1290.ch01 4/21/05 11:24 AM Page 1

CHAPTER 1

Commonly Used Frequency Measures in Health Care

KEY TERMS

- Variable
- Frequency distribution
- Rate
- Ratio
- Proportion
- Dichotomous variables
- Confounding factor
- Confounding variable
- Mortality rates
  - Crude death rate
  - Age-specific death rate (ASDR)
  - Age-adjusted death rate
  - Standard mortality ratio (SMR)
  - Race-specific death rate
  - Sex-specific death rate
  - Cause-specific death rate
  - Case fatality rate
  - Proportionate mortality ratio (PMR)
  - Maternal mortality rate
  - Neonatal mortality rate
  - Postneonatal mortality rate
  - Infant mortality rate
- Morbidity rates
  - Incidence rate
  - Prevalence rate
  - Point prevalence rate
  - Risk ratios
  - Relative risk
  - Odds ratio
  - Attributable risk
  - Kaplan Meier method
  - Kaplan-Meier survival analysis

LEARNING OBJECTIVES

At the conclusion of this chapter, you should be able to:

1. Define key terms.
2. Calculate measures of morbidity, mortality, and risk of disease for health care facilities and communities.
3. Identify variables that affect morbidity and mortality rates over time.

4. Adjust measures of morbidity and mortality by both the direct and indirect methods of standardization.
5. After adjustment, compare health care facility mortality/morbidity rates with community, state, and/or national rates.
6. Calculate risk of disease between groups.
7. Conduct survival analysis for tumor registries and clinical trials.

It is often said that hospitals and other types of health care facilities are data rich but information poor. There are many types of databases within the facility, many contained within the organization's information warehouse. Information warehouses contain both clinical and financial information. It is the job of the health information management professional to turn the data contained in these databases into information that can be used by physicians, administrators, and other interested parties. The health information management professional can become an invaluable member of the health care team by providing data that are presented in a meaningful way and by presenting data that have been analyzed to serve a specific medical or clinical need. Some typical questions might be:

- What are the top 25 medical and top 10 surgical diagnosis-related groups (DRGs) for inpatient discharges from our facility?
- Which medical/surgical services admit the most patients?
- Is the average length of stay (ALOS) for these DRGs significantly different from the national ALOS for these DRGs?
- How do our charges compare with national charges? How does our reimbursement compare with our costs?
- What geographical area does the health care facility serve?
- How many patients were admitted to the facility by payer? What is the number of inpatient service days by payer? What are the average charges by payer?
- How do lengths of stay (LOSs) compare by physician?
- How many patients acquired nosocomial infections?

In the course of this text we will answer these questions. We will learn how to use descriptive statistics to describe patient populations, how to analyze clinical data for significant differences and relationships, and how to present data in graphic form. Our goal is to collect, analyze, and interpret clinical information for both clinical and administrative health care decision makers. We will begin our discussion of clinical data analysis by reviewing morbidity and mortality measures that are often used to describe patient and community populations.

INTRODUCTION TO FREQUENCY DISTRIBUTIONS

In health care, we deal with vast quantities of clinical data. Since it is very difficult to look at data in raw form, data are summarized into frequency distributions. A **frequency distri-**

bution shows the values that a variable can take and the number of observations associated with each value. A **variable** is a characteristic or property that may take on different values. Height, weight, sex, and third-party payer are examples of variables.

For example, suppose we are studying the variable patient LOS in the pediatric unit. To construct a frequency distribution, we first list all the values that LOS can take, from the lowest observed value to the highest. We then enter the number of observations (frequencies) corresponding to a given LOS. Table 1–1 illustrates what the resulting frequency distribution looks like. Note that all values for LOS between the lowest and highest are listed, even though there may not be any observations for some of the values. Each column of the distribution is properly labeled; the total is given in the bottom row. We can also display a frequency distribution by categories into which a variable may fall. Table 1–2 shows a frequency distribution for the number of patients discharged from Critical Care Hospital by religion, a variable composed of categories. The proportion for each category is also displayed in the table. The sum of the proportions for each category is equal to 1.0. We will examine frequency distributions in greater detail in Chapter 4.

Table 1–1 Frequency Distribution for Patient Length of Stay (LOS), Pediatric Unit

LOS in Days	*No. of Patients*
1	2
2	2
3	0
4	6
5	6
6	11
7	6
8	5
9	3
10	1
Total	42

Table 1–2 Frequency Distribution of Number of Patients Discharged from Critical Care Hospital by Religion, July 20xx

Religion	*Number of Discharges*	*Proportion*
Protestant	422	0.48
Catholic	315	0.36
Jewish	20	0.02
Other	127	0.14
Total	884	1.00

RATIOS, PROPORTIONS, AND RATES

Variables often have only two possible categories, such as alive or dead, or male or female. Variables having only two possible categories are called dichotomous. The frequency measures used with **dichotomous variables** are ratios, proportions, and rates. All three measures are based on the same formula:

$$\text{ratio, proportion, rate} = x/y \times 10^n$$

In this formula, x and y are the two quantities being compared, and x is divided by y. 10^n is read as "10 to the nth power." The size of 10^n may equal, for example, 1, 10, 100, or 1,000, depending on the value of n:

$$10^0 = 1$$
$$10^1 = 10$$
$$10^2 = 10 \times 10 = 100$$
$$10^3 = 10 \times 10 \times 10 = 1{,}000$$

Ratios

In a **ratio**, the values of a variable, such as sex (x = female, y = male), may be expressed so that x and y are completely independent of each other, or x may be included in y. For example, the sex of patients discharged from a hospital could be compared in either of two ways:

Female/male or x/y
Female/(male + female) or $x/(x + y)$

In the first option, x is completely independent of y, and the ratio represents the number of female discharges compared to the number of male discharges. In the second option, x is a proportion of the whole, $x + y$. The ratio represents the number of female discharges compared to the total number of discharges. Both expressions are considered ratios.

How, then, would you calculate the female-to-male ratio for a hospital that discharged 457 women and 395 men during the month of July? The procedure for calculating a ratio is outlined in Exhibit 1–1.

Proportions

A **proportion** is a particular type of ratio. A proportion is a ratio in which x is a portion of the whole, $x + y$. In a proportion, the numerator is always included in the denominator. Exhibit 1–2 outlines the procedure for determining the proportion of hospital discharges for the month of July that were female.

Exhibit 1–1 Calculation of a Ratio: Discharges for July 20xx

1. Define x and y.
   x = number of female discharges
   y = number of male discharges
2. Identify x and y.
   $x = 457$
   $y = 395$
3. Set up the ratio x/y.
   457/395
4. Reduce the fraction so that either x or y equals 1.
   1.16/1

There were 1.16 female discharges for every male discharge.

Exhibit 1–2 Calculation of a Proportion: Discharges for July 20xx

1. Define x and y.
   x = number of female discharges
   y = number of male discharges
2. Identify x and y.
   $x = 457$
   $y = 395$
3. Set up the proportion
   $x/(x + y)$ $457/(457 + 395) = 457/852$
4. Reduce the fraction so that either x or $x = y$ equals 1.
   0.54/1.00

The proportion of discharges that were female is 0.54.

Rates

Rates are a third type of frequency measure. In health care, rates are often used to measure an event over time and are sometimes used as performance improvement measures. The basic formula for a rate is:

$$\frac{\text{No. of cases or events occurring during a given time period} \times 10^n}{\text{No. of cases or population at risk during same time period}}$$

or

$$\frac{\text{Total number of times something did happen} \times 10^n}{\text{Total number of times something could happen}}$$

In inpatient facilities, there are many commonly computed rates. In computing the Caesarean section rate, we count the number of Caesarean sections (C-sections) performed during a given period of time; this value is placed in the numerator. The number of cases or population at risk is the number of women who delivered during the same time period; this number is placed in the denominator. By convention, inpatient hospital rates are calculated as the rate per 100 cases ($10^n = 10^2 = 10 \times 10 = 100$) and are expressed as a percentage. The method for calculating the hospital C-section rate is presented in Exhibit 1–3.

Exhibit 1–3 Calculation of C-Section Rate for July 20xx

For the month of July, 23 C-sections were performed; during the same time period, 149 women delivered. What is the C-section rate for the month of July?

1. Define the variable of interest (numerator) and population or number of cases at risk (denominator).
   Numerator: total number of C-sections performed in July
   Denominator: total number of women who delivered in July, including C-sections
2. Identify the numerator and denominator.
   Numerator: 23
   Denominator: 149
3. Set up the rate.
   23/149
4. Divide the numerator by the denominator, and multiply by 100 ($10^n = 10^2$).
   $(23/149) \times 100 = 15.4\%$.

The C-section rate for the month of July is 15.4%.

POPULATION-BASED MORTALITY MEASURES

As the profession of health information management moves into integrated health care delivery systems and assumes more prominence in managed care organizations, it becomes more important to be familiar with community-based mortality and morbidity data. This type of information is often used in planning health services, such as number of inpatient facilities, type of outpatient facilities, and number or size of managed care plans for a given community, as well as for developing managed care contracts with hospitals and physicians.

