Cognitive Limitations and Methods for Improving Judgments: Implications for Establishing Medically Relevant Performance Goals

Neal V. Dawson, M.D.
Associate Professor of Medicine
Case Western Reserve University
MetroHealth Medical Center
Cleveland, Ohio

Abstract: During the past two decades, increasing amounts of information have become available about the performance of physicians for the common cognitive tasks of making diagnoses and estimating prognoses. These studies have demonstrated both excellent and poor performance by physicians. Among the reasons for less than optimal performance are cognitive limitations associated with the use of heuristics (rules of thumb) and the occurrence of cognitive biases. Obstacles to accurate probability estimation include three heuristics: availability (using the ease with which instances come to mind as a proxy for likelihood of occurrence), representativeness (pattern recognition), anchoring and adjustment (updating an initial estimate after additional information becomes available) and three cognitive biases: ego (self-serving probability estimates), hindsight (knowledge that the event occurred inflating the estimate that it would have occurred), and anticipated regret (allowing the undesirability of a diagnosis or outcome to alter the estimate of its likelihood of occurrence). Impediments to optimal information synthesis include confirmatory bias (tendency only to seek information that will confirm, rather than disconfirm, hypotheses), ignoring negative evidence (using abnormal but not normal findings to make a diagnosis or estimate prognosis), and framing (different ways to present the same information). Knowledge of cognitive limitations and development of techniques to assess components of judgmental accuracy (e.g., lens model analysis) have allowed specific methods for improving judgments to be identified and investigated. Such knowledge also should influence the research agendas of those who wish to design and test methods of providing interpretive guidance or influencing decision making based on laboratory test results.

In most clinical settings, much medical decision making occurs at an informal or intuitive level and includes such tasks as synthesizing information and estimating the likelihood of current unknowns (e.g., diagnoses) or future events (e.g., prognoses). Systematic errors in judgments (cognitive bias) have been documented in nonmedical and medical settings.[1,2,3] Cognitive limitations provide both challenges for current methods of decision making and opportunities for a systematic research effort into methods to enhance both medical judgments and outcomes.

In the sections that follow, I provide a) brief definitions and examples of specific impediments and obstacles to intuitive decision making and b) outline methods that have been used (or might be examined) to avoid or minimize the associated cognitive limitations.

Impediments to optimal information synthesis

Confirmatory bias is the tendency to seek evidence that can be used to confirm (but not disconfirm) hypotheses.[4,5] One can view such evidence as contributing to predictive value positive, rather than predictive value negative. Eddy[4] cites an article from the surgical literature where the author discusses how a "positive" mammogram can increase the likelihood of breast cancer (predictive value positive). The fact that a "negative" mammogram decreases the likelihood of breast cancer was not considered (predictive value negative). Confirmatory bias may also affect the interpretation of data. Walston[6] demonstrated that both medical students and practicing physicians used low-relevance information to support their own diagnoses.

Ignoring negative evidence is a phenomenon related to confirmatory bias. It represents the tendency to use abnormal but not normal findings in making judgments. Although both abnormal and normal findings should be used to make diagnoses efficiently, a study of practicing physicians demonstrates how they used abnormal, but not normal, findings in diagnosing pneumonia in outpatients.[7]

Framing, i.e., alternative ways of presenting the same information, can influence or even reverse medical decisions. McNeil et al[8] demonstrated how physicians' preferences for lung cancer treatment shifted between surgery and radiation therapy when data were presented as the probability of living as opposed to the probability of dying, when the treatments were specifically identified versus not identified and when life expectancy was provided rather than cumulative probability. Another medically relevant example of data presentation relates to the willingness of physicians to initiate therapy when study results are presented as relative risks as opposed to differences in absolute risks.[9-11] Mathematically equivalent rates may not be cognitively equivalent, a fact that has not been lost on pharmaceutical companies or research investigators.

Obstacles to accurate likelihood (probability) estimation

Heuristics: Familiar "rules of thumb" or other intuitive shortcuts may help simplify complex decision tasks but can lead to systematic errors in judgments.

