Could Our Pretest Probabilities Become Evidence Based?

A Prospective Survey of Hospital Practice

W. Scott Richardson, MD, Walter A. Polashenski, MD, Brett W. Robbins, MD

OBJECTIVE: We sought to measure the proportion of patients on our clinical service who presented with clinical problems for which research evidence was available to inform estimates of pretest probability. We also aimed to discern whether any of this evidence was of sufficient quality that we would want to use it for clinical decision making.

DESIGN: Prospective, consecutive case series and literature survey.

SETTING: Inpatient medical service of a university-affiliated Veterans' Affairs hospital in south Texas.

PATIENTS: Patients admitted during the 3 study months for diagnostic evaluation.

MEASUREMENTS: Patients' active clinical problems were identified prospectively and recorded at the time of discharge, transfer, or death. We electronically searched MEDLINE and hand-searched bibliographies to find citations that reported research evidence about the frequency of underlying diseases that cause these clinical problems. We critically appraised selected citations and ranked them on a hierarchy of evidence.

RESULTS: We admitted 122 patients for diagnostic evaluation, in whom we identified 45 different principal clinical problems. For 35 of the 45 problems (78%; 95% confidence interval [95% CI], 66% to 90%), we found citations that qualified as disease probability evidence. Thus, 111 of our 122 patients (91%; 95% CI, 86% to 96%) had clinical problems for which evidence was available in the medical literature.

CONCLUSIONS: During 3 months on our hospital medicine service, almost all of the patients admitted for diagnostic evaluation had clinical problems for which evidence is available to guide our estimates of pretest probability. If confirmed by others, these data suggest that clinicians' pretest probabilities could become evidence based.

KEY WORDS: pretest probability; differential diagnosis; evidence-based medicine.

J GEN INTERN MED 2003;18:203-208.

Many new resources are arriving to help the diagnostician. More diagnostic tests are becoming available, and more of the available tests are being subjected to

rigorous evaluation of their accuracy, precision, and utility in practice.¹ Parts of the clinical examination are also being studied rigorously for their characteristics as diagnostic tests.^{2,3} Frontline clinicians are gaining increasing access to high-quality evidence about diagnostic tests in textbooks^{4,5} and in systematic reviews.^{1,6,7} Using this evidence requires more than knowing a test's discriminatory power. Clinicians also need to estimate pretest probabilities for the disorders being considered.^{8,9} One would start with pretest probabilities for each disorder, combine them with the test's likelihood ratios or sensitivity and specificity, and then compare the resulting post-test probabilities to one's action thresholds to decide what to do next.^{10,11}

But where do these pretest probabilities come from? Some say clinicians can generate pretest probabilities on the basis of clinical experience, drawing on their memories of prior cases with the same clinical problem. Yet research has shown that clinicians' estimates of probability vary widely and are often inaccurate.¹²⁻¹⁵ This is because clinicians' memories are fallible and their thinking is subject to numerous biases.¹⁶⁻¹⁹ Clinicians tend to recall recent or striking numerator cases without the proper denominators, which leads to error in estimating probability. Therefore, by itself, clinical experience appears insufficient to guide accurate probability estimation.

Clinical care research represents another source of information about disease probability. In such research, investigators gather patients with a defined clinical problem, carry out a diagnostic evaluation, then record the diagnostic yield in terms of the frequency of underlying disorders found.^{20,21} Widely known examples include studies of patients with syncope^{22,23} and fever of unknown origin.^{24,25} If critical appraisal of this evidence suggests that it is valid, important, and applicable to their patients, clinicians can use the frequencies of disease from these studies as starting points for estimating pretest probabilities in their own patients. Clinicians would then need to adjust these probabilities up or down, depending on features of their patients or their practices.^{26,27}

But how often is this research evidence available to guide estimates of pretest probability? We knew of some examples, but we didn't know how often we could find it, and we found no published research of this question. Therefore, we surveyed our hospital medicine service to ascertain our patients' clinical problems, and then surveyed the medical literature for research evidence about the frequency of underlying diseases that cause these clinical problems. Our research aims were: 1) to measure the proportion of patients on our clinical service who presented with clinical problems for which research evidence was available to guide our estimates of pretest probability; and

Received from the Department of Internal Medicine, Wright State University School of Medicine (WSR), and the Three Owl Learning Institute (WSR), Dayton, Ohio; the General Medicine Unit, Rochester General Hospital (WAP), and the Departments of Medicine (WAP, BWR) and Pediatrics (BWR), University of Rochester, Rochester, NY.

