A RAND NOTE

Assessing Clinical Decision Making: Is the Ideal System Feasible?

Robert W. Dubois, Robert H. Brook

RAND

The research described in this report was supported by a grant from the U.S. Department of Health and Human Services.

This Note contains an offprint of RAND research originally published in a journal or book. The text is reproduced here, with permission of the original publisher.

Published 1991 by RAND
1700 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138

A RAND NOTE

N-3361-HHS

Assessing Clinical Decision Making: Is the Ideal System Feasible?

Robert W. Dubois, Robert H. Brook

Supported by the
U.S. Department of Health and Human Services

RAND

Robert W. Dubois
Robert H. Brook

Assessing Clinical Decision Making: Is the Ideal System Feasible?

While caring for patients, physicians make a variety of decisions. Can current methods adequately determine whether these decisions are correct? If not, what improvements are needed? This paper begins with a review of several explicit methods to assess physician decision making. It then describes a more comprehensive system that would use Bayesian logic to assess whether a physician responded appropriately to the needs of an individual patient. Although sophisticated branching logic may be theoretically desirable, it may not be feasible. The paper concludes by proposing an explicit, potentially practical method that would judiciously use branching logic.

While caring for patients, physicians make a variety of decisions. They choose workups, make diagnoses, and select management strategies. The appropriateness of these clinical decisions will strongly determine whether a patient receives high- or low-quality care. To assess the care provided by a physician, therefore, it is important to examine the following question: In a given clinical situation, did the physician make the proper decisions in terms of what he or she chose both to do and not to do?

Two general methods have been developed to answer this question. The implicit method uses expert clinicians to review a case and subjectively determine whether the physician made appropriate medical decisions and provided appropriate care. The explicit method uses specific process-of-care criteria to objectively assess a physician's decisions and actions. This paper explores the limits of the explicit approach to the assessment of quality.

We begin by reviewing the currently available explicit methods, which range from simple to sophisticated. We show that the more sophisticated methods can better define a patient's needs and more accurately determine whether a physician responded appropriately to them. If more complex systems function better than simple ones, then how complex would the "ideal" explicit system be? In a later section of this paper, we formulate and try to answer this question.

Explicit Review From the Past to the Present

Lembcke, in the 1950s, was the first to describe the explicit methodology for quality review.[1] His approach used objective criteria to evaluate a physician's activities. These criteria were developed by panels of clinical experts who specified the important features of care for a group of patients. With these criteria, a medical record could be objectively reviewed and the quality of care rated.

Unfortunately, the early applications of this explicit methodology could only crudely assess clinical descision making. In these examples, the physician could receive the same quality-of-care score whether he or she complied with critical or noncritical elements of care. For example, Hulka et al. used explicit review to examine the outpatient care provided to patients presenting with dysuria. All of the following criteria had equal weights:

☐ Document these historical elements: frequency,

Reprinted from *Inquiry*, Vol. 25, No. 1, Spring 1988, pp. 59–64.

nocturia, hematuria, symptom duration, abdominal/back pain, fever, prior urinary infection.

- ☐ Perform these physical examination elements: temperature, blood pressure, abdomen.
- ☐ Obtain these laboratory studies: urinalysis, urine culture.
- ☐ Provide this management: antibiotics, follow-up urinalysis.[2]

In this study, documenting the presence or absence of hematuria (minimally important criterion) had equal emphasis in the physician's overall quality score as obtaining a urine culture or prescribing antibiotics (very important criteria). Not surprisingly, with these simple criteria, Hulka et al. could not demonstrate a positive relationship between good process of care (adherence to explicit criteria) and good patient outcome. This failure may have reflected a methodology that did not fully capture the relative importance of the various clinical decisions made by health care providers.

In contrast, other explicit studies used weighted criteria to reflect the greater clinical importance of certain items over others. In Payne and Lyons's Hawaii Office Care Study, compliance with each history item for urinary tract infection increased the score by .5, whereas compliance with laboratory and therapeutic criteria increased the score by 2 and 3, respectively.[3] With weighted criteria, clinical errors in providing proper therapy had somewhat greater influence on the quality-of-care score (and presumably on the patient's outcome) than did errors in documenting historical features. Weighting of the criteria thus resulted in an explicit system that theoretically could more closely match clinical reality. Now, more important clinical issues (deciding on proper therapy) had greater influence on scores than did clinical issues of lesser importance (documentation of symptoms).

