
Clinical Versus Statistical Prediction: The Contribution of Paul E. Meehl

▼
William M. Grove

University of Minnesota, Twin Cities Campus

The background of Paul E. Meehl's work on clinical versus statistical prediction is reviewed, with detailed analyses of his arguments. Meehl's four main contributions were the following: (a) he put the question, of whether clinical or statistical combinations of psychological data yielded better predictions, at center stage in applied psychology; (b) he convincingly argued, against an array of objections, that clinical versus statistical prediction was a real (not concocted) problem needing thorough study; (c) he meticulously and even-handedly dissected the logic of clinical inference from theoretical and probabilistic standpoints; and (d) he reviewed the studies available in 1954 and thereafter, which tested the validity of clinical versus statistical predictions. His early conclusion that the literature strongly favors statistical prediction has stood up extremely well, and his conceptual analyses of the prediction problem (especially his defense of applying aggregate-based probability statements to individual cases) have not been significantly improved since 1954. © 2005 Wiley Periodicals, Inc. *J Clin Psychol* 61: 1233–1243, 2005.

Keywords: Paul E. Meehl; clinical psychology; clinical prediction; statistical prediction, actuarial prediction; data gathering; data combination; MMPI codebook

A major problem with reviewing a limited area pioneered by a genius is that all the important, original ideas have already been expounded. My aim in this article is not to present novelty, but rather to summarize Meehl's work on clinical versus statistical prediction. This controversy concerns the comparative accuracy of two ways of combining predictive data: clinicians' informal judgments versus statistical (mechanical) combination. The history of the controversy is outlined. Paul Meehl's remarkable contributions to understanding and resolving this controversy are discussed.

This article is dedicated to the memory of P. E. Meehl—mentor, colleague, and friend. I also gratefully acknowledge the assistance of Leslie J. Yonce and Martin Lloyd in the completion of this article.

Correspondence concerning this article should be addressed to: William M. Grove, Psychology Department, N438A Elliott Hall, 75 East River Road, Minneapolis, MN 55455-0344; e-mail: grove001@umn.edu

The clinical versus statistical prediction debate is central to the assessment and prediction of human (and organizational) behavior because of the following argument formulated by Meehl. I have elaborated it slightly, based on discussions with Meehl and on my own reasoning, the better to exhibit its inexorable logic.

1. Psychologists spend a good deal of time making explicit, or at least implicit, predictions. For example, “clinical judgments about the therapeutic potential of cases” are implicit predictions that a patient will do better with treatment *A* than they would if they instead got treatment *B* (bearing in mind that either *A* or *B* could, in fact, represent no treatment—watchful waiting). Decisions about personnel selection, admission of students to institutions of higher learning, civil commitment of the mentally ill, eligibility for parole—all these are more or less explicit predictions that one course of action will turn out better than another in the future, if not for the client, student, or inmate, then for the institution on behalf of which predictions are being made.
2. “A . . . distinction is that between the source or type of information employed in making predictions, and the manner in which this information is combined for predictive purposes.” (Meehl, 1954, p. 15)
3. “As for the combining method, by *mechanical* (or *statistical*) I mean that the prediction is arrived at by some straightforward application of an equation or table to the data. [Today, we would add “or computer program.”] . . . The defining property is that no judging or inferring or weighing [during combination] is done by a skilled clinician . . . By *non-mechanical* or *informal* methods of combining I mean those of any other sort” (Meehl, 1954, p. 15–16). Note that this way of framing the distinction suffices to make clinical and statistical prediction mutually exclusive. If actuarial predictions are synthesized clinically, this is clinical prediction; if clinical predictions are synthesized statistically, this is statistical prediction. There is no such thing as a true hybrid. This point has been repeatedly misunderstood, and in my opinion often tendentiously and misleadingly argued, by some psychologists. One example is the argument that because a clinician can integrate the output of a statistical prediction scheme and his or her own judgments, a hybrid exists. This is false because from a formal standpoint, the clinician is (as usual) synthesizing two or more potentially disparate items of information available to them (the statistical prediction, and one or more other items, which may or may not have entered into the statistical prediction). Clinicians must use their own judgment to deal with discrepancies in the data, e.g., a statistical prediction that says a sex offender will recidivate, and a prison psychologist’s statement that the prisoner likely will not reoffend. This is precisely the sort of thing Meehl called “clinical data combination.” Another example arises when broad clinical judgments form at least part of the inputs to a statistical prediction scheme. Once clinicians have enunciated their judgments (diagnoses, trait ratings, prognostications) and they have been quantified (even as crudely as “will reoffend” vs. “won’t”), the statistical formula requires no professional judgment to arrive at a prediction; this is exactly what Meehl meant by “statistical data combination.”¹