Address correspondence to Dr. Richardson: WSU Department of Internal Medicine, PO Box 927, Dayton, OH 45401 (e-mail: scott.richardson@wright.edu).

2) to discern whether any of this evidence was of sufficient quality that we would want to use it for clinical decision making.

METHODS

Clinical Setting

This survey was carried out on the inpatient general medical service at the Audie L. Murphy Memorial Veterans' Hospital in San Antonio, Texas. This hospital provides a mixture of secondary care for U.S. military veterans who reside in San Antonio and tertiary care for veterans who live anywhere in south Texas. This hospital's department of medicine averages approximately 5,000 admissions per year. The decisions to admit patients to the inpatient medical service are made either by clinicians in the hospital's emergency triage department or by referring clinicians in ambulatory care, but not by members of the clinical teams who provide inpatient care. During the time of the study, there were 5 teams providing inpatient general medical care. The assignment of patients to the 5 teams was made on the basis of a call rotation scheme, so that all patients admitted to the general medical service had a roughly equivalent chance (1 in 5) of being assigned to each team, with no preferential selection by clinical problem.

Clinical Team

The survey was conducted during the months of November, 1999, February, 2000, and March, 2000, when one of us (WSR) was the supervising attending physician. During each of these 3 months, the clinical team's composition was very similar: an attending with 20 years' experience as a general internist, 1 internal medicine resident in the second or third postgraduate year, 2 internal medicine residents in the first postgraduate year, and 2 to 4 medical students in the third or fourth years of a 4-year undergraduate program.

Clinical Problems

The research protocol and standardized data collection forms were approved by the institutional review board. For each patient admitted or transferred to our inpatient service during the study months, as part of routine clinical care, we prospectively identified any active clinical problems in order to plan our diagnostic evaluations. By *active clinical problem* we meant any symptom, symptom cluster, physical finding, test abnormality, clinical syndrome, or health condition that required diagnostic evaluation during the admission. This allowed us to distinguish a clinical problem, such as dyspnea, from the final diagnosis we might make, such as exacerbation of chronic obstructive airways disease. Also, by choosing presenting problems rather than only symptoms, we were able to include illnesses that required hospitalization and diagnostic evaluation, but of which our patients would not have been able to complain, such as delirium or coma. Hearing

patients' stories of illness and restructuring them as clinical problems is recommended by experts as an aid to differential diagnosis.²⁸⁻³¹ We identified any clinical problems that patients presented, whether of biological, psychological, or sociological origin.^{31,32}

At the time of each patient's discharge, transfer to another service, death, or at the end of the study months (if patient was still on service), the clinical team met to achieve consensus on and record whether the patient had had any active clinical problems that required diagnostic evaluation, and if so, what the patient's principal clinical problem was. By *principal* we meant the clinical problem that was entirely responsible, or, if there were several clinical problems, was most responsible for the patient's admission to hospital. We included every patient with any active clinical problem, even if the diagnosis seemed obvious. Only patients admitted solely for therapeutic interventions were excluded.

Searching

The first research aim was to measure the proportion of patients on the clinical service who presented with clinical problems for which research evidence was available to guide estimates of pretest probability. To find such evidence, we searched for reports of clinical care research about the frequency of underlying diseases that cause each of these clinical problems. One of us (WSR) carried out these searches by combining several methods. Electronic searches of MEDLINE from 1966 through June 2000 were carried out using text words and MeSH headings for the clinical problems, and using synonyms, explosion, and truncation where appropriate. No search filter or "hedge" has yet been published that combines methodologic terms into a search strategy to maximize the retrieval of high-quality articles of this type.³³ Therefore, we combined the resulting sets from the above clinical terms with a variety of terms we suspected could yield citations of this type of research. For text words, the following terms were used: *cause* or *causes*, *complications*, *consecutive*, *course*, *differential diagnosis*, *etiologies*, *frequency*, *outcomes*, and *presenting*. From the MeSH thesaurus, *Diagnosis*, *Differential*, and *exp Cohort Studies*, along with floating subheadings for complication and diagnosis were used. To maximize the yield, these methodologic terms were combined with a Boolean "OR"; then a Boolean "AND" was used to combine the results with the results from the above clinical terms.

