In *Proceedings of the 1955 invitational conference on testing problems* (pp. 136-141). Princeton, NJ: Eductaional Testing Service.

#038

1955 INVITATIONAL CONFERENCE

Clinical Versus Actuarial Prediction

PAUL E. MEEHL

I found Dr. Zubin's empirical data very stimulating; but since they illustrate the use of statistical method in typology and do not bear directly on the predictive efficiency question, I shall not comment upon them further. I am completely baffled by Dr. Zubin's main theme: that the clinical-actuarial issue is a pseudo-problem. I do not find anywhere in his paper a serious attempt at rigorously showing this, and it seems to me that he has clouded the issues by bringing in the interaction between the two methods in *research* work. This research interaction has never been disputed by anyone; all agree that clinicians do generate hunches and, on the other hand, that hunches in social science must usually be *tested* by statistical methods. But the title of this symposium is "Clinical vs. actuarial *prediction*," not "clinical vs. actuarial research-planning." I still maintain that given a finite set of data—tests or otherwise—on an individual patient, for whom a prediction is to be made, you can either hand the data to a clerk or you can hand them to a skilled clinician to think about. Surely this is a pragmatic distinction of real importance. Take a simple, concrete example. We have to decide whether a certain veteran is to be given intensive psychotherapy or not. This is a decision-problem which is being faced in clinics all over the country at this moment.

Does Dr. Zubin seriously assert that we cannot distinguish between these two operations: a naive clerk filling in the values of a regression equation, and 10 clinicians talking around a conference table? Since the latter costs from 10 to 30 times as much (VA rates), Dr. Zubin must have very different notions about economics from mine. *Of course*, the "context of discovery" displays both methods. In my book I emphasized Reichenbach's distinction between the two contexts not once but several times over. In the process of *constructing* a mechanical prediction system, the hunches of clinicians are usually valuable (not always!) and sometimes indispensable. Pick the variables any way you please—using Freudian theory, blind empiricism, or clairvoyance. You may use either "rational" combining functions or choose empirically by blind curve-fitting from a wide class of equations. You may study the hits and misses intensively and qualitatively, hoping to get further hunches as to how the combining function might be improved. At some point, however, you move from the research process to the practical setting; you are asked to apply the fruits of your cerebrations to a realistic prediction problem. *At that moment*, what do you propose to the clinic administrator? Do you give him a statistical table or equation? Or do you tell him to hire a clever psychologist who will think about the same data, *case by case*, and predict therefrom? The first of these solutions is, in daily practice, what I call actuarial, whatever its research history may be. The second solution is non-actuarial, even if actuarial information is part of the total data that the clinician has to "think about." Which of these two procedures has the greater success, the larger hit-frequency, in daily decision-making? This is no academic, hair-splitting question; it is a practical question of intense personal significance to the suffering patient and of great monetary importance to the taxpayer. I find in Dr. Zubin's paper no demonstration that the distinction between clinical and statistical prediction is spurious. Admittedly there are a few borderline methods. But in general, any genuinely *mixed* method is non-actuarial; because the defining property of the pure actuarial method is that it is unmixed. The existence of borderline methods which are difficult to classify does not abolish the distinction (although to believe that it does is one of the commonest of philosophical mistakes). We cannot say precisely how many whiskers it takes to constitute a beard. Any cutting point, as between 78 and 79 whiskers, is arbitrary and subliminal. But we do not conclude that there is no point in distinguishing or that a distinction cannot be made, between a man who is "clean-shaven" and a man who is "fully bearded." Dr. Zubin says the methods "complement each other." This sounds plausible and tolerant; but what does it actually mean? In some of the published studies the effect of allowing the clinician to adjust the actuarial prediction is a shrinkage in predictive efficiency. That seems to me to be a clear case not of complementation but of sabotage. It is senseless to speak of complementation when there are two procedures both purporting to do a specified task but one of these procedures in fact performs the task better than the other, and even better than some mixture of the two procedures will perform it. As to whether a really rock-bottom, epistemological distinction can be made, this is a question of great technical complexity. I would warn everyone against thinking it an easy question, disposable of by a few pleasantries (such as, "the methods complement each other"). Here is needed a thorough analysis using the technical tools of the logicians and mathematicians. I do not know where I stand on this one, and I have spent many hours discussing it with some of the ablest logicians and philosophers-of-science in the business.