¹This sharp demarcation does not prevent discovering that a particular subvariety of clinical prediction, e.g., our example of clinical recidivism prediction that is informed by the results of statistical-prediction scales, outperforms straight statistical prediction. Hence, nothing in terms of finding out how to make the best predictions is prejudged by making clinical and statistical prediction mutually exclusive, exhaustive categories.

4. "It is obvious . . . that the results of applying these two different [data combining] techniques to the same data set do not always agree. On the contrary, they disagree a sizeable fraction of the time. Now if a four-variable regression equation or . . . actuarial table tells the criminal court judge that this particular delinquent will probably commit another felony in the next 3 years and if a case conference or a social worker says he will probably not . . . the plain fact is that [the judge] cannot act in accordance with both of these incompatible predictions." (Meehl, 1986, p. 372)
5. The best prediction scheme is the one that produces the smallest error for each client. This might involve using clinical prediction in one case and statistical prediction in another case. However, to choose the data combining method on a case-by-case method, the overseer of predictions would have to know, in advance, which combination method would produce the best result for the individual case. Such information is, as far as I know always lacking. Note that dividing cases into classes, on the basis of before-the-prediction known characteristics of individuals or their situations, and using different data combination methods for different classes, is quite another matter and may prove eminently sound.
6. Barring such impossible-to-apply metarules as "Use the best data combination method for this particular case," the best prediction scheme, from the standpoint of Bayesian decision theory,² is one that maximizes expected utilities for the stakeholder on behalf of whom the prediction is made. Consider first yes–no (or succeed–fail) predictions (e.g., a patient will or won't soon attempt suicide, if not hospitalized at once). When relative costs of false positive versus false negative predictions are unknown, or agreement cannot be reached among stakeholders about their relative magnitudes, it is not irrational to treat the two kinds of cost as equal. Under such circumstances, the decision rule that minimizes expected costs is the one that maximizes the hit rate of predictions.³ For quantitative predictions (e.g., predicting number of publications in graduate school), minimizing the expected squared errors of prediction is a generally accepted rule in psychology. (It corresponds to the optimization criterion in multiple linear regression, for example.) For such a loss criterion, under- and overpredictions (corresponding to false negative and false positive yes–no predictions, respectively) of a given magnitude are treated equivalently.
7. Therefore, in the absence of reliable cost information, the best overall data combination method is the one with the best overall classification accuracy (hit rate) or best overall correlation with the criterion.
8. Overseers of prediction (i.e., clinicians making informal predictions, or users of a computer program making actuarial predictions) are obliged to maximize expected

²We set aside other decision rules, such as minimizing the maximum loss, because they produce such risk-averse decisions as to appear unattractive to most decision theorists.

³Expected disutility for such a decision equals $\Pr\{FP\} \times \text{cost}_{FP} + \Pr\{FN\} \times \text{cost}_{FN}$, where $\Pr\{FP\}$ is the probability of a false positive and $\Pr\{FN\}$ that of a false negative; the two costs are disutilities associated with each possible kind of misprediction. Without loss of generality, we can express this in terms of regrets (actual cost minus minimum potential cost, for a given state of nature; Blackwell & Girschick, 1954). Similarly, we can in turn express the decision problem in a metric for which $\text{regret}_{FN} = 1$. Then we want to minimize $\Pr\{FP\} \frac{\text{regret}_{FP}}{\text{regret}_{FN}} + \Pr\{FN\}$. When costs (regrets) are unknown and we take the default course (i.e., to act as if they are equal), then we want to minimize $\Pr\{FP\} + \Pr\{FN\}$ to maximize expected utility. This is the same as maximizing $\Pr\{TP\} + \Pr\{TN\}$ (true positives and negatives), which in turn, is the same as maximizing the hit rate.

utility for a relevant stakeholder.⁴ This is what it means to have a fiduciary responsibility (“Principle A, Beneficence and Nonmaleficence”; American Psychological Association [APA] Council of Representatives, 2002).