The availability heuristic occurs when a physician uses the ease with which diagnoses or outcomes are recalled as a proxy for the likelihood they will occur. Although common events may come easily to mind, other cases and occurrences may be easily remembered because of their rarity, uniqueness, or personal meaningfulness due to a physician's research interests, personal experiences or recency of their occurrence. Not all easily recalled instances are, in fact, common. Detmer and colleagues[12] asked surgeons to estimate the surgical mortality rate for the entire Surgical service. Surgeons from high mortality specialties (cardiovascular, neurosurgery, general surgery) estimated the overall mortality rate to be more than double that of the estimated rate by surgeons from low mortality services (plastic surgery, orthopedics, urology). A surgeon's own experiences would be expected to be more available than the experiences of others and seemed to exert a disproportionate effect on judgments about the mortality rate for the entire Surgery service.

Representativeness, or pattern recognition, is a method which uses resemblance as a quick way of assessing likelihood, i.e., the probability that "A" belongs to class "B" is directly related to the degree that "A" resembles "B." Pattern recognition is taught and commonly used in medicine but is not influenced by several factors that are known to affect actual likelihood: the prior probability of disease (or outcome), the fact that data from a small sample may be an unreliable estimator of the underlying (population) characteristic, the degree to which the event may be predictable, the likelihood of the event occurring by chance alone, and regression to the mean. For example, a single blood pressure reading may not be representative of a person's average blood pressure, and most patients with obesity, glucose intolerance and hypertension do not have Cushing's disease. Investigators need to obtain empirical evidence regarding the effects of this heuristic in medical decisions.

The anchoring and adjustment heuristic may be used by physicians in circumstances where an initial probability estimate is re-evaluated as new information becomes available. This describes the manner in which much of the diagnostic and prognostic information becomes available in medical settings, e.g., an initial impression is based on the history and physical examination, which is updated as routine laboratory and more specialized test results become available. Studies in nonmedical settings suggest that people tend to be too conservative as they adjust their initial estimate upward or downward, as if they were"anchored" to their initial estimate.[13]

<u>Cognitive bias</u>: These impediments to the accurate assessment of likelihood are not related to the use of cognitive short cuts or heuristics.

Ego bias occurs when estimates of probability are altered in a self-serving manner. Psychological research indicates that we tend to attribute our successes to skill and our failures to chance, i.e., "bad luck." In a study of estimated surgical mortality, Detmer et al[12] found that most surgeons estimated the mortality rate for their own patients to be lower than the mortality rate for the entire service. (This can be seen to be similar to the Lake Wobegon phenomenon where all children are above average.) Ego bias also may affect the confidence with which estimates are made.

Hindsight bias: knowledge that an event has occurred tends to inflate estimates that it would have occurred (compared with true a priori estimates). Hindsight bias has been shown to occur in clinicopathologic conferences.[14] A related phenomenon may affect judgments of physicians in quality improvement and malpractice reviews.[15]

Value induced bias is generally manifested as anticipated regret. This phenomenon can distort probability estimates when two steps in the judgment process are combined, i.e, when the likelihood estimate is influenced by the (un)desirability of the diagnosis or outcome. Inflation of probability estimates in medical settings related to value induced bias have been shown by Wallsten[6] and Poses.[16]

In the past few years, several research groups have been critical of the potential generalizability of the "heuristics and biases program" of Kahneman and Tversky.[17] For example, Lopes and Oden[18] have noted a) that the difference between "right" and "wrong" answers may be numerically rather small and b) that in certain instances heuristics seem to be properties of a general process of pattern recognition rather than individual "rules" that people apply to solve problems. Gigerenzer and colleagues[19,20] assert that internal problem representation, rather than general heuristics, drives probability assessments. They further note that presenting problems as frequencies, rather than probabilities, leads to a smaller proportion of respondents who seem to use heuristics.

