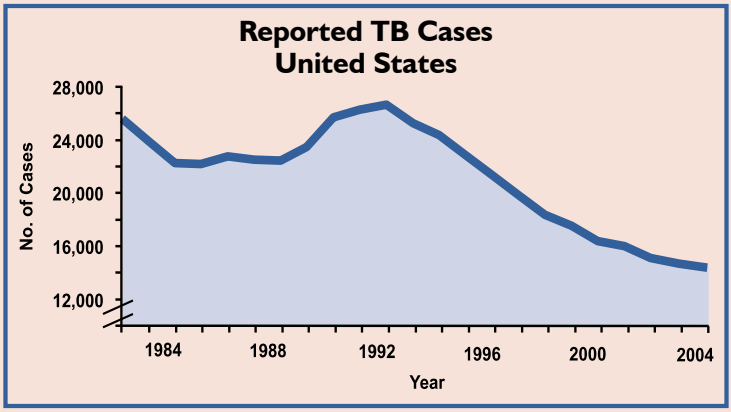
Basic Epidemiology for Tuberculosis Program Staff



New Jersey
Medical School
National
Tuberculosis
Center

A Founding Component of the International Center for Public Health

Basic Epidemiology for Tuberculosis Program Staff

Marian Passannante, PhD
Associate Professor
New Jersey Medical School & School of Public Health
Epidemiologist
New Jersey Medical School National Tuberculosis Center
University of Medicine and Dentistry of New Jersey
Newark, New Jersey

Nisha Ahamed, MPH, CHES
Training and Consultation Specialist
New Jersey Medical School National Tuberculosis Center
University of Medicine and Dentistry of New Jersey
Newark, New Jersey

Acknowledgements

We wish to thank the following individuals and groups who participated in drafting and reviewing this guide:

Donna Allis, PhD, RN
Program Manager for TB Control
Margaret Osborn, RN, BSN
Public Health Nurse
Snohomish Health District
Everett, Washington

Pete Denkowski, RN, MS
Director, Ben Franklin TB Program
Columbus Health Dept.
Columbus, Ohio

Kim Field, RN, MSN
TB Controller
Washington State Dept. of Health
Olympia, Washington

Vipra Ghimire, MPH, CHES
Health Education Coordinator
Virginia Dept. of Health
Richmond, Virginia

Myrene Couves
Manager
Capital Health Tuberculosis Clinic
Edmonton, Alberta

Bob Parker, MS
Epidemiologist
Thomas Jefferson Health District
Charlottesville, Virginia

Denise Cory
Director, Communicable Disease Control
Vanderburgh County Health Dept.
Evansville, Indiana

Thomas Navin, MD
Chief, Surveillance, Epidemiology, and Outbreak Investigations Branch
Kayla Laserson, ScD
Chief, Epidemiology and Evaluation Team, International Research and Programs Branch
Division of Tuberculosis Control and Elimination, Centers for Disease Control and Prevention
Atlanta, Georgia

Thomas Privett
TB Program Manager
Joanne Becker
Supervising Program Development Specialist
New Jersey Dept. of Health & Senior Services
Trenton, New Jersey

Linda Weldon, RN, BSN
Program Manager, Communicable Diseases
Henderson County Dept. of Public Health
Hendersonville, North Carolina

Marie Villa, RN
TB Program Manager
Oralia Zamora, RN
Nurse Supervisor
Tillman Health Center
TB Control Clinic, El Paso City-County Health & Environmental District
El Paso, Texas

Patsy Eddington
TB Case Manager
Diane McCracken, RN
Clinic Coordinator
Mary Spinner, RN, BSN
Interim Program Administrator
Diane Werning
Disease Investigation Specialist
Oklahoma City Health Dept.
Oklahoma City, Oklahoma

Beverly Ann Collins, RN, MS, CIC
Epidemiologist
Infection Control Dept.
University Hospital
Newark, New Jersey

Bart Holland, PhD
Associate Professor New Jersey Medical School & School of Public Health
University of Medicine and Dentistry of New Jersey
Newark, New Jersey

