Chapter 7

Prognosis

He, who would rightly distinguish those that will survive or die, as well as those that will be subject to disease a longer or shorter time, ought, from his knowledge and attention, to be able to form an estimate of all symptoms, and rationally to weigh their powers by comparison.

—Hippocrates 460–375 B.C.

KEY WORDS

Prognosis
Prognostic factors
Clinical course
Natural history
Zero time
Inception cohort
Stage migration
Event
Survival analysis
Kaplan-Meier analysis
Time-to-event analysis
Censored
Hazard ratios
Case series
Case report
Clinical prediction rules
Training set
Test set
Validation
Prognostic stratification
Sampling bias
Migration bias
Dropouts
Measurement bias
Sensitivity analysis
Best-case/worst-case analysis

When people become sick, they have a great many questions about how their illness will affect them. Is it dangerous? Could I die of it? Will there be pain? How long will I be able to continue my present activities? Will it ever go away altogether? Most patients and their families want to know what to expect, even in situations where little can be done about their illness.

Prognosis is the prediction of the course of disease following its onset. This chapter reviews the ways in which the course of disease can be described. The intention is to give readers a better understanding of a difficult but indispensable task—predicting patients’ futures as closely as possible. The objective is to avoid expressing prognoses with vagueness when unnecessary and with certainty when misleading.

Doctors and patients want to know the general course of the illness, but they want to go further and tailor this information to their particular situation as much as possible. For example, even though ovarian cancer is usually fatal in the long run, women with this cancer may live from a few months to many years, and they want to know where on this continuum their particular case is likely to fall.

Studies of prognosis are similar to cohort studies of risk. Patients are assembled who have a particular disease or illness in common, they are followed forward in time, and clinical outcomes are measured. Patient characteristics that are associated with an outcome of the disease, called prognostic factors, are identified. Prognostic factors are analogous to risk factors, except that they represent a different part of the disease spectrum, from disease to outcomes. Case-control studies of people with the disease who do and do not have a bad outcome can also estimate the relative risk associated with various prognostic factors, but they are unable to provide information on outcome rates (see Chapter 6).

DIFFERENCES IN RISK AND PROGNOSTIC FACTORS

Risk and prognostic factors differ from each other in several ways.
The Patients Are Different

Studies of risk factors usually deal with healthy people, whereas studies of prognostic factors are of sick people.

The Outcomes Are Different

For risk, the event being counted is usually the onset of disease. For prognosis, consequences of disease are counted, including death, complications, disability, and suffering.

The Rates Are Different

Risk factors are usually for low-probability events. Yearly rates for the onset of various diseases are on the order of 1/1,000 to 1/100,000 or less. As a result, relationships between exposure and disease are difficult to confirm in the course of day-to-day clinical experiences, even for astute clinicians. Prognosis, on the other hand, describes relatively frequent events. For example, several percent of patients with acute myocardial infarction die before leaving the hospital.

The Factors May be Different

Variables associated with an increased risk are not necessarily the same as those marking a worse prognosis. Often, they are considerably different for a given disease. For example, the number of well-established risk factors for cardiovascular disease (hypertension, smoking, dyslipidemia, diabetes, and family history of coronary heart disease) is inversely related to the risk of dying in the hospital after a first myocardial infarction (Fig. 7.1) (1).

Clinicians can often form good estimates of short-term prognosis from their own personal experience. However, they may be less able to sort out, without the assistance of research, the various factors that are related to long-term prognosis or the complex ways in which prognostic factors are related to one another.

CLINICAL COURSE AND NATURAL HISTORY OF DISEASE

Prognosis can be described as either the clinical course or natural history of disease. The term clinical course describes the evolution (prognosis) of a disease that has come under medical care and has been treated in a variety of ways that affect the subsequent course of events. Patients usually receive medical care at some time in the course of their illness when they have diseases that cause symptoms such as pain, failure to thrive, disfigurement, or unusual behavior. Examples include type 1 diabetes mellitus, carcinoma of the lung, and rabies. After such a disease is recognized, it is likely to be treated.

The prognosis of disease without medical intervention is termed the natural history of disease. Natural history describes how patients fare if nothing is done about their disease. A great many health conditions do not come under medical care, even in countries with advanced health care systems. They remain unrecognized because they are asymptomatic (e.g., many cancers of the prostate are occult and slow growing) and are, therefore, unrecognized in life. For others, such as osteoarthritis, mild depression, or low-grade anemia, people may consider their symptoms to be one of the ordinary discomforts of daily living, not a disease and, therefore, not seek medical care for them.

