Caring for Patients at the End of Life

Accurate Prognostications of Death
Opportunities and Challenges for Clinicians

JOANNE LYNN, MD, *and* JOAN M. TENO, MD, *Washington, DC;*
and FRANK E. HARRELL Jr, PhD, *Durham, North Carolina*

Linking survival time to an array of prognostic variables through a powerful statistical model can provide reliable, valid, and potentially useful information for patient care. We present a summary of the recently developed SUPPORT model [Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatment] for estimating survival time of seriously ill adult inpatients and illustrate the possible clinical use of such a model. This model is then translated into counseling. Clinicians are positioned to evaluate the relevance and validity of any model and to understand their persistent shortcomings.

(Lynn J, Teno JM, Harrell FE Jr: Accurate prognostications of death—Opportunities and challenges for clinicians, *In* Caring for Patients at the End of Life [Special Issue]. West J Med 1995; 163:250-257)

What tormented Ivan Illych most was the deception, the lie, which for some reason they all accepted, that he was not dying but was simply ill, and that he only need keep quiet and undergo a treatment and then something very good would result.

The Death of Ivan Illych
LEO TOLSTOY, 1886[1]

"Doc, if I am dying, don't use those machines on me!" This admonition is repeatedly heard by physicians. How are we to know that someone "is dying?" Most physicians are willing to forgo useless medical interventions when a patient is near death or to shape a plan of care so that it reflects an unavoidably grim prognosis. To accomplish that, physicians need a number of skills and tools, including accurate ways of estimating survival prospects. In this article we review current efforts in that regard and evaluate the challenges and opportunities they present to practitioners and patients.

Only a few decades ago, all physicians had to offer for prognostication was descriptions of the survival experience of a large group of persons who were defined by one characteristic: perhaps those newly diagnosed with a deadly illness or those who had reached a certain stage of such an illness. Such groups included a wide variety of patients, usually with no systematic way to sort them out. With the recent introduction of better computing and statistical tools, dramatically improved objective estimates of prognosis have become available. In general, objective estimates of prognosis rely on simultaneously modeling the relation of the risk of death with each of a number of patient and treatment characteristics. To be accurate, the modeling draws on the experience of many patients, usually numbering in the thousands. These models are proving to be accurate and stable across time and institutions.[2-6] Outcomes for groups of patients are used routinely in developing practice guidelines, comparing care systems, and describing physician practices. They are not often standard elements in the care of individual patients, however. Just as for any new drug or laboratory test, their usefulness should be critically evaluated, and practitioners need to understand what can be achieved with objective estimates of prognosis and what cannot.

Good decision making with patients relies on understanding a patient's likely outcomes with the alternative care plans that could be implemented and the patient's preferences among them.[7-13] Objective estimates of survival might prove helpful in understanding the likely outcomes. Some people are suspicious of computer-based estimates of survival solely from an unreflective distaste for impersonal machines taking on what seems to be an important task. An editor at *The Washington Post* called one system a "life and death" computer, giving it a sinister image ("The Life and Death Computer," January 5, 1992, C6). Physicians should set aside that image. The computer is only a tool for rapid information retrieval and analysis. The experience of previous patients is all that we ever have to draw on in forecasting the course of a current patient. Individual physicians are susceptible to a number of errors in estimating prognosis: bias from recent experience, overestimating or

From the Center to Improve Care of the Dying, George Washington University School of Medicine and Health Sciences, Washington, DC (Drs Lynn and Teno); and the Division of Biometry, Duke University Medical Center, Durham, North Carolina (Dr Harrell).

This project was supported in part by the Program on the Care of Critically Ill Hospitalized Adults, the Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatment (SUPPORT), the Robert Wood Johnson Foundation, and for Dr Harrell, grant No. HS07137 from the Agency for Health Care Policy and Research. The views expressed here are those of the authors and do not necessarily reflect those of the funders.

Reprint requests to Joanne Lynn, MD, Center to Improve Care of the Dying, George Washington University Medical Center, 1001 22nd St, NW, Ste 700, Washington, DC 20037.