Natalia Kurepina, PhD
Senior Scientist
PHRI TB Center at International Center for Public Health
Newark, New Jersey

Rajita Bhavaraju, MPH, CHES
Program Director, Education and Training
Chris Hayden, BA
Consultant, Evaluation Activities
Eileen Napolitano, BA
Deputy Director
Stefanie Napolitano, MPH
MPH Fieldwork
Nandini Selvam, MPH
Research Coordinator
Mark Wolman, MA, MPH
Program Manager, TB Control
New Jersey Medical School
National Tuberculosis Center
Newark, New Jersey

Information in this guide is drawn from a number of previously published sources, including those cited in the text and the reference list at the end of the guide. The authors are grateful for the use of these materials, particularly the CDC's Excellence in Curriculum Integration through Teaching Epidemiology (EXCITE) website. Available at: http://www.cdc.gov/excite/index.htm

Suggested citation:

New Jersey Medical School National Tuberculosis Center. *Basic Epidemiology for Tuberculosis Program Staff.* 2005: (inclusive pages)

Graphic Design: Dee Dee Hamm

Table of Contents

Table of Contents, continued

1.0 Introduction – Uses of Statistics and Epidemiology in Tuberculosis Control

Control of tuberculosis (TB) in the United States is an important public health responsibility. Effective TB control requires a complex system that merges elements of laboratory science, investigative work, public health, surveillance, and clinical care.

Epidemiology is the basic science of public health. An understanding of epidemiology is useful for all TB program staff, ranging from health care workers and public health representatives to TB program managers. The epidemiologic concepts presented in this guide will assist in analyzing and making practical use of data, assessing current and evolving trends in TB morbidity, identifying risk groups, and determining where to allocate staff and resources. Although not all TB program staff members are involved with all of these activities, a broad understanding of epidemiologic principles can assist all TB program staff in working toward effective TB control.

The first section of this guide (Chapters 2 through 5) provides a basic background and understanding of epidemiology for TB program staff. This section focuses on specific uses of epidemiology to assess and implement TB programs. The second section of the guide (Chapters 6 and 7) will explain epidemiologic terms and techniques that are used in research studies. This will assist TB program staff in reading and understanding TB-related articles in medical and public health journals. Awareness of new information about the epidemiology of TB and new research in TB transmission, diagnostics, and treatment can be very useful to TB program staff members in working to control TB within their program area.

This guide identifies and defines key concepts and terminology in epidemiology, and provides detailed examples and sample problems. Wherever possible, data and examples are drawn from existing epidemiologic studies related to TB. Further, the guide presents descriptions of how these concepts can be put to practical use by TB program staff. This guide is not intended to be a complete text on TB, but rather a reference that can be used to learn or review key concepts of epidemiology that will be useful in the overall effort to control TB in the United States.

Definitions and examples of selected statistical terms used in epidemiologic studies are underlined in the text and appear in Appendix I. In the online version of the guide, these terms are linked to the definitions in Appendix I.

2.0 What Is Epidemiology?

Definitions of epidemiology vary, but the one used in this guide is utilized by the Centers for Disease Control and Prevention (CDC).

Epidemiology

"The study of the distribution and determinants of health-related states in specified populations, and the application of this study to control health problems."

Available at: http://www.cdc.gov/excite/library/glossary.htm

A health-related state should be thought of in a very broad context, including the occurrence of infection, symptomatic disease, injury, disability (which are all aspects of morbidity or illness) and even death (ie, mortality). Epidemiology can also be described as the basic science of public health and is a discipline that helps explore and understand patterns of morbidity and mortality within and between populations, using statistical methods to clarify these patterns.

Epidemiology is an important part of TB control efforts because the information on patterns of infection and disease can assist in identifying people or groups of people at risk for TB, understanding how the disease is transmitted, prioritizing cases, and planning appropriate use of staff and resources.

3.0 Types of Epidemiology

Epidemiology is usually classified as: descriptive or analytic.