Example

Irritable bowel syndrome is a common condition that involves abdominal pain and disturbed bowel habits not caused by other diseases. How often do patients with this condition visit doctors? In a British cohort of 3,875 people without irritable bowel syndrome at baseline, 15% developed the syndrome over the next 10 years (2). Of these, only 17% consulted their primary care physician with related symptoms at least once in 10 years, and 4% had consulted in the past year. In another study, characteristics of the abdominal complaints did not account for whether patients with irritable bowel syndrome sought health care for their symptoms (3).
Even without national medical records, population-based studies are possible. In the United States, the Network of Organ Sharing collects data on all patients with transplants, and the Surveillance, Epidemiology, and End Results (SEER) program collects incidence and survival data on all patients with new-onset cancers in several large areas of the country, comprising 28% of the U.S. population. For primary care questions, in the United States and elsewhere, individual practices in communities have banded together into “primary care research networks” to collect research data on their patients’ care.

Most studies of prognosis, especially for less common diseases, are of local patients. For these studies, it is especially important to provide the information that users can rely on to decide whether the results generalize to their own situation: patients’ characteristics (e.g., age, severity of disease, and comorbidity), the setting where they were found (e.g., primary care practices, community hospitals, or referral centers), and how they were sampled (e.g., complete, random,
or convenience sampling). Often, this information is sufficient to establish wide generalizability, for example, in studies of community acquired pneumonia or thrombophlebitis in a local hospital.

**Zero Time**

Cohorts in prognostic studies should begin from a common point in time in the course of disease, called zero time, such as at the time of the onset of symptoms, diagnosis, or the beginning of treatment. If observation begins at different points in the course of disease for the various patients in a cohort, the description of their prognosis will lack precision, and the timing of recovery, recurrence, death, and other outcome events will be difficult to interpret or will be misleading. The term inception cohort is used to describe a group of patients that is assembled at the onset (inception) of their disease.

Prognosis of cancer is often described separately according to patients' clinical stage (extent of spread) at the beginning of follow-up. If it is, a systematic change in how stage at zero time is established can result in a different prognosis for each stage even if the course of disease is unchanged for each patient in the cohort. This has been shown to happen during staging of cancer—assessing the extent of disease, with higher stages corresponding to more advanced cancer, which is done for the purposes of prognosis and choice of treatment. Stage migration occurs when a newer technology is able to detect the spread of cancer better than an older staging method. Patients who used to be classified in a lower stage are, with the newer technology, classified as being in a higher (more advanced) stage. Removal of patients with more advanced disease from lower stages results in an apparent improvement in prognosis for each stage, regardless of whether treatment is more effective or prognosis for these patients as a whole is better. Stage migration has been called the “Will Rogers phenomenon” after the humorist who said of the geographic migration in the United States during the economic depression of the 1930s, “When the Okies left Oklahoma and moved to California, they raised the average intelligence in both states” (5).

**Follow-Up**

Patients must be followed for a long enough period of time for most of the clinically important outcome events to have occurred. Otherwise, the observed rate will understate the true one. The appropriate length of follow-up depends on the disease. For studies of surgical site infections, the follow-up period should last for a few weeks, and for studies of the onset of AIDS and its complications in patients with HIV infection, the follow-up period should last several years.

**Outcomes of Disease**

Descriptions of prognosis should include the full range of manifestations of disease that would be considered important to patients. This means not only death and disease but also pain, anguish, and the inability to care for one's self or pursue usual activities. The 5 Ds—death, disease, discomfort, disability, and dissatisfaction—are a simple way to summarize important clinical outcomes (see Table 1.2).

In their efforts to be “scientific,” physicians tend to value precise or technologically measured outcomes, sometimes at the expense of clinical relevance. As discussed in Chapter 1, clinical effects that cannot be directly perceived by patients, such as radiologic reduction in tumor size, normalization of blood chemistries, improvement in ejection fraction, or change in serology, are not clinically useful ends in themselves. It is appropriate to substitute these biologic phenomena for clinical outcomes only when the two are known to be related to each other. Thus, in patients with pneumonia, short-term persistence of abnormalities on chest radiographs may not be alarming if the patient's fever has subsided, energy has returned, and cough has diminished.