9. The principle of beneficence therefore generally requires psychologists to choose and use the prediction method that yields the most accurate predictions. This method may differ from class to class of prediction situations. “It is foolish, and I would say even immoral, for a trusted . . . expert . . . to employ a method which has a lower hit-frequency than another available method.” (Meehl, 1956b, p. 163)

When first presented, perhaps fewer psychologists found this argument compelling than would find it so today. It is worthwhile to examine work predating Meehl’s book to understand the context in which Meehl’s contribution entered the field.

How Meehl Found the Controversy, Circa 1954

The first empirical work known to me is Burgess’ 1928 study of parole failure prediction. There were theoretical debates, and there had been an APA symposium (chaired by E. L. Kelly, mentioned in Meehl, 1986). The principal protagonists of actuarial prediction were Lundberg (1926, 1941) and then Sarbin (1941, 1942, 1944), the latter beginning with his work at the Elgin State Hospital (where one of the early actuarial prediction studies was done). Lundberg and Sarbin were advancing a much stronger thesis than that actuarial prediction typically outperforms clinical prediction. Instead, Sarbin argued that, as Meehl paraphrased it, “the clinician is always predicting actuarially and from classes whether he knows it or not” (Meehl, 1954, p. 29). There had also been published a limited number of empirical comparisons of clinical and statistical prediction, almost all of them significantly methodologically flawed. The most comprehensive review (Sarbin, 1944) covered only four studies.

The essence of Sarbin’s theoretical position was that the clinician operates as an empiricist, gathering impressions of the past that allow prediction of the future. The clinician is forming, in essence, *S*–*R* bonds that link assessment data on the *S* side to predictions on the *R* side. All such bonds are inherently probabilistic and are based on the comparison of the current data with past data, making a prediction for the current case that is similar to that for a past case with similar data. According to Sarbin, single-case probability statements are unverifiable; hence, they are meaningless (at least according to Peirce’s empirical-conditions-of-verification theory of meaning). For example, suppose one says, “The probability that Professor *A* will attend the movies tonight is .7.” Now, either *A* will attend, or she will not attend; no sensible measure of attendance can yield the value .7, no matter what *A* does. Hence (on the Peircean theory), the statement is meaningless. Sarbin followed Reichenbach (1938) by identifying the rationality of single-case probability statements with their value for deciding on action. Clinicians simplify their cognitive task by (in principle) putting stimuli (assessment data) into equivalence classes and predicting like outcomes for members of the same class. Naturally, this is not supposed to be a complete rendition of what clinicians actually do. In fact, due to human cognitive limitations (e.g., on memory) they imperfectly approximate the behavior of an actuarial table. It was for this reason that Sarbin held that clinicians could not, in principle, outperform an actuarial table—the table engenders residual predictive error (to the

⁴We neglect situations where there are multiple stakeholders whose utility assessments conflict. As far as I know, this does not tilt the balance in favor of, or against, either clinical or statistical prediction.

extent that the table isn't a perfect predictor of behavior), whereas the clinician's head *also* engenders recollection error, misconstrual, and other errors.

Against this background, Meehl began thinking about the prediction problem, the nature and proper role of clinical inference, and formal prediction methods in the 1940s. After about a decade wherein he spent a substantial part of his working hours applying his capacious intellect to these problems, he published *Clinical Versus Statistical Prediction*. He had to offer it to three publishers to get it in print, because editors were sure it would not sell. (The first two publishers made clinical judgment errors—the book went through seven printings and, after being out of print for a time, is again selling well; Meehl, 1996. It has a phenomenally high ongoing citation rate, when one considers it is almost 50 years old.)

What Meehl Accomplished in His 1954 Book

Meehl made four major contributions in his famous little book (and subsequent related work). The first and most essential of these was to draw a sharp distinction between data gathering (e.g., interview, psychometrics, behavioral observation) from data combination, with a subsequent focus almost exclusively on the latter (Meehl, 1954, p. 18). He identified statistical, actuarial, or formal prediction with procedures that required no professional judgment; they could be carried out by a clerk (today, we would say by a computer program). Clinical, informal, or impressionistic prediction includes everything else.