Epidemiology

- **Descriptive epidemiology** concentrates on examining the distribution of diseases in the population in terms of person, (who gets the disease), place (where they get the disease) and time (when they get the disease).
- **Analytic epidemiology** is concerned with studying the relationship between risk factors and a disease.

Another way to think about descriptive epidemiology versus analytic epidemiology involves hypotheses, or tentative explanations for observations or scientific problems. Hypotheses are generated through descriptive epidemiology, while analytic epidemiology allows testing of those hypotheses to determine if they are likely to be correct or incorrect.

A. Descriptive Epidemiology

Descriptive epidemiologic data related to TB are collected through public health surveillance activities.

i. Public Health Surveillance

Public Health Surveillance

The systematic, ongoing collection, analysis, interpretation, and dissemination of health data. The purpose of public health surveillance is to gain knowledge of the patterns of disease, injury, and other health problems in a community so that we can work toward controlling and preventing them.

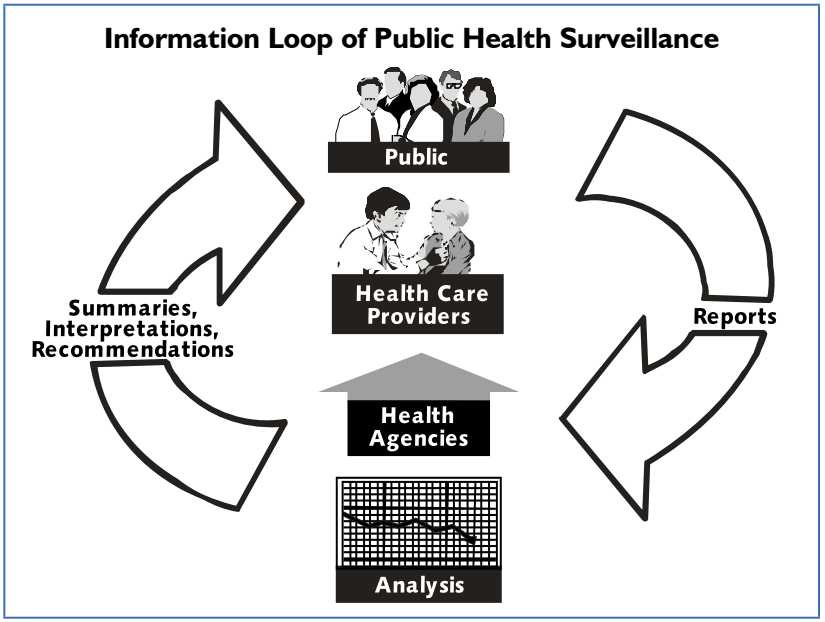
Available at: http://www.cdc.gov/excite/library/glossary.htm

Two types of public health surveillance are active and passive surveillance:

Active surveillance is a system in which the health department or other agency initiates the data collection activities. In TB control, targeted tuberculin skin testing (TST) by a health department among certain populations, such as persons living with HIV/AIDS, is an example of active surveillance for TB infection.

Passive surveillance is used when the health care provider is asked or required to report information to the health department. The CDC system for receiving reports of adverse effects associated with treatment is an example of passive surveillance.

Public health surveillance is an important part of an information feedback loop that links the public, health care providers, and health agencies. To complete the information loop detailed on the next page, data collected through both active and passive surveillance mechanisms should be summarized by the official health agency and then sent back to those who can make use of this information at the provider or program level. These data can be useful for public health education and interventions. Information from surveillance systems can also be used to generate public health recommendations that should then be disseminated to the general public. TB surveillance in the United States relies on both passive and active surveillance activities.