Many of the original heuristics and bias studies as well as those of their critical counterparts have been performed as pencil and paper experiments on college students. In contrast, many of the studies cited above have been performed in naturalistic medical settings. The importance of the use of heuristics and the occurrence of cognitive biases in medical settings is an empirical question about which we currently know too little. Although the medical examples cited above offer plausible arguments regarding the importance of heuristics and biases in medical settings, the frequency of occurrence and the magnitude of their effects are largely still unknown and will require ongoing research efforts.

Beyond the effect of heuristics and cognitive biases lie another set of challenges to clinical judgments. Methodologic considerations and inadequate feedback about prior judgments can greatly limit the opportunities to decipher the “true” predictive value of the data acquired in usual clinical practice. Spectrum and several forms of test related bias (verification bias, diagnostic review bias, test review bias and incorporation bias) can obscure the actual predictive characteristics of clinical data for the clinical observer.[2,21,22] In addition, ethical, cost, and pragmatic concerns often inhibit clinicians from performing the appropriate gold standard test. In many such circumstances, this can prevent feedback which could be used to recognize incorrect judgments and to appropriately alter perceptions of predictive information. An example of this phenomenon was recognized among physicians who were trying to distinguish outpatients with pneumonia from those who had other causes of acute cough.[23]

The accuracy of judgments can be partitioned into three general components: 1) prevalence (base rate), 2) discrimination (the ability to discern occasions when the event of interest will or will not occur), and 3) calibration (the ability to provide realistic probability estimates).

Over the past decade, an increasing number of common medical judgments have been scrutinized to determine their accuracy. Poses and colleagues[16] demonstrated that experienced physicians have difficulty predicting streptococcal pharyngitis in adult patients. These physicians had modest discrimination (receiver operating curve [ROC] area = .67) and a consistent tendency to overestimate the likelihood of strep (average estimate = 62%, actual prevalence = 8%). A similar tendency to overestimate occurrence rates has been demonstrated for physicians predicting pneumonia in outpatients.[24] When practicing physicians estimated the likelihood of pneumonia to be 90%, it was present in 20% of cases. In a similar study, Dawson and Speroff[25] also documented poor calibration and modest discrimination (ROC area = .73) by physicians predicting outpatient pneumonia. Tape and colleagues[26] demonstrated variability across three study sites in both accuracy and apparent physician use of clinical information for predicting outpatient pneumonia.

Tierney and co-workers[27] studied physicians’ judgments of probability of myocardial infarction among emergency room patients with chest pain. They demonstrated very good physician discrimination (ROC area = .87) and generally good calibration, except for

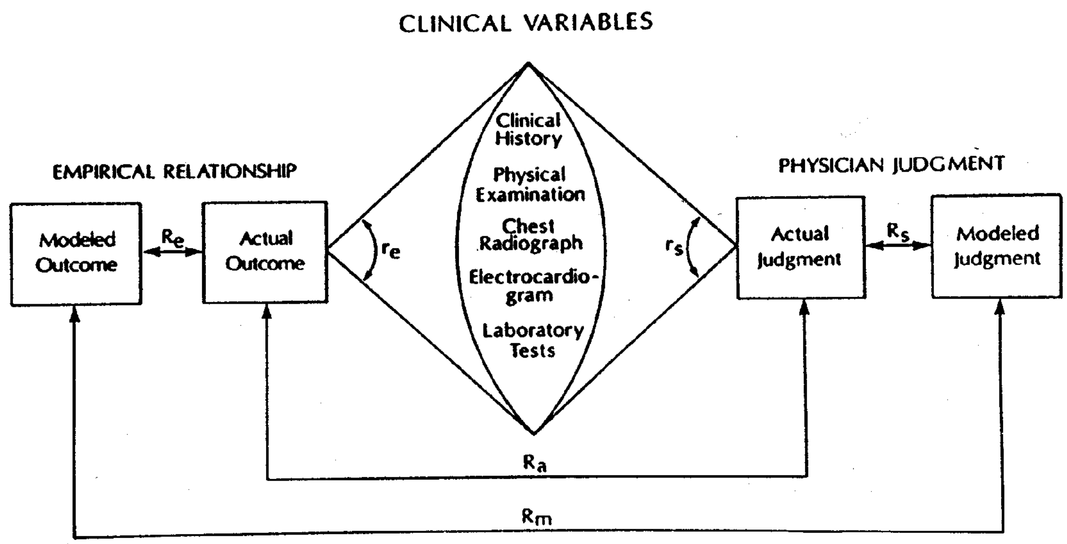

Figure 1. Lens model analysis of clinical variables.

