340.721 Epidemiologic Inference in Public Health I

PRE-Activity Questions: Measures of Disease Frequency

The Activities provide experience in applying epidemiologic methods, interpreting findings, and drawing inferences. Activities will be discussed during the LiveTalks. Students are expected to work with their assigned Course Group prior to the start of each LiveTalk.

Prior to each Activity, students are to complete the corresponding set of PRE-Activity Questions. Each set of PRE-Activity Questions consists of 10 graded multiple choice questions. The graded multiple choice questions are to be completed via CoursePlus by the date and time listed in CoursePlus. PRE-Activity Questions prepare you for a productive and collaborative experience during the Activities.

Expectations for the PRE-Activity Questions

1. *Individually, read and attempt to answer all PRE-Activity Questions.*
2. *"Meet" or communicate with fellow students discuss challenging concepts, questions and compare answers. You may refer to their course materials and are strongly encouraged to collaborate with fellow students to complete the PRE-Activity Questions.*
3. *PRE-Activity Questions are due to Courseplus by the date listed on the syllabus. Although group collaboration is encouraged to complete the PRE-Activity Questions, each student must individually submit the PRE-Activity Questions.* ***Without exception, no credit will be given for submitting the PRE-Activity Questions after the due date****. The lowest PRE-Activity grade will be dropped when calculating the overall course grade.*

Motivation
In this assignment, you will practice calculating measures of morbidity and mortality. As epidemiologists, we often wish to compare these measures in two populations. One issue we face when we make these comparisons is that differences we observe may be due, in whole or in part, to factors other than our primary exposure of interest. Some motivating questions to think about are: When comparing two populations can you directly compare the number of cases of disease? Under what situations is this comparison problematic?

This assignment corresponds to:
Lectures: Measures of Morbidity and Mortality; Comparing Mortality in Different Populations
Readings: Gordis text (5th ed.) Chapters 3 and 4

Introduction

This assignment focuses on measures used to summarize the frequency of disease occurring in a population, measures of association employed to assess risk factors for diseases, and measures of survival, and introduces methods to account for the influence of extraneous factors on measures of association.

Concepts covered:
- Mortality
- Crude rates of disease incidence
- Adjustment for extraneous factors (e.g., age)
- Direct adjustment method
- Indirect adjustment method
- Standardized Mortality Ratios (SMRs)

Learning Objectives:
1. Calculate crude mortality rates
2. Perform direct and indirect rate adjustment
3. Calculate Standardized Mortality Ratios (SMRs)
4. Describe advantages, limitations, and appropriate use of adjusted rates

In Questions 1-3 of this assignment, we will review how to calculate a mortality rate and think about the appropriate comparisons for research questions of the effect of a given factor on risk of mortality (or disease), independent of other factors. These questions will help to motivate the remainder of the assignment: the calculation of adjusted rates using direct and indirect methods.

Individuals are often exposed to multiple factors. If we are interested in quantifying the association between a given exposure and an outcome, we often wish to know what that relationship is *independent* of the other factors. For example, smoking, increasing levels of cholesterol, and increasing diastolic blood pressure are three well established risk factors for coronary heart disease (CHD). While an individual may exhibit one or more of these risk factors, it is important to be able to determine the independent (i.e., separate) effects of each of these risk factors. Let's begin by examining the relationship between these factors and CHD mortality in an important and well-known clinical trial, the Multiple Risk Factor Intervention Trial (MRFIT). MRFIT was a primary prevention trial to assess the effects of various interventions (including diet and exercise) on the prevention of CHD. Between 1973 and 1975, 361,662 men aged 35-57 without evidence of CHD were screened and measurements for serum cholesterol and blood pressure (among others) were taken. The vital status of study subjects was ascertained for the following six years.

The following table is taken from an article on the relationship between serum cholesterol and the risk of coronary heart disease mortality among MRFIT screenees that appeared by Stamler, *et al*, in JAMA, 256(20):2825, 1986.