Source: Public Health Surveillance – CDC slide set.
Available at: http://www.cdc.gov/epo/dphsi/phs/overview.htm

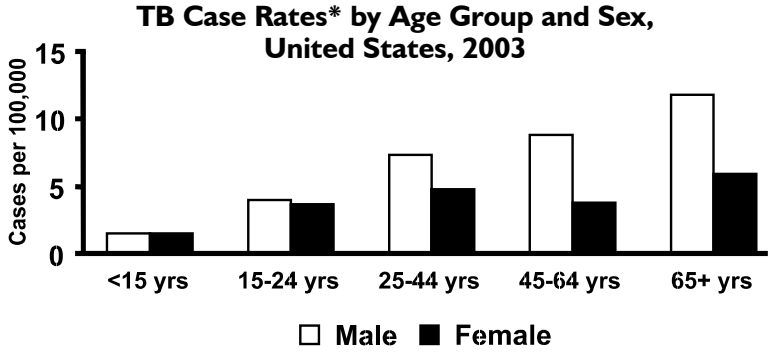
ii. Example of Descriptive Epidemiology: TIMS

The Tuberculosis Information Management System (TIMS) is one example of a public health surveillance system. TIMS is one of the main sources of descriptive data regarding TB in the United States. TIMS includes information on all cases of TB that have been reported to the Division of TB Elimination (DTBE) at the CDC. This information is reported to CDC by 50 states, the District of Columbia, the city of New York, Puerto Rico, and other jurisdictions in the Pacific and Caribbean.

Data on **person**, **place**, and **time** relating to TB in the United States are gathered using TIMS. These data are analyzed and published by the CDC annually and may be accessed through the CDC Website in the form of TB Surveillance Reports (available at: http://www.cdc.gov/nchstp/tb/surv/Surv.htm).

Person

The next figure presents the number of TB cases per 100,000 population in the United States that were reported to the CDC in 2003, by 2 characteristics that describe the **person**: age and sex.



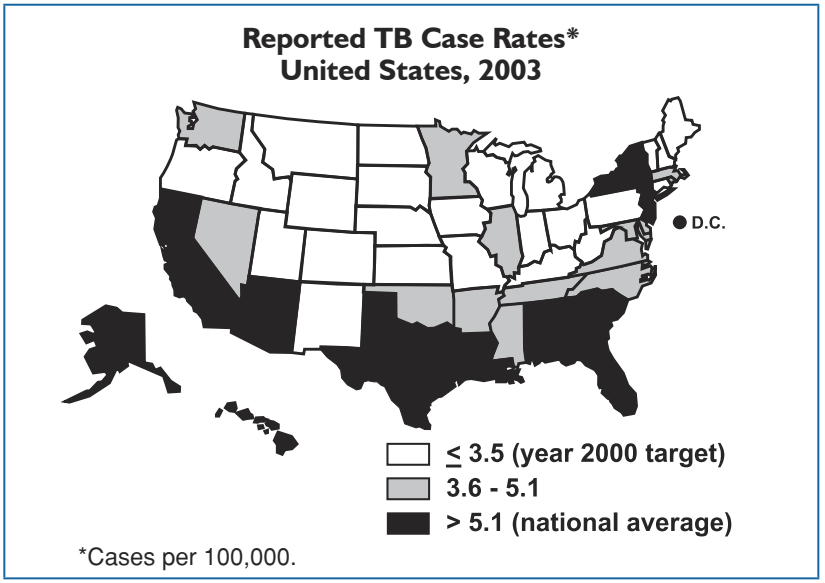
*Cases per 100,000.

Source: 2003 TB Surveillance – CDC slide set. Available at: http://www.cdc.gov/nchstp/tb/pubs/slidesets/surv/surv2003/default.htm

The number of TB cases per 100,000 population is also called the **TB case rate**. In this figure, the TB case rate is higher among men than among women for all age groups 15 years and older. The TB case rate is highest among those 65 and older. These data help to identify groups of people who *may* be at higher risk for developing TB.