Crude Death Rate

The **crude death rate** is a measure of the actual or observed mortality in a given population. Crude rates apply to a population without regard to characteristics of the population, such as the distribution of age or sex. The crude death rate is the starting point for further development of adjusted rates. It measures the proportion of a population that has died during a specific period of time, usually one year, or the number of deaths per 1,000 in a community for a given period of time. The crude death rate is calculated as follows (the midinterval population is the estimated population of a given community at the midpoint of the time frame under study):

$$\frac{\text{Total deaths during a given time interval} \times 10^n = \text{deaths per } 10^n}{\text{Estimated midinterval population}}$$

In calculating the crude death rate, the power of *n* is usually equal to the value that will result in a value greater than 1. This allows for easier interpretation of the rate—a death rate of less than 1 per 100 is not very meaningful. For example, the 2004 midyear population of Anytown, USA, is 1,996,355; 275 deaths occurred in 2004. The power of *n* that will result in a whole number is 4; $10^4 = 10 \times 10 \times 10 \times 10 = 10{,}000$. The crude death rate is calculated as follows:

$$(275 \times 10{,}000)/1{,}996{,}355 = 2{,}750{,}000/1{,}996{,}355 = 1.38 \text{ deaths per } 10{,}000$$

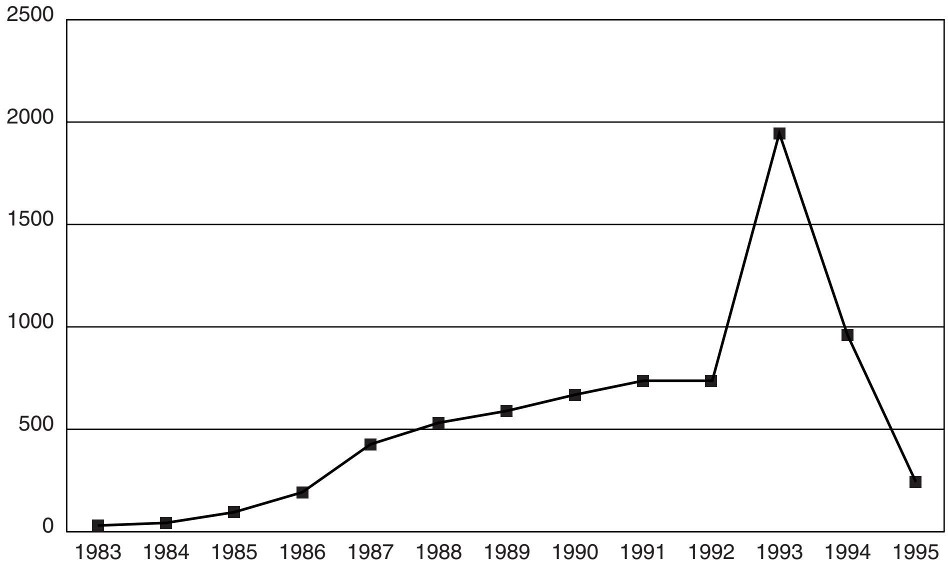
When analyzing crude death rates, or any type of rate, it is important to remember that these events do not occur in a vacuum. When analyzing any data set, we need to remember that the data do not stand alone, but reflect trends in the environment. Trends in death rates can be influenced by three variables: time, place, and person. Examples of time, place, and person variables are outlined in Exhibit 1–4. An example of how trended data may be affected by time, place, and person variables is presented in Figure 1–1. The line graph shows that the number of newly diagnosed acquired immune deficiency syndrome (AIDS) cases steadily increased from 1983 to 1992; then a rather dramatic increase occurred in 1993,

Exhibit 1–4 Variables Affecting Trends in Community Morbidity and Mortality

- Time
    - Transition from International Classification of Diseases, 9th Revision (ICD-9) to ICD-10 in coding of death certificates
    - Improvements in medical technology
    - Earlier detection and diagnosis of disease
- Place
    - Changes in environments
    - International and intranational differences in medical technology and the use of medical technology
    - Diagnostic practices of physicians
    - Variation in physician practice patterns by region
- Person
    - Age
    - Sex
    - Ethnicity
    - Social habits (smoking, diet, alcohol)
    - Genetic background
    - Emotional and mental characteristics

Figure 1–1 AIDS Cases Diagnosed in Ohio by Year, 1983–1995.

Source: Reprinted from *Prevention Monthly,* Vol. 19, No. 3, p. 6, 1996, Ohio Department of Health.



which was then followed by a return to previous levels in 1994 and 1995. What happened in 1993 that resulted in such a large increase in the number of newly diagnosed AIDS cases?

This is an example of how the time variable can affect the number of cases diagnosed. In 1993, the case definition of AIDS changed so that individuals who were human immunodeficiency virus (HIV) positive were designated as having full-blown AIDS at an earlier point in the progression of their disease. In 1993, the case definition was expanded to include HIV-positive cases with low CD4 counts, pulmonary tuberculosis, and recurrent pneumonia as AIDS qualifying conditions. The result was that a large number of HIV-positive individuals who already had one of these conditions suddenly qualified as AIDS cases.

Now let's return to our discussion of the crude death rate. Crude rates do not allow for valid comparisons across populations because of differences in the populations—primarily age. This is because age is the most important variable that influences mortality. To illustrate, let's compare two hypothetical crude mortality rates for the states of Arizona (10.9/1,000) and Alaska (4.4/1,000). The conclusion drawn from a comparison of the crude mortality rates is that the death rate is 148% higher in Arizona than in Alaska: $(10.9 - 4.4)/4.4$. However, the discrepancy is due largely to the age differences in the populations of Arizona and Alaska. In general, the population in Arizona is older than the population in Alaska. Without adjusting the rate, one might erroneously conclude that the Alaskan population was healthier than the population of Arizona. In this example, the comparison is confounded by age. **Confounding factor** is a general term used to describe the effect of a third variable on the estimate of risk of a health outcome.

Confounding occurs when a third factor related to outcome is differentially distributed across the levels (or categories) of a variable of interest. When this happens, we must take measures to separate the effect of the **confounding variable**—in this case, age—from the effect of the variable of interest. We can accomplish this by selecting subjects to be compared so that they are matched with respect to the confounding variables, or by using statistical adjustments during analysis to remove the effect of the confounding variable. For example, review the data in Table 1–3. Analysis of the data reveals that the overall crude rate is less for blacks than for whites but that the age-specific death rate for blacks is higher than the rates for whites in every age group. Why is there such a contradiction? It is because the 2001 population of the state of Georgia consisted of old whites and young blacks—33.7% of the white population was 24 years old or younger, and 43.1% of the black population was 24 years old or younger.

Table 1–3 Age-Specific Death Rates per 1,000 Population, State of Georgia, 2001

Race	*Crude Rate*	*< 1 Yr.*	*1–4 Yrs.*	*5–14 Yrs.*	*15–24 Yrs.*	*25–44 Yrs.*	*45–64 Yrs.*	*≥ 65 Yrs.*
White	8.15	6.25	0.42	0.18	0.92	1.49	6.53	51.25
Black	7.04	13.33	0.51	0.24	1.04	2.54	10.68	59.02

Source: United States Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), CDC On-line Database, wonder.cdc.gov.

Age-Specific Death Rates

In Table 1–3, we see the age-specific death rates (ASDR) for both whites and blacks. The ASDR is calculated as follows:

$$\frac{\text{No. of deaths in the age group of interest} \times 10^n}{\text{Estimated mid-period population in the age group of interest}}$$

Age-Adjusted Death Rates

Age-adjusted death rates are used when there are differences in the age distribution for the populations that are being compared. In Table 1–4, you can see that the population proportions for each age group vary slightly by race. For example, the proportion of whites that are older than age 65 is 0.115 (11.5%) and the proportion of blacks that are older than 65 is 0.064 (6.4%). When we adjust the crude rate for age, we are constructing a summary rate that is free of age bias. In Table 1–4, the ASDR for each age group is expressed as a percentage. There are two methods for adjusting the crude death rate—direct and indirect. We will first discuss the direct method of standardization.

Table 1–4 Calculation of Crude Death Rate, State of Georgia, 2001

Age	*(a) White Population*	*(b) Pop. Prop.*	*(c) Deaths*	*(d) ASDR (c/a) × 100*	*(e) Black Population*	*(f) Pop. Prop.*	*(g) Deaths*	*(h) ASDR (g/f) × 100*
<1	85,648	0.015	535	0.62%	43,727	0.018	583	1.33%
1–4	309,451	0.054	129	0.04%	163,909	0.067	83	0.05%
5–14	768,143	0.134	137	0.02%	444,244	0.181	108	0.02%
15–24	770,501	0.134	706	0.09%	404,438	0.165	420	0.10%
25–44	1,811,149	0.315	2,698	0.15%	793,495	0.324	2,015	0.25%
45–64	1,338,338	0.233	8,746	0.65%	442,005	0.180	4,719	1.07%
65+	660,428	0.115	33,847	5.13%	157,770	0.064	9,312	5.90%
Total	5,743,658	1.000	46,798	0.81%	2,449,588	1.000	17,240	0.70%

Crude Death Rate = 0.81/100 Crude Death Rate = 0.70/100

Source: United States Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), CDC On-line Database, wonder.cdc.gov.

