A reader's guide to the medical literature

A reader's guide to the evaluation of prognostic studies

Carl van Walraven, Paul C Hébert

Summary
In this article, we review guidelines which may be used to evaluate studies documenting prognosis. We describe a clinical problem involving the prognosis of a patient in an intensive care unit. An approach to the literature search is then outlined. The results of the literature search are described and criteria for the appraisal of articles describing prognosis and prognostic factors are discussed using one article as an example.

Keywords: literature evaluation, prognostic studies

Clinical problem

You are a general internist working in a busy mixed medical–surgical intensive care unit at a large community-based hospital. Early in the evening, the anaesthetist brings you a 65-year-old type II diabetic with chronic renal failure who just had an emergency abdominal aortic aneurysm repair because of a suspected aneurysm leakage. The patient was stable during the operation. After assessing and admitting the intubated patient, the nurse informs you that the rest of the patient's family is on their way to the hospital and would like to talk in the next couple of hours regarding the patient's chances of survival. Taking advantage of an almost empty unit with generally stable patients, you head off to the hospital library to research the problem of predicting prognosis in the intensive care unit

Box 1

Department of Medicine and Epidemiology, University of Ottawa, Ottawa, Canada
C van Walraven
PC Hébert

Correspondence to Dr PC Hébert, Department of Medicine, Rm LM14, Ottawa General Hospital, Ottawa, Ontario, Canada K1H 8L6

Dr Hébert is a career scientist with the Ontario Ministry of Health

Accepted 25 August 1995

The ability to predict outcomes of patients with particular diseases has tremendous importance for the clinician, patients and their families, and healthcare administrators. For the clinician, determining a patient's prognosis may aid in therapeutic decision making. A patient with an excellent prognosis may mandate an aggressive therapeutic approach, whereas a dismal prognosis for another patient may make a palliative therapeutic approach more suitable. Staging for many cancers, a form of prognostication, may be the most important factor in making treatment decisions. For example, patients diagnosed with non-small cell lung cancer without evidence of metastasis to the contralateral hilar nodes or beyond are candidates for resective surgery and have a better prognosis than patients with extrapulmonary metastases. Thus, the prognosis of patients may have a major influence on treatment options considered by physicians.

For patients and their families, relevant prognostic information may aid in treatment decisions in light of personal values, especially when radically different treatment options are all acceptable alternatives. In addition, a knowledge of the course of the illness may alleviate anxiety associated with disease, and allow families and patients to come to terms with the eventual outcome.

For healthcare administrators, an accurate appreciation of prognosis for particular diseases may permit comparisons between institutions in order to evaluate hospital-specific quality of care. For instance, if a particular hospital's mortality or length of hospital stay for patients with a specific illness significantly exceeds that of a large cohort study, an administrator may wish to search for explanations for this difference. In addition, accurate prognostication of patients may aid in financial planning and in the allocation of scarce resources.

Approaches to prognosis

There are a number of approaches in determining a patient's prognosis with a particular disease (table 1). Most of us often rely on our clinical experience to predict prognosis. In the clinical problem given in box 1, you may give the family a prediction of outcome based on your experience with patients having similar diseases. However, there are several limitations in using this approach, particularly if one's clinical experience is limited. Patients are extremely variable despite having the same disease. In addition, it is human nature to recall vividly the extreme cases rather than the common, perhaps more representative, cases. Thus, your impression of a patient's prognosis may be different following an atypical case. Finally, many similar cases are required before any confidence may be placed in predictions based on clinical experience alone. For example, assume that you've treated 10 patients with the same diagnosis and only one of them died. How confident are you when advising your patient's family that he had a 10% chance of dying? After calculating the standard error for this proportion (9.5%), multiplying it by 1.96, and then adding and subtracting the product from the original proportion (10%), the calculation provides confidence boundaries with an estimated mortality as low as 0% or as high as 29% (this is the 95% confidence interval). The upper and lower limits of this estimate are so large that the information is not clinically useful. In order to generate confidence intervals of $\pm 2.5\%$ for the previously mentioned 10% mortality, one would need to treat 144 patients with a similar condition.