ABBREVIATIONS USED IN TEXT

COPD = chronic obstructive pulmonary disease
SUPPORT = Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatment

underestimating the importance of a variable, or self-fulfilling prophesying, for example. A statistical model that relies on computer analysis can often weigh the elements more accurately and usefully than projections that rely on the average experience of mixed groups or clinicians' personal experience. If having better estimates proves useful, physicians should be as willing to use them as to use a better laboratory test or a more effective x-ray method.

What Can Be Accomplished With Objective Estimates of Survival?

The best that can be done with high-quality measures on a large population will not reveal the day that a person will die, and it probably will not limit the range of days to the next few, except in cases of persons barely surviving. This is because the pace of dying is affected by many factors over a substantial period of time. Prognosticating will always be inescapably difficult. Dying can be seen as a walk on a long tightrope. Measures of a walker's skill (the illness and the body's response), the gustiness of the wind (the rate of "external" events such as pneumonias and falls), and the nature of the assistance available (medical, nursing, and family care) will allow a prediction of how long the walk will go. But any such prediction is bound to be expressed in probabilities, as if the tightrope walker could start out many times, sketching out a curve of "time to fall" (time to death). An accurate prediction will entail expressing it as probabilities of attaining each successive meter on the rope (each successive time period). The likelihood of making it to each point is important and so is the variability around the estimate (a confidence interval around the survival curve).

Prognosticating those who are virtually certain to die or to live for the time interval of interest is generally obvious to a physician. The patients whose futures are difficult to discern are going to have less extreme prognoses. Models of prognosis are most important when the optimal care of a patient requires knowing whether survival likely lies above or below a treatment threshold. The same considerations apply as in evaluating the merits of a diagnostic test. The test is important only if its result would make it clear that a patient is beyond some individually determined treatment threshold. In the case of dying persons, the "treatment threshold" may be the survival rate at which a patient finds it worthwhile to endure more life-extending therapy or at which it becomes obvious that a patient's concern for spiritual solace has become paramount. We have shown, for example, that patients with a less than 1% likelihood of surviving for two months often have ventilators and other life-sustaining treatment withheld, indicating that this is often below the clinical threshold now in use for continuing aggressive care.[14]

Careful readers will have noted that we have defined a "treatment threshold" that grows from a patient's clinical situation and preferences. This is different from the use of statistical cutoff points below which the use of certain treatments is to be barred to constrain expenditures. The latter application of prognostic statistics is, and should be, controversial. Sustaining such a fixed cutoff point would require broad consensus on the trade-off between public welfare and individual loss. At the very least, statistical prognostication should be applied as fixed cutoff points only when the issue has been publicly disclosed and debated and when patients and families have had reasonable notice. This is too important a matter to be decided arbitrarily or secretly by insurers or managers, or even by physicians. Furthermore, the more acceptable a cutoff point may be, the less effect it will likely have on costs,[14] in part because current prognostication reflects current clinical care patterns that include patient-centered thresholds. In fact, if we arbitrarily cut off treatment for certain patients with bad prognoses, that change will alter the statistical prognostication itself by making bad prognoses worse from lack of treatment.

A prediction of survival can be expressed at least two ways: as a probability of surviving a certain period of time or as a median survival time. What is possible to say is that, for 100 people who are exactly like this person insofar as measured by the elements that are included in the model, the number expected to be alive at the target time is n, or the average or median survival time is t days. Each of these can be bounded by confidence intervals that reflect both the amount of evidence used to develop the prediction (sample size and follow-up duration) and the unexplained variability in outcome from patient to patient.

Practitioners must be familiar with the interplay between the prognosis for survival to a given time and the prognosis for median time to death. Sometimes a prognosis sounds almost encouraging when stated: "The patient has a 1% chance of making it for two months." The same patient would have a median survival time of one day.[14] Table 1 gives the equivalent statements for patients in the SUPPORT project (Study

TABLE 1.—*Relation of 2-Month Prognoses and Median Expected Date of Death for Patients With Acute Respiratory Failure or Multiple Organ System Failure With Sepsis in SUPPORT (n = 3,515)**

Prognosis at 2 Months, % Survival	*Median Days Until Death*
90	>4.6 yr
50	60
20	13
10	9
1	1

SUPPORT = Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatment

*Maximum follow-up, 4.6 years.