Payne and Lyons used criteria weights that varied from .5 to 3, or sixfold. As weights vary by greater amounts, certain items have more importance than others, reflecting their greater significance to patient care. Using criteria weights that varied 40-fold, Mates and Sidel examined the care of asthmatic adults in an emergency department.[4] For patients with "severe" episodes, documentation of certain historical features (i.e., recent fever, age at onset, duration of current attack) had very small weights (.25), whereas choosing the proper therapy (hydration, epinephrine, inhaled bronchodilators) had relatively large weights (10). Panels of clinical experts determined these weights as reflecting their relative clinical importance. They believed that the therapy decisions were far more important than those regarding the patient's history.

With this more complex system, Mates and Sidel demonstrated a small but statistically significant relationship between good patient care and good patient outcome. This observation suggests that a more complex explicit system (one that uses large weights) may reflect clinical reality more closely than a more simplified system.

Nevertheless, the explicit systems discussed thus far have a significant limitation: The same list of criteria applies to all patients with that condition. Yet, not all patients with urinary tract infections need the same diagnostic workup or management. For example, those experiencing a first episode may need care that differs from that of patients experiencing another in a long series of attacks. In addition, care for women with urinary tract infections will clearly differ from that for men with the problem. None of the three systems described above used these historical or demographic features in defining proper care. The same set of criteria applied to all cases. Weighting of the criteria would differentiate the more critical elements of care from the less critical ones. But it would not identify the needs of individual patients, and thus could not determine whether a provider responded appropriately to them. This important clinical feature requires the use of branching logic.

With branching logic, the proper history, physical examination, laboratory studies, and management may differ for groups of patients with the same diagnosis. One patient subgroup may need one type of care, whereas another subgroup may need a slightly different type of care. As an example, the Physician Performance Index for chronic cholecystitis developed by Payne and Lyons used two types of criteria. The first five, or mandatory, criteria applied to all cases:

- ☐ history of food intolerance
- ☐ history of previous attacks
- ☐ history of jaundice
- ☐ complete blood count
- ☐ urinalysis
- ☐ chest radiograph within the past one year

The last, or branched, criterion applied only if the patient was over age 50. In that instance the physician should perform an electrocardiogram (ECG).[5] Similarly, in the emergency department asthmatic study described above, the criteria for proper management included corticosteroids only if the patient showed no improvement in two to eight hours or if the patient had steroid dependence.[6] Both of these algorithmic branches led to different sets of explicit criteria that applied to different sets of patients. With branching, criteria can more closely match the clinical decisions made by physicians when they care for individual patients with individual needs.

Typical sets of explicit criteria, like the ones described above, have one or at most several branches embedded within them. This severely limits how precisely the criteria can match the real requirements of a given patient. Ideally, a set of criteria should have many branch points to reflect the needs of many different subgroups of patients.

The use of multiple branch points was formalized and popularized by Greenfield and his colleagues in their criteria maps. In one study, they designed a criteria map that dealt with the diagnosis and management of diabetic patients.[7] This map had 47 separate branch points, among them the following:

- ☐ If a woman has a vaginal discharge, she needs a pelvic examination.
- ☐ If a patient has new proteinuria, he or she needs a blood-urea-nitrogen (BUN), a serum creatinine, and a urine culture.
- ☐ If a patient has paresthesias, decreased vibratory sense, or decreased reflexes, he or she needs to have the skin examined for ulcers or infection.
- ☐ If a patient has new or worse chest pain, he or she needs an ECG.

These multiple branch points allow a physician's decisions to be reviewed using criteria that apply with greater specificity to a given patient. More than any of the methods previously described, the criteria map could more specifically assess the clinical decisions made by health care providers as they care for patients.

The criteria map approach has also been successfully applied to patients presenting to the emergency department with chest pain.[8] In that study, Greenfield et al. examined how often patients were inappropriately discharged home with a myocardial infarction or inappropriately admitted to the hospital with no significant disease. They also compared these patient outcomes with process scores derived from either simple criteria lists or from branching criteria maps. Using simple lists, they could not demonstrate any relationship between process and outcome. Using the branched and weighted criteria map, however, they found that good patient outcomes resulted from good patient care.