Direct Standardization

To age-adjust the crude death rates, we compare the two groups being studied to a standard population. We then apply the ASDRs for each group to this standard population. As an example, we will use the data in Table 1–4 to standardize the crude death rates for whites and blacks in the state of Georgia. The crude death rate for whites is 0.81 per 100, and the crude death rate for blacks is 0.70 per 100. To calculate the standardized rate, we first calculate the ASDR for each age group in the two populations. We then combine the populations for each age group. By multiplying ASDR for each group by the combined population, we can obtain the expected number of deaths for each group as if the population for each age group were the same. For example, for the age group from 1 to 4 years, we add 309,451 and 163,909 to obtain a total of 473,360. We then multiply the combined population total for each age group by the ASDR to obtain the expected number of deaths for each age group in each of the populations being compared. Thus, the groups are compared on an equal basis. The expected death rate for each population is calculated as follows:

Group	*Age Group*	*Total Population*	*ASDR*	*Expected No. of Deaths*
White	1–4	473,360	0.0004	189.3
Black	1–4	473,360	0.0005	236.7

After we have calculated the expected number of deaths for each age group in each population, we sum the expected number of deaths in each population group, as in Table 1–5. For whites the total number of expected deaths is 59,744.1, and for blacks the total is 77,209.7. The expected number of deaths for each population group is then divided by the

Table 1–5 Calculation of Adjusted Death Rate, Direct Standardization, State of Georgia, 2001

Age	*(a) Total Population*	*(b) ASDR Whites*	*(c) Expected No. Deaths (a × b)*	*(d) ASDR Blacks*	*(e) Expected No. Deaths (a × d)*
< 1	129,375	0.62%	802.1	1.33%	1,720.7
1–4	473,360	0.04%	189.3	0.05%	236.7
5–14	1,212,387	0.02%	242.5	0.02%	242.5
15–24	1,174,939	0.09%	1,057.4	0.10%	1,174.9
25–44	2,604,644	0.15%	3,907.0	0.25%	6,511.6
45–64	1,780,343	0.65%	11,572.2	1.07%	19,049.7
65+	818,198	5.13%	41,973.6	5.90%	48,273.7
Total	8,193,246	0.81%	59,744.1	0.70%	77,209.7
			0.73%		0.94%

Standardized Age-Adjusted Rate = 0.73%
59,744.1/8,193,246

Standardized Age-Adjusted Rate = 0.94%
77,209.7/8,193,246

Source: United States Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), CDC On-line Database, wonder.cdc.gov.

combined population. The result is that the standardized age-adjusted death rate for blacks is slightly higher (0.94%) than that for whites (0.73%).

Even though the standardized adjusted rate is not "real," it allows researchers to make better comparisons between groups. The crude rates indicate that the mortality rate is slightly higher for whites than for blacks, but the adjusted rates indicate that mortality among blacks is slightly higher than that for whites. Without adjustment, we would make the assumption that mortality was slightly higher in the white population. An adjusted rate informs us that this may not necessarily be the case.

Indirect Standardization

The indirect method of standardization is used when ASDRs are not available, or when the population that we wish to compare is small, as when we are comparing hospital inpatients to much larger populations. When using this method, we use standard rates obtained from some population and apply them to our population of interest. The basic steps for indirect standardization appear in Exhibit 1–5. In our hypothetical example, we compare 2002 Utah hospital discharges that resulted in death due to pneumonia to the number of hospital discharges that resulted in death due to pneumonia in Salt Lake County, Utah.

In our calculations in Table 1–6, we see that the overall mortality rate due to pneumonia in the state of Utah is 5.08% and that the mortality rate in Salt Lake County is 5.12% [(100 × 100)/1,953]. Salt Lake County had 1.5 more deaths than what was expected on the basis of the standard rates for the state of Utah; therefore, the expected mortality rate is 5.04% [(98.5 × 100)/1,953]. To make the comparison to the standard rates, we calculate a

Exhibit 1–5 Basic Steps for Indirect Standardization

1. Determine the standard mortality rates for pneumonia in the state of Utah for the age groups of interest.
2. Multiply the ASDR for the state of Utah (column c) times the number of county discharges in each age category to obtain the expected number of deaths for each category (columns c $\times$ d = column f) in Salt Lake County, Utah.
3. Sum the number of expected deaths.
4. Compute the standard mortality ratio (SMR), which compares the number of actual or observed deaths to the number of expected deaths. In Table 1–6, the number of actual or observed deaths is 100, and the number of expected deaths is 98.5.
5. Multiply the SMR by 100. The SMR is interpreted as a percentage lesser or greater than that of the standard population.

standard mortality ratio (SMR). The SMR compares the actual number of deaths in the group under study (Salt Lake County) to the expected number of deaths based on the standard population rates that were applied to the study group. For the data in Table 1–6, the SMR is calculated as:

$$\text{SMR} = \frac{\text{Observed death rate}}{\text{Expected death rate}} = \frac{0.0512}{0.0504}$$

$$= 1.016 \times 100 = 101.6\%$$

Table 1–6 Mortality Rates Due to Pneumonia (ICD-9-CM Codes 480–486) 2002, Ages 35+, State of Utah versus Salt Lake County, Utah

	State of Utah			*Salt Lake County, Utah*		
Age	*(a) Utah Discharges*	*(b) No. Deaths*	*(c) ASDR (b × 100)/a*	*(d) County Discharges*	*(e) Observed Deaths*	*(f) Expected Deaths (c × d)*
35–45	344	7	2.03%	151	3	3.1
45–54	533	9	1.69%	227	3	3.8
55–64	684	17	2.49%	237	5	5.9
65–74	1,071	53	4.95%	371	17	18.4
75+	2,542	177	6.96%	967	72	67.3
Total	5,174	263	5.08%	1,953	100	98.5
			Observed Death Rate	5.12%		
			Expected Death Rate	5.04%		
			SMR =	0.0512 / 0.0504	1.1016	

Source: Utah Inpatient Hospital Discharge Dataset, Utah Office of Health Care Statistics, www.health.state.ut.us.

If the calculated SMR is equal to 100, the number of observed deaths is the same as the number of expected deaths. If the SMR is greater than 100, the number of observed deaths is greater than the number of expected deaths. The interpretation of the SMR is that Salt Lake County's pneumonia death rate is 1% greater than that for the state of Utah. Stated another way, the death rate is 1% greater than what would be expected on the basis of the **mortality rates** due to pneumonia for the entire state of Utah.

In summary, rates are adjusted to remove the effect of the confounding factor for which the adjustment has been made—in this case, age. However, it is always necessary to calculate the crude rate because this represents the actual event. An adjusted rate is used for comparative purposes; adjusted rates do not reveal the underlying raw data that are shown by the crude rates.

Race- and Sex-Specific Death Rates

Mortality rates may be calculated for any variable of interest, such as race or sex, using the same basic formula specified for calculating the crude death rate. Historically in the United States, men have had higher mortality rates than women, but the gap may be narrowing. In 1995, the U.S. sex-specific rate was 9.2 per 1,000 for men and 8.6 per 1,000 for women. However, in 2001, the sex-specific death rate for men was 8.45 per 1,000 for men and 8.49 per 1,000 for women (Table 1–7).

Table 1–7 Sex-Specific Death Rates, United States, 2001

	Women			*Men*		
Age	*Population*	*Deaths*	*Rate/ 1,000*	*Population*	*Deaths*	*Rate/ 1,000*
Under 1 Year	1,968,011	12,091	6.14	2,057,922	15,477	7.52
1–4 years	7,491,412	2,208	0.29	7,841,553	2,899	0.37
5–9 years	9,861,089	1,366	0.14	10,347,035	1,727	0.17
10–14 years	10,199,195	1,561	0.15	10,711,245	2,441	0.23
15–19 years	9,847,662	3,789	0.38	10,423,650	9,766	0.94
20–24 years	9,630,499	4,500	0.47	10,080,924	14,197	1.41
25–34 years	19,698,788	12,926	0.66	20,116,087	28,757	1.43
35–44 years	22,675,474	33,510	1.48	22,464,812	58,164	2.59
45–54 years	19,971,971	63,217	3.17	19,256,395	104,848	5.44
55–64 years	13,160,005	99,181	7.54	12,155,918	144,958	11.92
65–74 years	10,020,545	189,379	18.90	8,301,935	241,581	29.10
75–84 years	7,585,929	361,187	47.61	4,996,556	340,742	68.20
85 years and over	3,127,729	447,998	143.23	1,320,580	217,533	164.73
Total	145,238,309	1,232,913	8.49	140,074,612	1,183,090	8.45

Source: United States Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), CDC On-line Database, wonder.cdc.gov.

It would be misleading to review the **sex-specific death rates** without review of the individual age-specific rates. Table 1–7 indicates that the death rate for men is higher for every age group. If we want to determine why the death rate of men is higher than that for women, we can compare causes of death by sex and age group. For example, in the combined age groups from 15 to 44 years, the death rate for men is higher than that for women because accidental death is the leading cause of death for men in these age groups. Sex-specific diseases may account for the differences in the death rates for other age groups, such as prostate cancer in men and breast cancer in women. Calculating the age-specific rates and the sex-specific rates can help us better understand what is taking place in the health care environment.

Cause-Specific Death Rates

The **cause-specific death rate** is the death rate due to a specified cause. It may be stated for an entire population or for any age, sex, or race. The numerator is the number of deaths due to a specified cause and the denominator is the size of the population at midyear. It is usually expressed in terms of a rate per 100,000 ($10^n = 10^5 = 100{,}000$). The formula is:

$$\frac{\text{Deaths assigned to a specified cause during a given time interval} \times 100{,}000}{\text{Estimated midinterval population}}$$

Table 1–8 presents the cause-specific death rates for males and females. The cause-specific death rate for pneumonia in the population aged 45 or older is 62.76 per 100,000 for women and 60.08 per 100,000 for men. While the overall cause-specific death rate for women is higher for women than for men, the cause-specific rates for each age group are consistently higher for men than for women. In reviewing the rates in Table 1–6, we can also see that the death rate increases with age for both men and women.

