Journal of Abnormal Psychology
2006, Vol. 115, No. 2, 192–194

Copyright 2006 by the American Psychological Association
0021-843X/06/$12.00 DOI: 10.1037/0021-843X.115.2.192

Meehl's Contribution to Clinical Versus Statistical Prediction

William M. Grove and Martin Lloyd
University of Minnesota, Twin Cities Campus

Paul E. Meehl's work on the clinical versus statistical prediction controversy is reviewed. His contributions included the following: putting the controversy center stage in applied psychology; clarifying concepts underpinning the debate (especially his crucial distinction between ways of gathering data and ways of combining them) as well as establishing that the controversy was real and not concocted, analyzing clinical inference from both theoretical and probabilistic points of view, and reviewing studies that compared the accuracy of these 2 methods of data combination. Meehl's (1954/1996) conclusion that statistical prediction consistently outperforms clinical judgment has stood up extremely well for half a century. His conceptual analyses have not been significantly improved since he published them in the 1950s and 1960s. His work in this area contains several citation classics, which are part of the working knowledge of all competent applied psychologists today.

Keywords: clinical prediction, actuarial prediction, statistical prediction, clinical judgment

Prediction is a central, indeed nearly ubiquitous, activity of psychologists. Many clinical decisions, such as treatment selection, depend on predictions. Psychologists, or at least applied psychologists, are, therefore, obliged to know as much as possible about how to make good predictions.

Paul E. Meehl saw this clearly. He focused on practical contexts in which predictions must be made immediately, based on then-available information. Meehl argued that in such contexts psychologists must choose between clinical and statistical prediction methods, and that they should use whichever method yields the most accurate predictions in the long run. His argument proceeded as follows:

1. There are various ways of gathering predictive data, such as interviews, direct observation, and psychometric tests. No matter how gathered, such data can be encoded or quantified. Encoded data can then be combined by a professional using trained judgment or mechanically (e.g., by a formula, actuarial table, or computer program). These are mutually exclusive and exhaustive classes of ways to combine data; the relative value of these two classes is a meaningful question, whether the data to be combined come from interviews, Rorschach protocols, Minnesota Multiphasic Personality Inventory-2 (MMPI-2) profiles, or behavior counts.

2. There is no true hybrid of these data combination methods; this point is too often misunderstood or erroneously denied. True, a clinician can be given, as one predictive datum to consider, the output of a statistical formula; a formula can include a variable representing a quantified clinician judgment. However, in the former situation, the final prediction depends on trained judgment, whereas in the latter it does not. Clinician judgments, for which statistical predictions were available as cues, can and should be studied to learn whether they are more accurate than clinical predictions made in the absence of such cues; likewise, mechanical predictions based in part on quantified judgments can be compared with mechanical predictions based solely on nonjudgmental data. The existence of these subtypes of clinical and mechanical prediction, respectively, does not make the clinical–statistical distinction meaningless, arbitrary, or lacking in practical interest.

3. When both clinical and statistical predictions are available for a given individual, they will not always agree, and then one cannot follow both. For a given prediction problem, it will surely be the case that the two data combination methods do not yield precisely the same accuracy of their predictions. Lacking clairvoyant knowledge of how the current case will turn out, the best way to maximize one's predictive accuracy is to use whichever data combination method yields the most accurate predictions in the long run.

4. Which data combination method is most accurate, for a given prediction problem, is a pragmatic question empirically answerable by running appropriately designed studies.

Logical arguments for and against mechanical prediction had been published since the 1920s, but Sarbin's (1944) review, the most comprehensive available when Meehl began work, covered only four studies. Meehl saw Sarbin's theoretical analysis, which argued a priori for the superiority of mechanical prediction, as not doing justice to the potential flexibility of clinical judgments.

William M. Grove and Martin Lloyd, Department of Psychology, University of Minnesota, Twin Cities.

This article is dedicated to the memory of P. E. Meehl—mentor, colleague, and friend. We thank Leslie J. Yonce for her help in completing this article.

Correspondence concerning this article should be addressed to William M. Grove, Department of Psychology, University of Minnesota, N218 Elliott Hall, 75 East River Rd, Minneapolis, MN 55455-0344. E-mail: grove001@umn.edu

(Sarbin postulated that the clinician essentially engages in the same kinds of weighting-and-adding processes used in statistical prediction formulas, but clinicians calculate less reliably and so are less accurate.) Meehl also had serious reservations about antiactuarial arguments published by some psychologists (e.g., Lundberg, 1941). He analyzed the controversy in a very short but powerful 1954 book, *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence*. It became a citation classic, going through seven printings by its original publisher and brought back into print in 1996 (Meehl, 1954/1996).