Thus far, we have described a methodologic evolution from systems with little complexity to ones with far greater sophistication. As these systems became more sophisticated, they could better identify the specific needs of specific patients and objectively assess the physician's responses to them. Moreover, these more complex systems could also more consistently demonstrate that good patient care leads to good patient outcomes. We now return to the question posed in the introduction: If a more complex explicit system can perform better both theoretically and empirically, how complex or sophisticated would the ideal system be?

Until now, the criteria maps represented the most complex method yet developed to retrospectively assess decisions made by providers and the care given to patients. With 47 separate branches, the diabetes criteria map seemed to characterize the most important technical aspects of patient care. It had a limitation, however. Each branch in the criteria map had no input from more distant branches. Thus, those branches that dealt with cardiovascular complications did not utilize the important information obtained from the branches that dealt with peripheral nerve complications and/or hyperlipidemia. For example, the importance of performing an ECG in a patient did not differ whether or not the patient had known peripheral neuropathy (impaired nerves). From the standpoint of clinical decision making, it is far more critical to pursue episodes of vague chest pain in patients with neuropathy than in patients without this diabetic complication. With nerve impairment, ischemic episodes (where the heart muscle receives an inadequate supply of oxygen) may not present in the usual manner with crushing, substernal pressure. Thus, in patients with nerve damage, even seemingly minor chest "aches" may represent an ominous process that needs urgent investigation and treatment.

Bayes's theorem states that the importance of a symptom (or a test result) in predicting the presence of disease and need for treatment will greatly depend on the prior probability of that disease. In other words, chest pain will more likely represent true ischemia if the patient has a higher likelihood of that disease (i.e., has known risk factors such as elevated cholesterol levels). In a nondiabetic, vague episodes of chest pain have less significance than those same episodes in a diabetic patient with hypercholesterolemia (elevated blood cholesterol level). In essence, patients with risk factors for heart disease need more thorough investigation of episodic chest pain than do patients without risk factors. A set of explicit criteria must differentiate these two groups of patients and place greater importance on the management of the former.

Despite the complexity that criteria maps already exhibit, they do not embody this critical theoretic construct. As presently designed, each branch in a criteria map specifies the proper care for that clinical issue in ignorance of data that were identified in other branches. Thus, the need to perform an ECG is not influenced by the patient's risk factors or other known diseases. In an ideal system, each branch should utilize whatever information is available – a straightforward concept. As we will show, a straightforward concept may not be straightforward to operationalize. In fact, this theoretically desirable system may not be feasible.

We will demonstrate these points with an example. In this example, a patient may have none, some, or all of the following five comorbidities:

- ☐ over age 65
- ☐ diabetes
- ☐ prior angina
- ☐ hypertension
- ☐ emphysema

The following five complications may ensue:

- ☐ tachycardia (rapid heart rate)
- ☐ chest pain
- ☐ fever
- ☐ renal failure
- ☐ anemia

In this example, the relative clinical importance of each complication definitely depends on the patient's comorbidities. Thus, prompt evaluation or treatment of chest pain in an otherwise healthy patient may have a weight of 1, representing an important but not a critical item. In a patient with diabetes, evaluation or treatment of that same episode may have a weight of 5, implying greater urgency. If the patient is diabetic and also over age 65, the weight might rise to 10. The weight could even jump to 25 for a patient over age 65 with diabetes and prior angina who experiences an episode of chest pain. In essence, the importance of the physician's decision to evaluate or treat an episode of chest pain will greatly depend on the patient's other comorbid conditions.

This example quickly becomes very complex as we place all of the potential scenarios on our "ideal" map. There are 32 ways to combine five comorbid conditions.[9] There are also 32 ways to combine the five potential complications. Because each of the 32 combinations of comorbid conditions could lead to any of the 32 combinations of complications, the final map will have 1,024 (32 × 32) potential scenarios. To judge the relative importance of these scenarios, each of them will need relative weights.

Despite the complexity of this map, we have made a very significant simplifying assumption – that is, that the order of complications does not matter. Thus, a patient who first experiences angina and later develops anemia does not differ from the patient who develops anemia first and then later experiences angina. This assumption, however, misstates clinical reality. A patient who experiences angina *after* developing anemia is quite different from a patient with the reverse scenario. Anemia implies a lowered ability of the blood to carry oxygen. This may well predispose the patient to subseqent episodes of angina that occur when the heart muscle has an inadequate supply of oxygen. Anemia that develops *after* the episode of angina most likely represents an unrelated phenomenon.

