Information Allergy

Frank E Harrell Jr

Department of Biostatistics
Vanderbilt University School of Medicine

useR!2010 NIST 21 July 2010

Information and Decision Making

What is **Information**?

- Messages used as the basis for decision-making
- Result of processing, manipulating and organizing data in a way that adds to the receiver's knowledge
- Meaning, knowledge, instruction, communication, representation, and mental stimulus

Value of Information

Judged by the variety of outcomes to which it leads

Optimum Decision Making

Requires the maximum and most current information the decision maker is capable of handling

`pbs.org/weta, wikipedia.org`

Some Important Decisions in Biomedical and Epidemiologic Research and Clinical Practice

- Pathways, mechanisms of action
- Best way to use gene and protein expressions to diagnose or treat
- Which biomarkers are most predictive and how should they be summarized?
- What is the best way to diagnose a disease or form a prognosis?
- Is a risk factor causative or merely a reflection of confounding?
- How should patient outcomes be measured?
- Is a drug effective for an outcome?
- Who should get a drug?

Information Allergy

Failing to Obtain Key Information Needed to Make a Sound Decision

- Not collecting important baseline data on subjects

Ignoring Available Information

- Touting the value of a new biomarker that provides less information than basic clinical data
- Ignoring confounders (alternate explanations)
- Ignoring subject heterogeneity
- Categorizing continuous variables or subject responses
- Categorizing predictions as "right" or "wrong"
- Letting fear of probabilities and costs/utilities lead an author to make decisions for individual patients

Prognostic Markers in Acute Myocardial Infarction

C-index: concordance probability $\equiv$ receiver operating characteristic curve or ROC area

Measure of ability to discriminate death within 30d

Markers	C-index
CK–MB	0.63
Troponin T	0.69
Troponin T > 0.1	0.64
CK–MB + Troponin T	0.69
CK–MB + Troponin T + ECG	0.73
Age + sex	0.80
All	0.83

Data from Ohman et al. [1996]

Inadequate Adjustment for Confounders

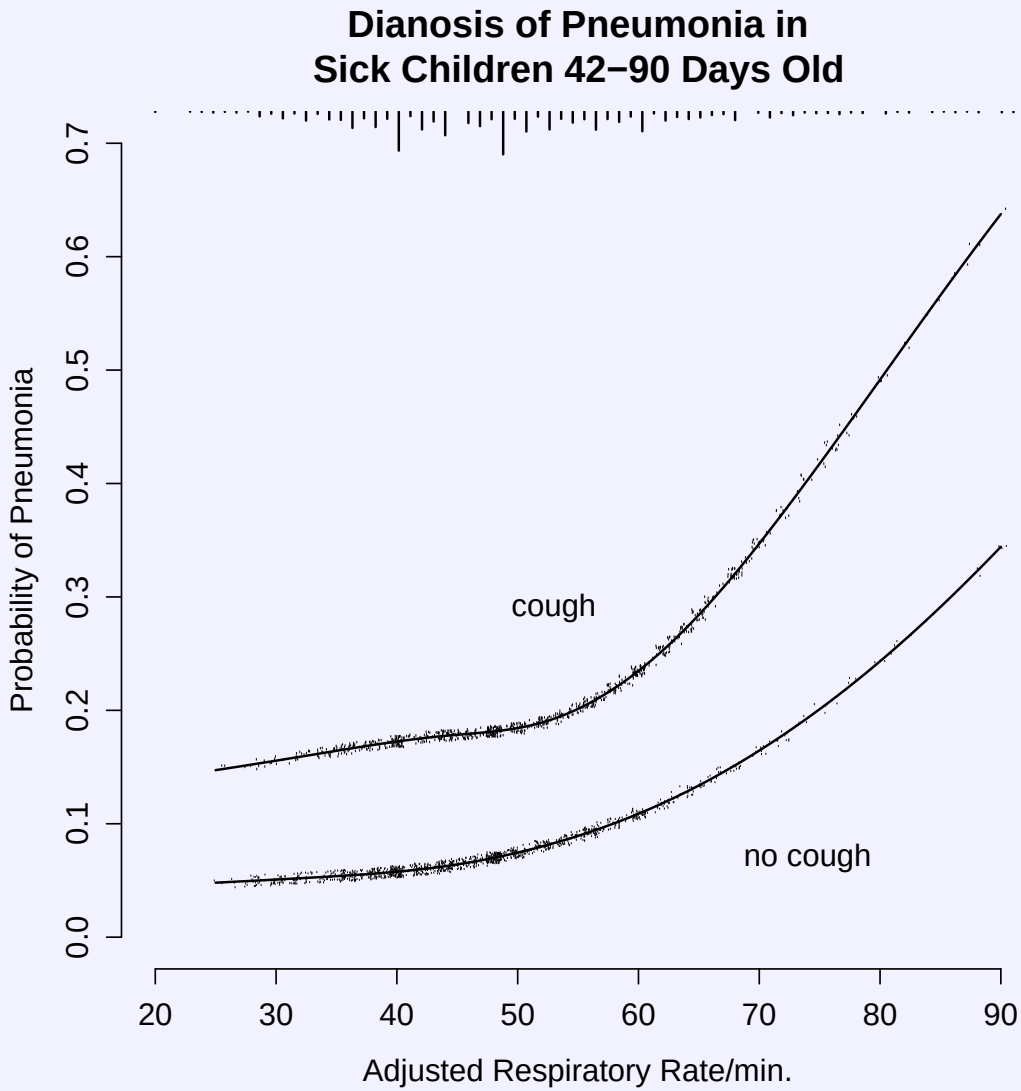
- Case-control study of diet, food constituents, breast cancer
- 140 cases, 222 controls
- 35 food constituent intakes and 5 confounders
- Food intakes are correlated
- Traditional stepwise analysis not adjusting simultaneously for all foods consumed $\rightarrow$ 11 foods had $P < 0.05$
- Full model with all 35 foods competing $\rightarrow$ 2 had $P < 0.05$
- Rigorous simultaneous analysis (hierarchical random slopes model) penalizing estimates for the number of associations examined $\rightarrow$ no foods associated with breast cancer

Greenland [2000] after Witte [1994]

Categorizing Continuous Diagnostic Variables

- Many physicians attempt to find cutpoints in continuous predictor variables
- Mathematically such cutpoints cannot exist unless relationship with outcome is discontinuous
- Even if the cutpoint existed, it has to vary with other patient characteristics, as optimal decisions are based on the overall probability of the outcome

Categorizing Diagnostic Variables, *cont.*



Harrell et al. [1998] and WHO

Cutpoints are Disasters

Prognostic Relevance of S-phase Fraction in Breast Cancer

19 different cutpoints used in literature

Cathepsin-D Content and Disease-Free Survival in Node-Negative Breast Cancer

12 studies, 12 cutpoints

ASCO guidelines: neither cathepsin-D nor S-phrase fraction recommended as prognostic markers

Holländer et al. [2004]