Key Features of Meehl's Book

Meehl made four major contributions to the clinical–statistical controversy in his book. First, he sharply distinguished data gathering from data combination, focusing on the accuracy of clinical versus mechanical methods for combining data. He identified statistical (actuarial, mechanical, formal, algorithmic) predictions as those requiring no professional judgment; they could be carried out by a clerk or computer. The clinical (informal, impressionistic, intuitive) method comprised all other sorts of predictions.

Meehl's second major advance was a convincing refutation of an often-repeated claim, namely that the clinical–statistical antithesis is artificial, because both methods can be used together. Implicit in this view is the idea that it is not necessary to choose between the two approaches. In 1986, Paul made this argument very pointedly:

> Some critics [of the 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. . .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 sizable 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 Paul's subtle analysis of clinical judgment. He was quite sympathetic to the clinician's potential for creative insight, not surprisingly because he was a practicing psychoanalytic psychotherapist. Six of the book's 10 chapters concerned topics such as "The Special Powers of the Clinician," "The Problem of the Logical Reconstruction of Clinical Activity," and "Remarks on Clinical Intuition." Meehl argued strongly that Sarbin's assertion, in effect that "the clinician is a second-rate substitute for a Hollerith machine" (Meehl, 1954/1996, p. 76), was erroneous.

Paul became identified in the field as a fervent proactuarial psychologist, especially by people who heard or read about his work but did not carefully read it for themselves. In fact, his book thoroughly analyzed inherent (i.e., irremediable) limitations of actuarial prediction accuracy. Paul's used what became known as the "broken-leg case" to explore this issue. It goes like this: We have observed that Professor A quite regularly goes to the movies on Tuesday nights. Our actuarial data support the inference "If it's a Tuesday night, then Pr{Professor A goes to movies} = .9." However, suppose we learn that Professor A broke his leg Tuesday morning; he's in a hip cast that won't fit in a theater seat. Any neurologically intact clinician will *not* say that Pr{goes to movies} = .9; they'll predict that he won't go. This is a "special power of the clinician" that cannot, in principle, be completely duplicated by even the most sophisticated computer program. That's because there are too many distinct, unanticipated factors affecting Professor A's behavior; the researcher cannot gather good actuarial data on all of them so the program can take them into account.

For several reasons, this argument does not prove that clinicians will generally outpredict statistical formulas. First, Meehl noted that broken legs are rare, so the clinicians avoiding error in such cases will not greatly increase their accuracy compared with statistical prediction. Second, the example involves a well-corroborated theory (namely, skeletal mechanics, which tells us how limbs—with or without casts on them—can and cannot be fitted into spaces), which predicts to near certainty that Professor A will not go to the movies. In contrast, psychological theories are very seldom this reliable; the advantage for a psychologist judge will decrease as the theory, which justifies making an exception to the statistical rule, becomes less dependable or exerts less influence on the behavior in question. In sum, broken-leg cases exist. They offer an opportunity for clinicians to be more than "second-rate Hollerith machines." However, the degree to which broken-leg cases actually allow clinicians to outpredict actuarial tables is an empirical matter. Moreover, it is very difficult to discern, in individual cases, whether one is justified in overriding a statistical prediction, when there appears to be an exceptional set of circumstances.

Meehl's chapter 8 reviewed 22 studies comparing clinical and statistical prediction. His box score, strongly favoring statistical prediction, was the fourth major point; many psychologists only know about this part of the book. (McNemar, 1955, pointed out a statistical error in one reviewed study, which Paul failed to catch, correction of which only strengthened the case in favor of statistical prediction.)

Meehl (1986) said he would, in hindsight, change at most 5% of the book. He thought that in 1954 he had overemphasized the advantage that configural (nonlinear, interactive) data-combining judgments might give clinicians over the actuary.

Meehl's Later Work

Meehl got so much right in 1954 that he did not have to publish substantially amended opinions on this subject. He encouraged applying actuarial prediction to clinical assessment in the 1950s. His student Hallbower (1955) developed the first actuarial code book for MMPI interpretation, assigning profiles to classes based on combinations of several high (and, more rarely, low) clinical scale scores. A manual giving probabilistic predictions about individuals having each code type was provided. Meehl discussed this work and its implications in "Wanted—A Good Cookbook." Many subsequent code book studies were published, based on larger samples (see, e.g., Gilberstadt & Duker, 1965; Marks & Seeman, 1963). This has been the most widely accepted application of actuarial prediction in clinical psychology.