A second approach would be to consult a local 'expert' or a textbook. This approach is convenient and instructive, but there are also a number of drawbacks (table 1). Clinical judgement and experience are best used in the interpretation of the literature and applying the findings to specific patients. Either approach alone has limitations. Coverage of prognosis in textbooks is often scanty, especially when the clinical scenario is very involved and complex. In our example, the patient suffers from diabetes mellitus and chronic renal failure, and required emergency surgery; these three conditions in combination are all likely

Table 1 Potential sources of information describing the prognosis of patients

Source	Advantage	Disadvantage
Expert	quick and convenient relevant to patient	consultant's experience only may be misleading and biased
Textbook	quick and convenient	limited discussion limited external validity
Background reviews	provide background and references	easily found with literature search selection of primary studies may be biased may be dated and incomplete
Systematic overview	provides background may avoid biased selection of studies answers a specific question systemic approach to evaluating evidence	may not review laboratory research methodology more important than expertise interpretative element may be limited
Meta-analysis	may avoid biased selection of studies answers a specific question using statistical techniques, may provide an answer to a clinical question where individual studies differ	approach is new and limitations not established analysis best used on randomised, controlled, studies
Primary studies	avoids selection bias reader may draw own conclusion	requires some critical appraisal skills may be more time consuming

to result in a worse prognosis. The difficulty is that they are all discussed in isolation in most texts, reviews, or books. Also, textbooks are often out-of-date and may not reflect current practice as they may not have incorporated new evidence, or may even be incorrect. Using cumulative meta-analyses, Antman and colleagues¹ demonstrated that both review articles and textbooks failed to mention or recommend thrombolytics for the treatment of acute myocardial infarction when evidence clearly documented a benefit for the therapy. Even if you find prognostic information for specific situations, it may not be clinically relevant to your patient because the experience described may be from a single, highly specialised centre. Also, many chapters in textbooks have single authors and the prognostic information given may be the author's experience and may not be shared by other 'experts'. Similar limitations are shared by the consultation approach. Many local experts are often authoritative (one who sounds good but has little substance) rather than an authority.² An authority may be defined as someone with sound judgement, who is up-to-date, and not only understands the literature but also its limitations.

Lastly, you may survey the literature to find articles that more accurately deal with the clinical problem at hand. This final approach, combined with critically assessing the articles compiled during a literature search, may seem daunting to some clinicians. In the next sections, we will describe a practical and efficient method of searching for the answer and critically assessing articles found.

Searching for the answer

While speaking with a senior colleague about your clinical conundrum, you are provided with a 1985 article published in *Critical Care Medicine* by Lemeshow *et al*³ which describes a model for predicting mortality in the intensive care unit. Wanting to find a more recent article, you use the hospital library's Medline CD ROM system. First you locate the 1985 article given to you by entering 'S Lemeshow' into the 'search by author' and then '*Critical Care Medicine*' into the 'search by journal'. You then proceed to join the two sets to find the articles which are common to both. Browsing through this set of five articles, you find the 1985 article along with the MeSH headings under which it has been classified. Using two of these headings (**Intensive Care Units** and **Probability**), you search the most recent 1990–94 CD-ROM to find an updated version of this article (the Mortality Probability Model-II, MPM-II) published in November 1993.⁴ Using the 'find similar articles' option, you generate a bibliography of recent articles regarding prognosis in the intensive care unit. Reviewing the abstracts, you feel that the MPM-II is most applicable to your patient. With this method, you used an older, possibly out-of-date article to find a more current study.

Assessing papers addressing prognosis

In studies describing prognosis, a collection of patients with a certain disease be assembled and their outcome over various lengths of time will be described. Often, these papers will attempt to show that certain attributes present in some of these patients modify the patient's risk, so-called prognostic factors.⁵ For example, from the MPM-II database a patient admitted to the intensive care unit with chronic renal failure has a risk of dying during that admission which is 10 times that of someone whose renal function was normal on admission to the unit.

Information regarding prognosis may be obtained from several types of studies. A *case series* is a description of a very select group of patients. Despite having the same disease, the patients may be quite dissimilar, having different disease severity, different stages of disease, and sometimes even different treatments. This type of information may be very useful in rare diseases or as an initial description of a disease or treatment. A *case-control study* is slightly more advanced, in that 'cases' of patients with a certain disease are compared retrospectively to other patients without the disease. Thus, a case series describes, while a case-control study compares. These types of studies document the potential outcomes from a given disease, but do not produce reliable estimates of the frequency with which these outcomes occur or estimates of the magnitude of risk from prognostic factors. This is because many patients with the disease may not be included in the study. Case-control designs are also not useful in very rare diseases or as a preliminary study. In a *cohort study*, a group of patients with the same disease is followed over time and outcomes are documented. If no interventions are performed on the patients, the natural history without treatment may be determined. Finally, the control arm of a *randomised controlled trial* may also describe the natural history of a disease since patients are given either a placebo or standard, accepted therapy. Each study design has advantages and disadvantages, as described in table 2. However, when attempting to establish a patient's prognosis, cohort studies should be considered as providing the most reliable information.