Extensive hand searches were also carried out of the following: the bibliographies of articles retrieved by electronic searches; the bibliographies of review articles and tutorials on each of the clinical problems; and the bibliographies of text chapters on each of the clinical problems that were found in textbooks of internal medicine, primary care and differential diagnosis, including both paper texts and an electronic textbook. The *Cochrane Library* and *Clinical Effectiveness* were not searched because of their focus on reviews of the effectiveness of

therapeutic maneuvers. Similarly, *Best Evidence* was not searched, because this resource did not begin including evidence about differential diagnosis until after the study period.²⁷ Wherever possible, the citations retrieved from hand searches were then found in MEDLINE to retrieve the title and abstract. Searches were not limited by language, although given the databases used, the great majority of citations were in English.

Article Selection

The titles and abstracts of the citations were reviewed for clues to the following: 1) the article was the report of original clinical care research about patients presenting with a specific problem, or was a systematic review of such articles; 2) the article described how the clinical problem was defined; 3) the article described how the patients were collected into the study sample; 4) the article described the diagnostic evaluation employed; and, 5) the article reported the frequency of disease etiologies found underlying the clinical problem. Only articles whose title and abstract persuaded us that the full article would contain all 5 of these elements were counted as qualifying citations. We retrieved for review the full text of 2 qualifying citations for each clinical problem, if 2 could be found. The full text was reviewed for the presence of the 5 elements sought in the titles and abstracts. Articles were considered confirmed if the full text did show all 5 elements. Editorials, commentaries, and narrative reviews that did not present original data, along with reports of single cases were excluded. Letters were included only if they described reports of original research in sufficient detail to meet the 5 criteria.

Article Appraisal

The second research aim was to discern whether any of this evidence was of sufficient quality to guide clinicians' estimates of pretest probability for diagnostic decision making. To examine this question, we retrieved for review the full text for 2 of the qualifying citations for each problem, if 2 could be found. If more than 2 qualifying citations per clinical problem were found, 2 were selected to be retrieved after considering the following: how recently the article was published; how similar the patients were to the team's actual patients; how large the study sample size was; and how feasible retrieval would be.

After these citations were confirmed, they underwent critical appraisal and ranking as to the strength of the evidence. The articles' validity was examined using published guides for appraising articles of this type.²⁰ These guides emphasize 5 aspects: the definition of the clinical problem, the representativeness of the patient sample, the credibility of the diagnostic criteria, the credibility of the diagnostic evaluation, and the completeness of outcomes (diagnostic or follow-up). Before appraising the articles, we had used these 5 attributes of validity to construct a hierarchy of levels of evidence for articles about disease probability for differential diagnosis, shown in Table 1,

designed to be congruent with a published hierarchy for other forms of evidence.³⁴ Also, before these ratings were done, we reckoned that evidence that would be ranked on the hierarchy as level 1, level 2, or level 3 would be sufficiently rigorous to be credible for guiding clinicians' estimates of pretest probability. As each article was appraised critically for these 5 aspects of validity, the overall methodologic strength was summarized by rating it on this hierarchy. We assigned articles to the level that best described the study's methods on the majority (3 or more) of the 5 attributes of validity.

Statistical Analysis

Descriptive methods were used to report the simple proportions of patients who had clinical problems for which qualifying citations could be found. We calculated 95% confidence intervals (95% CIs) for proportions using standard formulae.³⁵

RESULTS

During the 3 months of November, 1999, February, 2000, and March, 2000, our service admitted to the hospital 122 new patients with active clinical problems for which a diagnostic evaluation was carried out. Of the 122 patients, 44 had only one active clinical problem, while 40 had two problems, 23 had three problems, 9 had four problems, 5 had five problems, and 1 had six problems. Table 2 shows the 45 principal clinical problems that were identified in these 122 patients, arranged alphabetically.