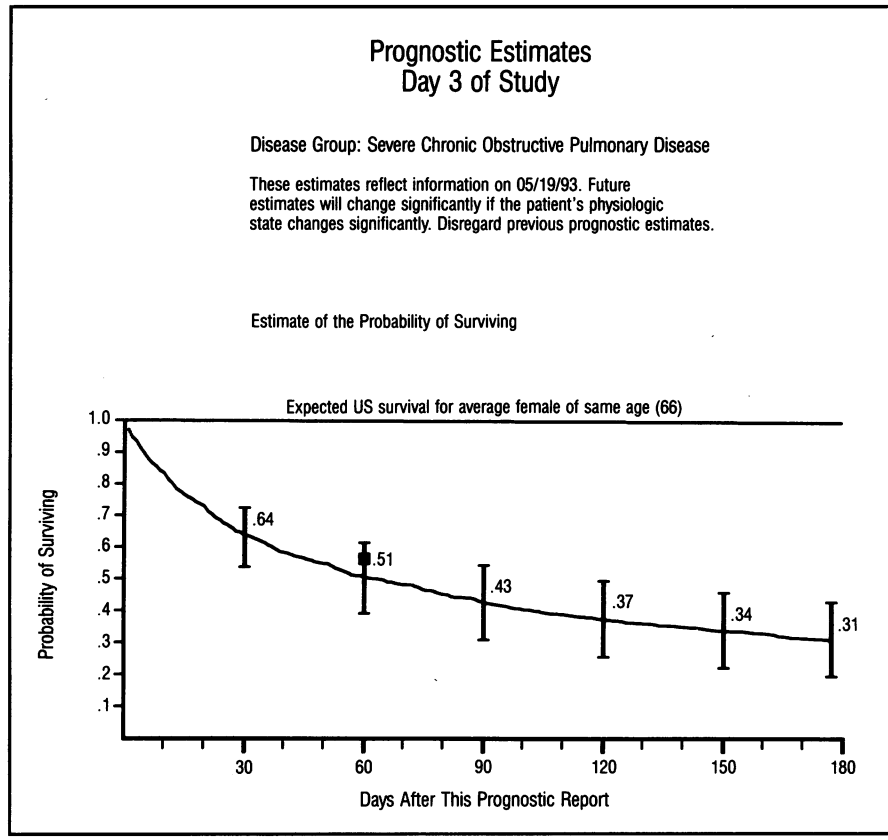

Figure 1.—This example of a prognostic estimate on the 3rd hospital day for a SUPPORT patient with severe chronic obstructive pulmonary disease shows the likelihood of survival on every day through 6 months, along with the 95% variance bounds and the effect of including an attending physician's prognosis as a data element. SUPPORT = Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatment, ■ = model estimate enhanced by physician's estimate is .57 (95% confidence interval, .46 and .66)

to Understand Prognoses and Preferences for Outcomes and Risks of Treatment)[6] who were enrolled with severe acute respiratory failure or multiple organ system failure with sepsis.

An optimal model requires including measures of variables that might affect survival, usually as drawn from clinical experience and preliminary research. These variables should be specified in advance of data collection and should have a discernible path by which to affect survival. An optimal model will be one generated from the population in which it will be applied. Applying models farther afield will always entail some uncertainties. We will return to an explication of how to assess prognostication models. A characterization of one such model will serve as an introduction.

The SUPPORT Survival Time Model

In the SUPPORT project, we developed a prognostic model (Figure 1) that predicts survival time of patients identified in hospitals with a defined level of one of these serious illnesses: acute respiratory failure, multiple organ system failure with sepsis or malignancy, coma, chronic obstructive pulmonary disease (COPD), congestive heart failure, cirrhosis, colon cancer, and lung cancer. The steepness of the curve is determined by the class of disease (chronic, acute, or malignant cancer). The overall placement of the curve is determined by the severity of the disease, as expressed in the 14 variables that are measured to generate a curve like this for each patient.