Ways to measure patient-centered outcomes are now used in clinical research. Table 7.1 shows a simple measure of quality of life used in studies of cancer.
There are also research measures for performance status, health-related quality of life, pain, and other aspects of patient well-being.

**DESCRIPTING PROGNOSIS**

It is convenient to summarize the course of disease as a single rate—the proportion of people experiencing an event during a fixed time period. Some rates used for this purpose are shown in Table 7.2. These rates have in common the same basic components of incidence: events arising in a cohort of patients over time.

### A Trade-Off: Simplicity versus More Information

Summarizing prognosis by a single rate has the virtue of simplicity. Rates can be committed to memory and communicated succinctly. Their drawback is that relatively little information is conveyed. Large differences in prognosis can be hidden within similar summary rates.

Figure 7.3 shows 5-year survival rates for patients with four conditions. For each condition, about 10% of the patients are alive at 5 years. However, the clinical courses are otherwise quite different in ways that are very important to patients. Early survival in patients with dissecting aneurysms is very poor, but if they survive the first few months, their risk of dying is much less affected by having had the aneurysm (Fig. 7.3A). Patients with locally invasive, non–small cell lung cancer experience a relatively constant mortality rate throughout the 5 years following diagnosis (Fig. 7.3B). The life of patients with amyotrophic lateral sclerosis (ALS, Lou Gehrig disease, a slowly progressive paralysis) and respiratory difficulties is not immediately threatened, but as neurologic function continues to decline over the years, the inability to breathe without assistance leads to death (Fig. 7.3C). Figure 7.3D is a benchmark. Only at age 100 years do people in the general population have a 5-year survival rate comparable to that of patients with the three diseases.

**Survival Analysis**

When interpreting prognosis, it is preferable to know the likelihood, on average, that patients with a given condition will experience an outcome at any point in time. Prognosis expressed as a summary rate does not contain this information. However, figures can show information about average time to event for any point in the course of disease. By event, we mean a dichotomous clinical outcome that can occur only once. In the following discussion, we take the common approach of describing outcomes in terms of “survival,” but the same methods apply to the reverse (time to death) and to any other outcome event such as cancer recurrence, cure of infection, freedom from symptoms, or arthritis becoming inactive.

**Survival of a Cohort**

The most straightforward way to learn about survival is to assemble a cohort of patients who have the

---

**Table 7.1**

<table>
<thead>
<tr>
<th>Performance Status</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Asymptomatic</td>
</tr>
<tr>
<td>1</td>
<td>Symptomatic, fully ambulatory</td>
</tr>
<tr>
<td>2</td>
<td>Symptomatic, in bed &lt;50% of the day</td>
</tr>
<tr>
<td>3</td>
<td>Symptomatic, in bed &gt;50% of the day</td>
</tr>
<tr>
<td>4</td>
<td>Bedridden</td>
</tr>
<tr>
<td>5</td>
<td>Dead</td>
</tr>
</tbody>
</table>


**Table 7.2**

<table>
<thead>
<tr>
<th>Rate</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>5-year survival</td>
<td>Percent of patients surviving 5 years from some point in the course of their disease</td>
</tr>
<tr>
<td>Case fatality</td>
<td>Percent of patients with a disease who die of it</td>
</tr>
<tr>
<td>Disease-specific mortality</td>
<td>Number of people per 10,000 (or 100,000) population dying of a specific disease</td>
</tr>
<tr>
<td>Response</td>
<td>Percent of patients showing some evidence of improvement following an intervention</td>
</tr>
<tr>
<td>Remission</td>
<td>Percent of patients entering a phase in which disease is no longer detectable</td>
</tr>
<tr>
<td>Recurrence</td>
<td>Percent of patients who have return of disease after a disease-free interval</td>
</tr>
</tbody>
</table>

*Time under observation is either stated or assumed to be sufficiently long so that all events that will occur have been observed.
condition of interest and are at the same point in the course of their illness (e.g., onset of symptoms, diagnosis, or beginning of treatment), and then keep them all under observation until all experience the outcome or not. For a small cohort, one might then represent these patients’ clinical course, as shown in Figure 7.4A. The plot of survival against time displays steps corresponding to the death of each of the 10 patients in the cohort. If the number of patients were increased, the size of the steps would diminish. If a very large number of patients were studied, the figure would approximate a smooth curve (Fig. 7.4B). This information could then be used to predict the year-by-year, or even week-by-week, prognosis of similar patients.