If the order in which complications occur does matter, these same five complications produce 326 permutations[10] (not merely 32 combinations), or ultimately a total of 10,432 unique clinical scenarios.

We must deal with another simplifying assumption: Patients can have far more than five different comorbid conditions and experience far more than five different complications. A map that uses 10 comorbidities and 10 complications would produce 1,048,576 different scenarios if order does not matter, and 10,100,839,424 if it does. This "ideal" map will also need a weight to reflect the relative importance of each branch. But even this model minimizes the number of significant comorbidities (more like 50) and complications (more like 20).

By relaxing several simplifying assumptions, we have quickly moved from a criteria map with 47 different branches to one that captures more than 10 billion different scenarios. This model would reflect the unique requirements of individual patients far better than any of the explicit systems described thus far. It would also better determine whether the physician had made proper clinical decisions and provided proper care. Unfortunately, this system is not currently feasible. First, there are more than 10 billion pathways through this model. Each one of these paths represents the proper management of one disease manifestation for one particular subgroup of patients. To function, this system would need to know the importance of one path or management strategy relative to another. This knowledge is not available. It is not even known how often a glucose level should be determined for diabetics in general, let alone for diabetics with coexistent renal disease or emphysema. To operationalize this "ideal" system, each of the 10 billion pathways needs weights that reflect its relative importance to patient care. Efficacy studies to define these weights have not been done.

Secondly, software development for a model of this size would be a monumental task. The program would require billions of steps and clinical weights for each one. Even if possible, this model represents just one type of patient. Other models would be needed to deal with myriads of other types of patients (e.g., obstetric, orthopedic, pediatric, and psychiatric patients).

We have tried to show that more sophisticated explicit models can better define the needs of an individual patient. In an ideal system, prior clinical history would strongly influence a patient's care. Yet, our "ideal" system seems completely infeasible. Have we, in essence, painted ourselves into a corner from which there appears to be no way out?

Our situation seems like that of a tree whose branches have grown so full that they will likely topple its base. In essence, our conceptual tree needs pruning. If possible, we must trim away unnecessary branches. For example, patients with arthritis and angina may not significantly differ from those with angina alone. When patients from either group present with chest pain, they may need similar workups or similar treatment. If true, this simplification could remove several of the branches from one part of the tree. In like manner, other combinations of comorbidities could be linked together, which would begin to shrink the size of the tree.

Another form of simplification uses temporal sequencing of complications. For patients with fever and hypokalemia (low serum potassium), the order in which they develop these complications probably does not matter. They need the same diagnostic workup and therapy whether they first develop fever and then hypokalemia, or vice versa. Selective removal of unnecessary branch points by either method would simplify the model and make it more practical to use.

This pruning process has a precedent. Pozen et al. examined a myriad of potential indicators for coronary artery disease in patients presenting to emergency departments with chest pain.[11] They hoped to find a small subset of criteria that could identify the patients most likely to have suffered a myocardial infarction. In fact, they found that seven key indicators could sufficiently differentiate patients presenting with myocardial infarctions from those presenting with less serious conditions. Additional parameters added little to their model. Similarly, other authors have identified the importance of pruning complex decision trees to make them manageable.[12]

Summary and Conclusion

In this paper, we describe a variety of explicit methods to assess the clinical decisions made by physicians and the care they provide to their patients. We try to show theoretically that the simple explicit methods cannot adequately capture the complex needs of individual patients. Empirically, it is difficult for these systems to demonstrate that good process of care leads to good outcomes. In contrast, the more complex explicit methods better identify the needs of important patient subgroups. They also more consistently demonstrate a relationship between process and outcome.

We then propose a system that would have an additional level of complexity. Using Bayesian analysis, an explicit system could use a patient's pertinent symptoms and findings to determine proper care. Unfortunately, this hypothetical system requires more than 10 billion pathways, and would be totally infeasible. We suggest a potential solution that involves pruning our massive logic tree to those elements that truly matter.