Place

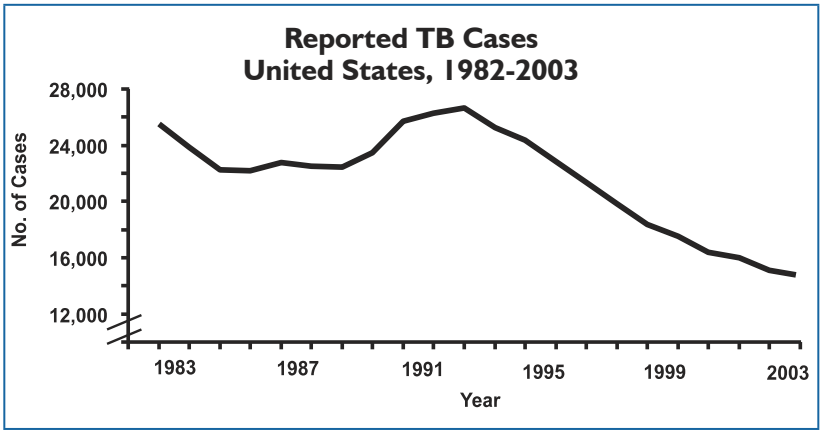
TB cases per 100,000 population are reported by state so that states with unusually high rates of TB can be identified. In the following figure, the shading indicates **places** (states) where TB cases per 100,000 people are near the Year 2000 target, as well as those that are above the 2000 target. This descriptive epidemiology can help identify areas where interventions to decrease the number of TB cases might be most valuable.



Source: 2003 TB Surveillance – CDC slide set. Available at: http://www.cdc.gov/nchstp/tb/pubs/slidesets/surv/surv2003/default.htm

Time

Finally, the next figure shows the changes in the number of TB cases over **time**.



Source: 2003 TB Surveillance – CDC slide set. Available at: http://www.cdc.gov/nchstp/tb/pubs/slidesets/surv/surv2003/default.htm

Analysis of the information contained in TIMS, collected through public health surveillance, allowed CDC to identify the resurgence and subsequent decline of TB cases in the United States.

In the 2002 TB Surveillance Summary, CDC reported that:

> "The resurgence of TB in the mid-1980s was marked by several years of slightly increasing case counts followed by a substantial rise for several years. The total number of TB cases peaked in 1992. From 1992 until 2002, the total number of TB cases decreased 5%-7% annually, and 2002 marks the tenth year of decline in the total number of TB cases reported in the United States since the peak of the resurgence. In 2002, a total of 15,075 TB cases were reported from the 50 states and the District of Columbia. This represents a 6% decrease from 2001 and a 43% decline from 1992."

A note of caution about rates versus actual numbers:

In the first two figures the data are presented as rates, while in the last figure the actual number of cases is presented on the vertical axis. Interpretation of the number of cases must be done cautiously since the number of cases of any disease may be affected by the entrance or exit of individuals from the population. Therefore, epidemiologists tend to compare rates over time, since rates take into account the size of the population.

For example, a county TB program may usually identify 20 new cases of TB in a county annually. However, in a particular year, 40 new cases were identified. From a clinical perspective this is important since a large number of additional cases must be treated. But how should this be interpreted from an epidemiologic perspective? What if the population in the county had doubled for some reason? In this situation, 20 additional cases might not be surprising. The only way to understand what is really happening in the community is to calculate the rates. The calculation and interpretation of rates will be discussed in more detail beginning on page 14.

iii. Using TIMS Data

Data on TB cases are collected using the Report of Verified Cases of Tuberculosis (RVCT) form (see Appendix II). This information can be used to provide the descriptive epidemiology of local and state TB programs. For example, a description of the sex, race, ethnicity, occupation, nationality, and place of residence of TB cases can be summarized for state or local areas from data collected through TIMS. Health information such as HIV status, history of substance use, prior diagnosis of TB, site of disease, smear and sputum culture results, initial drug regimen, initial and final drug susceptibility results, type of health care provider, and type of therapy received (directly observed therapy [DOT] vs self-administered) are all collected using TIMS.

TIMS also collects a large amount of information related to treatment outcomes that can be used to evaluate program performance and needs. For example, information on date of treatment initiation may be compared with date that therapy was completed to see how long, on average, it took for patients to complete therapy. A variety of program performance goals can be set by the state TB control program relating to these variables, allowing programs to assess how they are performing, using standardized measurements. Sample program performance goals can be found in Appendix III.