A second contribution was to recognize that, thus put, the clinical versus statistical prediction distinction is unavoidable. In 1986, he made explicit what was, to some extent, implicit in 1954:

Some critics [of my 1954 book] asked a question that Dr. Holt still asks, and which I confess I am totally unable to understand. Why should Sarbin and Meehl be fomenting this needless controversy? Let me state as loudly and as clearly as I can manage . . . that Sarbin and I did not artificially concoct a controversy . . . between two methods that complement each other and work together harmoniously. I think this is a ridiculous position when the context is the pragmatic context of decision making. You have two quite different procedures for combining a finite set of information to arrive at a predictive decision. . . . [These two procedures] disagree a sizeable fraction of the time. . . . The plain fact is that [a decision maker] cannot act in accordance with both of [two] incompatible predictions. Nobody disputes that it is possible to improve clinicians' practices by informing them of their track records actuarially. Nobody has ever disputed that the actuary would be well advised to listen to clinicians in setting up the set of variables. (Meehl, 1986, p. 372)

A third contribution was his subtle and engaging treatment of the problems and processes of clinical judgment. He was very sympathetic to the clinician, as one might expect. (Paul was a practicing psychotherapist who favored psychodynamic and rational-emotive methods.) Of the book's 10 chapters, the second through seventh concerned clinical judgment, with titles like "The Special Powers of the Clinician," "The Problem of the Logical Reconstruction of Clinical Activity," and "Remarks on Clinical Intuition." In the fifth chapter, he argued that Sarbin was mistaken in contending that clinicians can in principle, do no more than imitate actuarial tables. Paul colorfully paraphrased Sarbin's position as "the clinician is a second-rate substitute for a Hollerith machine" (Meehl, 1954, p. 76).

A central part of the discussion of limitations on actuarial prediction accuracy stems from Paul's insights about predicting behavior from class frequencies. This led to Paul's famous "broken leg case" example, which incidentally shows how actuarial prediction is

not limited to making judgments about groups of individuals. Suppose we observe Professor A, and find he regularly goes to the movies on Tuesday nights. Our actuarial table, built on what may be years of observation says, "If it's a Tuesday night, then $\Pr\{\text{Professor A goes to movies}\} = .9$." But suppose we also know that on Tuesday morning Professor A broke his leg; his hip cast will not allow him to fit into a theater seat. (This example was concocted circa 1954, before handicapped-accessibility laws!) Any human judge with a modicum of sense will *not* say that $\Pr\{\text{goes to movies}\}$ for that night equals .9, or anything like .9; they'll predict a probability in the neighborhood of zero. This is one of Paul's "special powers of the clinician."

As was typical for Paul, he identified this critical problem with actuarial prediction, and then he went further, demonstrating the central flaw in the critique as follows. First, the base rate of people breaking a leg is low, and so "broken leg cases" need not substantially decrease the accuracy of statistical predictions. Second, in the broken leg example, we have a highly reliable theory allowing us to predict clinically that Professor A has a very small probability of going to the movies; the theory rests on the physics of fitting a person with a hip cast into a 1954 movie seat. However, behavioral science theories are extremely seldom as well corroborated as those of physical mechanics. Third, the broken leg in Meehl's example reduces the probability of movie attendance to zero, or some figure close to it. By contrast, when rare events occur in applied psychology, the event seldom guarantees that a person will, or will not, engage in the behavior of interest. Indeed, human behavior is so multidetermined that even unusual events typically change the probabilities modestly, or at most moderately. In sum, it is easy to see that "broken leg cases" exist and offer an opportunity for the clinician to do what the formula cannot. However, as Paul always maintained, it is very difficult to know whether a given case is a bona fide broken leg case, i.e., whether the clinician should overrule the actuarial prediction for a particular individual, or follow it.

Another important part of Paul's third contribution was his rebuttal of Sarbin. Briefly, he pointed out that Sarbin did not "distinguish carefully between *how you get there* [i.e., to a certain judgment] and *how you check the trustworthiness* of your judgment" (Meehl, 1954, p. 31). Second, Sarbin mistook the form of predictions, by putting them in single-statement form, "Student X has one chance in 6 of meeting the standards of competition in the university", whereas they actually involve two statements, the first of which is the actual prediction, "The student X will not succeed at the university" (Meehl, 1954, p. 32). The second statement quantifies the uncertainty of the prediction: "The probability that the prediction is correct is five in six." (Note that I have used Quine corners to clarify the extents of statements, instead of nesting quotation marks.) Sarbin's way of collapsing the two statements had the effect of tendentiously making the clinician sound like an unreliable actuary; Paul rejected this conflation. Paul also criticized Sarbin's implicit position that there is a single true probability of student X succeeding, to which the clinician only approximates. Instead, on a frequentist view (Reichenbach, 1938; or even better, Kyburg, 1974) there are as many true probabilities as there are reference classes to which X belongs.