Table 1:

Quintiles of Serum Cholesterol, Diastolic Blood Pressure (DBP), Smoking Status, and Six-Year CHD Mortality Per Thousand Men at Risk for 356,222 Primary Screenees of MRFIT*

		DBP <90 mm Hg			DBP>90 mm Hg			Total		
Quintile	**Serum Cholesterol, mg/dL (mmol/L)**	**No. of Deaths**	**No. of Men**	**Mortality Rate Per 1,000 men at risk**	**No. of Deaths**	**No. of Men**	**Mortality Rate Per 1,000 men at risk**	**No. of Deaths**	**No. of Men**	**Mortality Rate Per 1,000 men at risk**
				Non-Smokers						
1	≤181 (⩽4.68)	47	35,741	1.6	36	9,612	3.7	83	45,353	2.1
2	182-202 (4.71-5.22)	82	34,553	2.5	51	11,599	4.0	133	46,152	2.9
3	203-220 (5.25-5.69)	87	31,939	2.7	80	12,839	5.6	167	44,778	3.5
4	221-244 (5.72-6.31)	126	30,431	3.8	94	14,500	5.6	220	44,931	4.4
5	⩾245 (⩾6.34)	188	26,996	6.4	200	16,930	10.7	388	43,926	8.0
Total		**530**	**159,660**	**?**	**461**	**65,480**	**6.4**	**991**	**225,140**	**4.3**
				Smokers						
1	⩽181 (⩽4.68)	82	20,017	5.2	31	5,002	6.3	113	25,019	5.4
2	182-202 (4.71-5.22)	95	19,675	5.5	60	5,977	10.0	155	25,652	6.7
3	203-220 (5.25-5.69)	128	18,812	7.3	100	6,397	15.5	228	25,209	9.5
4	221-244 (5.72-6.31)	186	19,119	10.2	127	7,533	16.6	313	26,652	12.1
5	⩾245 (⩾6.34)	250	18,907	13.3	208	9,643	21.4	458	28,550	16.0
Total		**741**	**96,530**	**8.4**	**526**	**34,552**	**15.1**	**1,267**	**131,082**	**10.3**
				All Men						
1	⩽181 (⩽4.68)	129	55,758	2.8	67	14,614	4.6	196	70,372	3.2
2	182-202 (4.71-5.22)	177	54,228	3.5	111	17,576	6.0	288	71,804	4.2
3	203-220 (5.25-5.69)	215	50,751	4.3	180	19,236	8.8	395	69,987	5.6
4	221-244 (5.72-6.31)	312	49,550	6.2	221	22,033	9.2	533	71,583	7.1
5	⩾245 (⩾6.34)	438	45,903	9.1	408	26,573	14.4	846	72,476	11.1
Total		**1,271**	**256,190**	**5.2**	**987**	**100,032**	**9.3**	**2,258**	**356,222**	**6.3**

* DBP Indicates Diastolic Blood Pressure; CHD, Coronary Heart Disease; and MRFIT, Multiple Risk Factor Intervention Trial. Analysis is Age Standardized.

Calculating and Comparing Mortality Rates

A mortality rate is defined as the number of deaths occurring in the population during a specified period of time:

$$Mortality\ rate = \frac{Total\ number\ of\ deaths\ during\ a\ specified\ time\ period}{Number\ of\ people\ in\ the\ population\ during\ that\ specified\ time\ period}$$

Question 1

Using the data in Table 1, calculate the overall mortality rate for non-smokers who had a diastolic blood pressure (DBP) <90 mm Hg.

a. 1.6 per 1,000 men at risk
b. 3.3 per 1,000 men at risk
c. 6.3 per 1,000 men at risk
d. 10.3 per 1,000 men at risk

(**Note:** Mortality rates presented in Table 1 are age-adjusted. We will be discussing direct adjustment in this assignment. Because the rates are age-adjusted, they will differ from what you would calculate using the formula provided.)