Citations that qualified as disease probability research evidence were found for 35 of the 45 principal clinical problems, or 78% (estimated 95% CI, 66% to 90%). Thus, 111 of the 122 patients, or 91% (estimated 95% CI, 86% to 96%), had principal clinical problems for which disease probability evidence was available. Altogether, 730 qualifying citations were found for 35 clinical problems, with a range from 2 to 49 citations per problem and a mean of 20.8 citations per problem (see Table 2). The full list of qualifying citations is available from the authors.

For the 35 clinical problems with qualifying citations, 2 citations for each problem were selected for retrieval, appraisal, and ranking, yielding 70 articles, or nearly 10% of the total. Of these 70, 69, or 98.6% (estimated 95% CI, 96% to 100%) were confirmed upon full-text review. After appraisal of the articles' validity, the 69 confirmed citations were ranked on the hierarchy of levels of evidence (Table 1), as shown in Table 3. Thus, 66 of these 69 confirmed citations, or 95.6% (estimated 95% CI, 91% to 100%), ranked as level 3 or higher, i.e., as being sufficiently rigorous to be credible enough to guide clinicians' estimates of pretest probability.

DISCUSSION

We surveyed our hospital medicine service for 3 months and found that most of the patients admitted for diagnostic

Table 1. Hierarchy of Levels of Evidence for Disease Probability Research

Level of Evidence	Disease Probability
1a	Systematic Review of Level 1 disease probability studies (with homogeneity).
1b	Individual cohort study or consecutive case series with most/all of: <ol style="list-style-type: none"> Clinical problem is clearly defined; Sample patients represent full spectrum of the clinical problem; Diagnostic criteria are explicit, credible; Diagnostic work-up is comprehensive and applied consistently to all patients; Final diagnosis is known in >80%, and status of all undiagnosed patients is known.
1c	"All or none" case series
2a	Systematic Review of Level 2 disease probability studies (with homogeneity).
2b	Individual cohort study or consecutive case series with: <ol style="list-style-type: none"> Clinical problem is well defined Patients show much of clinical spectrum; Diagnostic criteria appear credible; Diagnostic work-up is sensible, applied to most/all; Status (either final diagnosis or outcome) is known in >80%.
3a	Systematic Review of Level 3 disease probability studies (with homogeneity).
3b	Individual cohort study or consecutive case series with: <ol style="list-style-type: none"> Clinical problem is acceptably defined Clinical spectrum is restricted, but not worrisomely so; Diagnostic criteria may be credible; Diagnostic work-up appears acceptable; Status (either final diagnosis or outcome) is known in <80%.
4a	Systematic Review of Level 4 disease probability studies.
4b	Nonconsecutive case series, or with any of: <ol style="list-style-type: none"> Clinical problem poorly defined Spectrum of clinical problem is restricted or inappropriate; Diagnostic criteria are not credible; Diagnostic work-up not thorough or not applied to most; Status (either final diagnosis or outcome) is known in <50%.
5	Expert opinion, without explicit critical appraisal, based on unaided case memory, or on physiology, bench research, or "first principles".

evaluations had clinical problems for which disease probability research evidence was available to guide estimates of pretest probability. To our knowledge, this is the first ever survey of this question. The results of this survey are congruent with the results of prior studies showing that relatively high proportions of therapeutic decisions have good evidence available to inform those decisions.³⁶⁻³⁸

This study has some potential limitations to its validity. First, names were assigned to the patients' clinical problems that seemed most useful diagnostically, with no attempt to use prespecified standardized phraseology or definitions. It is possible that some of the patients' problems were framed imprecisely or in error. However, as can be seen in Table 2, the patients' presenting problems were framed with frequently used terms that are likely to be similar to others' usage, so we suspect that this potential for bias is minimal. Next, it is possible that the literature searches were incomplete, and that evidence about disease probability might have been missed. If this is true, i.e., if more powerful searches that included other databases might yield even more qualifying citations, then the true proportion of patients might be even higher than 91%.

Next, it is possible that research team's judgments about which citations qualified as disease probability evidence were in error, either as "false positives" (including citations that shouldn't have qualified) or "false negatives"

(excluding citations that should have qualified). However, the use of explicit qualifying criteria should have minimized the chance of false positive error. Also, in the 70 articles subjected to full-text review, appraisal, and ranking, only 1, or 1.4% (estimated 95% CI, 0% to 4%) was not confirmed as containing evidence about disease probability. If this low error rate is applied to all the qualifying citations, one can estimate that only 10 of the 730 would be false positives. False-negative errors, if present, would mean that even more disease probability evidence is available than was recognized, such that the true proportion of patients might be higher than 91%.