The variables included in the final model were the diagnosis, serum sodium level, temperature, respiratory rate, heart rate, oxygenation, creatinine level, mean blood pressure, bilirubin and albumin levels, Glasgow coma score, age, days in hospital before becoming eligible for SUPPORT, and having cancer as a co-morbidity. The model uses the most abnormal measured value during a given day. These variables are all available in the medical record and are commonly measured in hospital patients. The model generates a point estimate of the chances of surviving to each day and an estimate of the variation around that point (which is a function of the number of patients and the accord that they evidence in regard to the optimal weighting of each element). This model generates a unique curve for each patient on each reporting day, resting on that patient's specific vital signs, laboratory values, diagnoses, and experience.

Variables were allowed to take nonlinear forms and were entirely prespecified as being likely to have importance in prognosticating survival time. As distinct from clinicians, the model is not swayed by recent experience nor biased by clinical "rules of thumb." Instead, it is fitted directly from the data.

Clinicians might still find it less than satisfying. Certainly, patients' physicians know more than 14 things about their patients, and some of them are important in prognosis. Likewise, physicians know more about what treatment will be used, whereas the model must expect "usual" treatment.

In SUPPORT, the model was generally as accurate as were attending physicians, making fewer extreme errors but being a little less discriminating (see below for an explanation of these terms).[6] Most important, we found that adding the objective model to a physician's own estimate actually made a model that was much better than either alone. The statistical model builds on a database that is far larger than any one clinician's experience, and it allows no psychological errors. The physician knows more about any one patient, however, including some elements of the treatment approach that might affect prognosis. Of course, physicians are sometimes biased in various ways, including enthusiasm for new treatments. The model that predicts outcome most accurately is one that uses both physiology and physicians' estimates, and the best use of any model requires physicians' interpretations of the model's estimate.

Clinical Counseling

Obviously, the counseling of particular patients must take into account a host of emotional and social matters. Nevertheless, we will articulate here just the components of that counseling that arise from estimates of likely survival for a patient with severe COPD who is debating whether to forgo any future use of a mechanical ventilator. We do not put this forward as an optimal format, but only as legitimate and possibly useful elements. Any real discussion would have much more input from the patient, and the physician's words would be responsive to the patient's language, concerns, and interests. The following monologue would never properly be given in this format, but it illustrates the possible content.

Example

Mr Jones, you have been asking to know more about how the last stages of this disease are likely to unfold. I know you don't want to go back on a breathing machine if you are not likely to live long anyway. I can tell you how other patients with your disease have fared. You'll have to help me decide what that means for how we should handle the situation when you again become short of breath.

If we had 100 patients exactly like you, with your disease and age and all the same lab values, you should know that half would have died by the end of two months. However, there would be 10 who were still alive at six months, and 1 would make it a full year. On the other hand, 10 of our original 100 would have died in the next ten days.

Now, our best information is based on careful study of about 500 people with very bad lung disease, all treated at teaching hospitals. We could be more confident in these estimates with more study, of course, but I don't think we'll find that the time it takes to lose half of our original 100 will be in error by more than two weeks.

Evaluating a Model to Predict Survival and Applying It to Individual Patients

Using objective estimates for patient care raises different problems and possibilities than using them for research (such as to correct for disease severity in comparing groups) or for public policy (such as to monitor effects of a change in policy on survival). These latter uses are well established and rely on the model performing well in most cases and similarly in comparison groups. Its use in clinical medicine requires that the model perform well for particular patients. The latter use requires a generally more exacting standard and thoughtful clinicians.

Of course, clinicians need not stand alone in this evaluation. The first two questions below will usually require some evaluation of individual patients, but the last three are more generic. These might well be evaluated by peers in professional publications or by professional societies, such as the Clinical Efficiency and Assessment Program of the American College of Physicians, or by more local or regional practice groups.

The following five elements are central to determining whether a model developed to predict survival is applicable to a particular patient situation:

- Were the patients used to develop the model similar to the current patient?
- Was the survival end point the one that matters to this patient?
- Were the predictor variables reasonable, appropriate in number, measured reliably, handled well, and characterized adequately?
- How was accuracy of the model quantified?
- Has the model been validated in an unbiased fashion?

Were the Patients Used to Develop the Model Similar to the Current Patient?