Unfortunately, obtaining the information in this way is impractical for several reasons. Some of the patients might drop out of the study before the end of the follow-up period, perhaps because of another illness, a move from the study area, or dissatisfaction with the study. These patients would have to be excluded from the cohort even though considerable effort had been exerted to gather data on them until the point at which they dropped out. Also, it would be necessary to wait until all of the cohort’s members had reached each point in follow-up before the probability of surviving to that point could be calculated. Because patients ordinarily become available for a study over a period of time, at any point in calendar time, there would be a relatively long follow-up for patients who had entered the study first, but only brief experience with those who had entered more recently. The last patient who entered the study would have to reach each year of follow-up before any information on survival to that year would be available.

**Survival Curves**

To make efficient use of all available data from each patient in the cohort, survival analysis has been developed to estimate the survival of a cohort over time. The usual method is called Kaplan-Meier analysis, after its originators. Survival analysis can be applied to any outcomes that are dichotomous and occur only once during follow-up (e.g., time to
coronary event or to recurrence of cancer). A more general term, useful when an event other than survival is described, is time-to-event analysis.

Figure 7.5 shows a simplified survival curve. On the vertical axis is the estimated probability of surviving, and on the horizontal axis is the period of time from the beginning of observation (zero time).

The probability of surviving to any point in time is estimated from the cumulative probability of surviving each of the time intervals that preceded it. Time intervals can be made as small as necessary; in Kaplan-Meir analyses, the intervals are between each new event, such as death, and the preceding one, however short or long that is. Most of the time,
no one dies and the probability of surviving is 1. When a patient dies, the probability of surviving at that moment is calculated as the ratio of the number of patients surviving to the number at risk of dying at that point in time. Patients who have already died, dropped out, or have not yet been followed up to that point are not at risk of dying and are, therefore, not used to estimate survival for that time. The probability of surviving does not change during intervals in which no one dies, so it is recalculated only when there is a death. Although the probability at any given interval is not very accurate, because either nothing has happened or there has been only one event in a large cohort, the overall probability of surviving up to each point in time (the product of all preceding probabilities) is remarkably accurate. When patients are lost from the study at any point in time, they are referred to as censored and are no longer counted in the denominator from that point forward.

A part of the survival curve in Figure 7.5 (from 3 to 5 years after zero time) is presented in detail to illustrate the data used to estimate survival: patients at risk, patients no longer at risk (censored), and patients experiencing outcome events at each point in time.

Variations on basic survival curves increase the amount of information they convey. Including the numbers of patients at risk at various points in time gives some idea of the contribution of chance to the observed rates, especially toward the end of follow-up. The vertical axis can show the proportion with, rather than without, the outcome event; the resulting curve will sweep upward and to the right. The precision of survival estimates, which declines with time because fewer and fewer patients are still under observation, can be shown by confidence intervals at various points in time (see Chapter 11). Tics are sometimes added to the survival curves to indicate each time a patient is censored.

**Interpreting Survival Curves**

Several points must be kept in mind when interpreting survival curves. First, the vertical axis represents the estimated probability of surviving for members of the cohort, not the cumulative incidence of surviving if all members of the cohort were followed up.

Second, points on a survival curve are the best estimate, for a given set of data, of the probability of survival for members of a cohort. However, the precision of these estimates depends on the number of patients on whom the estimate is based, as do all observations of samples. One can be more confident that the estimates on the left-hand side of the curve are sound, because more patients are at risk early in follow-up. But on the right-hand side, at the tail of the curve, the number of patients on whom estimates of survival are based may become relatively small because deaths, dropouts, and late entrants to the study, so that fewer and fewer patients are followed for that length of time. As a result, estimates of survival toward the end of the follow-up period are imprecise and can be strongly affected by what happens to relatively few patients. For example, in Figure 7.5, only one patient was under observation at year 5. If that one remaining patient happened to die, the probability of surviving would fall from 8% to zero. Clearly, this would be a too literal reading of the data. Therefore, estimates of survival at the tails of survival curves must be interpreted with caution.

Finally, the shape of many survival curves gives the impression that outcome events occurs more frequently early in follow-up than later on, when the slope approaches a plateau. But this impression is deceptive. As time passes, rates of survival are being applied to a diminishing number of patients, causing the slope of the curve to flatten even if the rate of outcome events did not change.