Next, it is possible that the selection of 2 articles per problem for full-text review was biased. To minimize potential bias, an explicit list of criteria was used to guide the selection. In addition, these criteria were applied before the evidence was critically appraised, so the strength of the evidence was not yet known when these selections were made. It remains possible that some of the qualifying citations not appraised were reports of studies with less methodologic rigor or lower degrees of applicability to practice. The research goal was to determine whether at least some of the evidence was rigorous enough for clinical use, not whether all of it was.

Next, it is possible that the critical appraisal judgments and rankings on the levels of evidence were in error.

Table 2. Principal Clinical Problems of 122 Patients

Clinical Problem	Patients, <i>n</i>	Citations, <i>n</i>
Abdominal pain	9	20
Anemia	2	22
Ascites	2	13
Bradycardia	1	0
Chest pain	6	28
Constipation	1	6
Delirium	3	33
Diarrhea, acute	2	28
Dizziness	3	38
Dysphagia, odynophagia, or dyspepsia	4	12
Dyspnea (without suspected pneumonia)	16	8
Epistaxis	1	8
Facial paralysis	1	13
Facial swelling	2	0
Fever, acute and unspecified	1	0
Fever, postoperative (with wound pain)*	2	2
Fever, unknown origin	1	49
GI bleed, occult	2	12
GI bleed, upper	4	39
Groin drainage	1	0
Hemoptysis	4	31
Hyperkalemia	1	5
Hypernatremia	1	7
Hypertension, resistant	1	19
Hypoglycemia	1	10
Hypotension	1	0
Jaundice	1	7
Leg pain and swelling	6	7
Leg pain, without swelling or ulcer	1	0
Leg or foot ulcer	3	18
Lung mass	1	17
Lymphadenopathy	1	30
Musculoskeletal pain (includes back) [†]	5	27
Olecranon bursitis	1	0
Pleural effusion	1	19
Pneumonia, suspected [‡]	16	27
Raynaud's and arm swelling	1	24
Renal failure, acute	2	48
Seizure	2	37
Skin rash, diffuse	1	0
Skin ulcer of arm	1	0
Syncope	1	32
Urinary incontinence	1	24
Vomiting, recurrent	2	0
Weight loss	1	10
Totals	122	730

* Both were admitted within 1 week of operation.

[†] Most of these patients had musculoskeletal pain in more than one site, including back, arms, legs, and various joints.

[‡] These patients had 2 or more of the following: fever, cough, purulent sputum, dyspnea, and inspiratory chest pain.

The psychometric properties of the appraisal and ranking scales are not known, and this study was not designed to be able to measure them. Since appraisal and ranking involve making judgments without an available reference

standard, independent verification of the ratings cannot be provided. Nevertheless, efforts to minimize error were made, including the employment of a published set of critical appraisal criteria and a prespecified, explicit set of criteria for ranking. We spent considerable time before beginning this study learning how to use these criteria. Thus, we believe the appraisals and rankings were sufficiently robust, given the available methods.

This study also has some potential limitations to its precision and generalizability. First, as on many hospital services, the number of patients admitted to this clinical service for diagnostic evaluation was relatively small, which could have made the 91% proportion imprecise. However, even if the lower bound of the 95% CI (86%) were the correct estimate, this would still represent a substantial proportion of patients for whom evidence about disease probability could be found.

Next, the patients were male U.S. veterans, many of them elderly, who might have clinical problems that differ somewhat from those of other practices. As seen in Table 2, although all the study patients' clinical problems could have occurred in women, clinical problems that would be unique to women, such as breast mass or amenorrhea, were not encountered. The clinical problems shown in Table 2 are commonly seen in primary and secondary care, so the study results may be cautiously generalized to such settings. Other practice settings, particularly specialty referral practices, may be too different for our results to apply.