Finally, in his chapter on the logic of clinical activity, Meehl delivered his strongest argument against Sarbin's central thesis. He proposed an analogy that shows clinicians can in fact, be doing something *in principle different* from counting up frequencies in reference classes. The analogy involves a box repairperson (the "clinician") who is presented with an opaque box with push-on buttons on one side and lights on the other. One can press buttons and then observe what, if any, lights come on. The way these boxes are constructed is generally similar, but wiring details vary from box to box. Our repairperson knows this fact from examining many such boxes inside and out. Given all this, our

repairperson may press shrewdly chosen combinations of buttons, observe the results, and then concoct a structural-dynamic theory of how a particular box is wired. A few more presses may suffice to test the theory. If the repairman's theory is essentially correct, it will generally yield quite accurate predictions. Now, clinicians could at least sometimes be doing something of this sort when they examine clients and think about them. If they do something of this kind, they are not just counting up frequencies for *S-R* connections.

Many people who claim acquaintance with the book seem only to know about chapter 8, which systematically examines 20 empirical studies comparing clinical to statistical prediction. A box score approach was used, effectively counting tied performances between clinical and statistical prediction as "wins" for the actuary. This review was the fourth major accomplishment of Meehl in his book. It was also the locus of the only flat-out error in the book that is known to me. In discussing Hovey and Stauffacher's (1953) study of clinical versus actuarial MMPI interpretation, Paul did not detect a mistake (pointed out by McNemar, 1955) that the original authors made in computing the degrees of freedom on a test. When Meehl (1986) said he would change at most 5% of the book, he wasn't referring to unambiguous error. Instead, he meant that, in retrospect, it seemed he had overemphasized the advantage configural judgment would afford clinicians.

The book closes by making a point that depends on his earlier delineation of two uses of statistics: structural versus discriminative (Meehl, 1954, pp. 11–15). Statistics are used structurally when one models the nature and relationships of entities and events, e.g., when a factor analysis is conducted to explain the correlations between cognitive test scores, or when a prediction formula is developed. Statistics are used discriminatively to test whether, for example, one type of fertilizer gives better crop yields than another fertilizer, or clinician-generated predictions are more accurate than actuarial ones. Meehl points out that if the clinician eventually proves to be more accurate than the actuary is, this can only be demonstrated by the discriminative use of statistics. In that sense, "always the actuary will have the final word." (p. 138)

What Meehl accomplished in his 1954 book was remarkable. By my estimation, he gave himself a very tough act to follow. Because he got so much right so early, he did not have occasion to publish fundamentally new or substantially changed views on the subject, though he continued thinking and writing about this topic throughout his career.

Meehl's Work on Clinical Versus Statistical Prediction After 1954

From 1954 until late in his life, Paul collected numerous articles concerning clinical judgment and clinical versus statistical prediction, many of them sent to him by scholars around the country. Those that contained empirical comparisons between the two methods he summarized narratively on 66 tapes (originally onto magnetized belts—"dictabelts") for use in a second and enlarged edition of his book—which was, alas, never written. These summaries were the springboard for my own meta-analysis of studies concerning this controversy.

Meehl followed up the book with practical, if preliminary, first steps on the path suggested by the data. His student Hallbower (1955) had completed a dissertation concerning actuarial prediction from the MMPI by assigning profiles configurally to classes based on combinations of several high (and, more rarely, low) clinical scale scores; a manual giving probabilistic predictions about individuals having each of these "code types," but based on limited *Ns* for each type, was provided. Meehl discussed this work and its implications in *Wanted—A Good Cookbook* (Meehl, 1956a). This led to a cottage industry of researching and cataloging MMPI, and now MMPI-2, code types (Gilberstadt

& Duker, 1965; Gynther, Altman, & Sletten, 1973; Lachar, 1974; Marks, Haller, & Seeman, 1997; Marks & Seeman, 1963; Marks, Seeman, & Haller, 1974; McGrath & Ingersoll, 1999), the most thoroughgoing and successful application of actuarial prediction in clinical psychology.

