EBM NOTEBOOK

A clinician’s guide to journal articles about prognosis

ACP Journal Club and Evidence-Based Medicine provide summaries of studies on prognosis if they meet the criteria that appear in the Purpose and Procedures section of each issue. This note focuses on how clinicians can evaluate studies of prognosis for 3 key features: validity, importance, and applicability.

Guides have been published that can help the clinician critically appraise and apply evidence about prognosis to patients. (1, 2). We’ll expand upon these guides to decide if an article about prognosis is helpful, using the following patient as an example:

A 32-year-old woman at 21 weeks of gestation has a 2-year history of hypertension and had preeclampsia during a previous pregnancy. She isn’t currently receiving antihypertensive medication. Because of her history, she is concerned about her risk for preeclampsia and perinatal death. Together with the patient, we formulate a clinical question: “In a pregnant patient with chronic hypertension and a history of preeclampsia, what is the current risk for preeclampsia and perinatal death?” Using PubMed’s Advanced MEDLINE Search Internet service (http://www.ncbi.nlm.nih.gov/PubMed/medline.html), we enter the search terms “preeclampsia” and “hypertension” (major MeSH terms) and “risk factors” and find a recent relevant article (3). We appraise it to determine if its findings are valid.

Is this evidence valid? Was a defined, representative sample of patients assembled at a common (usually early) point in the course of their disease?

We use the phrase “usually early,” which implies an inception cohort (a group of people assembled at an early point in their disease), but clinicians may want information about prognosis in later stages of disease. In fact, one can look for a study that assembled patients at any point in their disease. However, if observations are made at different points in the course of disease for various patients in the cohort, the relative timing of outcome events would be impossible to interpret. For example, a recent cohort study of patients with cryptogenic fibrosing alveolitis included patients with this disorder who were already being followed at several hospitals (prevalent cases) as well as patients who were newly diagnosed (incident cases) during a prospective 18-month period of recruitment (4). Median survival for incident cases was 2.9 years; for prevalent cases, it was 9 years. The inclusion of prevalent cases leads to an overestimate of the median survival for patients, with the degree of overestimation dependent on the number of prevalent cases rather than on the natural history of the disease.

Investigators should also describe where study patients were recruited. Patients from tertiary care centres are more likely to have complicated and advanced cases of the target disorder and, therefore, may have much different prognoses than patients recruited from primary care settings.

In the article by Sibai and colleagues (5) we found for our patient, the study group consisted of 774 women with chronic hypertension and singleton pregnancies who were between 13 and 26 weeks of gestation. These women were involved in a randomised controlled trial (RCT) of low-dose aspirin versus placebo for the prevention of preeclampsia (5). They had chronic hypertension of varying duration and severity, and 50.9% were not receiving antihypertensive medication.

Was patient follow-up sufficiently long and complete?

Ideally, investigators should follow all patients until every patient recovers or has one of the other outcomes of interest or until the elapsed time of observation is of clinical interest to clinicians or patients. If follow-up is short, it may be that too few study patients will have the outcome of interest, and little information will be obtained to help us advise our patient. In the study by Sibai and colleagues, the women were appropriately followed until the end of their pregnancy, and the investigators were able to achieve a follow-up of almost 99% (763/774).

The fewer patients available for follow-up, the less accurate the estimate of the risk for the outcome. Losses to follow-up are sometimes unavoidable and unrelated to prognosis; and if the baseline demographics of these patients are similar to those followed up, we may feel reassured that the study results are robust. However, losses may occur with patients who are too ill (or too well) to be followed or who have died, and the failure to document these losses will threaten the validity of the study.

To assist us in judging whether follow-up is sufficient, we can consider the best- and worst-case scenarios. Consider a study that followed a group of 100 people for the occurrence of an event (e.g., death). 5 participants die, and 25 are lost to follow-up during the course of the study. The crude risk for death is 5/100, or 5%. The best-case scenario assumes that none of the 25 participants who were lost to follow-up may have died. If we consider the worst-case scenario, we assume that all 25 persons who were lost to follow-up may have died. If we consider the worst-case scenario, we assume that all 25 people died, and the risk for death is 25/100, or 30%. The best-case scenario assumes that none of the 25 people died; the risk for death therefore would be 5/100, or 5%. The best case of 5% doesn’t differ much from the observed 6.7%, but the worst case of 30% differs substantially. We would conclude that the follow-up was not sufficient in the latter case. As a quick check, follow-up of <80% is unacceptable, but doing a “sensitivity analysis” just described can help you to define the range of outcomes of patients lost to follow-up.