Despite these potential limitations, this survey suggests that credible research evidence about disease probability is available to guide clinicians' estimates of pretest probability for the clinical problems of a large majority of patients. If confirmed by others, these data suggest that clinicians' pretest probabilities can become evidence based. Clinicians using evidence to guide estimates of pretest

Table 3. Quality of Evidence Found in 70 Studies of Disease Probability

Citations	<i>n</i> (%)
Article not confirmed*	1 (1.4)
Articles confirmed*	69 (98.6)
Of these confirmed 69, quality was ranked as [†]	
Level 1	18 (26)
Level 2	23 (34)
Level 3	25 (36)
Level 4	3 (4)
Level 5	0
Level 1, 2, or 3	66 (95.6)

* An article was confirmed on full-text review if it met these 5 criteria:

1. reported research of patients presenting with a specified problem;
2. described how the clinical problem was defined;
3. described how the patient sample was assembled;
4. described the diagnostic evaluation employed;
5. reported the frequency of disease etiologies diagnosed.

[†] These are the numbers of citations rated at each level of the hierarchy of disease probability evidence shown in Table 1.

probability could avoid premature diagnostic closure by pursuing alternative hypotheses shown to be common in patients with a given clinical problem. At the same time, using such evidence wisely could lead clinicians to be more selective and to avoid "shotgun testing," thereby reducing the overuse of tests and conserving resources. Using evidence-based pretest probabilities could also help reduce the frequency and impact of diagnostic error.^{1,39,40} For clinical teachers, using evidence to guide estimates of pretest probability could change the teaching of differential diagnosis toward being simultaneously more rigorous and more practical. To realize these benefits, further research is needed into how best to search for this evidence, how best to synthesize it, and how best to bring it to the point of care.

The authors thank the many residents and students at the University of Texas Health Sciences Center at San Antonio who worked so hard and learned so cheerfully during the study months.