Applying predictive models to a new patient is safest when the patient is similar to the population used to make the model. The current patient is ordinarily separated from the latter at least by the passage of time. The patient also is usually not in the same care system, not in academic hospitals, or not of the same ethnic and genetic heritage. If the modeling stratifies patients into large groups (those with a certain extent of cancer), the current patient might be pushed into a group from which he or she is dissimilar because there were few patients with the particular combination that this patient exhibits. If the modeling comes from a randomized clinical trial, the entry criteria are often restrictive (many actual patients would have been excluded because they would not have consented to randomization). Some models have been tested or developed in enough environments that the

variation induced by disparate institutions is described, but that is uncommon.[4,15,16] More often, the practitioner will have to reflect thoughtfully on the likely differences between the population used to generate the prediction model and the current patient, considering both whether the disparities are so substantial that the model is irrelevant and, if not, what sort of corrections might be appropriate to apply if it is to be used.

Was the Survival End Point the One That Matters to This Patient?

If the clinical situation turns on survival for two months, a model to predict survival to hospital discharge may not answer the need. If the clinical situation turns on getting home, a model to predict survival for six months may be rather imprecise. The most useful models are those that predict survival over time or to the end points that a patient needs.

The outcomes of relevance need to be measured exceedingly accurately. A model that is constructed with a large number of persons lost to follow-up has serious risks of being unreliable, especially if the reasons for loss are not known. With the development of the National Death Index in the United States, this problem should be confined to the rare populations for whom personal identifiers are themselves difficult to obtain.

Were the Predictor Variables Reasonable, Appropriate in Number, Measured Reliably, Handled Well, and Characterized Adequately?

The most important decisions in formulating a prognostic model are selecting and measuring the predictor variables. First, which variables are tested as predictors of survival is critical in model development. Variables should reflect what is known of the cause and the course of an illness. In comparison with the array that a clinician thinks is likely to be important, the model may report a limited list. This often arises in analyses of data that were collected for some other purpose and that simply did not collect some key predictor variables. A model that predicts mortality from previous hospital utilization and "diagnosis-related group" for billing is obviously likely to perform less well for many clinical situations when compared with a model that also includes measures of physiologic function and disease severity. Sometimes variables indicating chronic physiologic dysfunction or deficient nutritional reserve (such as serum albumin level) are extremely important prognostic factors. Practitioners should again consider what elements of a clinical situation seem likely to be relevant to prognosticating survival and should be suspicious of a model-building process that did not test all of those elements.

Second, clinicians should ask whether the research that generated the model included too few patients to test the number of variables. Testing too many variables results in "overfitting," a situation in which the model works well only for the particular group that generated it. As an approximate rule of thumb, practitioners should be suspicious of any model that had fewer than 10 to 15 dead patients for every variable that is tested (counting nonlinear and interaction terms and counting every variable tested, even if it is not in the final model).[17] If a model is built on too small a database, clinicians cannot know what elements will be misleading.

Third, data on a large number of patients are likely to be missing. In retrospective studies, information is limited to what was written down at the time, and a missing laboratory value or vital sign is irreplaceable. Even in prospective studies, costs and ethical considerations may well limit how complete the data can be. How the missing data are treated can be important. Usually missing data are presumed to be normal, but this is potentially misleading because many patients who do not get those items measured would have evidenced substantial abnormalities. Sometimes missing data can be more accurately input from other variables, or a measure from a different respondent or test can be substituted. Clinicians should look for and evaluate a model's handling of missing data.

Finally, sometimes a predictor variable is used that is not readily measured in a particular patient. If it is important in the model and it cannot be estimated from measurements that are possible to do, then the model is not helpful to a particular patient's situation.

In sum, clinicians proposing to use a predictive model will need to see that the variables being used to predict are a reasonable array of the elements that should have a role, that they are measured well and handled well when missing, that the number of patients needed to develop the model was appropriate to support the model's complexity, and that the measures are ones that can be replicated in the current patient.

How Was Accuracy of the Model Quantified?

A model to predict survival has two major dimensions of accuracy.[18] First, do real patients have the expected survival experience? Second, does the model separate patients along a continuum from those dying quickly to those surviving a long time, or those patients having an event versus those not having that event? The first is termed "calibration" and the second "discrimination." A model should be well calibrated: of 100 patients (P) at any estimated likelihood of survival, $100 \times P$ would be expected to survive. This is ordinarily illustrated by dividing patients into groups along the range of predicted survival and plotting the proportion who survived within each group against the mean predicted outcome in that group (Figure 2). This curve not only shows the general match of prediction and performance but also the ranges of probability in which the model performs less well. Again, practitioners should pay attention to a model's overall calibration and whether a current patient is in a range in which performance is regularly biased.