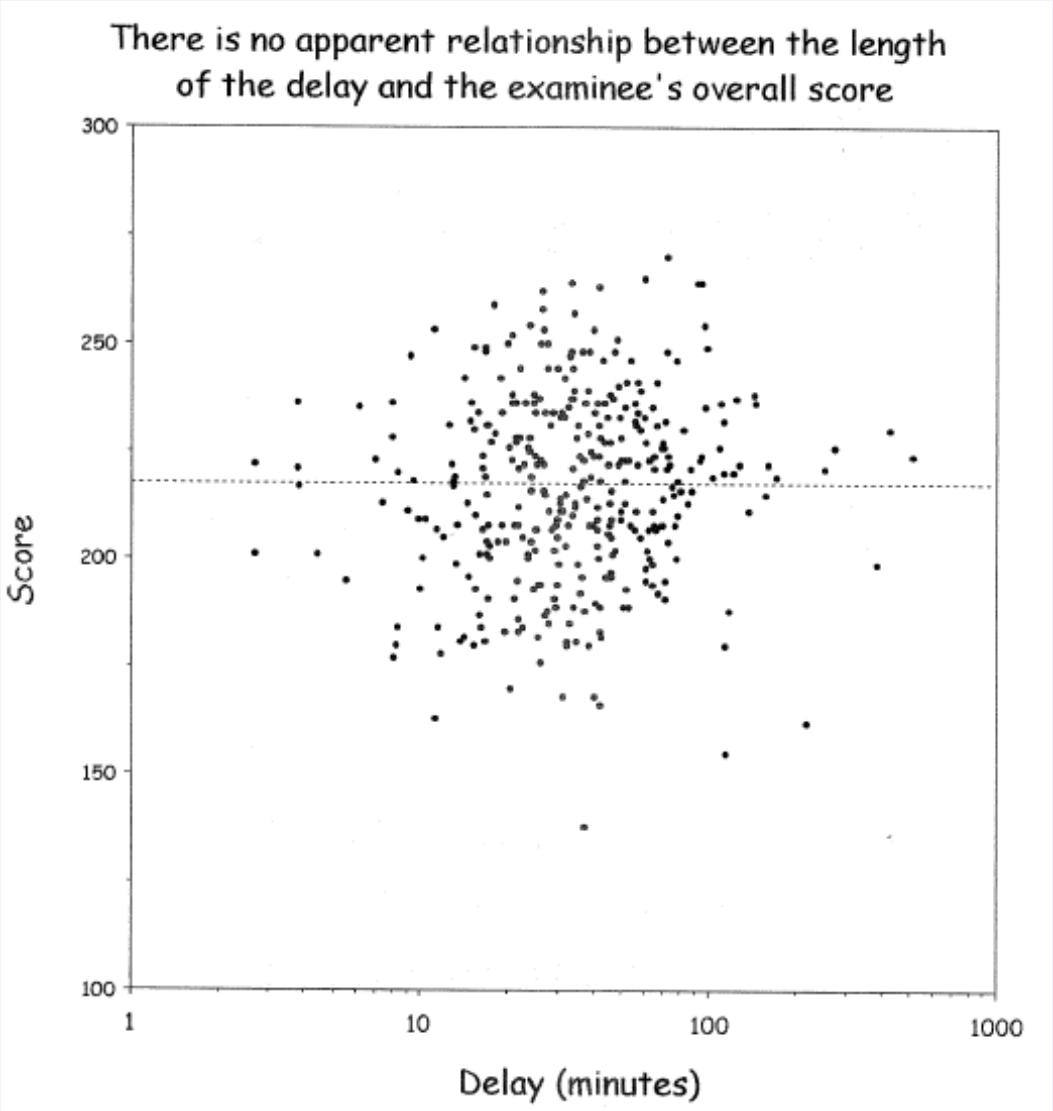
Cutpoints are Disasters, *cont.*

Cutpoints may be found that result in both increasing and decreasing relationships with **any** dataset with zero correlation

Range of Delay	Mean Score	Range of Delay	Mean Score
0-11	210	0-3.8	220
11-20	215	3.8-8	219
21-30	217	8-113	217
31-40	218	113-170	215
41-	220	170-	210

Wainer [2006]; See "Dichotomania" [Senn, 2005] and Royston et al. [2006]

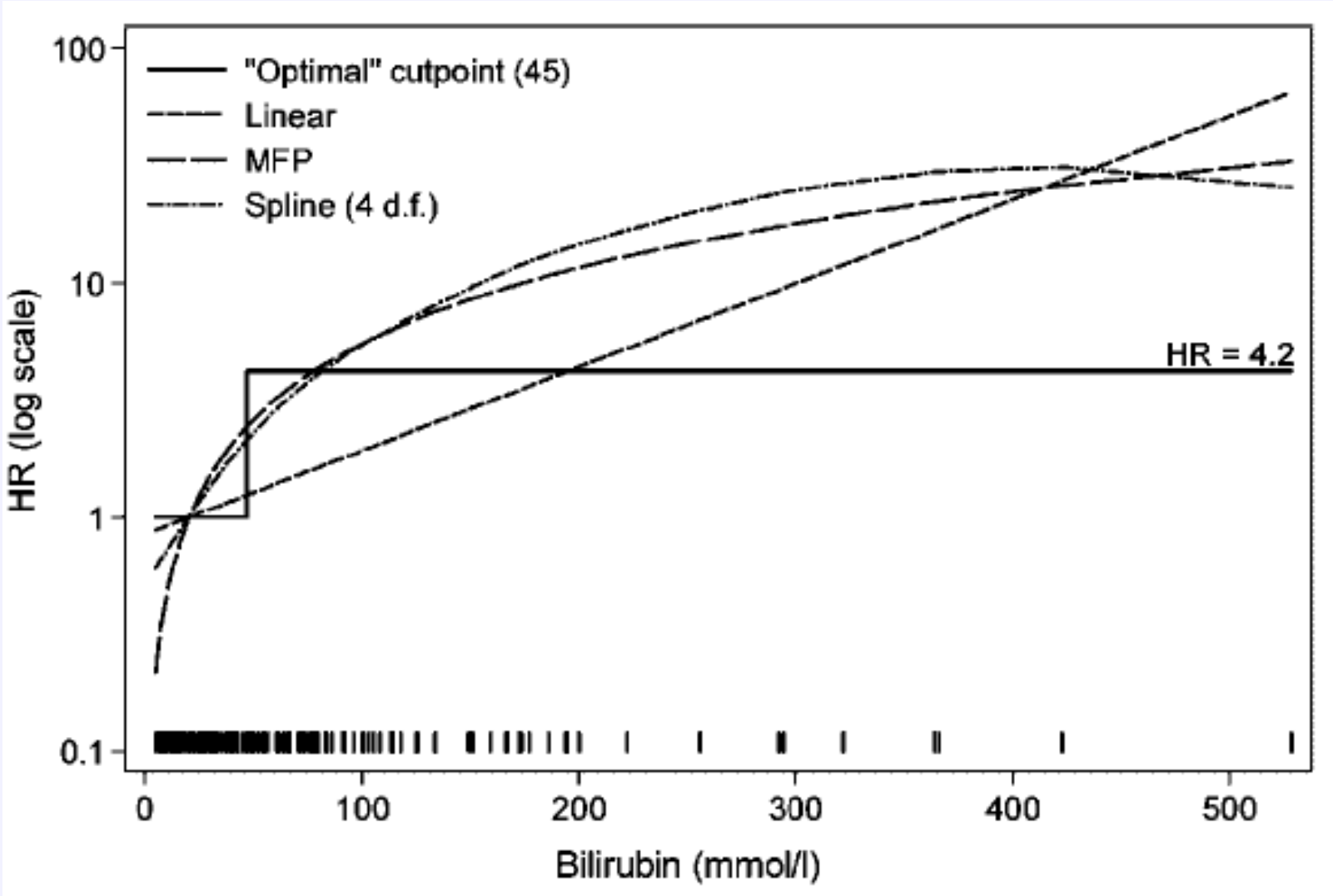
Data from Wainer [2006]



Lack of Meaning of Effects Based on Cutpoints

- Researchers often use cutpoints to estimate the high:low effects of risk factors (e.g., BMI vs. asthma)
- Results in inaccurate predictions, residual confounding, impossible to interpret
- high:low represents unknown mixtures of highs and lows
- Effects (e.g., odds ratios) will vary with population