Dr. Zubin quotes me as saying that the actuary, like the undertaker, has the final word; and he says he doubts this. He says that the actuary in turn has a clinician looking over *his* shoulder to "see where the formula fails." To which I must reply, so what? At this point, the clinician *thinks* he "sees" where the formula fails; but Dr. Zubin knows as well as I do that this is not the sort of thing you simply "see." We clinicians "see" a lot of things that are not so, if the verb "to see" is used as Dr. Zubin uses it. The context in which I make that remark about the actuary having the "final word" makes sufficiently clear what I mean by this. It is really no more complicated than the scientific principle that I assume we all share, namely, it is facts that check on theories and not the converse. That we will no doubt continue to make still further theories is irrelevant to this primacy of facts; with respect to a *given* theoretical or predictive claim, the facts do have the final word. I can therefore only recommend to Dr. Zubin that he re-read the passage from which he quotes, and ask him to show me specifically where the logic is defective. Jones says that he, using method J, can predict what will happen better than Smith using method S. If Dr. Zubin knows of some way to resolve such a disagreement besides keeping score on Jones and Smith, I should be fascinated to learn what it is. And keeping score—let's be clear about it—is an incurably actuarial process.

Now for Dr. McArthur. I gather he feels there is some kind of disagreement between us, at least with respect to the significance of the available empirical studies. It is perhaps foolish (and not in the symposium tradition!) to say of another scholar's paper: "I agree with everything he says." But I feel impelled to say something very like that about Dr. McArthur. And I don't suppose we can cook up a scientific fight if I insist upon agreeing with him. Let me here say something of a personal nature. I am deeply convinced that in my own therapeutic practice (which is about as psychoanalytically-oriented as one can be without labeling himself a "wild analyst") I do things daily which the best electronic computer cannot begin to do. If I didn't think this, I would feel pretty guilty taking $10 an hour from my clients. I don't see how anyone would even program a computer so as to make it use the raw data as I use them when I interpret a client's dream. It therefore bothers me that clinical psychologists seem to interpret my book as anti-clinical, and pro-statistician; actually, *by far* the larger part of the words in that little volume are devoted to refuting the Sarbin viewpoint. (If you doubt that, just count pages!) At Minnesota we are currently pre-occupied with designing experiments which *are* built to show forth the clinician's unique talents. And I am pretty convinced in advance what the outcome will be; it will be that when a clinician is allowed (quoting Dr. McArthur), to "use the data of his choice, make the analysis of his choice, and make the predictions of his choice," he will look pretty good; not merely better than the actuary, but—more importantly—capable of activities (e.g., open-ended predicting) which the actuary does not even pretend to try. So you see how close I am to the McArthur position. I, like him, believe that we clinikers do special, unique, unduplicable jobs of idiographic conceptualization, when Dr. McArthur's criteria are met by the task and its conditions. *Therefore* I want us clinikers to spend our high-cost time performing these kinds of tasks. Where do we get this time? Well, perhaps there are some other time-consuming activities which we clinikers currently engage in that do *not* meet the McArthur criteria, and in which, consequently, we are at a disadvantage. If the McArthur criteria are applied to perhaps 90% of the prediction tasks which are being daily *attempted* by working clinicians over the country, it is clear that they are not being met. The empirical studies I have surveyed (which now number over two dozen) exhibit a pretty uniform trend. It appears that in prognosis, *given the predictive conditions under which practicing clinicians usually have to operate*, the clinician is largely dispensable or positively adverse to predictive success. Dr. McArthur seems to depreciate the importance of these empirical studies because he sees, quite rightly, that they don't meet his criteria. This puzzles me, because I feel that they are grist for his (and my) clinical mill. (He is wrong about Sarbin, whose clinicians had at least an hour interview with the subjects.) These 25 studies lead me to say, in effect, "Good! Just as I thought, when you don't meet McArthur's criteria, the clinician is beat out by the clerk. So, let the clerk take over these kinds of coarse prognostic and diagnostic tasks. He does it cheaper, and he does it better. I will then occupy my third ear (and Tompkins' souped-up Mill's Methods) with therapy and research." *Part* of this research will be using both methods in a complementary way to develop an equation for the clerk to use. The Harvard Adult Development Study in which Dr. McArthur is engaged I classify as research. If he should propose utilizing the method he describes in the routine predictive tasks of working clinicians, then I will have to start asking him my usual mundane questions about hit-frequency and cost-accounting. Further, Dr. Zubin and I will turn over the McArthur "clinical-introspections" to a super-statistician, just to make sure that with this clinical help in the research context, the actuary is still unable to cook up a mechanical method which will compete with McArthur's clinicians. I, like Dr. McArthur, do not believe that he could; but this is an empirical question. Don't forget—most clinicians would not have expected the uniform trend of the 25 prognostic studies either. But in that non-optimal domain, it seems pretty clear that the clinician's confidence in himself is unjustified by the hard facts. The research task for those who believe, as Dr. McArthur and I do, in the unique clinical powers of the human brain, is to find out whether this belief is true, and in what contexts it is true in a *degree* great enough to be of practical importance.