By focusing on participants that are in a range that is considered normal for the other risk factors you can *separate* the effect for a single risk factor**. Participants that are non-smokers, have a diastolic blood pressure (DBP) below 90 mm Hg, and a serum cholesterol below 200 mg/dL are considered normal.** Therefore to examine the effect of cholesterol on mortality independent (i.e., separate from) smoking and DBP we would focus on the mortality rates for non-smokers with a DBP<90 mmHg: 1.6 (lowest quintile), 2.5 (2^{nd} quintile), 2.7 (3^{rd} quintile), 3.8 (4^{th} quintile) and 6.4 (5^{th} quintile) and see that mortality increases with increasing cholesterol category.

Question 2

Which numbers from Table 1 could be compared to describe the independent effects of diastolic blood pressure (DBP) level on the risk of death due to CHD?

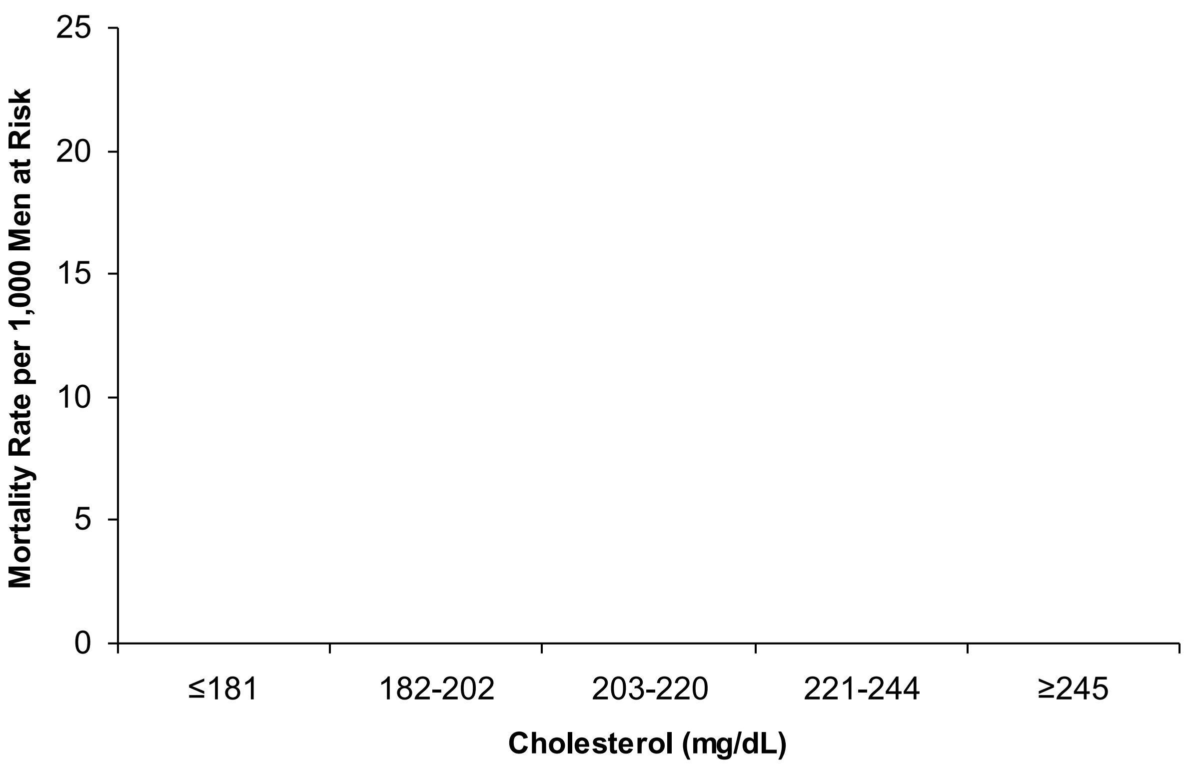
a. Mortality rate per 1,000 men at risk: 2.5 versus 4.0

b. Mortality rate per 1,000 men at risk: 5.5 versus 10.0

c. Number of deaths: 82 versus 51

d. Number of men: 95 versus 60

Plot on the graph below the mortality rate for each cholesterol level separately for the following four categories: Non-smokers with DBP less than 90; Non-smokers with DBP greater than 90; Smokers with DBP less than 90; Smokers with DBP greater than 90. Draw a line to connect the mortality rates across each cholesterol level for each category.