As with any estimate, Kaplan-Meier estimates of time to event depend on assumptions. It is assumed that being censored is not related to prognosis. To the extent that this is not true, a survival analysis may yield biased estimates of survival in cohorts. The Kaplan-Meier method may not be accurate enough if there are competing risks—more than one kind of outcome event—and the outcomes are not independent of each other such that one event changes the probability of experiencing the other. For example, patients with cancer who develop an infection related to aggressive chemotherapy and drop out for that reason may have had a different chance of dying of the cancer. There are other methods for estimating cumulative incidence in the presence of competing risks.

**IDENTIFYING PROGNOSTIC FACTORS**

Often, studies go beyond a simple description of prognosis in a homogenous group of patients to compare prognosis in patients with different characteristics, that is, they identify prognostic factors. Multiple survival curves, one for patients with each of the characteristics, are represented on the same figure where they can be visually (and statistically) compared.
The effects of one prognostic factor relative to the effects of another can be summarized from data in a time-to-event analysis by a hazard ratio, which is analogous to a risk ratio (relative risk). Also, survival curves can be compared after taking into account other factors related to prognosis so that the independent effect of just one variable is examined.

**CASE SERIES**

A case series is a description of the course of disease in a small number of cases, a few dozen at most. An even smaller report, with fewer than 10 patients, is called a case report. Cases are typically found at a clinic or referral center and then followed forward in time to describe the course of disease and backward in time to describe what came earlier.

Such reports can make an important contribution to understanding of disease, primarily by describing experiences with newly defined syndromes or uncommon conditions. The reason for introducing case series into a chapter about prognosis is that they may masquerade as true cohort studies even though they do not have comparable strengths.

**Example**

Physicians in emergency departments see patients with bites from North American rattlesnakes. These bites are relatively uncommon at any one place, making it difficult to carry out large cohort studies of their clinical course, so physicians must rely mainly on case series. An example is a description of the clinical course of all 24 children managed at a children’s hospital in California during a 10-year period (8). Nineteen of the children were actually injected with venom, and they were managed with the aggressive use of antivenin. Three had surgical treatment to remove soft-tissue debris or relieve tissue pressure. There were no serious reactions to antivenin, and all patients left the hospital without functional impairment.

Physicians caring for children with rattlesnake bites would be grateful for this and other case series if there were no better information to guide their care, but that is not to say that the case series provided a complete and fully reliable picture of snake-bite care. Of all children bitten in that region, some may have been doing so well after a bite that they were not sent to the referral center. Others might have been doing so badly that they were rushed to the nearest hospital or even died before reaching a hospital at all. In other words, the case series does not describe the clinical course of all children from
the time of snakebite (the inception) but rather a selected sample of children who happened to come under care at that particular hospital. In effect, case series describe the clinical course of prevalent, not necessarily a representative sample of incident cases, so they are “false” cohorts.

CLINICAL PREDICTION RULES

A combination of variables can provide a more precise prognosis than any of these variables taken one at a time. As discussed in Chapter 4, clinical prediction rules estimate the probability of outcomes (either prognosis or diagnosis) according to a set of patient characteristics defined by history, physical examination, and simple laboratory tests. They are “rules” because they are often tied to recommendations about further diagnostic evaluation or treatment. To make prediction rules workable in clinical settings, they depend on data that are available in the usual care of patients and scoring, the basis for the prediction, that has been simplified.

A clinical prediction rule should be developed in one setting and tested in others—with different patients, physicians, and usual care practices—to assure that predictions are good for a broad range of settings and not just for where it was developed, because it might have been the result of the particular characteristics of that setting. The data used to develop the prediction rule is called the training set, and the data used to assess its validity is called the test set, which is used for validation of the prediction rule.

Atrial fibrillation causes an increased risk of stroke. Clots form in the atria in the absence of regular, organized contractions and, if dislodged, can travel to the brain causing an embolic stroke. This complication can be prevented by anticoagulation but at the risk of bleeding. Clinicians and patients must weigh the risk of stroke in the absence of treatment against the risk of bleeding complications with treatment. To assist in this decision, a clinical prediction rule called CHADS2 has been developed and validated (9). A set of five readily available clinical observations are used to separate patients with atrial fibrillation into six risk strata with a 14-fold difference in risk of stroke (Table 7.3). The aggressiveness of anticoagulation is in turn tied to the CHADS2 score (10). For example, typical recommendations based on risk of stroke and effectiveness of anticoagulation are that patients with a CHADS2 score of 0 have no anticoagulation, those with a score of 1 are treated with aspirin or oral anticoagulation, and those with a score of at least 2 are treated with oral anticoagulation, the most aggressive treatment.