Meehl (1965) mentioned 51 studies known to him that compared clinical and statistical prediction, more than twice the size of the 1954 literature. These investigations almost uniformly confirmed Meehl's earlier conclusion: Statistical prediction essentially always worked at least as well, and usually worked better, than

clinical prediction. Meehl's (1986) "Causes and Effects of My Disturbing Little Book," delivered at the 40th anniversary symposium for his book, relates the genesis of Paul's interest in this problem and the trouble he had publishing his 1954 book (it was refused by two publishers because they confidently predicted it would not sell). Dawes, Faust, and Meehl (1989) briefly summarized the literature but gave most attention to reasons why few clinicians changed their practices in the face of ever-increasing, consistent evidence favoring statistical prediction. We surveyed a 10% random sample of American Psychological Association Division 12 (clinical) psychologists to learn how familiar they were with the controversy, their views on the matter, and their clinical practices. Of 183 responders (28% response rate), more than 15% had never heard of the controversy or had merely heard that it existed; only 42% had covered the controversy in detail during their training; 10% had not been taught that there were any available statistical prediction methods, let alone what they were or how to use them, and another 6% had only had the existence of such methods mentioned. These findings, supplemented by other reasons for not using statistical prediction discussed by Dawes et al., emphasize the need to improve training on this issue.

A meta-analytically based box score regarding 136 comparative studies (66 of them accumulated by Meehl) was reported by Grove and Meehl (1996). Of these studies, just eight notably favored clinical prediction; no study favoring the clinician replicated any other study.

Importance of Paul's Work Today

In my opinion, Meehl's dissection of prediction methods into two incompatible kinds (clinical and mechanical–algorithmic), when psychologists operate in specific practical prediction contexts, is a fundamental and invaluable insight. As Paul pointed out, there may well be reasoning processes that clinicians sometimes use that a formula, table, or computer program cannot precisely mimic. However, whether such reasoning actually helps clinicians dependably outperform statistical formulas and computer programs is an empirical question with a clear, convincing answer: No, for prediction domains thus far studied. The burden of proof is now squarely on clinicians' shoulders to show, for new or existing prediction problems, that they can surpass simple statistical methods in accurately predicting human behavior.

It remains important for applied psychologists in general and cognitive psychologists in particular, and not just clinical psychologists, to know Meehl's work on the clinical versus statistical prediction controversy. Industrial–organizational psychologists (including nonclinical applied psychologists in the military) are frequently faced with personnel selection problems that involve predicting future behaviors (e.g., work output, time lost from work, dishonest acts, advancement in the organization, being a good team player) from interviews, questionnaires, biographical data, and other data. Formally, this is exactly the same kind of data-combining problem faced by clinicians. Moreover, quite a few of the relevant studies in the literature (e.g., the meta-analysis) actually involve work-related behaviors. Similarly, educational and school psychologists have often been asked to predict who will succeed in schooling or training based on previous performances (e.g., grade point), test scores, and other information (e.g., recommendation letters). They are also frequently used to help select students deemed in need of individualized tutelage or nonmainstream education because of predicted failure in regular schooling; this typically involves integrating past performances, cognitive measures, and observed classroom behavior. Cognitive psychologists interested in human judgment, and especially those studying heuristics and biases, will find Meehl's work seminal. He thought long and hard not just about whether statistical formulas do or do not outpredict human judges but also about cognitive processes tending to help clinicians to perform well (e.g., recognizing novel patterns) or poorly (e.g., ignoring base rates, nonoptimal data combining procedures, inconsistent applications of a combining procedure).

Although Meehl's writings are filled with allusions to philosophy and logic, they are actually quite accessible to those who will read them with care. Very few other 50-year-old books in psychology are still being as frequently cited as Meehl's *Clinical Versus Statistical Prediction: A Theoretical Analysis and a Review of the Evidence* (1954/1996). One likely reason for this is that Paul was perhaps the most entertaining writer in clinical psychology, with a uniquely informal style (especially noticeable in articles written during the latter part of his career). His intellectual power bubbles up in nearly every paragraph, creating an exciting learning experience for the reader. His colloquial style and interesting anecdotes increase the attractiveness and accessibility of his work. The pleasure, as well as enlightenment, one can obtain from reading Meehl's work on clinical versus statistical prediction commend his work to a broad audience of behavioral scientists.

References

Dawes, R. M., Faust, D., & Meehl, P. E. (1989). Clinical versus actuarial judgment. *Science, 243,* 1668–1774.

Gilberstadt, H., & Duker, J. (1965). *A handbook for clinical and actuarial MMPI interpretation.* Philadelphia: Saunders.

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.

Hallbower, C. C. (1955). *A comparison of actuarial versus clinical prediction to classes discriminated by MMPI.* Unpublished doctoral dissertation, University of Minnesota.

Lundberg, G. A. (1941). Case studies versus statistical methods: An issue based on misunderstanding. *Sociometry, 4,* 379–383.

Marks, P. A., & Seeman, W. (1963). *Actuarial description of abnormal personality.* Baltimore, MD: Williams & Wilkins.

McNemar, Q. (1955). Review of clinical versus statistical prediction. *American Journal of Psychology, 68,* 510.

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.* Northvale, NJ: Jason Aronson. (Original work published 1954)

Sarbin, T. R. (1944). The logic of prediction in psychology. *Psychological Review, 51,* 210–228.

Received February 18, 2004
Revision received March 8, 2004
Accepted March 16, 2004 ■