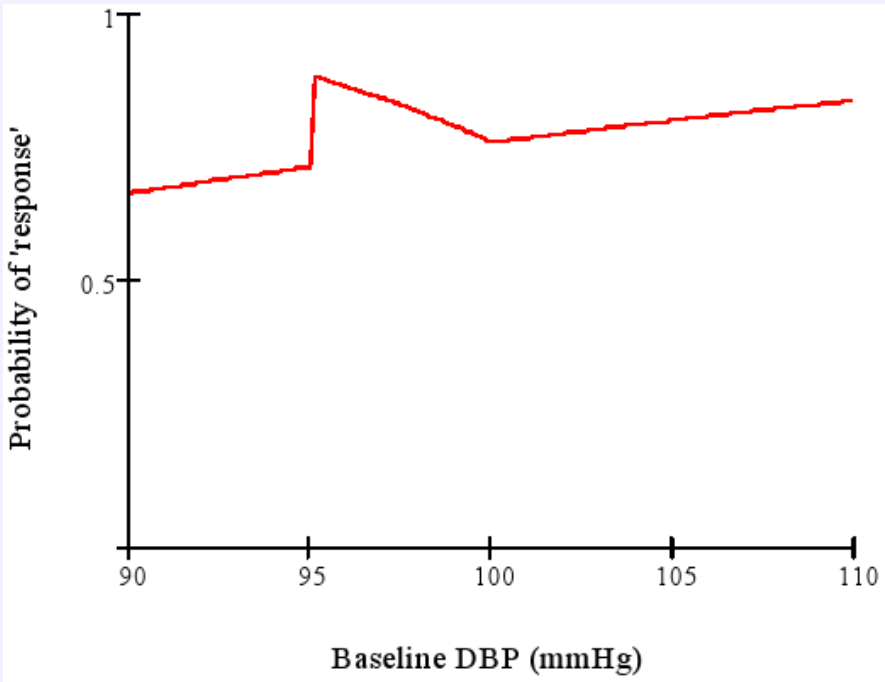
Dichotomization of Predictors, *cont.*



Royston et al. [2006]

Categorizing Outcomes

- Arbitrary, low power, can be difficult to interpret
- Example: "The treatment is called successful if either the patient has gone down from a baseline diastolic blood pressure of ≥ 95 mmHg to ≤ 90 mmHg or has achieved a 10% reduction in blood pressure from baseline."



Senn [2005] after Goetghebeur [1998]

Classification vs. Probabilistic Diagnosis

- Many studies attempt to classify patients as diseased/normal
- Given a reliable estimate of the probability of disease and the consequences of +/- one can make an optimal decision
- Consequences are known at the point of care, not by the authors; categorization **only** at point of care
- Continuous probabilities are self-contained, with their own "error rates"
- Middle probs. allow for "gray zone", deferred decision

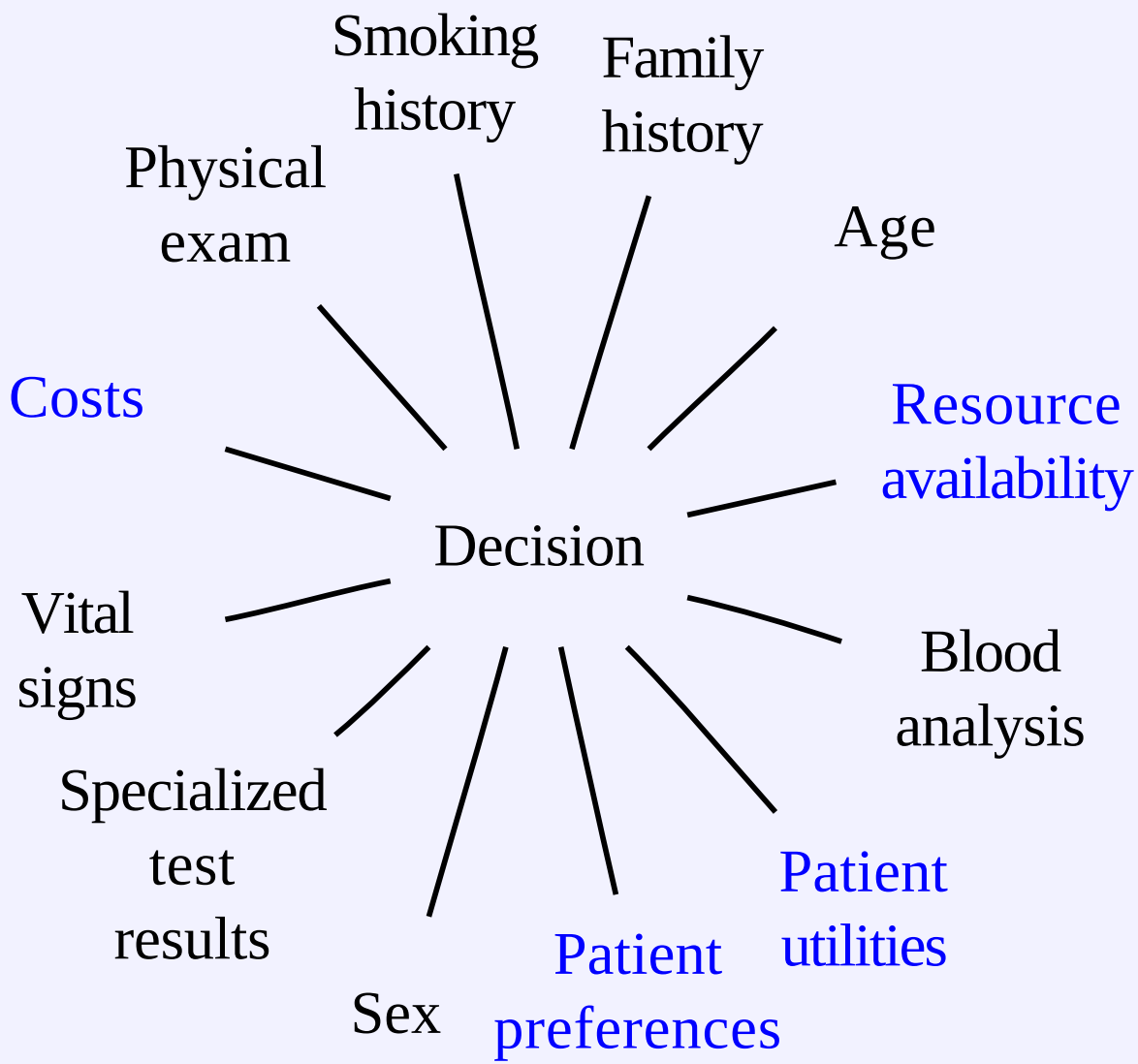
Patient	Prob[disease]	Decision	Prob[error]
1	0.03	normal	0.03
2	0.40	normal	0.40
3	0.75	disease	0.25