But we cannot develop even a well-pruned tree, since the relevant epidemiologic and clinical efficacy data do not yet exist. This requires further clinical research that could identify the aspects of clinical care that truly matter. Further research could answer the following types of questions: Which components of a physician's history and physical examination identify the information that most influences patient care? Which laboratory studies and treatment options influence patient outcomes? With these data, future explicit methods should be better able to assess those physician decisions that truly matter.

These future systems would use Bayesian logic and branching criteria, but they would use them judiciously. With the proper efficacy studies, future researchers could develop a carefully selected list of criteria that significantly influence patient outcomes, as opposed to one that includes all possible aspects of care. In this way, an explicit system could be both sophisticated in design and relatively simple to use.

Notes

The views expressed are the authors' own and are not necessarily those of the RAND Corporation.

1 P. Lembcke, "Medical Auditing by Scientific Methods: Illustrated by Major Female Pelvic Surgery," *Journal of the American Medical Association* 162 (Oct. 13, 1956): 646–655.

2 B. Hulka et al., "Peer Review in Ambulatory Care: Use of Explicit Criteria and Implicit Judgments," *Medical Care*, vol. 17, suppl. (March 1979).

3 B. Payne and T. Lyons, "Method of Evaluating and Improving Personal Medical Care Quality: Office Care Study" (Ann Arbor: University of Michigan School of Medicine, 1972).

4 S. Mates and V. Sidel, "Quality Assessment by Process and Outcome Methods: Evaluation of Emergency Room Care of Asthmatic Adults," *American Journal of Public Health* 71 (July 1981): 687–693.

5 T. Lyons and B. Payne, "The Quality of Physicians' Health-Care Performance," *Journal of the American Medical Association* 227 (Feb. 25, 1974): 925–928.

6 See Mates and Sidel (note 4).

7 S. Greenfield, C. Lewis, S. Kaplan, and M. Davidson, "Peer Review by Criteria Mapping: Criteria for Diabetes Mellitus," *Annals of Internal Medicine* 83 (December 1975): 761–775.

8 S. Greenfield, M. Nadler, M. Morgan, and K. Shine, "The Clinical Investigation and Management of Chest Pain in an Emergency Department: Quality Assessment by Criteria Mapping," *Medical Care* 15 (November 1977): 898–905; S. Greenfield, S. Cretin, and L. Worthman, "Comparison of Criteria Map to a Criteria List in Quality-of-Care Assessment for Patients With Chest Pain: The Relation of Each to Outcome," *Medical Care* 19 (March 1981): 255–272.

9 A patient could have one, two, three, four, five, or none of the comorbid conditions. There are five different ways that a patient could have a single comorbid condition, $C(5, 1)$, i.e., the number of combinations of five items taken one at a time. Similarly, there are 10 combinations of five items taken two at a time, $C(5, 2)$; 10 combinations of five items taken three at a time $C(5, 3)$; five combinations of five items taken four at a time, $C(5, 4)$; one combination of five items taken five at a time, $C(5, 5)$; and one combination of five items taken zero at a time, $C(5, 0)$. In total, there are 32 different combinations of the five comorbid conditions: $C(5, 0) + C(5, 1) + C(5, 2) + C(5, 3) + C(5, 4) + C(5, 5) = 1 + 5 + 10 + 10 + 5 + 1 = 32$.

10 This calculation is similar to that in note 9, except that permutations are used rather than combinations: $P(5, 0) + P(5, 1) + P(5, 2) + P(5, 3) + P(5, 4) + P(5, 5) = 1 + 5 + 20 + 60 + 120 + 120 = 326$ total permutations. Thus, 32 (combinations of comorbid conditions) × 326 (permutations of complications) = 10,432 unique clinical scenarios.

11 M. Pozen, R. D'Agostino, and J. Mitchell, "The Usefulness of a Predictive Instrument to Reduce Inappropriate Admissions to the Coronary Care Unit," *Annals of Internal Medicine* 92 (1980): 238.

12 M. Weinstein and H. Fineberg, *Clinical Decision Analysis* (Philadelphia: W. B. Saunders Co., 1980); H. Raiffa, *Decision Analysis: Introductory Lectures on Choices Under Uncertainty* (Reading, MA: Addison-Wesley Publishing Co., 1970).

RAND/N-3361-HHS