estimates in the mid range of probabilities (between 30 and 70%) where physicians tended to overestimate the likelihood of myocardial infarction. Very good discrimination (ROC areas = .83 - .90) and variable calibration have been demonstrated for physicians' judgments of in-hospital mortality for intensive care unit patients.[28,29] Good discrimination (ROC area = .78) and a general tendency to overestimate mortality also have been documented for physicians' judgments about longer term outcomes (2 and 6 month survival) for seriously ill hospitalized adults.[30,31] These patients, physician accuracy increased as physician confidence in their predictions increased.[32] This finding is in contrast to many nonmedical and medical studies which tend to show a lack of relationship between confidence levels and accuracy.[33]

A technique called lens model analysis (see Figure 1) has recently been introduced into medical studies and can help dissect out separate components of the judgment process.[34-37] It is designed to compare the relationships among clinical predictors (cues), and the outcome of interest (r_e), as well as cues and judgments made by physicians or others (r_s). The adequacy of the models of judgment (R_s) and outcome (R_e) can be assessed. In addition, the relationship between the outcome and judgment (R_a) and between the two models (R_m) can be compared. Speroff and coworkers[34] used it to examine why physicians have difficulty predicting hemodynamic status with noninvasive measures. They discovered that physicians seem to underutilize some important cues (30% of the explained variance came from data from the laboratory, chest radiograph and electrocardiogram whereas these data accounted for only 7% of the variance in physicians' judgments). In addition, physicians seem to place too much emphasis on other important cues (physicians placed too much weight on the clinical impression of the presence of congestive heart failure).

Hammond[38] has offered a theoretical model of decision making in situations where assessments may change over time. His theory of "dynamic tasks" asserts that the output from a task system will tend to stimulate a form of cognitive activity that lies on a continuum from calculation to intuition. The form of cognitive activity that is induced may (or may not) be compatible with the task system. He further argues that judgmental accuracy should be highest when the induced cognitive activity matches that part of the continuum that is appropriate for the task system (calculation, intuition or a combination). Hammond's formulation provides a specific structure against which future research in dynamic decision making can be tested.

Stewart and Lusk[39] recently have developed a method to assess judgmental accuracy for continuous outcome measures. Their decomposition of accuracy is explicitly linked to lens model analysis and allows specific methods for improving judgments to be identified and investigated (see Table 1). The first five components of prediction (rows 1-5) relate to judgment discrimination. The last two components (rows 6, 7) relate to calibration. Rows 3 through 7 are at least partially under the control of the judge. Columns A through N denote potential methods for improving judgments. Letters in individual cells indicate literature cited by the authors (n = nonmedical, b = both medical and non medical) that investigated a particular component of prediction. They also note areas that should be investigated (X).

In combination, the theoretical and analytic structures provided by Hammond[38] and Stewart and Lusk[39] provide a powerful construct within which the systematic development and evaluation of information systems and subsequent judgments by physicians can be evaluated.

References

1. Dawson NV, Arkes HR. Systematic errors in medical decision making: judgment limitations. Journal of General Internal Medicine 1987;2:183-7.

	C. Adapted from Stuart and Lusk (1994) **Method for Improving Judgements**																	
Component of Prediciton	**A***	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**	**P**	**Q**	**R**
1. Inherent (environmentally) predictability	**n****																	
2. Fidelity of information system		**n**																
3. Match between environment and judge			**n**	**n**	**b**	**b**												
4. Reliability of acquiring information							**X**	**X**	**X**									
5. Reliability of processing information										**n**	**b**	**n**	**b**	**b**				
6. Regression bias				**b**								**n**			**X**	**n**	**n**	**b**
7. Base rate bias				**b**											**X**	**n**		**b**