Question 3

What conclusions can you draw from these data about differences in the risk of death due to CHD by serum cholesterol level, smoking status and diastolic blood pressure level?

a. CHD mortality increases within increasing serum cholesterol level
b. CHD mortality is higher in smokers than in non-smokers
c. CHD mortality is highest in smokers with a DBP>90 and serum cholesterol ≥ 245 mg/dL
d. All of the above

Direct Age Adjustment

As epidemiologists, we often wish to compare incidence or mortality rates in two populations. One issue we face when we make these comparisons is that differences we observe may be due, in whole or in part, to factors other than our primary exposure of interest. In the next section of this assignment, we will illustrate this concept by comparing the incidence of breast cancer in white and black women. Because we know that older age is a risk factor for breast cancer, we will use the direct method of adjustment to account for differences in age in black and white women.

Presented below are statistics for breast cancer for white and black females in the United States for the five-year period 1998-2002, taken from the Surveillance, Epidemiology and End Results (SEER) program of the National Cancer Institute (2005). Refer to the appendix for a description of SEER.

TABLE 2:

Invasive Breast Cancer Incidence Rates in White and Black Females, United States, SEER, 1998-2002

	White Females			Black Females		
	Average Number of Cases per year[1]	Population (In 1,000's)[2]	Average annual incidence per 100,000 women at risk	Average Number of Cases per year[1]	Population (In 1,000's)[2]	Average annual incidence per 100,000 women at risk
All ages	34,165	114,734	**29.8**	3,701	18,645	_______

The average annual incidence per 100,000 women at risk in White Females is calculated for you. (Note that the Population is given in 1,000's, so 114,734 per 1,000 population is equal to 114,734 x 1,000 = 114,734,000. Note also that we are asked to provide the answer *per 100,000 women* at risk.)

$$\frac{34{,}165\ cases}{114{,}734{,}000\ population} = 0.000298 \times 100{,}000 = 29.8\ per\ 100{,}000\ women\ at\ risk$$

Question 4

What is the average annual incidence rate per 100,000 women at risk in black females?

a. 19.8 per 100,000 black females at risk
b. 25.4 per 100,000 black females at risk
c. 28.7 per 100,000 black females at risk
d. 29.8 per 100,000 black females at risk

---

1 SEER. Cancer Statistics Review 1975-2002: Table IV-3.
[Number of cases accumulated over 5 years divided by 5:
White women: 170,826 cases/5 years = 34,165.2 cases/year
Black women: 18,503 cases/5 years = 3700.6 cases/year]

2 Census P25-1130, Table 2: Projections of the Population by Age, Sex, Race and Hispanic Origin for the United States. As of July 1, 2000. [Midpoint population used to approximate average annual population]

The incidence rate calculated in Table 2 (Question 4) is called the *crude*, or *unadjusted* incidence rate, because it does not account for any other factors. We will now calculate an *age-adjusted* incidence rate so that we can compare the incidence of breast cancer in black and white women independent of age. We will use the direct method of adjustment to calculate the adjusted incidence rate in both groups. The direct method of adjustment uses a standard population to eliminate the effects of any extraneous factors, such as age, in the populations being compared.

Overview of direct adjustment

The direct method of adjustment allows us to estimate incidence (or mortality) rates that are comparable across populations. For example, differences in disease incidence rates between two countries might be attributed to differences in environmental exposures or lifestyle differences, but they may also be due simply to differences in the ages of the populations. Direct adjustment gives a **weighted average** of age-specific rates (i.e., the crude rates within each age stratum) to compensate for differences in age distributions between populations. The same weights, based on the age distribution of the standard population, are used for the populations being compared. The expected number of deaths is the number of deaths one would expect to observe if the study population had the same distribution of factors upon which you are standardizing (e.g., age, sex) as the standard population.

The expected count for each age, E_{age}, is obtained by multiplying the stratum-specific outcome incidence rate observed in the study population, A_{age}, by the number of subjects in that age category in the standard population, B_{age}. The expected counts, E_{age}, are then summed over all age categories to obtain the total expected number of deaths, E.