Drs. Humphreys and Sanford cleverly sent their papers to me after I had already dictated more than fifteen minutes of talk about Drs. Zubin and McArthur. But there is no point anyway in rephrasing their sound and insightful remarks, which is all that I could do. I have a disagreement here and there but it takes too long to develop most of these. I find myself unwilling to agree with Dr. Humphreys' view that we cannot expect to improve those clinical predictions that are based on brief exposure. There is evidence in the literature that people differ in their clinical talents; if we study the process carefully as Dr. McArthur and other researchers (such as Gage, Taft and the IPAR group) are doing, we *should* be able to tease out what is involved in doing it well. In 1944 I checked on the Multiphasic profiles of the patients I chanced to see walking down the hall of the psychiatric unit who appeared to me, at sight only, to be MMPI-psychopaths. During the year I spotted 13 such; in 12 cases I was right. If it were important enough, we could surely learn more about what I was responding to; it must be some fairly crude aspects of dress, appearance, and manner, since I have no psychic powers. And facts about dress, appearance, and manner, once made explicit, are presumably teachable. Dr. Humphreys refers to "pattern analysis" of test scores. Here is a big gap in our knowledge that will not be filled unless you statisticians quit telling us clinicians that Fisher or Hotelling or Rao and Slater solved this problem years ago. They did not. There is to my knowledge no convenient, practical, rigorous procedure for discovering the function and weighting the variables emerging from a many-score test like the Strong, the Multiphasic, or the Rorschach. I will here and now, in the presence of three or four hundred potential takers, offer to name several different clinical problems involving dichotomous criteria in which a Minnesota-trained eye can sort out Multiphasic profiles better than any of these methods. We are currently studying one such Multiphasic task—namely, the discrimination of psychosis from neurosis. I expect the discriminant function to excel the fledgling cliniker, but I expect the skilled cliniker to do still better. Better than all three (and a preliminary study shows this) will be an objective set of complex-pattern rules developed by Dr. Grant Dahlstrom and me. Why am I so confident, a priori, of this order? Because the student clinician follows a near-linear and unconfigured function, non-optimal weights, and low diurnal reliability for identical profiles. The discriminant function eliminates the unreliability and non-optimal weights. The skilled cliniker employs a configural function, and in the case of MMPI this is so important that the superimposed errors of non-optimal weights and unreliability do not wash out the configural gain. Finally, the objective pattern-criteria are configural and the decision is consistent from case to case. Non-optimal weights remain with us. With a 9-variable system, and no underlying theory to suggest a rational combining function, you would have 9 + 9 + 36 = 54 parameters to fit, if you went past the linear discriminant function to a second-degree expression (with the all-important cross-products). Think, dear brethren, of the sampling errors you would be packing into those 54 constants!

I think Dr. Sanford is right in suggesting that statisticians and clinicians are really interested in predicting different kinds of things. But I want to force this out into the open, because I insist that many working clinicians are blissfully misusing the clinical method to predict the actuary's kind of thing. One program that I am sure all five of us can agree to, and recommend to you as both stimulating and socially significant research, is the empirical study of the two methods of prediction under the various conditions set forth by the four speakers. For what kind of criterion, given what kinds of data, with how much exposure, in what sequence, and so on and on, can the clinician (what clinician?) excel the actuary? There is room for many more studies trying various combinations of conditions before we have the answer. And I should say "answers"; because it will hardly be a decision as to who wins. Rather we will have trustworthy information as to which predictive problem is best handled by which method. Here I would like to go into the tremendous matter of *form* versus *content*, which I now tend to see as the real nub of the business. But that would take all night, so it will have to wait for another time.

pdf by LJY, October 2015