### Table 7.3

<table>
<thead>
<tr>
<th>CHADS2 Points</th>
<th>Stroke Risk per 100 Person-Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.49</td>
</tr>
<tr>
<td>1</td>
<td>1.52</td>
</tr>
<tr>
<td>2</td>
<td>2.50</td>
</tr>
<tr>
<td>3</td>
<td>5.27</td>
</tr>
<tr>
<td>4</td>
<td>6.02</td>
</tr>
<tr>
<td>5–6</td>
<td>6.88</td>
</tr>
</tbody>
</table>

A clinical prediction rule should be developed in one setting and tested in others—with different patients, physicians, and usual care practices—to assure that predictions are good for a broad range of settings and not just for where it was developed, because it might have been the result of the particular characteristics of that setting. The data used to develop the prediction rule is called the training set, and the data used to assess its validity is called the test set, which is used for validation of the prediction rule.

The process of separating patients into groups with different prognosis, as in the previous example, is called prognostic stratification. In this case, atrial fibrillation is the disease and stroke is the outcome. The concept is similar to risk stratification (Chapter 4), where patients are divided into different strata of risk for developing disease.

BIAS IN COHORT STUDIES

In cohort studies of risk or prognosis, bias can alter the description of the course of disease. Bias can also create apparent differences between groups when differences do not actually exist in nature or obscure differences when they really do exist. These biases have their counterparts in case-control studies as well.
They are a separate consideration from confounding and effect modification (Chapter 5).

There are an almost infinite variety of systematic errors, many of which are given specific names, but some are more basic. They can be recognized more easily when one knows where they are most likely to occur in the course of a study. With that in mind, we describe some possibilities for bias in cohort studies and discuss them in relation to the following study of prognosis.

Sampling bias can also be misleading when prognosis is compared across groups and sampling has produced groups that are systematically different with respect to prognosis, even before the factor of interest is considered. In the Bell palsy example, older patients might have had a worse prognosis because they are the ones who had an underlying herpes virus infection, not because of their age.

Is this not confounding? Strictly speaking, it is not because the study is for the purpose of prediction, not to identify independent “causes” of recovery. Also, it is not plausible to consider such phenomena as severity of palsy or onset of recovery as causes because they are probably part of the chain of events leading from disease to recovery. But to the extent that one wants to show that prognostic factors are independent predictors of outcome, the same approaches as are used for confounding (Chapter 5) can be used to establish independence.

Migration Bias

Migration bias is present when some patients drop out of the study during follow-up and they are systematically different from those who remain. It is often the case that some members of the original cohort leave a study over time. (Patients are assured that this is their right as part of the ethical conduct of research on humans.) If dropout occurs randomly, such that the characteristics of lost patients are on average similar to patients who remain, then there would be no bias. This is so regardless of whether the number of dropouts is large or similar in the cohorts being compared, but ordinarily the characteristics of lost patients are not the same as those who remain in a study. Dropping out tends to be related to prognosis. For example, patients who are doing especially well or badly with their Bell palsy may be more likely to leave the study, as would those who need care for other illnesses, for whom the extra visits related to the study would be burdensome. This would distort the main (descriptive) results of the study—rate and completeness of recovery. If the study also aims to identify prognostic factors (e.g., recovery in old versus young patients), that also could be biased by patients dropping out, for the same reasons.

Migration bias might be seen as an example of selection bias because patients who were still in the study when outcomes were measured were selected from all those who began in the study. Migration bias might be considered an example of measurement bias because patients who migrate out of the study are no longer available when outcomes are measured.
Measurement Bias

**Measurement bias** is present when members of the cohort are not all assessed similarly for outcome. In the Bell palsy study, all members of the cohort were examined by a common protocol every month until they were no longer improving, ruling out this possibility. If it had been left to individual patients and physicians whether, when, and how they were examined, this would have diminished confidence in the description of time to and completeness of recovery. Measurement bias also comes into play if prognostic groups are compared and patients in one group have a systematically better chance of having outcomes detected than those in another. Some outcomes, such as death, cardiovascular catastrophes, and major cancers, are so obvious that they are unlikely to be missed. But for less clear-cut outcomes, including specific cause of death, subclinical disease, side effects, or disability, measurement bias can occur because of differences in the methods with which the outcome is sought or classified.