Table 1–8 Cause-Specific Mortality Rates, By Sex, Due to Influenza and Pneumonia (ICD-10 Codes J10–J18.9), Age 45+, United States, 2001

	Women			*Men*			
Age	*Population*	*Deaths*	*Rate/ 100,000*	*Population*	*Deaths*	*Rate/ 100,000*	*ASDR*
45–54 years	19,971,971	702	3.51	19,256,395	1,099	5.71	4.59
55–64 years	13,160,005	1,117	8.49	12,155,918	1,587	13.06	10.68
65–74 years	10,020,545	2,918	29.12	8,301,935	3,732	44.95	36.29
75–84 years	7,585,929	9,383	123.69	4,996,556	9,294	186.01	148.44
85 years and over	3,127,729	19,689	629.50	1,320,580	10,502	795.26	678.71
Total	53,866,179	33,809	62.76	46,031,384	26,214	56.95	60.08

Source: United States Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), CDC On-line Database, wonder.cdc.gov.

Case Fatality Rate

The **case fatality rate** or killing power of a disease measures the probability of death among the diagnosed cases of a disease. The higher the ratio, the more virulent the infection. It is most often used as a measure in acute infectious disease. The case fatality rate is not useful in chronic disease because such diseases have a longer and more variable course.

The formula for the case-fatality rate is:

$$\frac{\text{No. of deaths due to a disease during a given time interval} \times 100}{\text{No. of cases of the disease in the same time interval}}$$

Proportionate Mortality Ratio

The **proportionate mortality ratio** (PMR) describes the proportion of all deaths for a given time interval that are due to a specific cause. Each cause is expressed as a percentage of all deaths, and the sum of all the causes is 1.00 (100%). The PMR is not a mortality rate, since the denominator is all deaths, not the population in which the deaths occurred. Its formula is:

$$\frac{\text{No. of deaths due to a disease during a given time interval} \times 100}{\text{No. of deaths from all causes in the same time interval}}$$

The PMR is often used to make comparisons between and within age groups and occupational groups, as well as for the general population. The PMR for pneumonia appears in Table 1–9.

Maternal Mortality Rate

The **maternal mortality rate** measures deaths associated with pregnancy. Pregnancy often places a woman at risk for medical problems that would not usually be encountered in the nonpregnant state, such as hemorrhage or toxemia of pregnancy. Pregnancy also complicates chronic conditions such as diabetes mellitus and heart disease. In some women, pregnancy precipitates gestational diabetes. The maternal mortality rate is calculated only for deaths that are related to pregnancy; thus, if a pregnant woman is killed in an automobile accident, the death is not considered a pregnancy-related death.

The numerator is the number of deaths assigned to causes related to pregnancy during a given time period; the denominator is the number of live births reported during the same period. Because the maternal mortality rate is usually very small, it is usually expressed as the number of deaths per 100,000 live births.

Table 1–9 Proportionate Mortality Ratios for Influenza and Pneumonia (ICD-10 Codes J10–J18.9), United States, 2001

Age	*Influenza and Pneumonia Deaths*	*Total Deaths*	*PMR/100*
0–4 years	411	32,675	1.26
5–9 years	46	3,093	1.49
10–14 years	46	4,002	1.15
15–19 years	66	13,555	0.49
20–24 years	115	18,697	0.62
25–34 years	339	41,683	0.81
35–44 years	983	91,674	1.07
45–54 years	1,801	168,065	1.07
55–64 years	2,704	244,139	1.11
65–74 years	6,650	430,960	1.54
75–84 years	18,677	701,929	2.66
85 years and over	30,191	665,531	4.54
Total	62,029	2,416,003	2.57

Source: United States Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), CDC On-line Database, wonder.cdc.gov.

Rates of Infant Mortality

There are three rates of infant mortality, all of which are based on age. Of the three, the infant mortality rate is the most commonly used measure for comparing health status between nations. All three rates are expressed in terms of the number of deaths per 1,000.

Neonatal Mortality Rate

The neonatal period is defined as the period from birth up to but not including 28 days of age. The numerator is the number of deaths of infants under 28 days of age during a given time period; the denominator is the total number of live births reported during the same period. The **neonatal mortality rate** may be used as an indirect measure of the quality of prenatal care and/or the mother's prenatal behavior (e.g., tobacco, alcohol, and drug use).

Postneonatal Mortality Rate

The postneonatal period is the time period from 28 days of age up to but not including one year of age. The numerator is the number of deaths among children from age 28 days up to but not including one year of age during a given time period; the denominator is the total number of live births reported less the number of neonatal deaths during the same period. The **postneonatal mortality rate** is often used as an indicator of the quality of the infant's home environment.

Infant Mortality Rate

In effect, the **infant mortality rate** is a summary of the neonatal and postneonatal mortality rates. The numerator is the number of deaths among children under one year of age; the denominator is the number of live births reported during the same period. Table 1–10 provides a summary of these rates.

Table 1–10 Frequently Used Mortality Measures

Measure	*Numerator (x)*	*Denominator*	*10^n*
Crude death rate	Total no. of deaths reported during given time interval	Estimated midinterval population	1,000 or 10,000
Cause-specific death rate	Total no. of deaths due to a specific cause during a given time interval	Estimated midinterval population	100,000
Proportionate mortality ratio	Total no. of deaths due to a specific cause during a given time interval	Total no. of deaths from all causes during the same time interval	100 or 1,000
Case fatality rate	Total no. of deaths assigned to a specific disease during a given time interval	Total no. of cases of the disease during the same time interval	100
Neonatal mortality rate	No. of deaths under 28 days of age during a given time interval	No. of live births during the same time interval	1,000
Postneonatal rate	No. of deaths from 28 days up to and not including one year of age, during a given time interval	No. of live births during the same time interval less neonatal deaths	1,000
Infant mortality rate	No. of deaths under one year of age during a given time interval	No. of live births during the same time interval	1,000
Maternal mortality rate	No. of deaths assigned to pregnancy-related causes during a given time interval	No. of live births during the same time interval	100,000

FREQUENTLY USED MEASURES OF MORBIDITY

Some commonly used measures to describe the presence of disease in a community or a specific location, such as a nursing home, are incidence and prevalence rates. Disease can be illness, injury, or disability, and measures can be further elaborated into specific measures of age, sex, race, or other characteristics of a particular population.

Incidence Rate

The **incidence rate** is the commonly used measure for comparing frequency of disease in populations. Populations are compared using rates instead of raw numbers because rates adjust for differences in the size of the populations. The incidence rate expresses the probability or risk of illness in a population over a period of time. The formula for calculating the incidence rate is:

$$\frac{\text{Total no. of new cases of a specific disease during a given time interval} \times 10^n}{\text{Total population at risk during the same time interval}}$$

For the incidence rate, the denominator represents the population from which the case in the numerator arose, such as a nursing home, school, or company. For 10^n, a value is selected so that the smallest rate calculated results in a whole number.

Prevalence Rate

The **prevalence rate** is the proportion of persons in a population that have a particular disease at a specific point in time, or over a specified period of time. The formula for calculating the prevalence rate is:

$$\frac{\text{All new and preexisting cases of a specific disease during a given time interval} \times 10^n}{\text{Total population during the same time period}}$$

Incidence and prevalence rates are often confused. The rates differ based on which cases are included in the numerator. The numerator of the incidence rate is *new cases* occurring during a given time period; the numerator of the prevalence rate is *all cases* present during a given time period. In comparing the two, you can see that the incidence rate includes only individuals whose illness began during a specified period of time, whereas the numerator for the prevalence rate includes all individuals ill from a specified cause, regardless of when the illness began. A case is counted in prevalence until the individual recovers. Exhibit 1–6 presents an example of incidence and prevalence rates in a nursing home.

At times we may be interested in tracking prevalence rates more closely—for example, tracking *Klebsiella pneumoniae* on a daily basis. We can do this by calculating the **point**

Exhibit 1–6 Calculation of Incidence and Prevalence Rates of *Klebsiella pneumoniae* at the Manor Nursing Home, Month of January

At Manor Nursing Home, 10 new cases of *Klebsiella pneumoniae* occurred in January. For the month of January there were a total of 17 cases of *Klebsiella pneumoniae*. The facility had 250 residents during January.

What are the incidence and prevalence rates for *Klebsiella pneumoniae* during January?

Incidence Rate

1. Identify the variable of interest (numerator) and population at risk (denominator).
   Numerator: Total no. of new cases of *Klebsiella pneumoniae* in January
   Denominator: Total no. of nursing home residents in January
2. Identify the numerator and denominator.
   Numerator: 10
   Denominator: 250
3. Set up the rate.
   10/250
4. Divide the numerator by the denominator and multiply by 100 ($10^n = 10^2$).
   $(10/250) \times 100 = 0.04 = 4.0\%$

The incidence rate for *Klebsiella pneumoniae* for the month of January is 4.0%.

Prevalence Rate

1. Identify the variable of interest (numerator) and population at risk (denominator).
   Numerator: Total no. of cases of *Klebsiella pneumoniae* in January
   Denominator: Total no. of nursing home residents in January
2. Identify the numerator and denominator.
   Numerator: 17
   Denominator: 250
3. Set up the rate.
   17/250
4. Divide the numerator by the denominator and multiply by 100 ($10^n = 10^2$). $(17/250) \times 100 = 0.068\%$

The prevalence rate for *Klebsiella pneumoniae* for the month of January is 6.8%.

prevalence rate. The point prevalence rate is the number of cases of a specific disease at a specific point in time. The point prevalence rate is more narrow in its time frame than the general prevalence rate. Table 1–11 displays the point prevalence rates for each day during one week in January.