The adjusted incidence rate is calculated as the sum of the age-specific expected counts divided by the number of subjects in the standard population.

Age Category	Outcome Incidence Rate in Study Population (Incident cases / no. at risk)	Number of Subjects in Standard Population	Age-Specific Expected Count
1	A_1	B_1	$E_1 = A_1 \times B_1$
2	A_2	B_2	$E_2 = A_2 \times B_2$
3	A_3	B_3	$E_3 = A_3 \times B_3$

$$\text{Adjusted Incidence Rate} = \frac{E_1 + E_2 + E_3}{B_1 + B_2 + B_3}$$

As a first step in adjustment, we will calculate the age-specific risk for both white and black females (see Table 3).

(In Table 3, don't forget to fill in the average annual incidence rate per 100,000 women at risk for black women – from Table 2, the answer you calculated for Question 4).

TABLE 3:

Invasive Breast Cancer Incidence Rates in White and Black Females by Age, United States, SEER, 1998-2002

	White Females				Black Females			
Age Group (years)	Average Number of Cases per year[3]	Population (In 1,000's)[4]	Percent of Total Population	Average annual incidence per 100,000 women at risk	Average Number of Cases per year[3]	Population (In 1,000's)[4]	Percent of Total Population	Average annual incidence per 100,000 women at risk
0 – 19	0	30,057	26	0	0	6,147	33	0
20 – 29	118	14,081	12	0.8	30	2,722	15	1.1
30 – 39	1,525	16,848	15	9.0	281	2,917	16	9.6
40 – 49	5,630	17,627	15	31.9	795	2,737	15	29.0
50 – 59	8,081	13,334	12	60.6	872	1,740	9	50.1
60 – 69	5,488	9,123	8	60.2	792	1,156	6	68.5
70 – 79	8,300	8,137	7	102.0	615	778	4	79.0
80+	5,023	5,527	5		316	448	2	
All ages	34,165	114,734	100		3,701	18,645	100	
Average annual incidence rate per 100,000 women at risk (from Table 2):				29.8				

The average annual incidence rate per 100,000 women at risk by age group are provided for you in Table 3 for both white females and for black females.

Make sure you can replicate the answers! A sample calculation is below:

$$Risk\ for\ \ White\ Females\ aged\ 20-29\ years = \frac{118\ cases}{14{,}081{,}000\ population} = 0.0000084 \times 100{,}000$$
$$= 0.8\ per\ 100{,}000\ women\ at\ risk$$

[3] SEER. Cancer Statistics Review 1975-2002: Table IV-3. [Number of cases accumulated over 5 years/5 years]
[4] Census P25-1130, Table 2: Projections of the Population by Age, Sex, Race and Hispanic Origin for the United States. As of July 1, 2000. [Midpoint population used to approximate average annual population]

Question 5

Calculate age-specific average annual incidence rate of breast cancer per 100,000 women at risk for white and black females aged 80+ years in Table 3.

a. White females: 60.6; Black females: 68.5
b. White females: 90.9; Black females: 19.8
c. White females: 60.6; Black females: 79.0
d. White females: 90.9; Black females: 70.5

Next, we will apply the age-specific incidence rates to a standard population (see Table 4).

First, we create a hypothetical standard population:

> In this example, in each age strata, we create a hypothetical standard population by summing the total number of white and black females in that age-specific stratum.
>
> These calculations have been provided for you in Table 4. Make sure you can replicate the calculations! A sample calculation is below:

$$\boldsymbol{Standard\ Population\ Age\ Group\ 20-29} = 14{,}081{,}000\ white\ women + \ 2{,}722{,}000\ black\ women$$

Once we have created the hypothetical population in each age strata for white females and for black females, we apply the observed race-specific incidence rates for a given age-stratum to the hypothetical population in that age stratum. These calculations have been provided for you in Table 4. Make sure you can replicate the calculations! A sample calculation is below:

$$\boldsymbol{Expected\ number\ of\ cases\ for\ white\ females\ age\ 20-29} = \frac{0.8\ cases}{100{,}000\ women} \times 16{,}803{,}000\ women$$
$$= 134.42\ cases$$

134.42 is the expected number of cases for the standard population for the 20-29 age stratum if the standard population had the same incidence of breast cancer as what we observed for the white females.