A model's discrimination is its ability to separate those who die (or die early) from those who survive (or live a longer time). A general index of discrimination is

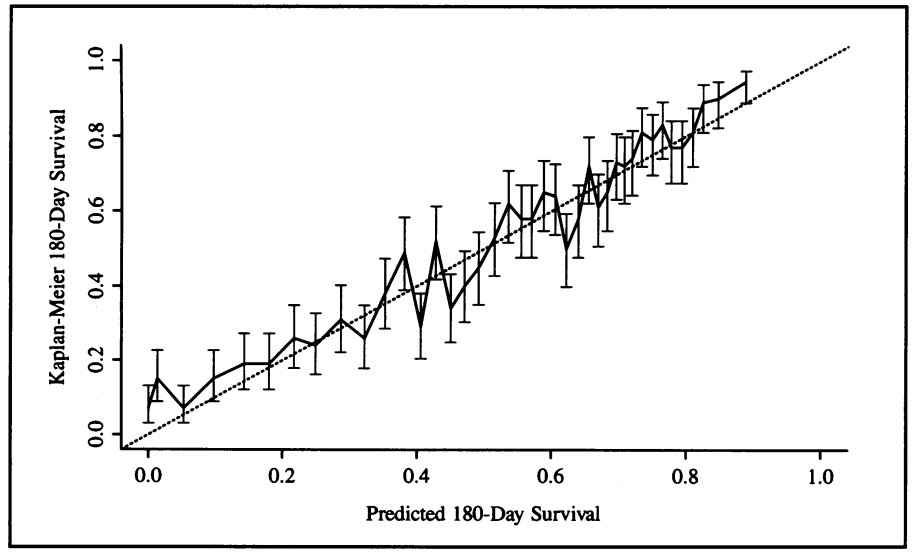

Figure 2.—The calibration curve for the SUPPORT [Study to Understand Prognoses and Preferences for Outcomes and Risks of Treatment] prognostic model shows a close fit across range of prognoses (n = 4,028; 1,889 died) (from Knaus et al,[6] reprinted with permission).

the *c* index, where *c* stands for concordance. All possible pairs of patients for whom survival times can be ordered are examined, and the proportion of such pairs for which the predicted survival and observed survival are concordant (that is, in the same direction) is computed.[18] The *c* index can be easily converted to a Somers' D rank correlation index that quantifies the association between predicted and observed survival, using the equation $D = 2(c - 0.5)$. When the outcome is dichotomous (such as hospital death), the *c* index is the same as the area under a "receiver operating characteristic curve."[19] A model that has no discrimination ability has $c = 0.5$ and $D = 0$, whereas a model that perfectly separated high- and low-risk patients has $c = D = 1$. A model is typically deemed to be clinically useful if *c* is 0.8 or higher ($D \geq 0.6$) for binary outcomes. For continuous outcomes, lower indices are still worthwhile.

The difficulty in assessing a model's predictive accuracy is to avoid allowing overfitting to account for the model's performance (with respect to both calibration and discrimination). This is also addressed under the discussion of whether a model has been validated in an unbiased way.

For some clinical uses, the measure of importance is not the generalized performance of the model but its performance in regard to a specific threshold. If the threshold for subjecting a patient to an additional onerous but possibly valuable treatment is that the patient be more than 50% likely to be alive in two months, then the clinician needs to understand the model's discrimination around that threshold. If the threshold for stopping a ventilator is that the patient has no more than a 1-in-20 chance of surviving for a month, then what matters is the model's ability to discriminate at that end point, rather than its generalized performance. Such specific measures are only sometimes available in the literature but often could be generated from the original data or approximated from what is published.

Has the Model Been Validated in an Unbiased Way?

The most authoritative way to prove merit in a model is that it maintains its calibration and discrimination when it is applied to a different population that is similar to the one that generated the model. This test simultaneously accounts for problems with data and overfitting and enhances a practitioner's confidence that the model can be used in settings other than the original.