In 1957, Paul again addressed the vexing “broken leg problem” in “When Should We Use Our Heads Instead of the Formula?” He established four points about his “broken leg” example that *differ markedly* from typical occasions when a clinician is moved to countervail an actuarial prediction:

1. A broken leg is “an objective fact, determinable with high accuracy”;
2. the “correlation [of a broken leg] with relative immobilization is near perfect, based on a huge *N*, and attested by all sane men . . .”;
3. Interaction effects between a broken leg and various other factors that influence going to the movies “are conspicuously lacking”—hence, the broken leg fact can be considered in isolation; and
4. “the prediction [that a hip-casted professor will not go to the movies] is mediated without the use of any doubtful theory . . .” (Meehl, 1957, p. 85)

These conditions are not met when Dr. Smith decides, based on his “gut” feeling, that Ms. Jones has a markedly hysteroid personality and so won’t do well in insight-oriented psychotherapy, despite an actuarial prediction that she will. Alas, Meehl did not give a convincing analysis of when, if ever, clinicians are justified in departing from actuarial predictions under less than “broken leg” circumstances. Meehl emphasized that this is not, primarily, a conceptual problem. It is mostly a matter of the rarity of exceptions, which make us reasonably want to “break the actuarial rules,” and the consequent paucity of good outcome data on clinicians’ success rates when they do break those rules.

In 1965 (“Seer over Sign: The First Good Example”), Paul mentioned, but did not fully present, 51 studies on clinical versus statistical prediction, which tended toward the same conclusion he reached in 1954. The main point of this article, however, was his belief that he had found a clinical prediction notably superior to actuarial schemes. This time the evidence came from two studies, both involving the detection of homosexual orientation from the Thematic Apperception Test. However, as with Hovey and Stauffacher (1953) before, this sanguine view was subsequently corrected (Goldberg, 1968). It seems that Paul neglected to notice, or failed to give sufficient weight to, the fact that across two studies, no clinical judge showed sustained performance superior to the actuary. The clinicians in Study I who outperformed statistical prediction showed inferior accuracy in Study II, and vice versa. As a result, sampling error is a plausible explanation for any apparent superiority of clinical prediction.

Meehl’s 1986 “Causes and Effects of My Disturbing Little Book” article, written for a symposium on the 40th anniversary of his original book, relates a bit of history of Meehl’s beginning interest in this topic, and the considerable difficulty he had in publishing it. It is here that he explicitly refutes Zubin’s (1956) and Holt’s (1958) claims that this is a false controversy, or based on a misunderstanding. Zubin (1956), for example, stated, “the distinction between actuarial and clinical prediction is heuristic rather than basic” (p. 627). Zubin also states that from his chosen point of view, “the question of whether the actuarial approach is superior to the clinical is tantamount to asking whether the sperm is more important than the ovum” (p. 627). The point of view of which Zubin speaks is the advancement of our knowledge about the prediction of human behavior. But this is not the problem Meehl tried to solve; his problem was to pick the right tool for the practical task at hand. If by “heuristic” Zubin meant “pragmatic,” then he and Meehl

were, this far, in agreement. When presented with a current need to make predictions for a series of cases, clinical and statistical prediction methods do not cooperate like a ratchet wrench and its socket set, let alone an ovum and a sperm. Instead, they compete like screwdriver and pry bar, as ways to open a can of paint.

I quoted Meehl's answer to such claims earlier in establishing the controversy's importance. Suppose one grants, as I would not, that the distinction between clinical and statistical classes of prediction methods is in some ways conventional, rather than objectively deriving from two crucial, linked facts—nonmechanical prediction schemes require subjective judgments, and as such always lack complete reproducibility. Even granting *ex hypothesi* that the distinction is not written in stone, this would not diminish the importance of answering the question: "For a given prediction problem, which of these 'conventionally' distinguishable methods yields at present the more accurate predictions?" Meehl made much the same point about the non-artificiality of the controversy in Grove and Meehl (1996).

An article in *Science* by Dawes, Faust, and Meehl (1989) referred to an in-process meta-analysis of comparison studies of clinical versus statistical prediction. That meta-analysis (Grove, Zald, Lebow, Snitz, & Nelson, 2000) summarized results from 136 studies concerning the prediction of human health and behavior. Dawes et al. (1989) speculate as to why so few practitioners seem to have changed practice habits in the face of such an increasingly massive and consistent body of evidence. This same topic was treated in Grove and Meehl (1996), wherein a preliminary analysis of effect sizes from the meta-analysis was reported. (Meehl wrote well over half of this paper, but we flipped a coin to establish the authorship order.) Meehl also very carefully analyzed numerous objections to the meta-analytic results, or to the imperative that actuarial prediction be used much more often. Paul's final noteworthy writing on the topic was his new forward to the 1996 reprint edition of his 1954 book (Meehl, 1954/1996, pp. v–xii). There, he refuted several "legends" about him and his views, which try (or tend) to undercut the perception that he approached this issue with an open mind.