For a summary of morbidity measures, see Table 1–12.

Table 1–11 Point Prevalence Rates of *Klebsiella pneumoniae* for the Manor Nursing Home, Week of January 3

	Sun.	*Mon.*	*Tues.*	*Weds.*	*Thurs.*	*Fri.*	*Sat.*
No. of cases	10	12	14	13	15	16	16
No. of residents	250	250	250	250	250	250	250
Point Prevalence rate	4.0%	4.8%	5.6%	5.2%	6.0%	6.4%	6.4%

Table 1–12 Frequently Used Measures of Morbidity

Measure	*Numerator*	*Denominator*
Basic formula for computing rates	No. of events occurring during a given time interval	No. of cases or population at risk during the same time interval
Incidence rate	Total no. of new cases of a specific disease during a given time interval	Total population at risk during the same time interval
Prevalence rate	All new and preexisting cases of a specific disease during a given time interval	Total population during the same time interval
Relative risk	Risk for exposed group	Risk for unexposed group
Relative risk using incidence rates	Incidence rate for group of primary interest	Incidence rate for comparison group
Attributable risk	Risk for exposed group minus risk for unexposed group	Risk for exposed group

RELATIVE MEASURES OF DISEASE FREQUENCY

Risk Ratio/Relative Risk

Relative risk (RR) is a ratio that compares the risk of disease or other health event between two groups. What we are comparing is the actual risk of illness between the two groups. In calculating relative risk, we are using the actual rates of illness for each group to make the comparison. The two groups may be differentiated by demographic variables, such as sex or race, or by exposure to a suspected risk factor.

The group of primary interest is labeled as the exposed group, and the comparison group is labeled the unexposed group. The exposed group is placed in the numerator, and the unexposed group is placed in the denominator:

$$\frac{\text{Risk for exposed group}}{\text{Risk for unexposed group}}$$

A risk ratio of 1.0 indicates that the risk is identical in both groups; a risk ratio greater than 1.0 indicates that the risk is greater for the numerator group; and a risk ratio of less than 1.0 indicates that the risk is less for the numerator group.

As an example, we can compare the risk of death due to malignancies in men versus women in Michigan in 2001. First, the collected data are summarized in a two-by-two table. Two-by-two refers to two variables, each with two categories, as shown in Table 1–13.

Table 1–13 Relative Risk of Death Due to Malignancies, Women versus Men Aged 65+, State of Michigan, 2001

	Death Due to Pneumonia		
Sex	*Yes*	*No*	*Total*
Men	7,153 (a)	21,507 (b)	28,660 (a + b)
Women	6,565 (c)	28,890 (d)	35,455 (c + d)

Risk of illness among men:
$a/(a + b)$ = 7,153/(7,15321,507) = 0.2496
Risk of illness among women
$c/(c + d)$ = 6,565/(6,565 + 28,890) 0.1852
Risk ratio, men to women: 0.2496/1852 = 1.34
Thus, the risk of death due to malignancy among men aged 65+ is 1.3 times greater than the risk of death due to malignancy in women in the same age group.

Source: United States Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), CDC On-line Database, wonder.cdc.gov.

To determine the risk of death among men, we compare the total number of men who died from malignancies (a = 7,153) to the total number of men in the group of interest ($a + b$ = 7,153 + 21,507). The same procedure is followed to determine the risk of death due to pneumonia among women. The two ratios are then compared to determine the RR of death due to malignancies among men as compared to women. A summary of these calculations appears in Table 1–13. Note that the RR in each group is somewhat high, 25.0% and 18.5% respectively. Deaths due to malignancies were the second leading cause of death in the state of Michigan in 2001.

Instead of using the risk ratios to compare risks between groups, we can use actual rates to make the same comparisons. In Table 1–14, hypothetical mortality rates are used to com-

Table 1–14 Lung Cancer Data

Cigarettes/Day	*Death Rate/1,000/Year*
0	0.07
1–14	0.57
15–24	1.39
25+	2.27

Rate ratios:
1–14 cigarettes/day to nonsmokers:
0.57/0.07 = 8.1
15–24 cigarettes/day to nonsmokers:
1.39/0.07 = 19.9
25+ cigarettes/day to nonsmokers:
2.27/0.07 = 32.4

Thus, the risk is 8.1 times greater for those who smoke 1 to 14 cigarettes per day than for nonsmokers; 19.9 times greater for those who smoke 15 to 24 cigarettes per day than for nonsmokers; and 32.4 times greater for those who smoke 25 cigarettes per day than for nonsmokers.

Source: Adapted from *Principles of Epidemiology: An Introduction to Applied Epidemiology and Biostatistics,* p. 95, 1992, United States Department of Health and Human Services, Public Health Service.

pare the risk of death due to lung cancer by number of cigarettes smoked per day. Using the same procedure, we can compare the risk of stroke between men who smoke and men who do not smoke. In this example, we are trying to determine if there is a greater risk of stroke among men who smoke than among men who do not smoke. The statistic is called "relative risk using incidence rates" and is calculated as:

$$\frac{\text{Incidence rate for group of primary interest}}{\text{Incidence rate for comparison group}}$$

The data for this example are presented in Table 1–15. Note that these ratios represent only RR, or the possibility of acquiring an illness, in comparison to another group.

Odds Ratio

The **odds ratio** (OR) is another relative measure of occurrence of illness. The odds in favor of a particular event are defined as the frequency with which the event occurs divided by the frequency with which it does not occur. Estimates of RR and the OR are both used to measure the strength of the association between exposure and disease. The OR is an estimate of

Table 1–15 Twelve-Year Risk of Stroke Among Male Smokers and Nonsmokers

	Stroke		
Smokers	*Yes*	*No*	*Total*
Yes	171	3,264	3,435
No	117	4,320	4,437
Total	288	7,584	7,872

Risk of stroke among smokers:
171/3,435 = 0.049
Risk of stroke among nonsmokers:
117/4,437 = 0.026
Risk of male smokers to male nonsmokers:
0.049/0.026 = 1.88
Thus, the risk of stroke is 1.88, or almost two times greater in men who smoke than men who do not smoke.

RR. It is calculated from data obtained from retrospective studies where actual incidence rates are not calculated.

To calculate the OR, a two-by-two table is first constructed as shown in Table 1–16. Exhibit 1–7 displays the calculation of the odds ratio using the data from Table 1–15. The results indicate that the odds of having a stroke is 1.93 times greater in men who smoke than in men who do not smoke.

Table 1–16 Two-by-Two Table for Odds Ratio

	Disease	
Risk Factor	*Cases*	*Non-cases*
Present	*a*	*b*
Absent	*c*	*d*

Odds Ratio = $(a \times d)/(b \times c)$, where a = number of persons with disease and with exposure of interest, b = number of persons without disease and with exposure of interest, c = number of persons with disease but without exposure of interest, and d = number of persons without disease and without exposure of interest.
$a + c$ = total persons with disease (cases)
$b + d$ = total persons without disease (controls)

Exhibit 1–7 Procedure for Calculating Odds Ratio (OR)

$$OR = (a/b) \div (c/d) = \frac{(a \times d)}{(b \times c)}$$

$$OR = \frac{171 \times 4{,}320}{3{,}264 \times 117} = 1.93$$

The probability of having a stroke is 1.93 times greater in men who smoke than in men who do not smoke.

The interpretation of the OR is similar to that for RR. If the exposure is not related to the diagnosis, the OR will equal 1; if the exposure is positively related to the disease, the OR will be greater than 1; and if the exposure is negative, the OR will be less than 1. We could also apply this same ratio, or any others, to the acute care setting. An outcomes evaluator learns that patients on the surgical unit were exposed to the *E. coli* bacterium. Data were collected for two weeks to determine if the odds for obtaining *E. coli* infection were greater for patients on the surgical units than for patients hospitalized on the medical unit. The data are displayed in Table 1–17. As you can see from the calculations for the OR, the odds or probability of obtaining an *E. coli* infection is 2.68 times greater for a patient hospitalized on the surgical unit than for a patient hospitalized on a medical unit.

Table 1–17 *E.Coli* Infections of Medical and Surgical Patients

	Nosocomial Infection		
Hospital Unit	*Yes*	*No*	*Total*
Surgical Unit	20	628	648
Medical Unit	10	842	852
Total	30	1,470	1,500

The odds ratio is calculated as follows:
OR = $(a \times d)/(b \times c) = (20 \times 842)/(10 \times 628) = 2.68$

When the health outcome is uncommon, the OR approximates the RR. Using the same data from Table 1–17, we can determine the RR as follows:

Risk of infection on surgical unit: $20/648 = 0.031$

Risk of infection on medical unit: $10/852 = 0.012$

Risk of infection on surgical unit compared to medical unit: $0.031/0.012 = 2.58$

As you can see, the results for both the OR and the RR are similar: 2.68 and 2.58, respectively.