Probabilities, Odds, Number Needed to Treat, and Physicians

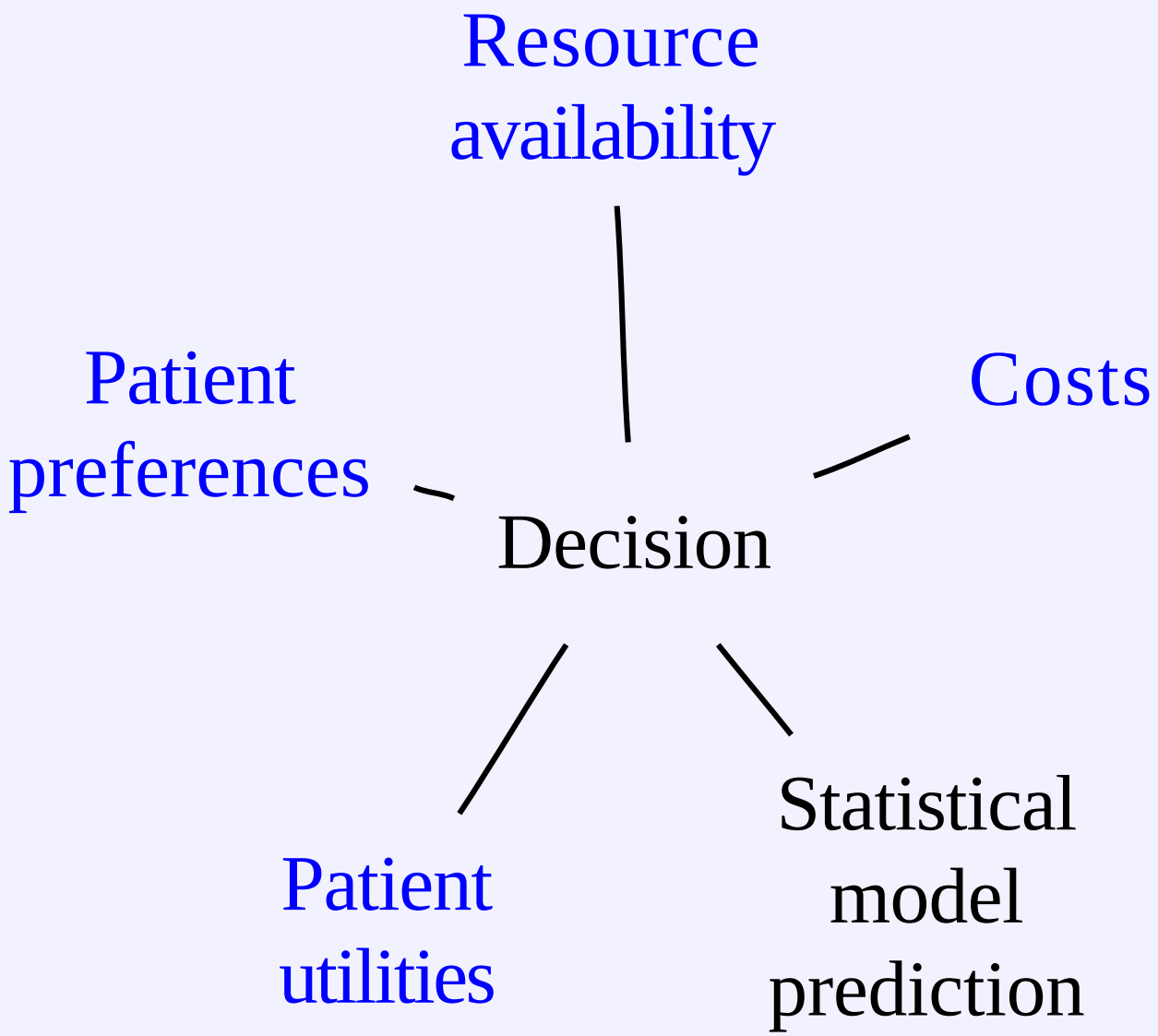
Number needed to treat. The only way, we are told, that physicians can understand probabilities: odds being a difficult concept only comprehensible to statisticians, bookies, punters and readers of the sports pages of popular newspapers.

Senn [2008]

Some Components of Optimal Clinical Decisions



Statistical Models Reduce the Dimensionality of the Problem *but not to unity*



Problems with Classification of Predictions

- Feature selection / predictive model building requires choice of a scoring rule, e.g. correlation coefficient or proportion of correct classifications
- Prop. classified correctly is a discontinuous **improper scoring rule**
  - Maximized by bogus model (example below)
- Minimum information
  - low statistical power
  - high standard errors of regression coefficients
  - arbitrary to choice of cutoff on predicted risk
  - forces binary decision, does not yield a "gray zone" $\rightarrow$ more data needed
- Takes analyst to be provider of utility function and not the treating physician
- Sensitivity and specificity are also improper scoring rules

Example: Damage Caused by Improper Scoring Rule

- Predicting probability of an event, e.g., Prob[disease]
- $N = 400$, 0.57 of subjects have disease
- Classify as diseased if prob. > 0.5

Model	C Index	χ^2	Proportion Correct
age	.592	10.5	.622
sex	.589	12.4	.588
age+sex	.639	22.8	.600
constant	.500	0.0	.573

Adjusted Odds Ratios:

age (IQR 58y:42y)	1.6 (0.95CL 1.2-2.0)
sex (f:m)	0.5 (0.95CL 0.3-0.7)

Test of sex effect adjusted for age $(22.8 - 10.5)$:
$P = 0.0005$

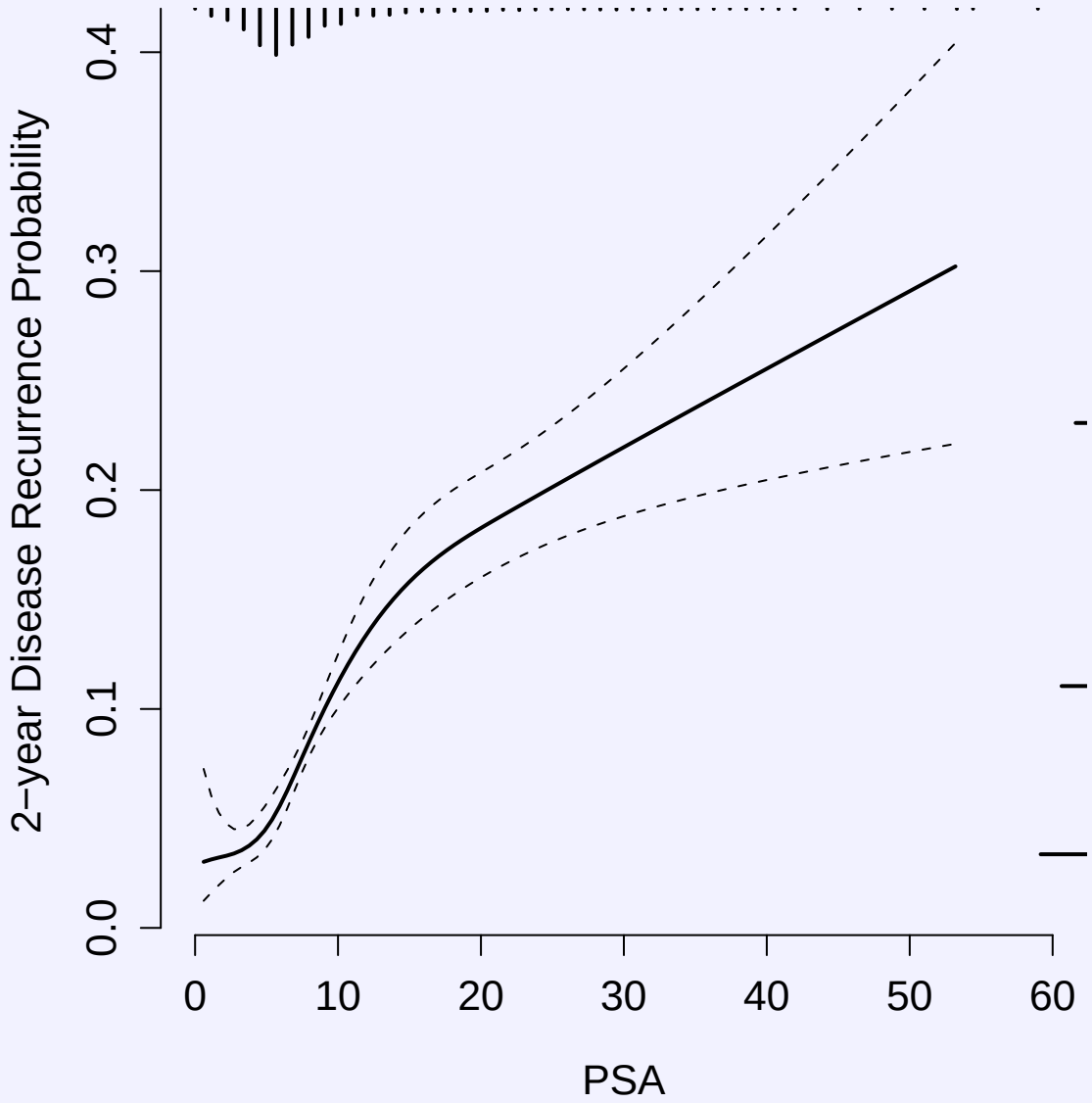
Hazards of Classification Accuracy, *continued*