REFERENCES

1. Knottnerus JA, ed. *The Evidence Base of Clinical Diagnosis*. London: BMJ Books; 2002.
2. Sackett DL. A primer on the precision and accuracy of the clinical examination. *JAMA*. 1992;267:2638-44.
3. Simel DL, Rennie DR. The clinical examination—an agenda to make it more rational. [Editorial] *JAMA*. 1997;277:572-4.
4. Ball CM, Phillips RS, eds. *Evidence-based On-call: Acute Medicine*. Edinburgh: Churchill Livingstone; 2001.
5. Black ER, Bordley DR, Tape TG, Panzer RJ, eds. *Diagnostic Strategies for Common Medical Problems*, 2nd ed. Philadelphia: American College of Physicians; 1999.
6. Irwig L, Tosteson ANA, Gatsonis C, et al. Guidelines for meta-analyses evaluating diagnostic tests. *Ann Intern Med*. 1994;120:667-76.
7. Deeks JJ. Systematic reviews of evaluations of diagnostic and screening tests. In: Egger M, Smith GD, Altman DG, eds. *Systematic Reviews in Health Care: Meta-analysis in Context*. London: BMJ Books, 2001:248-82.
8. Sackett DL, Straus SE, Richardson WS, Rosenberg W, Haynes RB, eds. *Evidence-Based Medicine: How to Practice and Teach EBM*, 2nd ed. Edinburgh Churchill Livingstone; 2000.
9. Guyatt GH, Rennie D, eds. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. Chicago: AMA Press; 2002.
10. Pauker SG, Kassirer JP. The threshold approach to clinical decision making. *N Engl J Med*. 1980;302:1109-17.
11. Gross R. *Making Medical Decisions: An Approach to Clinical Decision Making for Practicing Physicians*. Philadelphia: American College of Physicians; 1999.
12. Dolan JG, Bordley DR, Mushlin AI. An evaluation of clinicians' subjective prior probability estimates. *Med Decis Making*. 1986;6:216-23.
13. Bobbio M, Detrano R, Shandling AH, et al. Clinical assessment of the probability of coronary artery disease: judgmental bias from personal knowledge. *Med Decis Making*. 1992;12:197-203.
14. Bobbio M, Fubini A, Detrano R, et al. Diagnostic accuracy of predicting coronary artery disease related to patients' characteristics. *J Clin Epidemiol*. 1994;47:389-95.
15. Lyman GH, Balducci L. The effect of changing disease risk on clinical reasoning. *J Gen Intern Med*. 1994;9:488-95.
16. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. *Science*. 1974;185:1124-31.
17. Dawson NV, Arkes HR. Systematic errors in medical decision making: judgment limitations. *J Gen Intern Med*. 1987;2:183-7.
18. Bornstein BH, Emler AC. Rationality in medical decision making: a review of the literature on doctors' decision-making biases. *J Eval Clin Pract*. 2001;7:97-107.
19. Redelmeier DA, Ferris LE, Tu JV, Hux JE, Schull MJ. Problems for clinical judgment: introducing cognitive psychology as one more basic science. *Can Med Assoc J*. 2001;164:358-60.
20. Richardson WS, Wilson MC, Guyatt GH, Cook DJ, Nishikawa J, for the Evidence-Based Medicine Working Group. Users' guides to the medical literature. XV. How to use an article about disease probability for differential diagnosis. *JAMA*. 1999;281:1214-9.
21. Kroenke K. Studying symptoms: sampling and measurement issues. *Ann Intern Med*. 2001;134:844-53.
22. Kapoor WN. Evaluation and outcome of patients with syncope. *Medicine (Baltimore)*. 1990;69:160-75.
23. Getchell WS, Larsen GC, Morris CD, McAnulty JH. Epidemiology of syncope in hospitalized patients. *J Gen Intern Med*. 1999;14:677-87.
24. Petersdorf RG, Beeson PB. Fever of unexplained origin: report on 100 cases. *Medicine (Baltimore)*. 1961;40:1-30.
25. Larson EB, Featherstone HJ, Petersdorf RG. Fever of undetermined origin: diagnosis and follow-up of 105 cases. 1970-1980. *Medicine (Baltimore)*. 1982;61:269-92.
26. Richardson WS. Where do pretest probabilities come from? [EBM Note] *Evidence-Based Medicine*. 1999;4:67-68.
27. Richardson WS, Glasziou P, Polashenski WA, Wilson MC. A new arrival—evidence about differential diagnosis [Editorial]. *ACP J Club*. 2000;133:A11-A12.
28. Baroness JA, Carpenter CCJ, eds. *Differential Diagnosis*. Philadelphia: Lea and Febiger; 1994.
29. Higgs J, Jones M, eds. *Clinical Reasoning in the Health Professions*. Oxford: Butterworth-Heinemann; 1995.
30. Glass RD. *Diagnosis: A Brief Introduction*. Melbourne: Oxford University Press; 1996.
31. Cox K. *Doctor and Patient: Exploring Clinical Thinking*. Sydney: UNSW Press; 1999.
32. Engel GL. The need for a new medical model: a challenge for biomedicine. *Science*. 1977;196:129-36.
33. McKibbon KA, Eady A, Marks S. *PDQ Evidence-Based Principles and Practice*. Hamilton, Ontario: Decker; 1999.
34. Ball C, Sackett DL, Phillips B, Haynes RB, Straus SE. Levels of Evidence, Centre for Evidence-Based Medicine. Available at: <http://www.jr2.ox.ac.uk/cebm/docs/levels.html>. Accessed May 30, 2002).
35. Altman DG. Confidence intervals [Appendix]. In: Sackett DL, Straus SE, Richardson WS, Rosenberg WMC, Haynes RB, eds. *Evidence-Based Medicine: How to Practice and Teach EBM*, 2nd ed. New York: Churchill Livingstone; 2000:233-43.
36. Ellis J, Mulligan I, Rowe J, Sackett DL. Inpatient general medicine is evidence based. *Lancet*. 1995;346:407-9.
37. Gill P, Dowell AC, Neal RD, Smith N, Heywood P, Wilson AE. Evidence-based general practice: a retrospective study of interventions in one training practice. *BMJ*. 1996;312:819-21.
38. Booth A. What proportion of health care is evidence based? Resource guide. Available at <http://www.shef.ac.uk/~scharr/ir/percent.html>. Accessed January 16, 2003.
39. Kassirer JP, Kopelman RI. Cognitive errors in diagnosis: instantiation, classification and consequences. *Am J Med*. 1989;86:433-41.
40. Bordage G. Why did I miss the diagnosis? Some cognitive explanations and educational implications. *Acad Med*. 1999;74(10S):138-43.