These performance measures can be reported annually or more frequently using the cohort review process. A description of implementation of the cohort review process in a state TB control program is available in the article "Implementing Cohort Review in Washington State." Available at: http://www.cdc.gov/nchstp/tb/notes/TBN_4_03/highlights_state_local.htm

B. Analytic Epidemiology

While descriptive epidemiologic data (by person, place, and time) are used to create surveillance summaries or annual reports, analytic epidemiology is often used in the medical literature when researchers are trying to assess the value of a new drug regimen compared with an established one, or to identify factors that might predict adherence to treatment or the development of drug resistance.

An excerpt from an article that appeared in *Morbidity and Mortality Weekly Report* in 1999 illustrates this point.

> To identify risk factors for P-MDRTB, a case-control study was conducted in February 1999 of never-treated, smear- and culture-positive pulmonary TB patients reported during October 1995-October 1998. A case of P-MDRTB was defined as culture-confirmed MDRTB in a patient; controls were patients with culture-confirmed drug-susceptible TB.
>
> … compared with controls, case-patients were significantly more likely to have a history of homelessness (23% versus 5%; OR=3.1; 95% CI=1.1-8.8; p=0.04).

Source: Primary multidrug-resistant tuberculosis—Ivanovo Oblast, Russia, 1999. *Morb Mortal Wkly Rep*. 1999;48:661-664.

In this study, the researchers were interested in identifying risk factors for primary drug-resistant tuberculosis (P-MDRTB). They found that when comparing P-MDRTB cases with a comparison group (also called a control group) who had culture-confirmed drug-susceptible TB, **"case-patients were significantly more likely to have a history of homelessness."** This is an example of an *analytic epidemiologic* study because the purpose of the study was to identify "risk factors" for P-MDRTB. More information on the interpretation of the other measures in this abstract, odds ratios and *p*-values, will be presented in Chapter 7 of this guide.

4.0 Key Concepts in Epidemiology

As in any other field, epidemiology has its own language or terms that are used to describe events that relate to disease occurrence and outcomes. For example, epidemiology involves the study of morbidity and mortality.

Epidemiology Involves the Study of...

- Morbidity – Disease; any departure, subjective or objective, from a state of physiological or psychological health and well-being.

  and

- Mortality – Death

Available at: http://www.cdc.gov/excite/library/glossary.htm

There are various **measures** that can be used to describe morbidity and mortality.

A. Morbidity

Morbidity may be endemic or epidemic. An endemic health condition is one that can be thought of as "usual" or "background" occurrence in a population, while epidemic occurrence can be thought of as "unusual" occurrence. When an epidemic occurs in many parts of the world, it is often referred to as pandemic. Finally if the occurrence of a health condition continues to occur at a very high rate, it may be called hyperendemic. These terms are all relative to the situation in a particular geographic region, so TB may be endemic in one country and epidemic in another. Finally, the word outbreak is often used interchangeably with epidemic.

The most common way to express morbidity or disease occurrence is by calculating incidence rates and prevalence ratios. Unlike the examination of cases alone, measures of incidence and prevalence allow comparison across populations and time periods while adjusting for the fact that the number of people in the population may have changed over this same time period.

i. Incidence

Incidence
Incidence is one measure of morbidity:

A rate that measures the frequency with which a health problem, such as a new injury or case of illness, occurs in a population. In calculating incidence, the numerator is the number of new cases occurring in the population during a given period of time, and the denominator is the total population at risk during that time.

Available at: http://www.cdc.gov/excite/library/glossary.htm

The incidence rate formula appears below:

Incidence Rate

$$\frac{\text{\# of NEW cases of disease during a specified time period}}{\text{Population at risk of disease during the same time period (also measured as person-time)}} \times 1{,}000$$

An incidence rate is calculated by taking the number of new cases of disease during a particular time period (the numerator, or top number) and dividing that number by the population at risk of disease during that time period (the denominator, or bottom number). Ideally, individuals who are not at risk of developing the disease would be subtracted from the denominator of the rate prior to doing these calculations. However, in most instances this is not possible, so the total population is used as the denominator instead. This measurement is sometimes called **cumulative incidence**.

