■ EDITORIAL

The Patient's Question—Unanswered

Mikel Aickin, PhD

Because you are reading this, the chances are that you are a health professional, probably a physician. Whatever your relationship to health care, I want you to take off your professional hat for a moment, and imagine that you are a patient who has just received a diagnosis. Your condition is potentially serious in the future, but not automatically so. Your immediate issue is to decide between therapy A and therapy B. Whichever you choose, you will have to commit to it for a considerable time without knowing any direct benefit, in the hope of future benefit. As you start to discuss this choice with your physician, you have one main question in mind. Now stop reading—and before going on figure out what that question is.

The Patient's Question. My answer is "which treatment should I choose?," but obviously the real question behind that is "which treatment will give me more benefit?" Until doctors become seers, this is not an answerable question. But a very closely related question that could be answered is "what is the probability that I will do better on A than on B?" My purpose here is to argue that we could be answering this kind of question with existing knowledge, and that it provides a far richer and more useful way to understand medical research results than the ways that we customarily use now.

I call this The Patient's Question (TPQ): what is the probability I will do better on A than B? When I bring this up with colleagues they usually say that I'm not really raising any fundamental issue here, because we can compute this probability from studies in the literature, if we want to. I have yet to find anyone, however, who can actually do this. Some start out by thinking maybe the reported p-value is the answer, but that isn't even close. Others say that you can see how far apart the measures of benefit were in the A-treated-group and the B-treated-group, take into account the standard deviations of the estimates, and use those somehow. But when it comes to the computation I asked for, they can't resolve the "somehow" part. When the outcome is a yes/no type (such as: recovered or not, went into remission or not, had improved condition or not) then some are convinced that the fractions of successes in the two groups answer the question. But they can't do the conversion to answer TPQ either, and there are mathematical reasons why they can't. I have yet to see any publication in the biomedical research literature that presented an answer to TPQ.

The Perfect Experiment. There is a methodological reason why we don't answer TPQ, and considering that reason can lead us to some potential strategies for finding answers. The basic idea is somewhat philosophical, but for those who don't want to go down that path, they should recognize that philosophical beliefs can have considerable impact on real events. Think about what a perfect medical experiment would be. We would give the patient A and then record what happens. We then turn back the clock to the pretreatment time, and now give the patient B without changing anything else in the world, and record what happens. Surely the difference between what happened in these two cases is the causal effect of one treatment relative to the other. No doubt the causal effect is what we are looking for, and it completely answers TPQ. And even more surely, the perfect experiment is impossible.

The critical point about the perfect experiment is that the causal effect is measured within an individual patient. This explains why randomized clinical trials (RCTs) as they are now analyzed cannot, even in principle, answer TPQ. Half of the participants are given A, but never get B, and the other half are given B, but never get A. Each person in the study provides only half the information of the perfect experiment. When we compare A-responses with B-responses we are comparing across patients, not within patients. It is a mathematical fact that from the marginal distributions of two random variables you cannot in general compute the probability that one is greater than the other. The RCT provides the marginal probabilities of A-successes and B-successes in the patient population, but it cannot compute the within-patient probability of doing better on one than the other. The answer to TPQ is beyond the RCT.

The Almost-Perfect Experiment. My solution to this problem would be to carry out the almost-perfect experiment. I would match two patients on as many factors as I could, trying to include all of those that are important for eventual success or failure on any given therapy. Even though there would inevitably be residual differences between them, therapeutically they would be as alike as I can achieve. I would then take the position that these highly matched patients shared an A-response and a B-response

Mikel Aickin, PhD, is a Research Professor in the Department of Family and Community Medicine at the University of Arizona in Tucson, AZ. E-mail: maickin@earthlink.net.

(or nearly so). Measuring an A-response on one of them is the same as measuring it on both, and the same for the B-response. By giving one A and the other B, I obtain (an estimate of) the within-patient causal effect of treatment. If I do this over enough pairs, then I can compute the fraction of times in which the A-response was better than the B-response, the fraction where the B-response was better than the A-response, and the fraction where they were for all intents and purposes equivalent. I can then answer TPQ, and even better, I can give the probability that it won't make any essential difference whether they take A or B. By way of contrast, in the standard cohort-based approach differences between the two partners are computed and then summarized over the entire sample, thus discarding the (admittedly approximate) within-person information.