Attributable Risk

The **attributable risk** (AR) is a measure of the impact of a disease or other causative factor on a population. With this calculation, we assume that the occurrence of the disease in a group not exposed to the risk factor represents the baseline or expected risk for that disease; any risk above that level in the exposed group is attributed to exposure to the risk factor. Basically, the assumption is that the disease will occur in some individuals even without exposure to a given risk factor. The AR measures the additional risk of illness as a result of an individual's exposure to the risk factor. With AR, we attempt to answer the question, "How much of the disease that occurs can be attributed to a certain exposure?" and subsequently, "How much of the risk of disease can we prevent if we eliminate the exposure to the risk factor in question?"

$$\frac{(\text{Risk for exposed group}) - (\text{risk for unexposed group}) \times 100}{\text{Risk for exposed group}}$$

Using the lung cancer data from Table 1–14, we calculate the attributable proportion as outlined in Exhibit 1–8.

Exhibit 1–8 Calculation of Attributable Proportion

1. Identify the exposed group rate. Lung cancer death rate for smokers of 1–14 cigarettes per day = 0.57 per 1,000 per year
2. Identify the unexposed group rate.

   0.07 per 1,000 per year
3. Calculate the attributable proportion.

   $$\frac{0.57 - 0.07}{0.57} \times 100 = 87.7$$

The conclusion from the calculation of the attributable proportion is that 87.7% of the lung cancer cases are due to or attributed to smoking 1 to 14 cigarettes per day. Approximately 12% (1.00 − 0.877) of the cases in this group would have occurred without exposure to the risk factor—in this case, cigarettes. By carrying out the calculations for the remaining two groups, we can see that the AR increases with the number of cigarettes smoked per day.

$$\text{AR 15–24 cigarettes/day} = [(1.39 - 0.07)/1.39] \times 100 = 95.0\%$$
$$\text{AR 25+ cigarettes/day} = [(2.27 - 0.07)/2.27] \times 100 = 96.9\%$$

Approximately 5% and 3%, respectively, of the individuals in these two groups would have acquired the disease regardless of whether or not they smoked cigarettes.

KAPLAN-MEIER SURVIVAL ANALYSIS

Many individuals within the health information management profession are employed in tumor registries or in the capacity of assisting researchers in analyzing data from clinical trials. In clinical trials, the researcher is interested in determining whether a specific medical or surgical intervention improves survival for a particular condition. A major criterion in measuring the success of a clinical trial is the survival time of individuals undergoing the experimental treatment. In survival analysis we are examining the survival rates as a result of a clinical trial involving a medical or surgical intervention. A major problem in conducting survival analysis is that patients may be lost to follow up or some may be **censored**. A censored patient is one who for some reason is unable to complete the study.

There are several methods for analyzing survival rates, but we will limit the discussion to the **Kaplan Meier method**, a type of life table analysis, since it is most often used in analysis of data collected from clinical trials. **Kaplan-Meier survival analysis** requires a dichotomous outcome such as survival/death or improvement/no improvement.

The major reason for using the Kaplan Meier method is that it takes into account some of the problems commonly encountered when conducting prospective studies. The Kaplan Meier method compensates for subjects who are lost to follow-up or who are unable to complete the study. To conduct an accurate survival analysis, we need to know:

- the reason for patients' withdrawal from the study (i.e., death, loss to follow-up, or censorship)
- the date of withdrawal from study (i.e., date of death, date patient last seen alive or lost to follow-up, or date withdrawn from study)

When survival time is censored, the subject is alive at the time of analysis, or was alive at the time last seen. Survival times tagged with a "+" indicate that they are censored. Table 1–18 presents some hypothetical data for 10 patients in a clinical trial for treatment of bladder cancer. The survival times, in months (column 1), for each patient are rank ordered from lowest to highest.

Each row in Table 1–18 represents an interval. The first row is the first study interval. An interval is a death-free time period. So row 1, column 6, represents a death-free time period of less than 23 months. This is interpreted as meaning that the probability (p_x) of surviving up to but less than 23 months is 1.000 (10/10). The p_x of the first interval is always 1.000 because the first death ends the first interval. The occurrence of a death ends one death-free interval and begins another.

Table 1–18 Hypothetical Data on Survival Times for Bladder Cancer Patients

(1) Survival Time Mo.	*(2) No. Living Prior to Subject's Death*	*(3) No. Living After Subject's Death*	*(4) # Lost to Follow-Up*	*(5) p_x*	*(6) Interval for p_x (Mo.)*	*(7) p_x at End of Interval*
—	—	—	—	1.000	0 to <23	1.000
23	10	9	—	0.900	23 to <34	0.900
34	9	8	—	0.889	34 to <37	0.800
37	8	7	—	0.875	37 to <41	0.700
40+			1	—	—	—
41	6	5	—	0.833	41 to <42	0.583
42	5	4	—	0.800	42 to <43	0.466
43	4	3	—	0.750	43 to <45	0.350
45	3	2	—	0.667	45 to <47	0.233
47	2	1	—	0.500	47 to <48	0.117
48+	1	1	1	1.000	>48	0.117

Column 1 in Table 1–18 indicates the survival time, in months, for each subject. Two patients were lost to follow-up, as indicated by "+" — one at 40 months and one at 48 months. Patients lost to follow-up are not included in the calculations of survival rates. Columns 2, 3, and 4 indicate the number surviving before and after each death and the number lost to follow-up during that interval. Column 5 is the proportion of patients surviving the interval and is obtained by dividing the proportion surviving from the beginning of the interval—from the time of the previous death to just before the next death. For example, for the interval "23 to <34," 10 patients were alive at the start of the interval, and 9 were alive at the end. To obtain p_x, divide 9 by 10 to obtain 0.900.

Column 6 is the death-free period—that is, the time of the last death to the time of the next death. Column 7, p_x, is the proportion of subjects surviving from the beginning of the study to the end of the interval. The p_x is obtained by multiplying the p_x values of all the intervals up to and including the row of interest. For the survival time of 34 months, p_x is obtained by multiplying $1.000 \times 0.900 \times 0.889 = 0.800$. Based on the calculations in Table 1–18, the probability of surviving 48 months is 0.117.

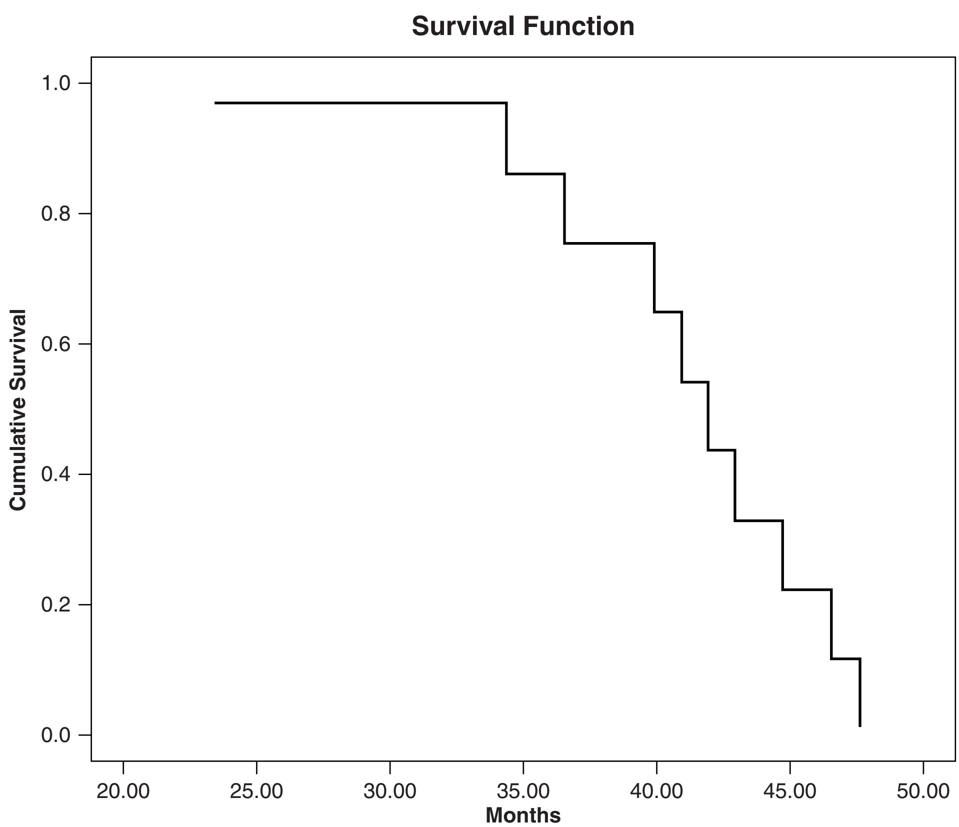
We can use SPSS (Statistical Package for the Social Sciences) to conduct the Kaplan-Meier survival analysis. SPSS is a microcomputer statistical package that we will use throughout this text to solve statistical problems. For the Kaplan-Meier survival analysis, two columns on the data sheet need to be completed. The first column indicates the survival time, in months, for each case; the second column indicates whether the survival time is censored. This can be accomplished by assigning "1" for uncensored survival times and "2" for censored survival times under the "Define Variable" selection. An example of the SPSS data sheet appears in Exhibit 1–9.

Exhibit 1–9 SPSS Data Sheet for Survival Data

Survival Time (Mo.)	*Status*
23.00	Uncensored
34.00	Uncensored
37.00	Uncensored
40.00	Censored
41.00	Uncensored
42.00	Uncensored
43.00	Uncensored
45.00	Uncensored
47.00	Uncensored
48.00	Censored

After completing the data sheet, select "Survival" and then "Kaplan Meier" under the "Statistics" menu. The output, including the survival graph, appears in Figure 1–2. Note that the SPSS printout provides only the p_x (cumulative survival)—the probability of surviving to the end of the interval.