**n = nonmedical studies cited, usually weather forecasting and psychology

b = both medical and nonmedical studies cited

x = no studies found specifically designed to improve judgements in these areas (although reliability of acquiring information has been shown to be a problem
in medical and nonmedical studies)

*A) Research to find new predictors
B) Develop better measures of true predictors
C) Train judge about environmental system
D) Experience with specific judgement problem
E) Cognitive feedback biases
F) Train judge to ignore non-predicitve cues
G) Develop clear definitions of cues
H) Training to improve cue judgements
I) Improve information displays
J) Replace judge with a model
K) Combine several independent judgements
L) Require justification for judgements
M) Decompose the judgment task
N) Mechanical combination of cues
O) Statistical training
P) Feedback about judgment
Q) Search for discrepant information
R) Statistical correction for bias

Table 1. Method to assess judgmental accuracy for continuous outcome measures
(Adapted from Stewart and Lusk)

2. Dawson NV. Physician judgment in clinical settings: methodological influences and cognitive performance. *Clin Chem.* 1993;39:1468-80.

3. Hershberger PJ, Park HM, Markert RJ, Cohen SM, Finger WW. Development of a test of cognitive bias in medical decision making. *Academic Medicine.* 1994;69:839-42.

4. Eddy DM. Probabilistic reasoning in clinical medicine: problems and opportunities. In: Kahneman D, Slovic P, Tversky A, eds. Judgment Under Uncertainty: Heuristics and Biases. New York: Cambridge University Press, 1982:249-67.

5. Wolf FM, Gruppen LD, Billi JE. Differential diagnosis and the competing hypothesis heuristic: a practical approach to judgment under uncertainty and Baysian probability. *JAMA.*1985;253:2858-62.

6. Wallsten TS. Physician and medical student bias in evaluating diagnostic information. *Medical Decision Making* 1981;1:145-64.

7. Christensen-Szalanski JJ, Bushyhead JB. Physicians' misunderstanding of normal findings. *Medical Decision Making*. 1983;3:169-75.

8. McNeil BJ, Pauker SG, Sox HC, Tversky A. On the elicitation of preferences for alternative therapies. *N Engl J of Med.* 1982;306:1259-62.

9. Forrow L, Taylor WC, Arnold RM. Absolutely relative: how research results are summarized can affect treatment decisions. *Amer J of Med.* 1992;92:121-4.

10. Naylor CD, Chen E, Strauss B. Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness? *Ann of Intern Med.* 1992;117:916-21.

11. Bucher HC, Weinbacher M, Gyr K. Influence of method of reporting study results on decision of physicians to prescribe drugs to lower cholesterol concentration. *Brit Med J.* 1994;309:761-4.

12. Detmer DE, Fryback DG, Gassner K. Heuristics and biases in medical decision making. *J of Med Educ.* 1978;53:682-3.

13. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. *Science.* 1974;185:1124-31.

14. Dawson NV, Arkes HR, Siciliano C, Blinkhorn R, Lakshmanan M, Petrelli M. Hindsight bias: an impediment to accurate probability estimation in clinicopathologic conferences. *Medical Decision Making*. 1988;8:259-64.

15. Caplan RA, Posner KL, Cheyney FW. Effect of outcome on physician judgments of appropriateness of care. *JAMA* 1991;265:1957-60.

16. Poses RM, Cebul RD, Collins M, Fager SS. The accuracy of experienced physicians' probability estimates for patients with sore throats. *JAMA.* 1985;254:925-9.

17. Kahneman D, Slovic P, Tversky A (eds). Judgment Under Uncertainty: Heuristics and Biases. New York: Cambridge University Press; 1982.

18. Lopes LL, Oden GC. The rationality of intelligence. Poznan Studies in the Philosophy of the Sciences and the Humanities 1994;21:199-223.

19. Gigerenzer G, Hell W, Blank H. Presentation and content: the use of base rates as a continuous variable. *J of Exper Psych.: Human Perception and Performance.* 1988;14:513-25.

20. Gigerenzer G. Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa). In Wright G, Ayton P, eds. Subjective Probability. New York: John Wiley and Sons; 1994:129-161.