There are no statistical difficulties with my solution, and in fact the statistical analysis is even easier than it is in most biomedical research, because it only involves observed fractions (of times when A is better than B, and so on). I don't believe I would spend too much time testing statistical hypotheses about the benefit probabilities, because the real issue is whether there is enough research behind them to be accurate, and that is a fairly simple statistical question.

Ambiguous Probabilities. There is, however, another large and lurking question about probabilities. Again I apologize to those who think that philosophy is irrelevant, because I think it is critical to understanding probability. When we apply a probability to a person (there is a 60% chance you will do better on A than B) we are always implicitly placing that person in the context of a population, in which the probability is as stated (60% in this case). In common parlance, probabilities are used as if they were personal characteristics, like height and weight, which do not need any external reference to be valid. But clearly this is untrue. For any given person, there may well be multiple populations I can consider them to be from, and each might (usually does) have a different probability for the event in question. Thus, if I know your age, I may say that you have an 80% chance of doing better on A, but if I then in addition learn that you are a smoker, I would say that you have a 50% chance of doing better on A, and so on with other health-related characteristics. There are profound ambiguities in the way we ordinarily talk about probabilities.

For this reason, the research program I want to encourage would provide the patient with benefit probabilities based on samples of other patients that are as much like him/her as possible. I want to be able to say that *your* probability of doing better on A is 60%, to mean that it really pertains in a specific way to *you,* and it is not simply a mass-statistic computed from a general and poorly characterized population, that you are in some vague sense a part of. I believe that making benefit probabilities patient-specific is a critical component of patient-centered biomedical research, and if we could put it into place we would be practicing much better medicine. Patient-centered analysis is novel ground, and therefore difficult to contemplate, but it is a needed counterbalance to the emphasis in biomedical science on patient populations instead of patients as individuals.

"Better" is in the Eye of the Beholder. Another lurking issue has to do with what "better" means. Here again I have been puzzled by the responses of health professionals to this issue. Some seem to believe that I must make one single definition of "better" and then stick to it throughout my analysis (by analogy with RCTs, presumably). Others believe that "better" cannot include issues like patient-preferences, costs, attitudes toward risk, evaluation of disability or side-effects, and so on. The reason I am mystified is because it seems obvious that we can use multiple definitions of "better" in a single study. In fact, it would be highly interesting to know if the benefit probabilties shift wildly, or remain essentially constant, as we vary the definition of "better." And it is even clearer that some of our definitions of "better" can encompass harms in addition to benefits. In pain syndromes, for example, we could include both pain reduction and reduced drug taking as parts of the definition, thus accommodating patients who value the latter as much as the former. Indeed, from my perspective the false forced-choice of a single definitive endpoint in RCTs is one of the several reasons why they do not serve the interests of clinical medicine.

Real Clinical Research. Because TPQ seems to me to be so obviously important, I have thought about why we did not use this concept in coming to our current conceptualization of biomedical research. I have two theories. One is that we humans seem to be attracted by "binary thinking." Things are either right or wrong, up or down, black or white—or if not, we are still better off thinking that way. Thus therapies should either work or not, be effective or not, be recommended or not. The second theory is that the manufacturers of therapies (mostly pharmaceuticals) need some kind of yes/no decision with regard to their products, both for regulatory approval as well as marketing leverage. The RCT is admirably fashioned for both binary thinking and promoting drugs, and that is why it has become dominant in biomedical research. Somewhere along the way the patients and their question got lost. In my opinion, it would be a good thing if we tried to recover real clinical research by putting all of our focus on answering TPQ. ❖

> **Patient-centered analysis is novel ground, and therefore difficult to contemplate, but it is a needed counterbalance to the emphasis in biomedical science on patient populations instead of patients as individuals.**