Figure 1–2 SPSS Output for Kaplan-Meier Survival Analysis



Survival Analysis for MONTHS (survival time in months)

Time	*Status*	*Cumulative Survival*	*Standard Error*	*Cumulative Events*	*Number Remaining*
23.00	Uncensored	0.9000	0.0949	1	9
34.00	Uncensored	0.8000	0.1265	2	8
37.00	Uncensored	0.7000	0.1449	3	7
40.00	Censored	0.6000	0.1549	4	6
41.00	Uncensored	0.5000	0.1581	5	5
42.00	Uncensored	0.4000	0.1549	6	4
43.00	Uncensored	0.3000	0.1449	7	3
45.00	Uncensored	0.2000	0.1265	8	2
47.00	Uncensored	0.1000	0.0949	9	1
48.00	Censored	0.0000	0.0000	10	0

Number of Cases: 10 Censored: 0 (.00%) Events: 10

Survival Time	*Standard Error*	*95% Confidence Interval*
Mean: 40.00	2.32	(35.45, 44.55)
Median: 41.00	1.58	(37.90, 44.10)

SPSS provides a summary of the number of cases included in the analysis, including the number of censored cases. The confidence intervals for the mean and median survival times also are provided. (We will discuss confidence intervals in Chapter 5.) The graph depicts the cumulative survival rate for the group under study. Time, in months, is displayed on the *x*-axis, and proportion surviving is displayed on the *y*-axis.

CONCLUSION

In this chapter, we have discussed rates, ratios, and proportions in the form of mortality and morbidity rates and RR. Facility-based morbidity and mortality rates can be compared with community, state, and national rates after adjustment. We can adjust rates using either the direct or the indirect method. Crude rates are important for internal analysis or other non-comparative purposes.

We also reviewed various ratios that are used to measure frequency of disease. Using the various risk ratios and the OR, we can compare risk of certain diseases and causes of morbidity between groups.

Last, we discussed one method commonly used for survival analysis—the Kaplan Meier method. Survival analysis is a tool often used in tumor registries and when analyzing results of clinical trials.

ADDITIONAL RESOURCES

Gail, M.H. and Benichou, J., eds. 2000. *Encyclopedia of epidemiologic methods.* Wiley Reference Series in Biostatistics. Hoboken, NJ: Wiley Publishers.

Gordis, L. 1996. *Epidemiology.* Philadelphia: W.B. Saunders.

Jekel, J.F., et al. 1996. *Epidemiology, biostatistics and preventive medicine.* Philadelphia: W.B. Saunders.

Morton, R., et al. 1989. *A study guide to epidemiology and biostatistics.* Gaithersburg, MD: Aspen Publishers, Inc.

Anderson, K.N., Anderson, L.E., and Glanze, W.D. *Mosby's medical, nursing, and allied health dictionary.* 4th ed. 1994. St. Louis, MO: C.V. Mosby.

Norman, G.R., and Streiner, D.L. 2000. *Biostatistics: The bare essentials.* Ontario, Canada: Hamilton.

Ohio Department of Health. 1996. *Prevention Monthly* 19, no. 3:6.

Pagano, M., and Gauvreau, K. 2000. *Principles of biostatistics.* Pacific Grove, CA: Brooks Cole Publishing.

Ratelle, S. 2001. *Preventive medicine and public health: PreTest self-assessment and review.* New York: McGraw-Hill/Appleton & Lange McGraw-Hill.

U.S. Department of Health and Human Services (US DHHS), Public Health Service. 1992. *Principles of epidemiology: An introduction to applied epidemiology and biostatistics.*

U.S. Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS), Office of Analysis, Epidemiology, and Health Promotion (OAEHP), Compressed Mortality File (CMF) compiled from CMF 1968–1988 Series 20, No. 2A 2000, CMF 1989–1998, Series 20, No. 2E 2003 and CMF 1999–2001, Series 20, No. 2G 2004 on CDC WONDER on-line database. wonder.cdc.gov.

U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for HIV, STD, and TB Prevention, Division of HIV/AIDS Prevention, AIDS Public Information Data Sets, CDC WONDER on-line data base. wonder.cdc.gov.

Utah Inpatient Hospital Discharge Data Set. http://hlunix.ex.state.ut.us/had.

Appendix 1–A

Exercises for Solving Problems

KNOWLEDGE QUESTIONS

1. Define the key terms listed at the beginning of this chapter.
2. Describe the differences and similarities between rates, ratios, and proportions.
3. Outline the procedure for age-adjusting crude mortality rates by the direct standardization method.
4. Describe the differences between the direct and indirect standardization methods of adjusting mortality and morbidity rates.
5. Describe the differences between neonatal mortality rate, postneonatal mortality rate, and infant mortality rate.
6. Describe the difference between incidence and prevalence rates.

MULTIPLE CHOICE

For questions 1 and 2, refer to the following table:

Age Group	*Population*	*No. of Deaths*
< 30	15,000	20
30–65	17,000	55
> 65	6,000	155

1. What is the crude mortality rate?
   a. 230
   b. 6.1 per 1,000
   c. 8.6 per 1,000
   d. 6.1 per 10,000

2. The age-specific death rate for the over-65 age group is:
   a. 155
   b. 25.8 per 1,000
   c. 1.55 per 10,000
   d. 25.8 per 10,000

PROBLEMS

1. Review the hypothetical data on deaths in the MICU in Table 1–A–1 and answer the questions that follow:
   a. What is the ratio of male deaths to female deaths?
   b. What proportion of the patients who died were admitted from the Emergency Department? What proportion were transfers from other hospitals?
   c. The total number of patients discharged from DRG 475 was 61. What is the case fatality rate for DRG 475?
   d. The total number of patients discharged from DRG 483 was 51. What is the case fatality rate for DRG 483?
   e. What is the relative risk of death for patients discharged from DRG 475 compared to discharges from DRG 483?

Table 1–A–1 Critical Care Hospital, Deaths in the MICU by DRG

DRG	*DRG Title*	*Adm. Source*	*Gender*	*LOS*
001	Craniotomy Age >17 W Cc	SNF	Male	2
014	Intracranial Hemorrhage & Stroke W Infarct	Other	Male	3
014	Intracranial Hemorrhage & Stroke W Infarct	Emerdept	Female	3
020	Nervous System Infection Except Viral Meningitis	Other	Female	15
075	Major Chest Procedures	Hospital	Male	6
105	Cardiac Valve & Oth Major Cardiothoracic Proc W/O Card Cath	Hospital	Female	23
123	Circulatory Disorders W Ami, Expired	Hospital	Male	7
123	Circulatory Disorders W Ami, Expired	Other	Male	1
123	Circulatory Disorders W Ami, Expired	Other	Male	4
123	Circulatory Disorders W Ami, Expired	Emerdept	Male	5
172	Digestive Malignancy W Cc	Emerdept	Male	1
172	Digestive Malignancy W Cc	Physician	Male	1
188	Other Digestive System Diagnoses Age >17 W Cc	SNF	Female	1
191	Pancreas, Liver & Shunt Procedures W Cc	Hospital	Male	9
202	Cirrhosis & Alcoholic Hepatitis	Physician	Female	15
202	Cirrhosis & Alcoholic Hepatitis	Other	Male	1
202	Cirrhosis & Alcoholic Hepatitis	Emerdept	Male	24

continued

DRG	*DRG Title*	*Adm Source*	*Gender*	*LOS*
205	Disorders Of Liver Except Malig,Cirr,Alc Hepa W Cc	Physician	Male	20
331	Other Kidney & Urinary Tract Diagnoses Age >17 W Cc	Physician	Male	44
357	Uterine & Adnexa Proc For Ovarian Or Adnexal Malignancy	Physician	Female	24
416	Septicemia Age >17	Hospital	Female	4
416	Septicemia Age >17	Other	Male	2
449	Poisoning & Toxic Effects Of Drugs Age >17 W Cc	Other	Male	1
473	Acute Leukemia W/O Major O.R. Procedure Age >17	Physician	Male	5
475	Respiratory System Diagnosis With Ventilator Support	Physician	Male	25
475	Respiratory System Diagnosis With Ventilator Support	Physician	Female	1
475	Respiratory System Diagnosis With Ventilator Support	Hospital	Female	1
475	Respiratory System Diagnosis With Ventilator Support	Other	Female	1
475	Respiratory System Diagnosis With Ventilator Support	Hospital	Female	21
475	Respiratory System Diagnosis With Ventilator Support	Other	Male	5
475	Respiratory System Diagnosis With Ventilator Support	Hospital	Female	8
475	Respiratory System Diagnosis With Ventilator Support	Emerdept	Female	10
475	Respiratory System Diagnosis With Ventilator Support	Clinic	Male	13
475	Respiratory System Diagnosis With Ventilator Support	SNF	Female	1
475	Respiratory System Diagnosis With Ventilator Support	Emerdept	Male	5
475	Respiratory System Diagnosis With Ventilator Support	Clinic	Female	4
475	Respiratory System Diagnosis With Ventilator Support	Emerdept	Male	3
475	Respiratory System Diagnosis With Ventilator Support	Other	Female	12
475	Respiratory System Diagnosis With Ventilator Support	Other	Female	5
483	Trac W Mech Vent 96+Hrs Or Pdx Except Face, Mouth & Neck Dx	Hospital	Female	30
483	Trac W Mech Vent 96+Hrs Or Pdx Except Face, Mouth & Neck Dx	Hospital	Male	19
483	Trac W Mech Vent 96+Hrs Or Pdx Except Face, Mouth & Neck Dx	Hospital	Male	22
483	Trac W Mech Vent 96+Hrs Or Pdx Except Face, Mouth & Neck Dx	Physician	Female	46
483	Trac W Mech Vent 96+Hrs Or Pdx Except Face, Mouth & Neck Dx	Other	Female	28

2. Review the data in Table 1–A–2 and answer the questions that follow.
   a. What is the case fatality rate for AIDS for the years 1981 through 1995?
   b. The midyear population for the state of Ohio in 1994 was 11,140,950. What is the incidence rate for AIDS for 1994?