Measurement bias can be minimized in three general ways: (i) examine all members of the cohort equally for outcome events; (ii) if comparisons of prognostic groups are made, ensure that researchers are unaware of the group to which each patient belongs; and (iii) set up careful rules for deciding if an outcome event has occurred (and follow the rules). To help readers understand the extent of these kinds of biases in a given study, it is usual practice to include, with reports of the study, a flow diagram describing how the number of participants changed as the study progressed and why. It is also helpful to compare the characteristics of patients in and out of the study after sampling and follow-up.

**Bias from “Non-differential” Misclassification**

Until now, we have been discussing how the results of a study can be biased when there are systematic differences in how exposure or disease groups are classified. But bias can also result if misclassification is “non-differential,” that is, it occurs similarly in the groups being compared. In this case, the bias is toward finding no effect.

**Example**

When cigarette smoking is assessed by simply asking people whether they smoke, there is substantial misclassification relative to a gold standard such as the presence or absence of cigarette products in saliva. Yet in a cohort study of cigarette smoking and coronary heart disease (CHD), misclassification of smoking could not be different in people who did or did not develop CHD because the outcome was not known at the time exposure was assessed. Even so, to the extent that smoking is incorrectly classified, it reduces whatever differences in CHD rates in smokers and non-smokers that might have existed if all patients had been correctly classified, making a “null” effect more likely. At the extreme, if classifying smoking status were totally at random, there could be no association between smoking and CHD.

**BIAS, PERHAPS, BUT DOES IT MATTER?**

Clinical epidemiology is not an error-finding game. Rather, it is meant to characterize the credibility of a study so that clinicians can decide how much to rely on its results when making high-stakes decisions about patients. It would be irresponsible to ignore results of studies that meet high standards, just as clinical decisions need not be bound by the results of weak studies.

With this in mind, it is not enough to recognize that bias might be present in a study. One must go on to determine if bias is actually present in the particular study. Beyond that, one must decide whether the consequences of bias are sufficiently large that they change the conclusions in a clinically important way. If damage to the study’s conclusions is not very great, then the presence of bias is of little practical consequence and the study is still useful.

**SENSITIVITY ANALYSIS**

One way to decide how much bias might change the conclusions of a study is to do a **sensitivity analysis**, that is, to show how much larger or smaller the observed results might have been under various assumptions about the missing data or potentially biased measurements. A **best-case/worst-case analysis** tests the effects of the most extreme possible assumptions but is an unreasonably severe test for the effects of bias in most situations. More often, sensitivity analyses test the effects of somewhat unlikely values, as in the following example.
Chapter 7: Prognosis

Review Questions

Read the following and select the best response.

7.1. For a study of the risk of esophageal cancer in patients with Barrett esophagus (a precancerous lesion), which of the following times in the course of disease is the best example of zero time?
A. Diagnosis of Barrett esophagus for each patient
B. Death of each patient
C. Diagnosis of esophageal cancer for each patient
D. Calendar time when the first patient is enrolled in the study
E. Calendar time when no patient remains in the study

7.2. A cohort study describes the recurrence of seizure within 1 year in children hospitalized with a first febrile seizure. It compared recurrence in children who had infection versus immunization as an underlying cause for fever at the time of the first seizure. Some of the children were no longer in the study when outcome (a second seizure) was assessed at 1 year. Which of the following would have the greatest effect on study results?
A. Why the children dropped out
B. When in the course of follow-up the children left the study
C. Whether dropping out was related to prognosis
D. Whether the number of children dropping out is similar in the groups

7.3. A cohort study of prostate cancer care compares rates of incontinence in patients who were treated with surgery versus medical care alone. Incontinence is assessed from review of medical records. Which of the following is not an example of measurement bias?
A. Men were more likely to tell their surgeon about incontinence.
B. Surgeons were less likely to record complications of their surgery in the record.

Example

Poliomyelitis has been eradicated in many parts of the world but late effects of infection continue. Some patients develop post-polio syndrome (muscle weakness, pain, and fatigue) many years after the original infection. To describe the rate of post-polio syndrome, investigators identified a cohort of 939 patients who had poliomyelitis in the 1950s (12). Up to 40 years later, they were able to get information about symptoms for 551 members of the cohort, and in them the rate of post-polio syndrome was 137 of 551 (25%). Understandably, after so many years, 388 patients could not be followed-up.