When calculating incidence rates, a multiplier of 1,000 is used. This allows expression of the rate as the number of cases per 1,000 people in a population. Since the numbers are often quite small, using the multiplier allows for easier understanding of the rate.

TB Case Rates

A special type of incidence rate used to describe the epidemiology of TB is the TB case rate.

TB Case Rate

$$\frac{\text{Number of TB cases that occur during a specified time period}}{\text{Population at risk during that time period}} \times 100{,}000$$

Note: cases are verified cases of TB. If TB recurs or if more than 12 months has elapsed since the person was discharged or lost to follow-up, then the person is counted as a new case.

The numerator of the TB case rate refers to cases that are "new" cases, based on the CDC's definition of a new case. The denominator is the population during that time period. So, the TB Case Rate is clearly an incidence rate. The only difference between these 2 formulas is the multiplier (100,000 instead of 1,000) used to generate the rates. The explanation for this is that, when calculating incidence rates for any one cause (or disease), the rates tend to be small (compared with an overall morbidity rate for all causes), so a larger multiplier, such as 100,000, is used to make the numbers easier to understand. To be consistent with published data, TB case rates should be calculated per 100,000.

In epidemiology the definition of what constitutes a case (also known as the case definition) is a very important concept, since comparison of case rates can only be useful if those who are calculating the rates are using the same definition. The CDC case definition for TB is standardized so that a case rate from one area of the country will be measuring the same thing as a case rate from another area, and will, therefore, be comparable.

ii. Prevalence

A second measure of disease occurrence is prevalence.

Prevalence Ratio

$$\frac{\text{Total \# of (new and old) cases of disease during a time period (or at one point in time)}}{\text{Total (usually mid-period) population during the same time period}} \times 1{,}000$$

The numerator of a prevalence ratio includes all current cases (both new and old) during a specified time period divided by the total population during that same time period. Prevalence ratios may be calculated for a period of time (called period prevalence) or for a particular point in time (called point prevalence).

iii. Comparison of Incidence Rates and Prevalence Ratios

Incidence rates and prevalence ratios provide different types of information. ***Incidence rates provide an estimate of risk for developing a disease***. This information is useful for clinicians to estimate the risk that a patient has developed a particular infection or disease (such as TB), as well as for policy makers wishing to identify geographic locations or population groups that may be identified as "high risk."

In contrast, ***prevalence ratios provide a measure of how many people have been infected*** (both new and old infection) as well as the proportion of the population with a particular disease and, therefore, a measure of the burden of disease in the population. This information would be useful for decision makers who allocate resources. The next box provides a review of how these measures are calculated and used.

Measures of Morbidity: Incidence Rate vs Prevalence Ratio	
Incidence Rate	**Prevalence Ratio**
Numerator • *New* cases during a time period	**Numerator** • *New* and *old* cases at one point in time or during a time period
Denominator • Population at risk or person-time* • Excludes pre-existing cases during a specified time period	**Denominator** • Total population • At one point in time or during a time period
Use • Estimate of risk	**Use** • Burden of disease

* Sometimes epidemiologists can actually estimate something called person-time (the number of people multiplied by the length that they were studied). Person-time means that if 1 person was studied for 2 years and another was studied for half a year, then in total they would have been studied for 2.5 person-years. Person-time provides a more precise estimate of the time that a person was at risk for developing the disease. This is more likely to be done in small studies than for population rates. When person-time is used in the denominator of an incidence rate, then the rate is called *incidence density*.

iv. Sample Calculations: Incidence and Prevalence

Surveillance reports generated from TIMS data (available at: http://www.cdc.gov/nchstp/tb/surv/Surv.htm) can be used along with US Census data to calculate the incidence rate (also known as the case rate). According to the surveillance report, 16,377 cases of TB were reported to the CDC in 2000.