Michiels et al. [2005]	Aliferis et al. [2009]
% classified correctly	C-index
Single split-sample validation	Multiple repeats of 10-fold CV
Wrong tests (censoring, failure times)	Correct tests
5 of 7 published microarray studies had no signal	6 of 7 have signals

Value of Continuous Markers

- Avoid arbitrary cutpoints
- Better risk spectrum
- Provides gray zone
- Increases power/precision

Prognosis in Prostate Cancer

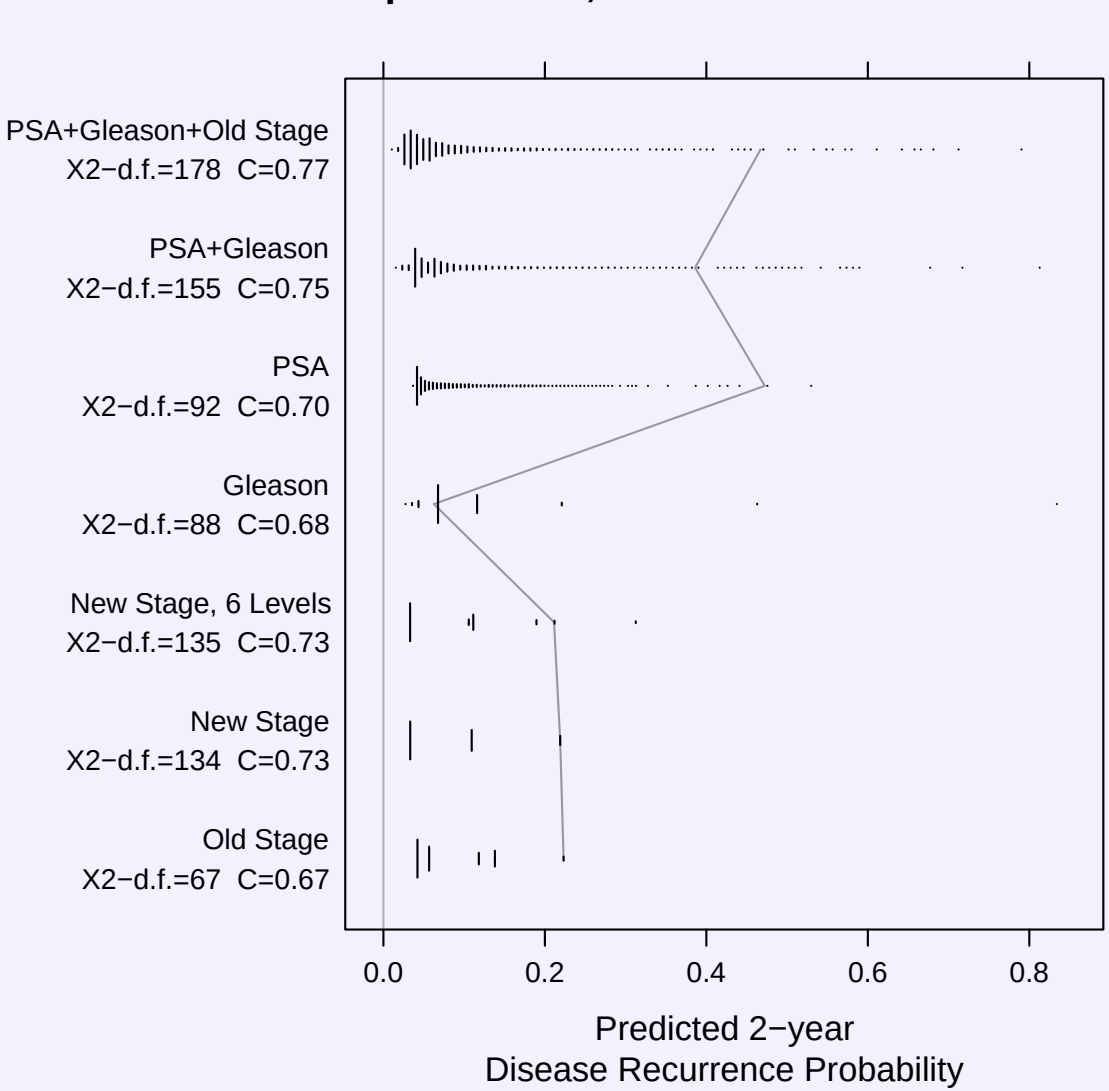


Data courtesy of M Kattan from JNCI 98:715; 2006

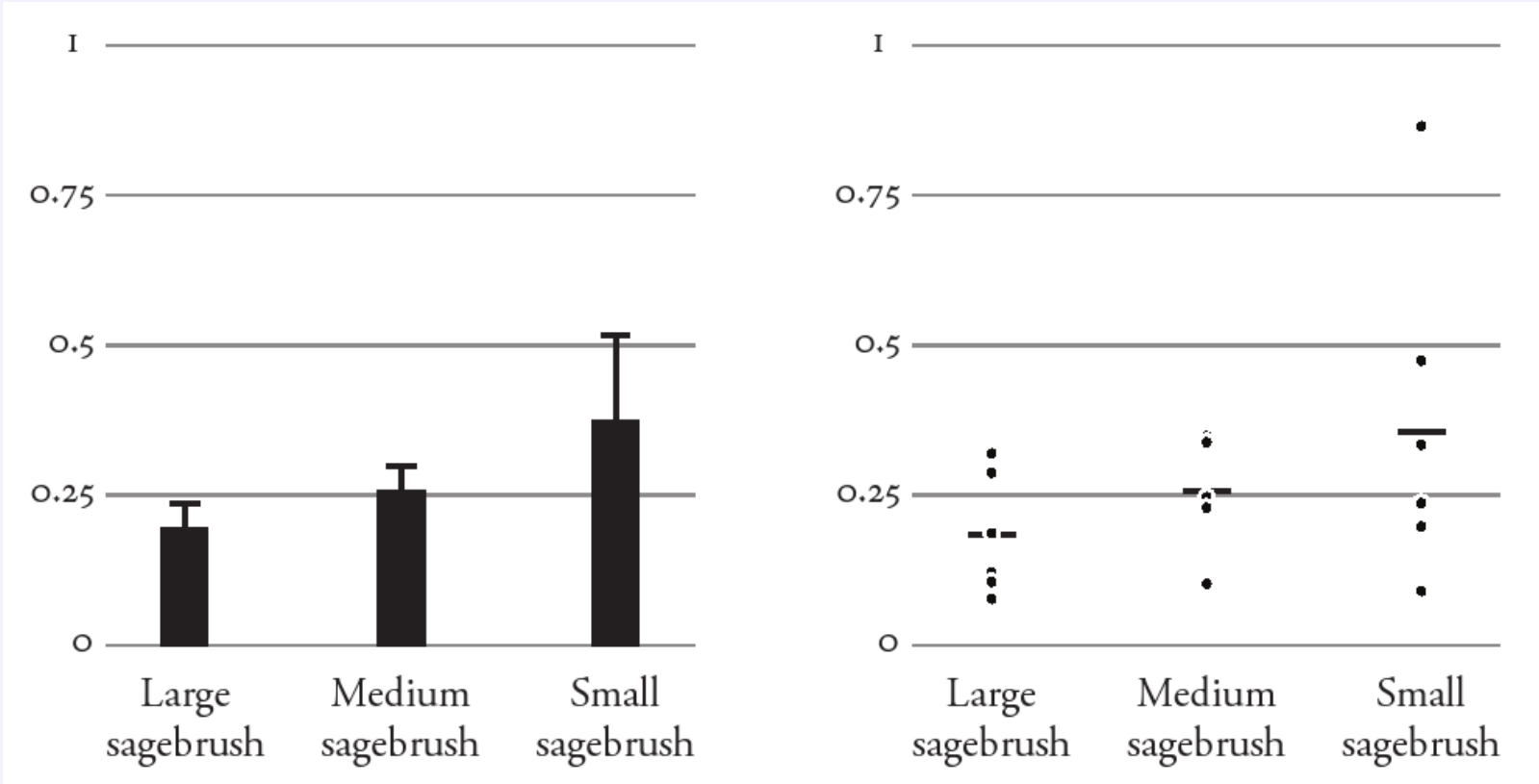
Horizontal ticks represent frequencies of prognoses by new staging system