Clearly, Paul's views of major issues in the clinical versus statistical prediction controversy changed little over the years. No great theoretical breakthroughs, radically improved prediction methods, or momentarily different empirical outcomes emerged after 1954. However, this did not have the effect of making Meehl's ideas "old hat" or reducing his contribution to an historical footnote. Indeed, to this day, applied and cognitive psychologists find his writings fresh, insightful, unusually clear and careful, and engagingly presented.

Importance of Paul's Ideas and Conclusions Today

As just noted, none of Paul's "disturbing little book" needs to be retracted and little even needs significant modification. Over time, it became clear that in 1954, he was somewhat too sanguine about clinicians' abilities to process configurations of (nonlinear interactions among) assessment data. Nevertheless, he was prescient about how the studies would turn out, extrapolating rather accurately from 20 mostly not very good studies, to over 160 often well-conducted comparisons.

In 1954, there were few prediction problems whose features were likely to enable clinical predictions to shine. A half-century later, clinicians still operate at a decided disadvantage in psychology (and to a slightly smaller extent, in medicine as well). There are two main reasons for the enduring intractability of clinical prediction. First, we lack well-corroborated, comprehensive nomothetic (let alone idiographic) theories of how the

mind works, which allow accurate theory-mediated individual predictions. Second, the cost of clinical judgments is a bigger issue than ever.

In contrast, the application of statistical prediction has grown, albeit slowly and perhaps more outside of mainstream psychology than within it. First, actuarial MMPI interpretation has been conspicuously successful since the 1960s. Second, the prediction of criminal recidivism, one of the earliest applications, has likewise grown (e.g., Hanson & Thornton, 2000). Third and finally, the use of automated decision rules and aids in medicine has been adopted with moderate enthusiasm as one aspect of "Evidence Based Medicine" (Elstein et al., 1996).

It remains important for applied psychologists in general and cognitive psychologists in particular, not just clinical psychologists, to familiarize themselves with Meehl's thinking on the clinical versus statistical prediction debate. Although his writings are peppered with references to philosophy of science and formal logic, they are quite accessible to any competent student of psychology who will take the trouble to read them with care. Until the day when a more exact, well-corroborated psychological theory of clinical judgment becomes available, his analyses (and even his speculations) will remain valuable. Unless and until the very unlikely happens, and a body of evidence emerges that overturns the massive set of published studies favoring statistical prediction, his substantive conclusions regarding the *prima facie* best approach to predicting human behavior will stand indefinitely. The sheer intellectual pleasure to be had by reading Meehl's uniquely styled, engaging presentations of the fruits of his considerable intellect will, I am sure, never dissipate.

References

- American Psychological Association Council of Representatives. (2002). Ethical principles of psychologists and code of conduct. Retrieved August 27, 2003, from <http://www.apa.org/ethics/code2002.html>
- Blackwell, D., & Girschick, M.A. (1954). *Theory of games and statistical decisions*. New York: Wiley.
- Burgess, E.W. (1928). Factors determining success or failure on parole. In A.A. Bruce (Ed.), *The workings of the indeterminate sentence law and the parole system in Illinois*. Springfield: Illinois State Board of Parole.
- Dawes, R.M., Faust, D., & Meehl, P.E. (1989). Clinical versus actuarial judgment. *Science*, 243, 1668–1774.
- Elstein, A.S., Friedman, C.P., Wolf, F.M., Murphy, G., Miller, J., Fine, P., et al. (1996). Effects of a decision support system on the diagnostic accuracy of users: A preliminary report. *Journal of the American Medical Informatics Association*, 3, 422–428.
- Gilberstadt, H., & Duker, J. (1965). *A handbook for clinical and actuarial MMPI interpretation*. Philadelphia: Saunders.
- Goldberg, L.R. (1968). Seer over sign: The first "good" example? *Journal of Experimental Research in Personality*, 3, 168–171.
- Grove, W.M., & Meehl, P.E. (1996). Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: The clinical-statistical controversy. *Psychology, Public Policy, and Law*, 2, 293–323.
- Grove, W.M., Zald, D.H., Lebow, B.S., Snitz, B.E., & Nelson, C. (2000). Clinical versus mechanical prediction: A meta-analysis. *Psychological Assessment*, 12, 19–30.
- Gynther, M.D., Altman, H., & Sletten, I.W. (1973). Replicated correlates of MMPI two-point code types: The Missouri actuarial system. *Journal of Clinical Psychology Monographs*, 29, 263–289.