Critical appraisal of prognostic studies

The criteria for the critical appraisal of prognostic studies may be divided into two criteria regarding how the patients were assembled and criteria regarding how the outcomes were described (box 2). These criteria are best applied to cohort designs.

Criteria for assessing a prognostic study
<ul style="list-style-type: none">• was an 'inception cohort' assembled?• was patient surveillance complete and long enough?• were blind and objective outcomes used?• was the outcome adjusted for other factors?

Box 2

Was an 'inception cohort' assembled?
It is essential that all patients in a prognostic study: a) meet a definition of disease which is reproducible, clinically relevant, easy to use, and accurate; b) should be either 'disease free' or at an easily identified point early in the illness; c) and have comparable stages in the natural history of the disease that is being studied. This collection of similar patients is called an inception cohort. The inability to assemble a proper inception cohort may significantly influence the description of prognosis.

Table 2 Types of primary studies used to assess prognosis

Study design	Description	Advantage	Disadvantage
Case report/series	observations of patient or a group with disease	initial description very useful in rare disease inexpensive and quick	very selective prone to bias
Case-control	patients compared to controls	useful in defining risk or prognostic factors very useful in rare disease inexpensive and quick	prone to selection bias limited generalisability
Retrospective cohort study	patients with same disease identified in the past and followed forward in time	describes natural history and incidences of outcomes efficient and inexpensive	prone to losses to follow-up prone to biases very difficult in rare diseases
Prospective cohort study	patients with same disease followed forward in time	describes natural history and incidences of outcomes	time consuming and expensive very difficult in rare diseases
Randomised controlled study	patients in the control group of a randomised controlled study	similar to cohort	similar to cohort

Definitions
<i>Inception cohort</i> : a group of patients who are disease-free or at an easily identified point early in the illness when enrolled in a study
<i>Observational or measurement biases</i> : systematic errors in the assessment of measurements or outcomes
<i>Internal validity</i> : results and conclusions of a study which are free of bias, thus accurately approximating the truth for patients in the study
<i>External validity</i> : results and conclusions of a study which are free of bias and apply to patients outside the study

Box 3

Difficulties with the proper identification of the patient population may lead to a misclassification of patients resulting in misleading results. For example, if a study addressing the prognosis of patients admitted to the intensive care unit with acute renal failure followed all patients admitted with a serum urea of > 20 mmol/l, the study’s conclusions would be weakened since some patients enrolled would not truly have acute renal failure. Elevated serum urea may be secondary to other causes such as dehydration or gastrointestinal bleeding rather than the more ominous diagnosis of acute renal failure. This may result in the study underestimating the true effect of acute renal failure on mortality in the intensive care unit. This example illustrates how inadequate definitions of disease in the inception cohort may lead to unreliable results.

Enrolling patients at different stages of the natural history of their disease can also result in inconsistencies. Imagine a study designed to describe the prognosis of patients admitted to the intensive care unit with an exacerbation of chronic obstructive pulmonary disease (COPD) without documenting the patients’ baseline pulmonary function. Patients in the cohort with mild decreases in FEV₁ would be expected to recover quickly. In contrast, patients with severely limited flow rates, may be expected to have a protracted intensive care unit course. Thus, any conclusions describing COPD patients’ prognosis would depend on the ‘mix’ of patients in this study. Specifically, an overly optimistic prognosis would be noted if a large number of patients with early COPD were studied. Therefore, assembling a group of patients at different stages of disease may result in inappropriate study conclusions.

Inaccurate results may also occur when investigators enroll patients referred from other centres. Since complicated cases are apt to be referred to specialised centres, the description of prognosis arising from these institutions may be gloomier than the prognosis of patients in community hospitals. In critically ill patients, the reverse may also occur since sicker patients in the referring centre may be too ill to transfer. The difficulties noted above are examples of *referral-filter bias*, where selection at each stage of the referral process results in a group of patients that are different from the overall population of patients with a specific disease.⁶ Referral filter may not be misleading if referral patterns are well described and similar to your own. Therefore, not only should all patients in a prognostic paper truly have the same disease and be at an early stage of that disease, but the referral patterns should also be described.