In the medical literature, the accuracy of a model is often assessed on the same set of patients who were used to find the important variables and to estimate the regression coefficients (the weights to be assigned to each predictor variable). This results in an inflated estimate of accuracy because of overfitting: fitting spurious associations in the data that are not likely to be replicated in future data. In extreme cases, a model with nine predictor variables measured in ten patients will perfectly predict the outcomes of those ten patients, no matter which predictors are used.

The first method of validation that a clinician should look for is an internal validation that corrects for overfitting. The most common approach is cross-validation, but bootstrapping has been shown to be much more precise. In bootstrapping, a large number of samples are taken (with replacement) from the original patients. For each of these bootstrap samples, a model is fitted using the same strategy that was used to fit the "final" model as reported. The resulting new model is then evaluated in both the bootstrap sample and in the original patient sample and the average disparities used to adjust the estimates of accuracy in the "final" model.[20,21]

Although internal validation shows that predictive accuracy does not come from overfitting, being sure that the model can be applied to new populations requires external validation. By estimating the calibration and discrimination accuracy of the model on a new patient series from another time or geographical location, a check is obtained on the entire process of variable definitions, measurement methods, patient entry criteria, and the fit of the model. A well-constructed model should lose no more than a few percentage points in its *c* index or Somers' D when it is applied to a new group. Losing much more should be a warning that the original model may be too closely tied to the original population, variable definitions, or measurement methods to be used in other settings.

Persisting Limitations on Objective Estimates of Survival

The most striking limitation is that existing models do not often account for variations in treatment. Instead, patients are assumed to be treated "in a typical way," and neither actual treatment nor choices about possible treatment (such as decisions to order no resuscitation) are entered into the formula. This is not because these decisions are thought to have no effect. Rather, it is difficult to discern how to include as predictive elements those aspects of care that regularly happen after a prediction is made. What might be wanted is to have two predictions: one for patients with a hospice-style supportive care plan and one for patients with a fully aggressive, life-extending approach on the first hospital day, for example. But these are changeable behaviors, rather unlike the immutability of age. Behaviors might well be changed after receiving an estimate of prognosis, and those changes in behaviors might alter the estimate itself. This possible "feedback" of treatment choices on estimates and predictor variables has been evaded in prognostic models to date by excluding elements that are clearly under human control. Nevertheless, the treatment plan probably does make a difference, at least sometimes. Practitioners will have to bear in mind that prognostic models to date reflect current practices, whatever they are, and a patient who is pursuing more, or less, aggressive care may have a different survival likelihood. If current treatment plans are biased to the disadvantage of some group (for instance, on the basis of age), that bias will affect predictions of survival as if the alterations resulted from physical processes rather than behavior.

Prognostic models are obviously more limited in the diversity of predictors than is actual clinical practice. Diagnostic labels reflect a major oversimplification, for example. A patient who has an unusual case by virtue of its cause or course will likely have a different experience than the usual person with this diagnosis, which is all that the model can estimate. We can, of course, develop gradually more precise categories and estimators, but there will always be a limit and thereby the need for physicians to note that a given patient's course is expected to deviate from the usual.

More troubling to clinicians is that many prognoses will forever be "intermediate," in the sense that the prognosis is neither so bad nor so good that it makes a decision obvious. We may succeed in gradually finding ways of enhancing discrimination of our models, but the natural variability of humans suggests that there will be a limit and that many people will be in a muddled middle ground of prognosis at a time when clarity in either direction would facilitate decision making substantially. Rather than being a failing of the modeling, this is a reflection of the finitude of what is possible. Once a model is constructed, though, application is so inexpensive that the marginal contribution of the information can be small in many cases and still be worth generating.

Whereas prognostic models are gradually more capably assembled and more adequately validated and reported, few have had their use evaluated in practice. Until that is done, we have to acknowledge that it is uncertain whether better prognostic information will improve outcomes for patients or society.

A Practical Approach

Physicians who regularly work with a population at risk of dying and for whom accurate models of survival are available should certainly know the characteristics of these models and generally for whom they will be useful. Such a physician will also have to develop skill in conveying the essential information to patients and families, many of whom will have difficulty understanding the uncertainties involved. But such a physician will have garnered a uniquely powerful tool. The model is largely immune from the errors that physicians are prone to make.[22] It "remembers" cases completely accurately, and it weighs the predictors with cold precision. These correctives are of value to physicians and patients. When a model's predictions are substantially different from those of a physician, a search for the reason can be illuminating. When a patient seeks the information, it can provide a solid anchor to what otherwise might be free-floating conjecture about a person's life span.

