

CLINICAL  
versus STATISTICAL  
PREDICTION

*A THEORETICAL ANALYSIS  
AND A REVIEW OF THE EVIDENCE*

by Paul E. Meehl  
revisited 71 years later by Ben Recht

IDIOGRAPHIC *versus* NOMOTHETIC

# PREDICTON PROBLEMS

- Given an application with LSAT score, undergraduate grades, and letters of recommendation, who should be admitted into law school?
- Given a record of behavior, should a jailed person be released on parole?
- Should you hospitalize a patient who is clinically depressed to prevent suicide?
- Should a person who doesn't respond to antidepressant prescriptions be given shock therapy?

# PREDICTON and STATISTICS

- Meehl (1954): Take the smallest reference class for which you have stable frequencies.  
(Cites Reichenbach 1938)

prediction = function of average outcome of group

$$\frac{1}{n_G} \sum_{i \in G} y_i$$

# PREDICTON and STATISTICS

- Meehl (1954): Take the smallest reference class for which you have stable frequencies. (Cites Reichenbach 1938)
- M. S. Bartlett (1951): "Insofar as things, persons, are unique or ill-defined, statistics are meaningless and statisticians silenced; in so far as things are similar and definite ... they can be counted and new statistical facts are born."
- Bradford Hill (1952): "...the most frequent and the most foolish criticism of the statistical approach in medicine is...each patient is 'unique' and so there can be nothing for the statistician to count. But if this is true it has always seemed to me that the bottom falls out of the clinical approach as well as the statistical. If each patient is unique, how can a basis for treatment be found in the past observations of other patients?"
- "[Clinicians] base their 'method of choice' upon what they have seen happen before -whether it be in only two or three cases or in a score."

# PREDICTON and THE INDIVIDUAL

- Allport (1942) “A fatal non-sequitur occurs in the reasoning that if 80% of the delinquents who come from broken homes are recidivists, then this delinquent from a broken home has an 80% chance of becoming a recidivist. The truth of the matter is that this delinquent has either 100% certainty of becoming a repeater or 100% certainty of going straight. If all the causes in his case were known, we could predict for him perfectly (barring environmental accidents). His chances are determined by the pattern of his life and not by the frequencies found in the population at large. Indeed, *psychological causation is always personal and never actuarial.*” (Emphasis added)

IDIOGRAPHIC *versus* NOMOTHETIC

# Is everything statistical prediction?



# Problem Setup

## Actions

- The *goal* is predicting the outcome of interventions.
- Small list of possible outcomes.

## Data

- Can be anything: interviews, life history data, a mental test, other biometrics.
- “Machine readable” is the only requirement

# What is Actuarial Prediction?

*mechanical prediction = algorithm*

*actuarial prediction = machine learning  
= statistical prediction  
= pattern recognition  
= artificial intelligence*

- Find the best features
- Build actuarial tables
- Assume future  $\sim$  past
- Maximize outcomes

*actuarial  $\subset$  mechanical*

# What is clinical prediction?

- Anything “informal” made by a human specialist.
- Whatever process occurs in a person’s head.
- Clinical rules are those made by clinicians based on intuitive assessments of data.
- Not formalizable as algorithms.

# The clinical-statistical conundrum

*Given a decision problem with a small set of possible outcomes and an appropriate, fixed collection of data, do actuarial rules or clinical judgment provide more accurate judgments about the future?*

- Note the requirements!
- Well-posed questions, with measurable, discrete answers.
  - If granted admission, will a person succeed in law school? If released from prison, will a person recidivate? If a depressed person isn't hospitalized, will they commit suicide? If a person receives shock therapy, will their depression be relieved?
- Well-posed, machine readable data.

# Evidence

- Burgess (1928). Psychiatrists were inferior to a crude mechanical rule for predicting recidivism.
- Sarbin (1943): A competent statistical clerk can make predictions as well as a highly trained clinical worker.
- Meehl (1986): “There is no controversy in social science that shows such a large body of qualitatively diverse studies coming out so uniformly in the same direction as this one.”
- Grove (2000), Ægisdóttir (2006) metaanalyses found  $d=0.12$  in favor of statistical prediction.

# Why does the actuary always win?

- How do you *evaluate* decisions about individuals?
- Take the outcomes and report the average correctness...

$$\text{hit rate} = \frac{1}{N} \sum_{i=1}^N \mathbb{1}[\text{prediction}_i = \text{outcome}_i]$$

maximize  $\mathbb{E}_p [R(a, x)]$

minimize Regret

*Decisions about individuals are only evaluated on average.*

*If you ask “what is best?” and then evaluate “on average,” the fix is in.*

Metrical Determinism

# Simple model: Brier Scores

- What confidence minimizes Brier Score *on unseen data*?
- Assume a set of *future* observations  $\{(x_i, y_i)\}$ .

$$\text{BS} = \frac{1}{N} \sum_{i=1}^N (p(x_i) - y_i)^2 = \sum_{G \in \mathcal{G}} \frac{n_G}{N} \left\{ \frac{1}{n_G} \sum_{x_i \in G} (p_G - y_i)^2 \right\}$$

$$p_G^* = \frac{1}{n_G} \sum_{x_i \in G} y_i$$

*optimal prediction is  
frequency in the future*

Golden Rule: Use past rates to predict future rates.