In the MPM-II, the inception cohort was very well defined, enrolling all patients admitted to the intensive care unit with few exceptions. Diagnostic criteria were precise and clinically appropriate. Also, only 10% of the patients were referred from other hospitals, with the bulk of patients admitted from hospital emergency rooms, wards and recovery rooms. Content that the study had a well-defined and appropriate inception cohort, we proceed to assess how patient outcomes were determined.

Was patient surveillance complete and long enough?

Prognostic factors detected in studies may precede the outcomes of interest by varying lengths of time, from days to years. It is important that studies have appropriate follow-up periods so that they may manifest the outcomes of interest. This is especially important for chronic disease where outcomes occur late or when events are rare. In the MPM-II, patients who did not die or get discharged from the hospital, were followed for a minimum of two months after admission. Because of the acuity of the illnesses that are treated in the intensive care unit, the prognostic factors of interest for these studies usually exert their influences quickly. Thus, a follow-up time of two months is appropriate.

In addition to following patients for an adequate length of time, it is also essential that the study’s surveillance of the patients be sufficiently complete. Patients who are not accounted for are said to be ‘lost to follow-up’. Patients are not usually considered lost to follow-up for trivial reasons. The more people who are considered lost to follow-up in a study, the less reliable is the prognosis observed. For example, imagine that a study examining mortality of patients admitted to the intensive care unit with coma enrolled 100 patients and reported a 50% mortality during hospital admission. If the study only determined the outcome of 80 patients, with the other 20 patients being lost to follow-up for various reasons, the true incidence of mortality may be as low as 40% (if all of the 20 ‘lost’ patients lived) or as high as 60% (if all of the ‘lost’ patients died). This example illustrates that results from studies not reporting the number of patients who were lost to follow-up should be questioned. When losses to follow-up are reported, the potential impact on study outcomes should be determined by recalculating mortality rates assuming the ‘lost’ patients either all lived or died (as illustrated above). This is called the best case/worst case scenario and will allow you to determine whether the results of the study should be disregarded.

The MPM-II enrolled every patient admitted to the intensive care unit at various centres during a prespecified time. All patients were included in the analysis except for 1.6% of the original group. Because of this small proportion of losses to follow-up, you believe that study is complete enough to proceed with an assessment of the analysis.

Were blind and objective outcomes used?

Clinical medicine maybe subject to *observational biases*. Inconsistent and possibly biased assessments of outcomes may give misleading results. It is important that all outcomes used in a study are clearly defined, clinically relevant, and reproducible. As an example, a study examining whether stroke patients develop more pneumonias would need to ensure that 'pneumonia' is precisely defined using objective criteria. The diagnosis should be determined without knowledge of the patients' clinical status. Objectivity may be ensured by blinding both the patient and the physician. *Blinding* is important, especially with more subjective outcomes. An investigator's decision about whether or not to label a haziness on a chest X-ray as a 'pneumonia' may be influenced by the knowledge of the patient's clinical status. Additionally, it is desirable that more than one person be used to interpret test results involving significant judgement (*inter-rater reliability*).

In MPM-II, the primary outcome measure was all-cause mortality which is less prone to observational biases. While mortality is very important, other outcomes may also be relevant. One limitation of the PMP-II is the lack of indication of the patients' functional status at the end of the study period.

Was the outcome adjusted for other factors?

Once overall outcomes have been described, it is often important to identify prognostic factors. Before authors claim that certain factors predict outcomes or claim that certain factors are related to outcomes, adjustments in baseline characteristics presently known to be prognostic factors should be made. These differences may influence outcomes more than the prognostic factors proposed by the study. For instance, a study comparing mortality rates following an acute myocardial infarction in teaching and nonteaching hospitals would need to make adjustments in many baseline characteristics such as age, location of the infarct, as well as treatments such as thrombolytics, aspirin, β -blockade of the two populations. Once appropriate adjustments have been made, conclusions may be drawn about the prognosis of patients following an acute myocardial infarction from each category of hospital. Stratification and multivariate statistical methods may be used to make the necessary adjustments.