Modification of AJCC staging by Roach *et al.* [2006]

Prognosis in Prostate Cancer, *cont.*



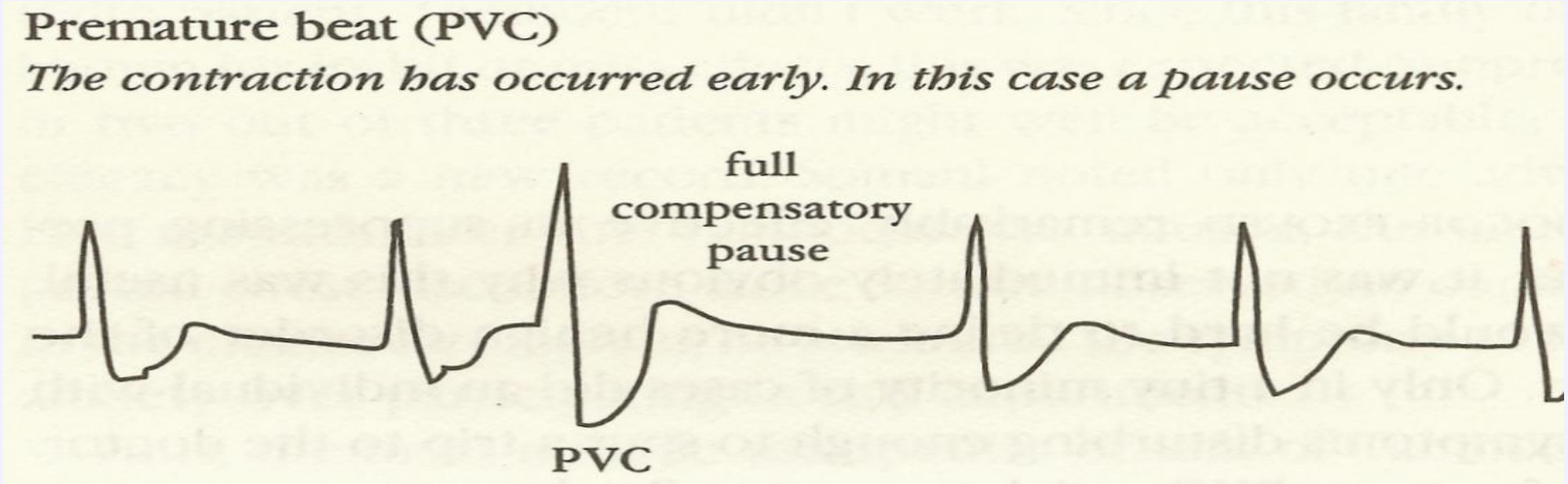
Visual Numeric Information: Covering and Uncovering



Consequences of Ignoring Information

Ignoring Information Can Kill: Cardiac Anti-arrhythmic Drugs

- Premature ventricular contractions were observed in patients surviving acute myocardial infarction
- Frequent PVCs ↑ incidence of sudden death



Moore [1995, p. 46]

Arrhythmia Suppression Hypothesis

Any prophylactic program against sudden death must involve the use of anti-arrhythmic drugs to subdue ventricular premature complexes.

Bernard Lown
Widely accepted by 1978

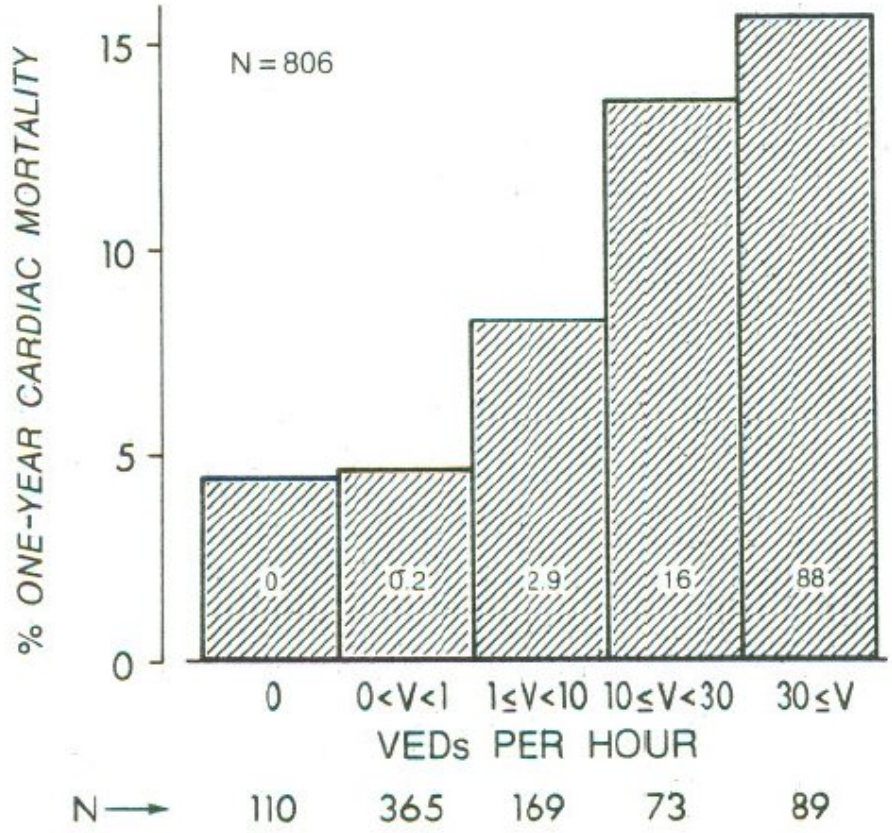


Figure 2. Cardiac Mortality Rate in Five Categories for Frequency of Ventricular Ectopic Depolarizations (VEDs) Determined by 24-Hour Holter Recording before Discharge.

