two of the available markers for reasons other than scientific rigor. Let’s say I get funded to do an intervention trial with folate and calcium: this is what John Baron is doing right now – he’s giving folate or aspirin to people who have had index polyps removed. He has chosen as his primary endpoint adenoma recurrence. I would argue that this has the drawback that it’s so proximate to cancer itself that he has to have several hundred people in his trial and it has to go on for 5 years. However, polyp recurrence is the only validated intermediary marker of colon cancer. If you look at the national polyp trial, there are enough data now to confirm this concordance between polyp recurrence and cancer itself. Everything else is just frosting on the cake right now; there’s no other truly validated marker. But John Baron is going to use that
Workshop

Participant: And even then, when he gets negative results, you don’t know if it’s because it wasn’t long enough.

Dr. Mason: By the way, I’m not being critical of John Baron!

Participant: I think John fully agrees with you, but his point is that adenoma is the only endpoint that’s been validated; there’s no other validated marker for what he’s looking at.

Dr. Mason: But he’s also playing safe, because he’s now written a supplement to his grant, and he is using a number of these other intermediary markers just in case polyp recurrence doesn’t work out to be a suitable endpoint of itself.

Participant: Can you not already say that p53 is a inadequate marker in this situation because it needs a biopsy and because it’s later than the recurrence of the polyp?

Dr. Hursting: I would think that p53 might be a valid intermediary marker in the study but it would not be a practical one for the reasons that you allude to: it’s even more distal than adenoma recurrence, and therefore you are losing almost all the advantages that you are trying to obtain from an intermediary marker. Yes, I agree.

Participant: But this study also includes colonoscopy to prove if there are polyps or if there are no polyps. So the gold standard will be colonoscopy, to give the endpoint of new formation of polyps.

Dr. Mason: Correct. That’s what he’s doing. The primary endpoint of his study is colonoscopically proven adenoma recurrence.

Participant: And then you also need biopsies for your biomarkers.

Dr. Mason: You’re right. John Baron obviously has some misgivings about using that as the sole intermediary marker, because he went to the effort to write a supplementary grant to look at other markers that are more proximate in the whole process. Are there any other reasons why he should have bothered to write that supplementary grant?

Dr. Bostick: I think part of it is to elucidate mechanisms of action. I don’t know which biomarkers he’s actually measuring, but I think we often have ideas about how, let’s say, folate might reduce the risk of colon cancer and using several biomarkers may help us to understand where in the folate metabolism pathway the intervention might be working. If you could show that you can modulate both the intermediary marker and the adenoma, you would be more confident that this is a good biomarker for the disease.

Dr. Mason: That’s an excellent point. Sometimes it’s worthwhile adding a biomarker solely to help elucidate a mechanism, not necessarily because you just want to use the shotgun approach.

Dr. Bostick: Biomarker ensembles really shouldn’t be used to provide a shotgun approach. Biomarkers should revolve around mechanisms of possible causation; for example, in colon cancer there’s probably an APC pathway and a mismatch repair pathway; with ulcerative colitis there’s p53 and maybe others. When you try to elucidate that, then maybe in the validation process you’ll discover that you don’t need to measure all of the markers to measure a particular mechanism, but there are one or two that do a particularly good job, while maybe there are others that do a good job of summarizing another mechanism. So I think more thought needs to be given to what kind of biomarker you’re using.

Dr. Mason: Maybe when you do a trial you should have not only a kind of clinical monitoring committee but a biomarker monitoring committee that actually decides for you what are the appropriate biomarkers you should be incorporating into your trial. I say that tongue in cheek, but it would be nice to have something like that.

I can only speak from my own personal experience, but in developing the concept for a new biomarker what has usually happened with me is that I do a basic experiment that looks at mechanism, and I hypothesize that this process is involved in carcinogenesis, and then I say wouldn’t it be fun to pull out our human samples from the freezer. That is exactly what happened in the studies I presented today showing the difference in p53 strand breaks in the
Workshop

colons of people who harbor adenomas versus controls. We were doing these studies in rats, showing that folate deficiency creates p53 strand breaks, and I just suggested to my postdoctoral student that he pulled out the human samples to look for the same thing. And lo and behold, that’s what we found.

Participant: Let me carry this a little further if I may. You saw something in the laboratory that looked exciting, so you then looked at human cancer tissue and you found that what you saw in the experiment model also occurred in human tissue. But that’s just the first step. The next step if you want to validate it as a clinical marker is to ask the questions, do I want it as a screening marker? Or do I want it as a diagnostic marker? The sensitivity and specificity will be different depending on the use of the marker. The marker you found might be a normal event, like carcinoembryonic antigen, which I spent my first 10 years in cancer research studying. It turned out to be a nonspecific colonic antigen, but the amount of information we got from it about normal metabolism was amazing. It is an expression marker of normal tissue, but it could be used as a marker where you have mucosal changes.

The reason we gave Dr. Albanes a hard time earlier this afternoon was that he’s doing a very expensive trial for cancer of the prostate in 100 plus centers in the United States, and they do not know what marker they want to use. Unfortunately I was a member of the panel that reviewed the whole protocol and I insisted that we know what biomarker they intend to use as an intermediate marker before we spent 48 million dollars. They cannot answer that question, which is why the whole study’s been delayed for a year, to try to define what proliferation markers they should be looking at. There must be a committee to look at this, and even then they may not come up with the right answer. Maybe by the time they finish the study 15 years from now, what we are talking about today may not be ideal; they may come up with something else. At least the tissue, the blood, the lymphocytes, and so on have been frozen properly so that they can go back and look at it.

Dr. Mason: Thank you. We’ll try not to make the same mistakes, but history has a way of repeating itself.