21. Ransohoff DF, Feinstein AR. Problems of spectrum and bias in evaluating the efficacy of diagnostic tests. *N Engl J of Med.* 1978;299:926-30.

22. Begg CB. Biases in the assessment of diagnostic tests. *Stat in Med.* 1987;6:411-23.

23. Bushyhead JB, Christensen-Szalanski JJ. Feedback and the illusion of validity in a medical clinic. *Medical Decision Making*. 1981;1:115-23.

24. Christensen-Szalanski JJ, Bushyhead JB. Physicians' use of probabilistic information in a real clinical setting. *J of Exper Psych.* [Human Perceptions] 1981;4:928-35.

25. Dawson NV, Speroff T. Validated decision aid for outpatient pneumonia [Abstract]. *Clin Res.*1989;37:774.

26. Tape TG, Heckerling PS, Ornato JP, Wigton RS. Use of clinical judgment analysis to explain regional variations in physicians' accuracies in diagnosing pneumonia. *Medical Decision Making*.1991;11:189-97.

27. Tierney WM, Fitzgerald J, McHenry R, Roth BJ, Pstay B, Stump DL, Anderson FK. Physicians' estimates of the probability of myocardial infarction in emergency room patients with chest pain. *Medical Decision Making*.1987;6:12-7.

28. Poses RM, Bekes C, Copare FJ, Scott WE. The answer to "What are my chances, Doctor?" depends on who is asked: prognostic agreement and inaccuracy for critically ill patients. *Crit Care Med.* 1989;17:827-33.

29. McClish DK, Powell SH. How well can physicians estimate mortality in a medical intensive care unit? *Medical Decision Making*. 1989;25-32.

30. Knaus WA, Harrell FE, Lynn J, Goldman L, Phillips RS, Connors AF, Dawson NV, Fulkerson WJ, Califf RM, Desbeins N, Layde P, Oye RK, Bellamy PE, Wagner DP for the SUPPORT Investigators. The SUPPORT prognostic model: prediction of survival for seriously ill hospitalized adults. *Ann of Intern Med.* 1995;122:190-202.

31. Arkes HR, Dawson NV, Speroff T, Harrell FE, Alzola C, Phillips RS, Goldman L, Fulkerson W, Califf R, Desibiens N, Oye RK, Knaus W, Connors AF, and the SUPPORT Investigators. The covariance decomposition of the probability score and its use in evaluating prognostic estimates. *Medical Decision Making*. 1995;15:120-31.

32. Connors AF, Dawson NV, Speroff T, Arkes H, Knaus WA, Harrell FE, Lynn J, Teno J, Goldman L, Califf R, Fulkerson W, Oye R, Bellamy P, Desbiens N and the SUPPORT Investigators. Physicians' confidence in their estimates of the probability of survival: relationship to accuracy [Abstract]. *Medical Decision Making*. 1992;12:336.

33. Dawson NV, Connors AF, Speroff T, Kemka A, Shaw P, Arkes HR. Hemodynamic assessment in the critically ill: "Is physician confidence warranted?" *Medical Decision Making.* 1993;13:258-66.

34. Speroff T, Connors AF, Dawson NV. Lens model analysis of hemodynamic status in the critically ill. *Medical Decision Making*. 1989;9:243-52.

35. Wigton RS. Use of linear models to analyze physicians' decisions. *Medical Decision Making*. 1988;8:241-52.

36. Schwartz S, Griffin T. Medical Thinking: The Psychology of Medical Judgment and Decision Making. New York: Springer Verlag, 1986.

37. Hursch C, Hammond KR, Hursch J. Some methodological considerations in multiple-cue probability studies [Review]. *Psych Rev.* 1964;71:42-60.

38. Hammond KR. Judgment and decision making in dynamic tasks. *Information and Decision Technologies.* 1988;14:3-14.

39. Stewart TR, Lusk CM. Seven components of judgmental forecasting skill: implications for research and the improvement of forecasts. *J of Forecasting* 1994;13:579-99.