Were outcome criteria applied in a blinded fashion?

We would like to see that the investigators defined explicit criteria for each outcome of interest and described whether they applied these criteria without knowledge of the risk factor.
In Sibai and colleagues' study, women with a previous history of preeclampsia had an increased risk for this condition with their current pregnancy. We might ask that, since these data arose from a study of low-dose aspirin for the prevention of preeclampsia, did the use of aspirin have an effect on the risk for preeclampsia, and, if so, was this adjusted for? The authors tell us, however, that the risk for preeclampsia was the same in both the aspirin and placebo groups. Women with preeclampsia did have an increased risk for perinatal death, and this difference remained significant after adjustment for several potential prognostic factors (including maternal age, previous history of preeclampsia, duration of hypertension and use of antihypertensive therapy).

If a prognostic study identifies a prognostic factor for the first time, it could be the result of a chance difference in its distribution between patients with different prognoses. The initial patient group in which it was identified as a prognostic factor is called a training set. If investigators search for multiple potential prognostic factors in the same data set, it is likely that a few would be found on the basis of chance alone. Therefore, ideally, we should look to see if it is confirmed as a prognostic factor in a second, independent patient group called a test set. But this method isn't often used in prognostic studies. *ACP Journal Club* and *Evidence-Based Medicine* do not require it for inclusion as a prognostic study unless the authors are proposing their findings as a clinical prediction guide for singling out subgroups of patients who should be treated differently. We consider the presence of a test set to be a minor criterion for assessing validity, and we do not believe that its absence in the current study threatens the study's validity. In summarising our appraisal of Sibai and colleagues' study, we were sufficiently satisfied with the validity of the results to now consider whether they are important.

**Is this valid evidence important?**

**How likely are the outcomes over time?**

Our patient is interested in her chance of developing preeclampsia, given that she had it with her first pregnancy. From the study we found, the best estimate is that she has a 32% risk for preeclampsia (compared with a 23% risk without a history of preeclampsia). She was also interested in the risk for perinatal death if she develops preeclampsia, and we can tell her that this risk is doubled (from 4% for women who do not have preeclampsia to 8% for those who do).

**How precise are the prognostic estimates?**

In addition to determining the magnitude of the risk, we need to look at the precision of the estimate, which is best done by looking at its 95% confidence interval (CI). If survival over time is the outcome we and our patient are interested in, then data on earlier follow-up periods should be examined because they usually include results from more patients than do later periods. Therefore, survival curves are more precise (i.e., they have a narrower CI) earlier in follow-up.

In the study by Sibai and colleagues, 32% of 181 women with previous preeclampsia developed the condition with their current pregnancy. We can use these 2 numbers to calculate the 95% CI around our risk estimate: 25% to 38% (2). This CI is wider than that for women without a history of preeclampsia because more women were in the latter group (risk estimate 23%, CI 20% to 26%). The width of the CI helps us decide whether the results of the study are important. For example, if our study only had 10 women with a history of preeclampsia and 32% of these developed preeclampsia with their current pregnancy, the CI would be 15% to 49%—a result that is too wide to be useful to us or our patient.
Can we apply this valid and important evidence about prognosis to our patient's care? Is our patient so different from the study patients that we cannot use the study results? Rather than considering whether the study patients are similar to our own patient (generalisability of the study), we suggest considering whether our patient is so different from those in the study that it isn’t helpful to us (applicability). For most differences, the answer to this question will be “no,” and we will then need to consider what prognostic factors our patient has that might alter the estimate of her prognosis. Our patient is sufficiently similar to those included in the study (she has chronic hypertension and a history of preeclampsia) that we decide that the results of the study can be applied to her.

Will this evidence make a clinically important impact on our conclusions about what to tell or offer our patient? In deciding whether we can apply the results of this study to our patient, we need to consider whether its results lead us to select or avoid a particular course of action. If, for example, a study demonstrated a good prognosis for patients with a particular target disorder who didn’t receive treatment, we would probably not recommend therapy in our discussions with the patient. In this case, the high rate of preeclampsia in patients like ours leads us to decide (with our patient) on a policy of close monitoring during her pregnancy for signs of preeclampsia.

ACP Journal Club and Evidence-Based Medicine, and their electronic database, Best Evidence, provide prescreening of articles on prognosis for assembly of an inception cohort and at least 80% follow-up, as well as providing sufficient details of the studies that meet these criteria for readers to complete the appraisal task efficiently. If you are looking for original articles that have not been screened for methods, the guides outlined in this note will identify the studies worthy of attention.

Sharon E. Straus, MD
Finlay A. McAlister, MD

References