(In Table 4, don't forget to add in the average annual incidence per 100,000 women at risk for both white and black women aged 80+ years that you calculated for Question 5.)

TABLE 4:

Age-Adjustment of Invasive Breast Cancer Incidence Rates in White and Black Females, United States, SEER, 1998-2002

		White Females		Black Females	
Age Group (years)	Standard Population (In 1,000's)	Average annual incidence per 100,000 women at risk	Expected Number of Cases for White Females	Average annual incidence per 100,000 women at risk	Expected Number of Cases for Black Females
0 - 19	36,204	0	**0.0**	0	**0.0**
20 - 29	16,803	0.8	134.42	1.1	184.83
30 - 39	19,765	9.0	1,778.85	9.6	1,897.44
40 - 49	20,364	31.9	6,496.12	29.0	5,905.56
50 - 59	15,074	60.6	9,134.84	50.1	7,552.07
60 - 69	10,279	60.2	6,187.96	68.5	7,041.12
70 - 79	8,915	102.0	9,093.30	79.0	7,042.85
80+					
Total					
Age-adjusted average annual incidence rate per 100,000 women at risk:					

Question 6

Using data from Table 3, complete Table 4 by calculating the standard population, expected number of cases for white females, and expected number of cases for black females for the 80+ year age group.

a. Standard population: 316; Expected cases white females: 90.90; Expected cases black females: 448.21
b. Standard population: 5,023; Expected cases white females: 29.80; Expected cases black females: 33,836.25
c. Standard population: 5,527; Expected cases white females: 9,093.30; Expected cases black females: 70.50
d. Standard population: 5,975; Expected cases white females: 5,431.28 ; Expected cases black females: 4,212.38

Finally, calculate the age-adjusted incidence rates, using the data in Table 4 and the answers you calculated for Question 6:

- For white females, divide the total number of expected cases for white females by the total standard population.
- For black females, divide the total number of expected cases for black females by the total standard population

Question 7

What is the age-adjusted average annual incidence rate per 100,000 women at risk for white females and for black females?

a. White females: 25.8; Black females: 23.3
b. White females: 28.7; Black females: 25.4
c. White females: 29.8; Black females: 19.8
d. White females: 30.0; Black females: 26.6

Question 8

From Table 2, calculate the ratio of crude incidence rates comparing white females to black females. Then, using data from Table 4, calculate the ratio of the age-adjusted incidence rates among white females compared to black females using the combined population of white and black females as the standard population.

.

a. Crude ratio: 0.66; Age-adjusted ratio: 0.08
b. Crude ratio: 1.04; Age-adjusted ratio: 0.78
c. Crude ratio: 1.10; Age-adjusted ratio: 5.60
d. Crude ratio: 1.51; Age-adjusted ratio: 1.13

Indirect Adjustment

Overview of indirect adjustment

Indirect adjustment enables the calculation of a relative measure of association when the outcome incidence rate is available for an exposed group, but not for a suitable comparison group. The relative measure of mortality calculated when death is the outcome of interest is the standardized mortality ratio (SMR).

The SMR is defined as:

$$\frac{\text{Observed number of deaths over a specified time period}}{\text{Expected number of deaths over a specified time period}} \times 100$$

The expected number of deaths is the number of deaths one would expect to observe if a standard population had the same distribution of factors upon which you are standardizing (e.g. age, sex) as the exposed population under study.

The expected count for each age category, E_{age}, is obtained by multiplying the mortality rate observed in the standard population, A_{age}, by the number of subjects in that age category in the study population, B_{age}. The expected counts, E_{age}, are then summed over all age categories to obtain the total expected number of deaths, E.