Table 1–A–2 AIDS Cases in Ohio 1981–1995

Year of Diagnosis	*Total No. of New Cases*	*Cases Dead*
1981	2	2
1982	7	7
1983	27	25
1984	58	56
1985	120	113
1986	211	198
1987	401	374
1988	540	482
1989	631	537
1990	682	577
1991	763	644
1992	775	587
1993	1935	908
1994	947	259
1995	259	63

Source: Department of Health HIV/AIDS Surveillance Program, Columbus, OH, www.odh.state.oh.us.

3. Review the data in Table 1–A–3 and answer the questions that follow.
   a. What is the male-to-female ratio for AIDS in Ohio? In the United States?
   b. Out of the total number of AIDS cases in Ohio, what proportion are women? Of the total cases in the United States, what proportion are women?
   c. What proportion of the total AIDS cases in Ohio are ages 30 to 39? What proportion in the United States are ages 30 to 39?
   d. Calculate the proportion of AIDS cases in Ohio by race. Calculate the proportion of AIDS cases in the United States by race.
   e. How do the preceding ratios and proportions, Ohio versus United States, compare?

Table 1–A–3 Ohio AIDS Cases by Age, Race, and Sex, as of June 30, 2003; U.S. AIDS Cases 1981–1999

Demographics	*Total Ohio*	*Total U.S.*
Age		
<13	96	8,718
13–19	72	3,725
20–24	331	25,904
25–29	776	97,676
30–39	4,686	329,066
40–49	5,362	190,087
50–64	2,254	68,196
65+	217	10,002
Subtotal	13,794	733,374
Race/Ethnicity		
White	6,943	318,354
Black	5,742	272,881
Hispanic	642	133,703
Other	74	7,479
Unknown	393	957
Subtotal	13,794	733,374
Sex		
Male	10,766	609,329
Female	2,634	124,045
Unknown	394	
Subtotal	13,794	733,374

Source: Ohio HIV/AIDS Statistical Summary, HIV Infection and AIDS Cases Diagnosed through June 2003, Ohio Department of Health, www.odh.state.oh.us

US DHHS, Public Health Service, CDC, National Center for HIV, STD, and TB Prevention, AIDS Public Information Data Set, CDC WONDER On-line Database, wonder.cdc.gov

4. Complete the columns in Table 1–A–4.
   a. Compute the age-specific death rates for whites and blacks.
   b. Compute the 2001 overall crude death rate for the state of California and the crude death rates for whites and blacks.
   c. Compute the 2001 age-adjusted death rates for whites and blacks in the state of California using the standardized method.
   d. Is there a difference between the age-adjusted mortality rates for whites and blacks? If so, explain the reason for the discrepancy.

Table 1–A–4 Age-Specific Mortality Rates, State of California, 2001

Age	*(a) White Pop.*	*(b) Deaths*	*(c) White ASDR*	*(d) Black Pop.*	*(e) Deaths*	*(f) Black ASDR*	*(g) Comb. Pop. Total*	*(h) Expected No. of Deaths Whites (g × c)*	*(i) Expected No. of Deaths Blacks (g × f)*
<1	428,238	2,131		33,774	435		462,012		
1–4	1,565,447	413		170,587	80		1,736,034		
5–9	2,120,923	291		240,189	45		2,361,112		
10–14	2,084,668	311		244,031	55		2,328,699		
15–19	1,929,503	1,129		208,006	185		2,137,509		
20–24	1,916,977	1,569		186,458	274		2,103,435		
25–34	4,123,447	3,399		373,455	644		4,496,902		
35–44	4,318,242	7,394		415,178	1,258		4,733,420		
45–54	3,554,132	13,766		304,914	2,339		3,859,046		
55–64	2,201,539	18,939		176,743	2,647		2,378,282		
65–74	1,531,032	33,192		11,657	3,517		1,542,689		
75–84	1,119,160	59,115		62,592	3,998		1,181,752		
85+	395,512	57,200		20,542	2,906		416,054		
Total	27,288,820	198,849		2,448,126	18,383		29,736,946		

Source: United States Department of Health and Human Services, Centers for Disease Control and Prevention (CDC), CDC On-line Database, wonder.cdc.gov.

5. Calculate the odds ratio for the data in Table 1–13. Interpret the results.

6. At Critical Care Hospital, the complication rate for hip replacement surgery is 8.96%. The relevant statistics appear in Table 1–A–5. The administrative staff at the hospital is concerned that the hospital complication rate does not compare favorably with the overall complication rate of all patients with hip replacement surgery in the county. The complication rate for the county is 5.5%. The county complication rate for patients age 65 or older is 8.0%; for those under age 65, the complication rate is 3.0%. Using the indirect method of standardization, calculate the complication rate for the hospital that has been adjusted for age.

Table 1–A–5 Critical Care Hospital, Hip Replacement Surgery

Age Group	*No. of Patients*	*No. of Patients with Complications*	*Complication Rate*
≥ 65	170	17	10.00%
< 65	42	2	4.76%
Total	212	19	8.96%

7. The overall mortality rate for patients who have had a cerebrovascular accident (CVA) is 15.8% at CGH. You have been asked to compare the hospital's mortality rate to that of the state. Using the data provided in Table 1–A–6, calculate the age-adjusted death rate and the standard mortality ratio (SMR) for the hospital, using the indirect method of standardization. Explain the results.

Table 1–A–6 Mortality Rates for CVAs, State versus City General Hospital

Severity of Illness	*State Mortality Rate*	*Hospital Discharges for CVA*	*Observed Deaths*	*Expected Deaths*
1	4.2	55	2	
2	5.9	116	8	
3	7.8	195	20	
4	20.9	147	29	
5	34.6	62	32	
		575	91	

INTERNET ACTIVITY

An important skill for the health information management professional is the ability to search the Internet for information. This can be particularly useful when one is searching for comparative information. This activity is designed to provide experience working with an on-line interactive database and to provide experience analyzing and summarizing the results of data queries.

Instructions

1. The Utah Department of Health has an on-line interactive database that is available for public use. The database is constructed from the Uniform Hospital Discharge Data Set (UHDDS). Information on DRGs and ICD-9-CM codes (International Classification of Diseases, 9th revision, Clinical Modification) can be obtained through queries. The public data set contains data for the years since 1992. The website address ishlunix.hl.state.ut.us/.
2. Once at the site, click "Descriptive Statistics." This should take you to the Utah Hospital Discharge Query System. The Utah External Injury Data System will also be accessed.
3. Answer the questions that follow. An alternative website is the Centers for Disease Control and Prevention's data sets at http://wonder.cdc.gov.

Questions

1. For the diagnosis of acute myocardial infarction, ICD-9-CM category 410:
   a. Prepare a bar graph that displays the number of deaths due to AMI by year, 1998 through 2002.
   b. Prepare a line graph that displays the number of deaths by gender for the years 1998 through 2002. What are your conclusions?
   c. Prepare a table that displays the number of deaths due to AMI by age group. Use the table to prepare a bar graph of the same information.
   d. Prepare a bar graph that displays the average length of stay by gender for the years 1992 through 2002. What are your conclusions after reviewing the data?
   e. Prepare a line graph that displays the median charges by year, 1992 through 2002. What does the graph indicate?

2. How many patients with coronary atherosclerosis, ICD-9-CM category 414, had a coronary artery bypass graft (CABG) procedure, ICD-9-CM procedure category 36?
   a. Prepare a line graph that displays both the total number of discharges with a principal diagnosis of coronary atherosclerosis and the total number with the CABG procedure.
   b. Construct a bar graph that displays average length of stay, by year and gender, for patients with coronary atherosclerosis and CABG procedure for the years 1998–2002.
   c. Construct a bar graph that displays the number of CABG procedures, by gender, for the years 1998–2002. What are your conclusions?

3. Determine the number of patient discharges with pathological fractures, ICD-9-CM code 733.1, by year, 1998 through 2002, and by gender. You are interested in patients aged 65 years and over. Prepare a line graph displaying the number of discharges by year and by gender. Discuss your findings.

4. In table form, how many patients were discharged, by year, 1998 through 2002, and by gender, with malignant neoplasms of the trachea, bronchus, and lung? Use the selection option that is available on the database. Prepare a bar or line graph that displays the percentage of patients, by gender, who expired from these illnesses.

5. For ICD-9-CM code 185, for the years 1998 through 2002:
   a. Prepare a bar graph or pie chart, by third-party payer, of men, aged 45 and older, discharged with a diagnosis of prostate cancer.
   b. Prepare a bar graph that displays the number of men, by age group, discharged with prostate cancer.
   c. Discuss your findings.

6. For patients discharged with pneumonia during the years 1998 through 2002 (use the selection option that is available on the database):
   a. Prepare a table that reports the average length of stay for patients discharged with pneumonia by year and by gender. Include only patients who are aged 65 years and older.
   b. Prepare a bar or line graph to display your results.
   c. Discuss your findings.