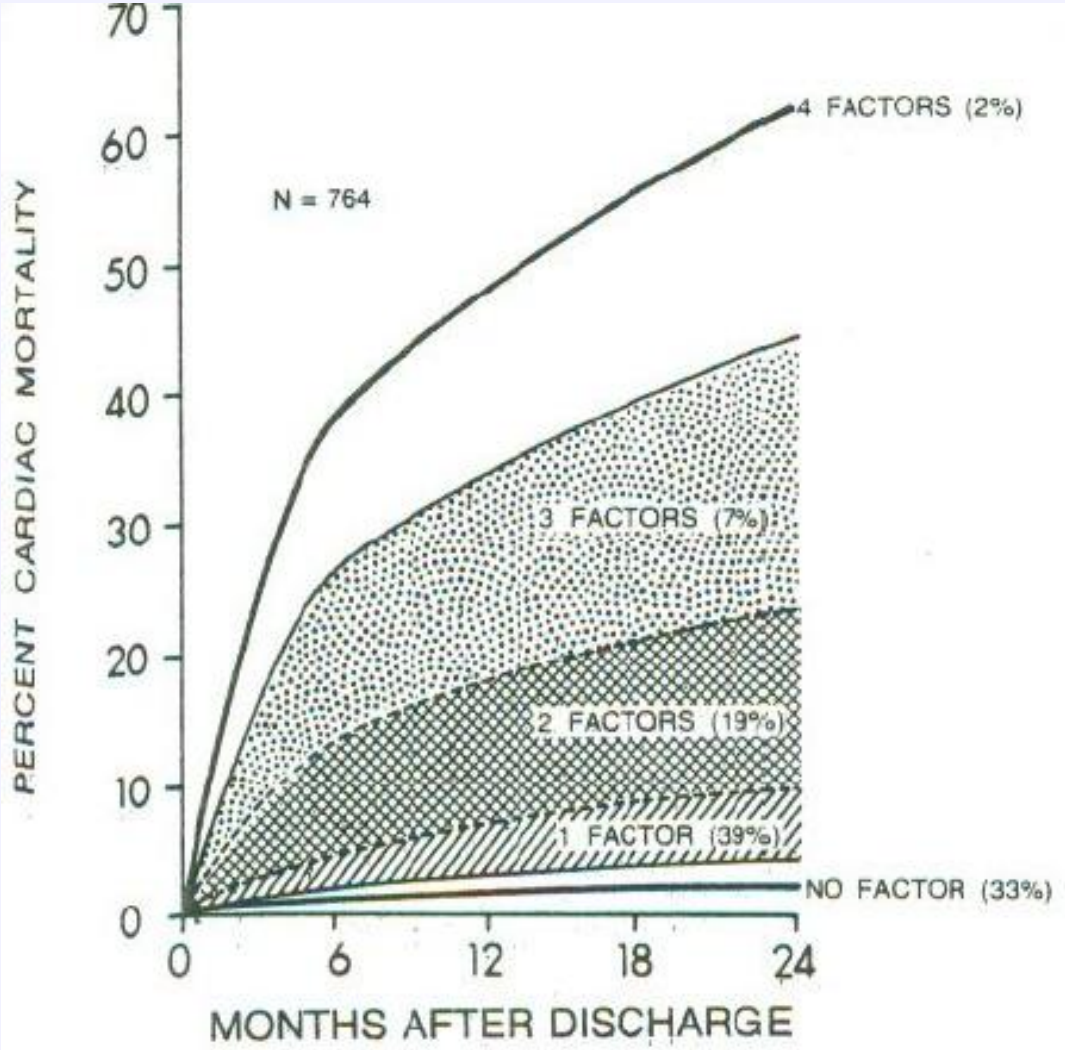
N denotes the number of patients in the total population and in each category. Of 819 patients with Holter recordings, 13 were lost to follow-up during the first year after hospitalization. The numbers within each of the boxes denote the median frequency of ventricular ectopy.

Moore [1995, p. 49]; Multicenter Postinfarction Research Group [1983]

Are PVCs Independent Risk Factors for Sudden Cardiac Death?

Researchers developed a 4-variable model for prognosis after acute MI

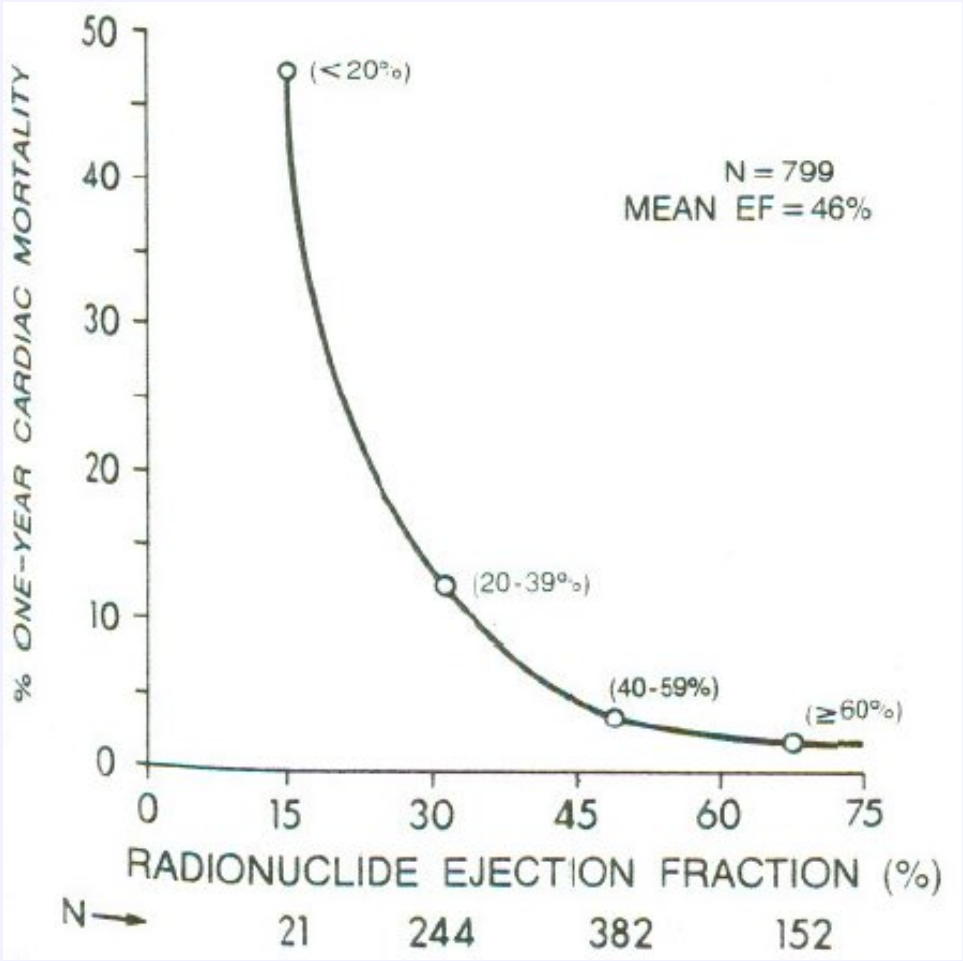
- left ventricular ejection fraction (EF) < 0.4
- PVCs > 10/hr
- Lung rales
- Heart failure class II,III,IV



Multicenter Postinfarction Research Group [1983]

Dichotomania Caused Severe Problems

- EF alone provides same prognostic spectrum as the researchers' model
- Did not adjust for EF!; PVCs ↑ when EF$<$ 0.2
- Arrhythmias prognostic in isolation, not after adjustment for continuous EF and anatomic variables
- Arrhythmias predicted by local contraction abnorm., then global function (EF)