These are important and substantial uses that take predictive models out of the research and policy environment and put them into clinical practice. These models are not dictators, but only tools. They will be most helpful when well understood and thoughtfully applied, and their use warrants careful evaluation.

Hippocratic writers characterized the physician's role thus: "Declare the past, diagnose the present, foretell the future; practice these arts."[23] We should strive to do so; a contemporary Ivan Illych deserves to know his peril.

REFERENCES

1. Tolstoy L: The Death of Ivan Illych and Other Stories. New York, NY, New American Library, 1960

2. Feinstein AR: An additional basic science for clinical medicine—I. The constraining fundamental paradigms. Ann Intern Med 1983; 99:393-397

3. Lemeshow S, Teres D, 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; 270:2478-2486

4. Knaus WA, Wagner DP, Draper EA, et al: The APACHE III prognostic system—Risk prediction of hospital mortality for critically ill hospitalized adults. Chest 1991; 100:1619-1636

5. Knaus WA, Draper EA, Wagner DP, Zimmerman JE: APACHE II: A severity of disease classification system. Crit Care Med 1985; 13:818-829

6. Knaus WA, Harrell FE, Lynn J, et al: The SUPPORT prognostic model: Objective estimates of survival for seriously ill hospitalized adults. Ann Intern Med 1995; 122:191-203

7. President's Commission for the Study of Ethical Problems in Medicine and Biomedical and Behavioral Research: Deciding to Forego Life-Sustaining Treatment. Washington, DC, Government Printing Office, 1983 March

8. Guidelines on the Termination of Life-Sustaining Treatment and the Care of the Dying: A Report by the Hastings Center. New York, NY, Hastings Center, 1987

9. The Coordinating Council on Life-Sustaining Medical Treatment Decisions by the Court: Guidelines for State Court Decision Making in Authorizing or Withholding Life Sustaining Medical Treatment: A Project of the National Center for State Courts. Williamsburg, Va, West Publishing, 1991

10. *Cruzan v Director, Missouri Department of Health*, 110 SCt 284 (1990) 2855

11. Council on Ethical and Judicial Affairs: Guidelines for the appropriate use of do-not-resuscitate orders. JAMA 1991; 265:1868-1871

12. American Thoracic Society Bioethics Task Force: Withholding and withdrawing life-sustaining therapy. Ann Intern Med 1991; 115:478-485

13. Society of Critical Care Medicine Ethics Committee: Consensus statement on the triage of critically ill patients. JAMA 1994; 271:1200-1203

14. Teno JM, Murphy D, Lynn J, et al: Prognosis-based futility guidelines: Does anyone win? J Am Geriatr Soc 1994; 42:1202-1207

15. Damiano AM, Bergner M, Draper EA, Knaus WA, Wagner DP: Reliability of a measure of severity of illness: Acute Physiology and Chronic Health Evaluation II. J Clin Epidemiol 1992; 45:93-101

16. Knaus WA, Wagner DP: Selection bias and the relationship between APACHE II and mortality (Letter). Crit Care Med 1990; 18:793-794

17. Harrell FE Jr, Lee KL, Califf RM, Pryor DB, Rosati RA: Regression modelling strategies for improved prognostic prediction. Stat Med 1984; 3:143-152

18. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA: Evaluating the yield of medical tests. JAMA 1982; 247:2543-2546

19. Bamber D: The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. J Math Psychol 1975; 12:387-415

20. Efron B, Gong G: A leisurely look at the bootstrap, the jackknife, and cross-validation. Am Stat 1983; 37:36-48

21. Efron B, Tibshirani R: An Introduction to the Bootstrap. New York, NY, Chapman and Hall, 1993

22. Redelmeier DA, Shafir E: Medical decision making in situations that offer multiple alternatives. JAMA 1995; 273:302-305

23. Hippocrates: History, *In* Epidemics, T. Cadell (Trans). London, 1780