Multivariate analysis may be used to consider simultaneously the effects of a number of different variables on the outcome of interest. Logistic regression and multiple linear regression are two of the more common examples of multivariate analysis used in research. In prognostic studies, they serve three functions: to determine if potential factors significantly influence the patient's prognosis; to document the strength of the association; and to predict outcomes based on a number of factors simultaneously. While the inner workings of multivariate analyses remain a mystery to most clinicians, Sackett *et al*² remind us to keep two rules of thumb in mind regarding the use of multivariate analysis in prognostic studies. First, whenever an analysis produces a predictive model, it is important that it be tested on a second set of patients to determine the validity of the model outside of the computer or the group of patients with which it was developed. Second, when investigators report on a number of variables in multivariate analysis to decide which potential prognostic factors are important, the study should have at least 10 patients with the outcome of interest for every potential prognostic factor tested.

The MPM-II used logistic regression techniques to establish which clinical variables influenced 30-day mortality. Because more than 12 000 patients were included in the model development, the authors were able to test for many potential prognostic factors. The final model from the first cohort of patients was subsequently validated in a second large cohort.

Do the study results apply to your patient?

The MPM-II is a well-designed, well-described and *internally valid* study of prognosis applicable to a variety of patients admitted to the intensive care unit. Before the study's results are applied to your patient, similarities between the study patients and your own should be assessed. That is, the *external validity* of the study must be established. The closer an average patient in the study resembles your patient, the more confidence you may have that the conclusion of a study are applicable. If differences between an average study patient and your

Key points
<ul style="list-style-type: none">• an understanding of a patient's prognosis may aid the patient and physician in making sound clinical decisions• the most reliable prognostic information is provided by well-conducted prospective cohort studies• the use of an inception cohort with complete follow-up and well-described objective outcomes will provide the most accurate prognosis• a patient should be similar to the average patient in the study to best reflect the individual's prognosis

Box 4

patient are clinically significant, then the study results may not be valid. The same guideline may also apply to patients who are being compared to an average patient in a well-described subgroup of a study.

The MPM-II enrolled a large number of critically ill patients from many types of intensive care units, including mixed medical–surgical units in community hospitals similar to your own. Therefore, you believe the results may be used to establish a prognosis for your patient. From the MPM-II results you conclude that your patient has a 36% predicted mortality. When you meet with the patient's family, you explain that results from large studies suggest that patients similar to their loved one have a 60–70% chance of surviving their hospital stay. You believe that aggressive treatment is still warranted, and that the next few days is the period of highest risk. Even though this study may be superior to clinical experience alone in describing the prognosis of critically ill patients, the results should be used with some degree of caution. For example, predictions from this study alone should not be used in the decision to withdraw care or restrict admission to a special care facility. The predicted outcome from a study may have a significant range of error when applied to individual patients. Predictions for individual patients are most accurate when a very large group of patients either all survive or all die. However, most diseases have a spectrum of outcomes leading to a complex relationship between prognosis and prognostic factors. Thus, clinical judgement should always be exercised when interpreting prognostic information.

1 Antman EM, Lau J, Kupelnick B, *et al.* A comparison of results of meta-analysis of randomized control trials and recommendations of clinical experts. *JAMA* 1992; 268: 240–8.

2 Sackett DL, Haynes RB, Guyatt GH, Tugwell P. *Clinical epidemiology: a basic science for clinical medicine*. Boston: Little, Brown and Company, 1991.

3 Lemeshow S, Teres D, Pastides H, *et al.* A method for predicting survival and mortality of ICU patients using objectively derived weights. *Crit Care Med* 1985; 13: 519–525.

4 Lemeshow S, Klar J, Avrunin JS, Gehlbach SH, Rapoport J. Mortality probability models (MPM-II) based on an international cohort of intensive care unit patients. *JAMA* 1993; 270: 2478–86.

5 Laupacis A, Wells G, Richardson WS, Tugwell P. User's guides to the medical literature: V. How to use an article about prognosis. *JAMA* 1994; 272: 234–7.

6 Sackett DL, Whelan G. Cancer risk in ulcerative colitis: scientific requirements for the study of prognosis (editorial). *Gastroenterology* 1980; 78: 1632.

7 Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada. How to read clinical journals. III. To learn the clinical course and prognosis of disease. *Can Med Assoc J* 1989; 124: 869.

8 Fletcher RH, Fletcher SW, Wagner EH. *Clinical epidemiology – the essentials*. Baltimore: Williams and Wilkins, 1982.