Age Category	Mortality Rate in Standard Population (incident cases / no. at risk)	Number of Subjects in Study Population	Expected Count	Observed Count
1	A_1	B_1	$E_1 = A_1 \times B_1$	O_1
2	A_2	B_2	$E_2 = A_2 \times B_2$	O_2
3	A_3	B_3	$E_3 = A_3 \times B_3$	O_3

$$\text{SMR} = \frac{O_1 + O_2 + O_3}{E_1 + E_2 + E_3} \times 100$$

You would like to study mortality from all causes among former mercury miners. You assemble a cohort of 6,000 male miners, abstracting data on their job history from employment records. You exclude all workers with less than 1 year of work on the job, leaving you with 5,742 subjects. You ascertain vital status for all subjects using the National Death Index. You have no information on non-exposed subjects.

Table 6 provides the expected numbers of deaths among your study population, broken down into categories of age. The expected numbers of deaths for miners aged 25-34 years and for miners aged 35-44 years if they had the same mortality experience as the U.S. population are provided for you.

Table 6: Standardized Mortality Ratio (SMR) among a hypothetical cohort of miners			
Age Category	**U.S. All-cause annual mortality rate per 100,000 population**	**Miner population (number of exposed subjects)**	**Expected number of deaths**
25-34 years	138.6	2297	3.18
35-44 years	255.2	1723	4.40
45-54 years	542.8	1148	
55-64 years	1230.7	574	
		Total	

Sample calculation:

$$Expected\ number\ of\ deaths\ in\ 25-34\ year\ olds = 2{,}297\ miners \times \frac{138.6}{100{,}000\ population} = 3.18$$

Question 9

Complete Table 6 by calculating the expected number of deaths for the age category ‘45-54 years’ and for the age category ‘55-64 years’.

a. 45-54 years: 0.47; 55-64 years: 2.11
b. 45-54 years: 6.23; 55-64 years: 7.06
c. 45-54 years: 14.13; 55-64 years: 3.12
d. 45-54 years: 20.87; 55-64 years: 26.00

Question 10

The total number of observed deaths among miners was 26. Given this information, and the information in Table 6, calculate the SMR for the entire cohort.

a. 0.80
b. 1.25
c. 1.88
d. 5.13

(Note: You may also see the SMR expressed as a percentage multiplied by100.)

APPENDIX

SEER: A Source for U.S. Cancer Rates.

The following description of SEER can also be found on the National Cancer Institute's Web site (http://seer.cancer.gov/about/)

The Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute is the most authoritative source of information on cancer incidence and survival in the United States. Case ascertainment for SEER began on January 1, 1973, in the states of Connecticut, Iowa, New Mexico, Utah, and Hawaii and the metropolitan areas of Detroit and San Francisco-Oakland. In 1974-1975, the metropolitan area of Atlanta and the 13-county Seattle-Puget Sound area were added. In 1978, 10 predominantly black rural counties in Georgia were added, followed in 1980 by the addition of American Indians residing in Arizona. Three additional geographic areas participated in the SEER program prior to 1990: New Orleans, Louisiana (1974-1977); four counties in New Jersey (1979-1989); and Puerto Rico (1973-1989). The National Cancer Institute also began funding a cancer registry that, with technical assistance from SEER, collects information on cancer cases among Alaska Native populations residing in Alaska. In 1992, the SEER Program was expanded to increase coverage of minority populations, especially Hispanics, by adding Los Angeles County and four counties in the San Jose-Monterey area south of San Francisco.

Geographic areas were selected for inclusion in the SEER Program based on their ability to operate and maintain a high quality population-based cancer reporting system and for their epidemiologically significant population subgroups. The population covered by SEER is comparable to the general U.S. population with regard to rates of poverty and education. The SEER population tends to be somewhat more urban and has a higher proportion of foreign-born persons than the general U.S. population.

Updated annually and provided as a public service in print and electronic formats, SEER data are used by thousands of researchers, clinicians, public health officials, legislators, policymakers, community groups, and the public. The SEER Program provides cancer incidence, mortality, and survival data in an annual cancer statistics review, in monographs on relevant topics, through the SEER Web site (http://seer.cancer.gov), in various specially-developed software packages (e.g., SEER*Stat, SEER*Prep), and in a public-use data file available on CD-ROM. SEER records have also been linked with other data systems in order to meet research needs.