- Hallbower, C.C. (1955, March). A comparison of actuarial versus clinical prediction to classes discriminated by MMPI. Unpublished doctoral dissertation, University of Minnesota.
- Hanson, R.K., & Thornton, D. (2000). Improving risk assessments for sex offenders: A comparison of three actuarial scales. *Law and Human Behavior*, 24, 119–136.
- Holt, R.R. (1958). Clinical and statistical prediction: A reformulation and some new data. *Journal of Abnormal and Social Psychology*, 56, 1–12.
- Hovey, H.B., & Stauffacher, J.C. (1953). Intuitive versus objective prediction from a test. *Journal of Clinical Psychology*, 9, 349–351.
- Kyburg, H.E., Jr. (1974). *The logical foundations of statistical inference*. Dordrecht, Holland: D. Reidel.
- Lachar, D. (1974). *The MMPI: Clinical assessment and automated interpretation*. Los Angeles: Western Psychological Services.
- Lundberg, G.A. (1926). Case work and the statistical method. *Social Forces*, 5, 61–65.
- Lundberg, G.A. (1941). Case studies versus statistical methods: An issue based on misunderstanding. *Sociometry*, 4, 379–383.
- Marks, P.A., Haller, D.L., & Seeman, W. (1997). *Actuarial use of the MMPI: Adolescents and adults*. Cambridge: Oxford University Press.
- Marks, P.A., & Seeman, W. (1963). *Actuarial description of abnormal personality*. Baltimore: Williams & Wilkins.
- Marks, P.A., Seeman, W., & Haller, D.L. (1974). *The actuarial use of the MMPI with adolescents and adults*. Baltimore: Williams & Wilkins.
- McGrath, R.E., & Ingersoll, J. (1999). Writing a good cookbook: I. A review of MMPI high-point code system studies. *Journal of Personality Assessment*, 73, 149–178.
- McNemar, Q. (1955). Review of clinical versus statistical prediction. *American Journal of Psychology*, 68, 510.
- Meehl, P.E. (1954). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence*. Minneapolis: University of Minnesota.
- Meehl, P.E. (1956a). Wanted—A good cookbook. *American Psychologist*, 11, 263–272.
- Meehl, P.E. (1956b). Symposium on clinical and statistical prediction: The tie that binds. *Journal of Counseling Psychology*, 3, 163–173.
- Meehl, P.E. (1957). When shall we use our heads instead of the formula? *Journal of Counseling Psychology*, 4, 268–273.
- Meehl, P.E. (1965). Seer over sign: The first good example. *Journal of Experimental Research in Personality*, 1, 27–32.
- Meehl, P.E. (1986). Causes and effects of my disturbing little book. *Journal of Personality Assessment*, 50, 370–375.
- Meehl, P.E. (1996). *Clinical versus statistical prediction: A theoretical analysis and a review of the evidence* (pp. v–xii). Lanham, MD: Rowan & Littlefield/Jason Aronson. (Original work published 1954)
- Reichenbach, H. (1938). *Experience and prediction*. Chicago: University of Chicago.
- Sarbin, T.R. (1941). Clinical psychology—Art or science? *Psychometrika*, 6, 391–400.
- Sarbin, T.R. (1942). A contribution to the study of actuarial and individual methods of prediction. *The American Journal of Sociology*, 48, 593–602.
- Sarbin, T.R. (1944). The logic of prediction in psychology. *Psychological Review*, 51, 210–228.
- Zubin, J. (1956). Clinical versus actuarial prediction: A pseudo-problem. In N. Sanford, C.C. McArthur, J. Zubin, L.G. Humphreys, & P.E. Meehl (Eds.), *Proceedings of the Conference on Testing Problems* (pp. 625–637). Princeton, NJ: Educational Testing Service.

Copyright of Journal of Clinical Psychology is the property of John Wiley & Sons Inc.. The copyright in an individual article may be maintained by the author in certain cases. Content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.