Statistical judgment can be considerably better than expert judgment.

And yet...

- One can still screw up statistical prediction: which features, which data?
- Require constant updates: The future is not the same as the past!!
  - Tech companies retrain every day. Medical risk assessments stay static for decades.
- Human costs: erosion of expertise, decision fatigue, complacency.
- Impossible to have statistics for every decision problem!

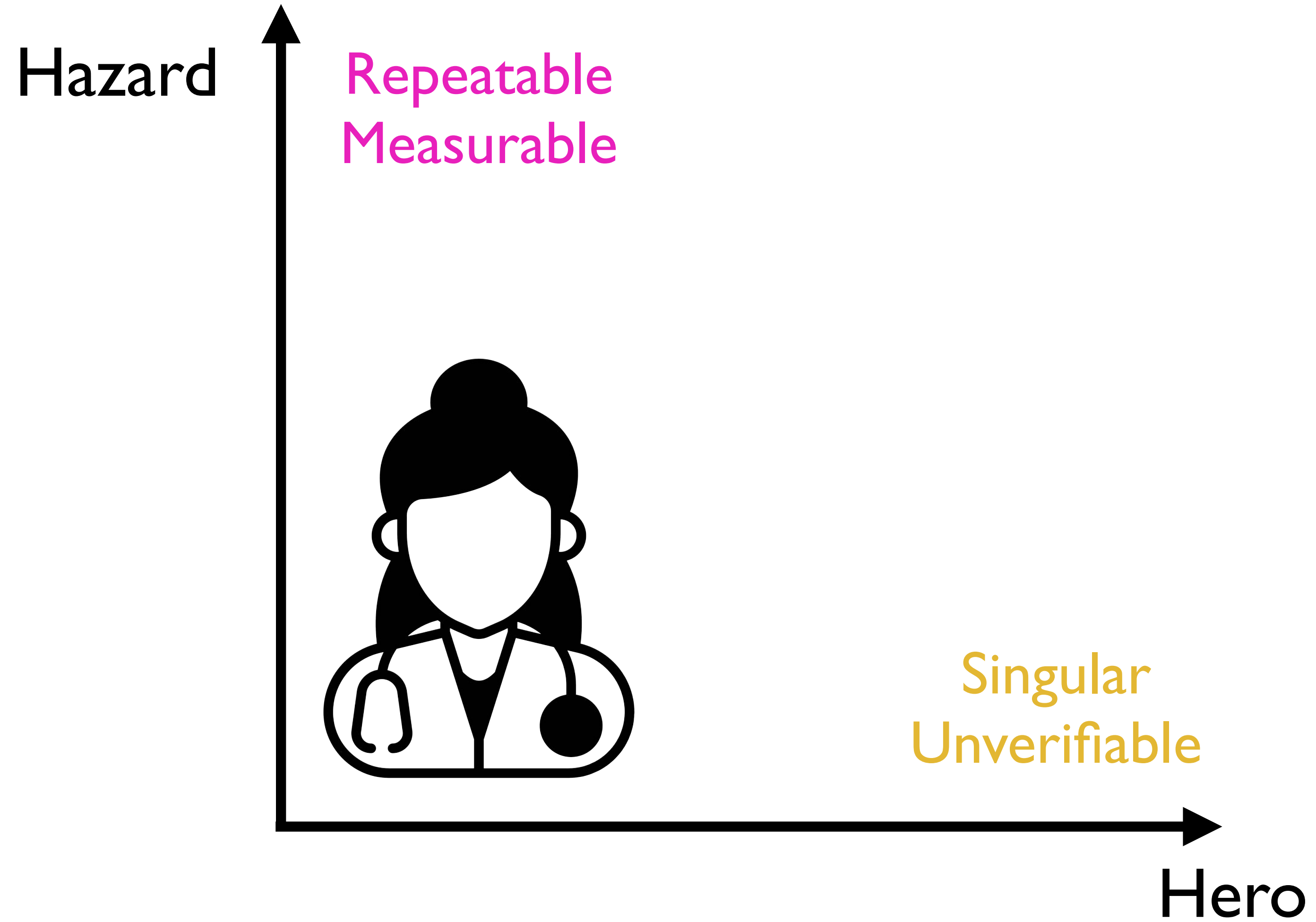
*If we're going to rely on statistical prediction, then we need data-scientific expertise, agility to adapt, ability to measure.*

- When statistical is better than clinical: multiple-choice questions about simple actions from machine-readable data evaluated on average. This characterization is useful!
- Metrical Determinism
- And yet... what are the  $x, y, f, d, g, c$ ???
- “Our arithmetic is useless unless we are counting the right things.” (Bartlett)
- The computable is that which we know how to compute.

# Hazard or Hero?



# Hazard or Hero?



“If a clinician says, ‘This one is different’ or ‘It’s not like the ones in your table,’ ‘This time I’m surer,’ the obvious question is, ‘Why should we care whether you think this one is different or whether you are surer?’ Again, there is only one rational reply to such a question. We have now to study the success frequency of the clinician’s guesses when he asserts that he feels this way. If we have already done so and found him still behind the hit frequency of the table, we would be well advised to ignore him. Always, we might as well face it, the shadow of the statistician hovers in the background; always the actuary will have the final word.”