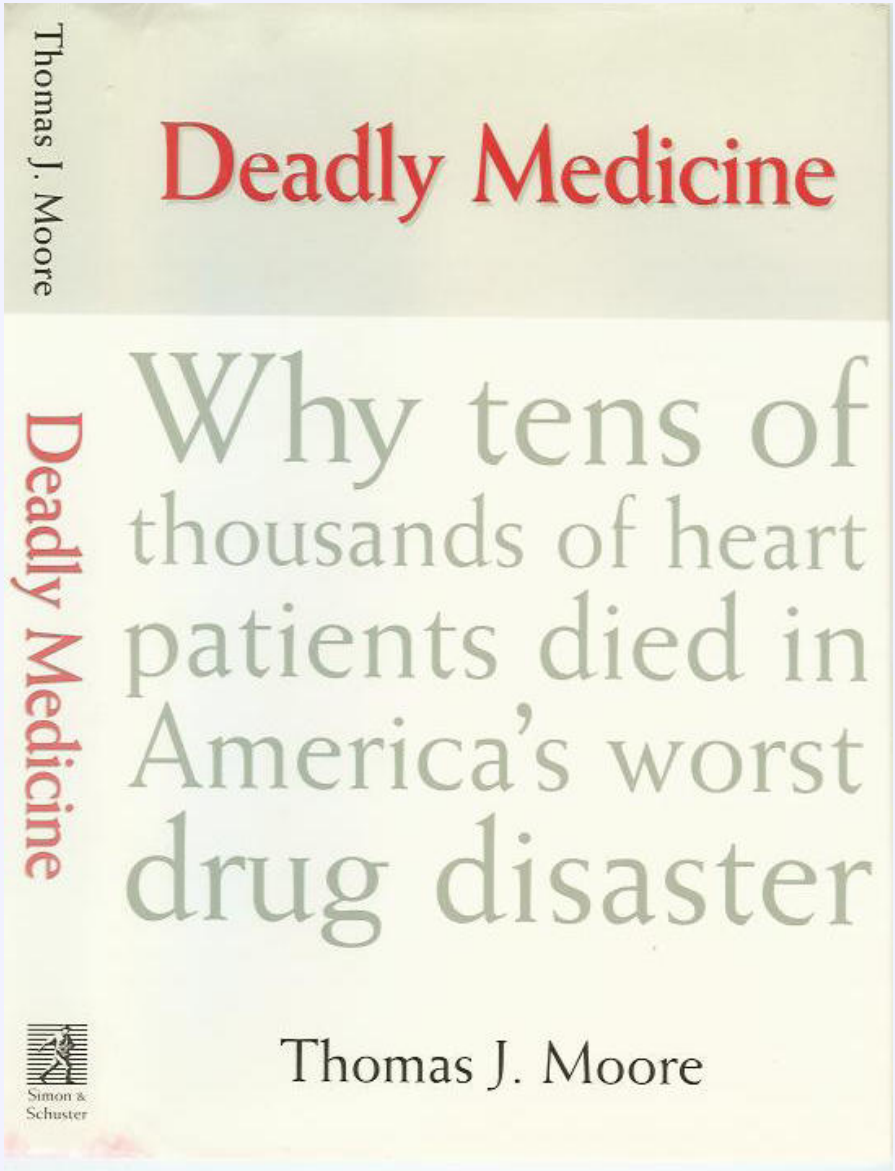
Multicenter Postinfarction Research Group [1983]; Califf et al. [1982]

CAST: Cardiac Arrhythmia Suppression Trial

- Randomized placebo, moricizine, and Class IC anti-arrhythmic drugs flecainide and encainide
- Cardiologists: unethical to randomize to placebo
- Placebo group included after vigorous argument
- Tests design as one-tailed; did not entertain possibility of harm
- Data and Safety Monitoring Board recommended early termination of flecainide and encainide arms
- Deaths $\frac{56}{730}$ drug, $\frac{22}{725}$ placebo, RR 2.5

CAST Investigators [1989]

Conclusions: Class I Anti-Arrhythmics



Moore [1995, p. 289, 49]; D Graham, FDA

Estimate of excess deaths from Class I anti-arrhythmic drugs: 24,000–69,000

Estimate of excess deaths from Vioxx: 27,000–55,000

Arrhythmia suppression hypothesis refuted; PVCs merely indicators of underlying, permanent damage

Information May be Costly

When the Missionaries arrived, the Africans had the Land and the Missionaries had the Bible. They taught how to pray with our eyes closed. When we opened them, they had the land and we had the Bible.

Jomo Kenyatta, founding father of Kenya;
also attributed to Desmond Tutu

Information May be Dangerous

Information itself has a liberal bias.

The Colbert Report, 28Nov06

References

C. F. Aliferis, A. Statnikov, I. Tsamardinos, J. S. Schildcrout, B. E. Shepherd, and F. E. Harrell. Factors influencing the statistical power of complex data analysis protocols for molecular signature development from microarray data. *PLoS One*, 4(3):e4922, 2009. PMID 19290050.

R. Bordley. Statistical decisionmaking without math. *Chance*, 20(3):39–44, 2007.

W. M. Briggs and R. Zaretzki. The skill plot: A graphical technique for evaluating continuous diagnostic tests (with discussion). *Biometrics*, 64:250–261, 2008.

R. M. Califf, R. A. McKinnis, J. Burks, K. L. Lee, V. S. Harrell FE Jr., Behar, D. B. Pryor, G. S. Wagner, and R. A. Rosati. Prognostic implications of ventricular arrhythmias during 24 hour ambulatory monitoring in patients undergoing cardiac catheterization for coronary artery disease. *Am J Card*, 50:23–31, 1982.

CAST Investigators. Preliminary report: Effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. *NEJM*, 321(6):406–412, 1989.

S. Greenland. When should epidemiologic regressions use random coefficients? *Biometrics*, 56:915–921, 2000.

F. E. Harrell, P. A. Margolis, S. Gove, K. E. Mason, E. K. Mulholland, D. Lehmann, L. Muhe, S. Gatchalian, and H. F. Eichenwald. Development of a clinical prediction model for an ordinal outcome: The World Health Organization ARI Multicentre Study of clinical signs and etiologic agents of pneumonia, sepsis, and meningitis in young infants. *Stat Med*, 17:909–944, 1998.

N. Holländer, W. Sauerbrei, and M. Schumacher. Confidence intervals for the effect of a prognostic factor after selection of an 'optimal' cutpoint. *Stat Med*, 23:1701–1713, 2004.

S. Michiels, S. Koscielny, and C. Hill. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet*, 365:488–492, 2005.

T. J. Moore. *Deadly Medicine: Why Tens of Thousands of Patients Died in America's Worst Drug Disaster*. Simon & Shuster, New York, 1995.

Multicenter Postinfarction Research Group. Risk stratification and survival after myocardial infarction. *NEJM*, 309:331–336, 1983.

E. M. Ohman, P. W. Armstrong, R. H. Christenson, C. B. Granger, H. A. Katus, C. W. Hamm, M. A. O'Hannesian, G. S. Wagner, N. S. Kleiman, F. E. Harrell, R. M. Califf, E. J. Topol, K. L. Lee, and the GUSTO-IIa Investigators. Cardiac troponin T levels for risk stratification in acute myocardial ischemia. *NEJM*, 335:1333–1341, 1996.

P. Royston, D. G. Altman, and W. Sauerbrei. Dichotomizing continuous predictors in multiple regression: a bad idea. *Stat Med*, 25:127–141, 2006.

S. Senn. *Statistical Issues in Drug Development*. Wiley, Chichester, England, second edition, 2008.

S. J. Senn. Dichotomania: an obsessive compulsive disorder that is badly affecting the quality of analysis of pharmaceutical trials. In *Proceedings of the International Statistical Institute, 55th Session*, Sydney, 2005.

A. J. Vickers. Decision analysis for the evaluation of diagnostic tests, prediction models, and molecular markers. *Am Statistician*, 62(4):314–320, 2008.

H. Wainer. Finding what is not there through the unfortunate binning of results: The Mendel effect. *Chance*, 19(1):49–56, 2006.

Information Allergy

Frank E Harrell Jr
Department of Biostatistics
Vanderbilt University

Information allergy is defined as (1) refusing to obtain key information needed to make a sound decision, or (2) ignoring important available information. The latter problem is epidemic in biomedical and epidemiologic research and in clinical practice. Examples include

- ignoring some of the information in confounding variables that would explain away the effect of characteristics such as dietary habits
- ignoring probabilities and "gray zones" in genomics and proteomics research, making arbitrary classifications of patients in such a way that leads to poor validation of gene and protein patterns
- failure to grasp probabilitistic diagnosis and patient-specific costs of incorrect decisions, thus making arbitrary diagnoses and placing the analyst in the role of the bedside decision maker
- classifying patient risk factors and biomarkers into arbitrary "high/low" groups, ignoring the full spectrum of values
- touting the prognostic value of a new biomarker, ignoring basic clinical information that may be even more predictive
- using weak and somewhat arbitrary clinical staging systems resulting from a fear of continuous measurements
- ignoring patient spectrum in estimating the benefit of a treatment

Examples of such problems will be discussed, concluding with an examination of how information–phobic cardiac arrhythmia research contributed to the deaths of thousands of patients.