If the missing members of the cohort had different rates of post-polio syndrome from those who were followed-up, how much could they have biased the observed rate of post-polio syndrome? If patients who were not followed up were twice as likely to get post-polio syndrome as those who were, the true rate would have been \((137 + 194)/939 = 35\%\). (That is, all 137 patients known to have the syndrome plus 50\% of those who were not followed up divided by all members of the original cohort.) If the missing patients were half as likely to get the syndrome, the true rate would have been \((137 + 48)/939 = 20\%\). Thus, even with an impossibly large difference in post-polio syndrome rates in missing patients, the true rate would still have been in the 20\% to 35\% range, a useful approximation for clinical use.

More or less extreme differences could have been assumed for the missing patients to explore how “sensitive” the study result were to missing members of the cohort. Sensitivity analysis is a useful way of estimating how much various kinds of bias could have affected the results of studies of all kinds—cohort and case-control studies of risk or prognosis, the accuracy of diagnostic tests, or clinical trials of the effectiveness of treatment or prevention.
C. Men who got surgery were more likely to have follow-up visits.
D. Chart abstractors used their judgment is deciding whether incontinence was present or not.
E. Rates of incontinence were higher in the men who got surgery.

7.4. A clinical prediction rule has been developed to classify the prognosis of community-acquired pneumonia. Which of the following is most characteristic of such a clinical prediction rule?
A. Calculating a score is simple.
B. The clinical data are readily available.
C. Multiple prognostic factors are included.
D. The results are used to guide further management of the patient.
E. All of the above.

7.5. A study describes the clinical course of patients who have an uncommon neurologic disease. Patients are identified at a referral center that specializes in this disease. Their medical records are reviewed for patients’ characteristics, treatments, and their current disease status. Which of the following best describes this kind of study?
A. Cohort study
B. Case-control study
C. Case series
D. Cross-sectional study
E. A randomized controlled trial

7.6. A study used time-to-event analysis to describe the survival from diagnosis of 100 patients with congestive heart failure. By the third year, 60 patients have been censored. Which of the following would not be a reason for one of these patients being censored?
A. The patient died of another cause before year 3.
B. The patient decided not to continue in the study.
C. The patient developed another disease that could be fatal.
D. The patient had been enrolled in the study for less than 3 years.

7.7. Which of the following kinds of studies cannot be used to identify prognostic factors?
A. Prevalence study
B. Time-to-event analysis
C. Case-control study
D. Cohort study

7.8. Which of the following best describes the information in a survival curve?
A. An unbiased estimate of survival even if some patients leave the study
B. The estimated probability of survival from zero time
C. The proportion of a cohort still alive at the end of follow-up
D. The rate at which original members of the cohort leave the study
E. The cumulative survival of a cohort over time

7.9. Which of the following is the most appropriate sample for a study of prognosis?
A. Members of the general population
B. Patients in primary care in the community
C. Patients admitted to a community hospital
D. Patients referred to a specialist
E. It depends on who will use the results of the study.

7.10. Investigators wish to describe the clinical course of multiple sclerosis. They take advantage of a clinical trial, already completed, in which control patients received usual care. Patients in the trial had been identified at referral centers, had been enrolled at the time of diagnosis, and had met rigorous entry criteria. After 10 years, all patients had been examined yearly and remained under observation, and 40% were still able to walk. Which of the following most limits the credibility of this study?
A. Inconsistent zero time
B. Generalizability
C. Measurement bias
D. Migration bias
E. Failure to use time-to-event methods
7.11. Many different clinical prediction rules have been developed to assess the severity of community-acquired pneumonia. Which of the following is the most important reason for choosing one to use?

A. The prediction rule classifies patients into groups with very different prognosis.
B. The prediction rule has been validated in different settings.
C. Many variables are included in the rule.
D. Prognostic factors include state-of-the-science diagnostic tests.
E. The score is calculated using computers.

7.12. In a study of patients who smoke and developed peripheral arterial disease, the hazard ratio for amputation in patients who continued to smoke, relative to those who quit smoking, is 5. Which of the following best characterizes the hazard ratio?

A. In this study, it is the rate of continued smoking divided by the rate of quitting.
B. It can be estimated from a case-control study of smoking and amputation.
C. It cannot be adjusted for the presence or absence of other factors related to prognosis.
D. It conveys information similar to relative risk.
E. It is calculated from the cumulative incidence of amputation in smokers and quitters.

7.13. In a time-to-event analyses, the event:

A. Can occur only once
B. Is dichotomous
C. Both A and B
D. Neither A nor B

Answers are in Appendix A.

REFERENCES