The Census Bureau offers a quick way to find population data through its Website (available at http://www.census.gov). At the Census Bureau Website, go to **American FactFinder**, a product that can be easily accessed and provides population level data (for the nation, by state, county, and census-tract). According to the Census Bureau, the total population in the United States in 2000 was 281,421,906.

Incidence Rate

Using the incidence rate formula:

$$\frac{\text{16,377 new cases of TB}}{\text{281,421,906}} \times 100{,}000 = \text{5.8 per 100,000 population in the United States in 2000}$$

The CDC published this rate for the total United States, but this same method could be used to calculate a rate for a state or local area.

Prevalence Ratio

An example of a study that allowed for the calculation of TB prevalence follows:

In a study in New York City from 1994 to 2001, researchers wanted to determine the prevalence of latent TB infection (LTBI) among New York City Department of Health and Mental Hygiene employees. The investigators collected baseline tuberculin skin test (TST) positivity data:

Total # of employees tested: 1,658

# of employees TST-positive: 600

$$\text{Prevalence of TST positivity} = \frac{\text{Total \# of employees with positive test}}{\text{Total \# of employees}} \times 1{,}000$$

$$= \frac{600}{1{,}685} \times 1{,}000$$

$$= \text{361.8 per 1,000 employees}$$

Source: Cook S, Maw KL, Munsiff SS, Fujiwara PI, Frieden TR. Prevalence or tuberculin skin test positivity and conversions among healthcare workers in New York during 1994-2001. *Infect Control and Hosp Epidemiol*. 2003;24:807-813. Data reprinted here with permission.

Distinguishing Incidence From Prevalence

It is important to note that the employees who had a positive TST result during this baseline survey could be either incident (new infection) cases or old infections. If this survey were repeated in this same group a year later and new TST-positive cases appeared, then the researchers could calculate the incidence of TB infection in this group. For example, if during a 1-year period following the baseline survey, a certain number of new infections were identified among these employees, the incidence rate would be calculated as follows:

$$\text{Total \# of employees at risk of developing TB infection} \quad 1{,}658 - 600 = 1{,}058$$

\# of employees with new TST-positive results = A

$$\text{Incidence rate of TST positivity} = \frac{\text{\# of new employees with positive TST}}{\text{\# of at-risk employees}} \times 1{,}000$$

$$= \frac{A}{1{,}058 \text{ employees}} \times 1{,}000$$

Sample Problems: Incidence and Prevalence

Suppose that a county TB controller would like to know how many people currently living in a local homeless shelter are TST positive. After receiving the appropriate approval and consent from the members of the shelter, she has a trained nurse plant the TSTs and read the results. Of 100 homeless shelter residents, 40 had a positive TST. As it turns out, all 100 residents remained in this shelter for the next year at which time only those who did not have an initial positive TST result were tested again. Among these 60 residents, 20 had a positive test result.

Calculate:

A. The baseline prevalence of TB infection at this homeless shelter.

B. An estimate of the risk of developing TB infection in this population.

Answers to sample problems appear in Appendix IV.

A note of caution about morbidity data:

The quality of morbidity data is not as high as mortality data since disease (as compared with death) is more subjective and is only recorded if a person seeks care and the information about that care is recorded. If a person never seeks care, that person's information would be missing from incidence rates and prevalence ratios. This also means that the most severe cases of disease are more likely to be reported than less severe cases. This may suggest that a very large proportion of cases will die from a disease or infection when, in fact, the less severe cases are just not being reported. In addition, some states require reporting of particular diseases and others do not have this requirement.

Finally, since no test is perfect, the validity of the test that is used to make the disease diagnosis may affect the quality of the morbidity measure. Validity will be discussed in further detail in the following section.

v. Measuring Test Validity

Validity indicates how well a test measures what it is supposed to be measuring. The following measures are used to describe how well a test performs: sensitivity, specificity, positive predictive value, and negative predictive value of the test.

The formulas for all 4 measures are seen below:

Test Validity			
	Disease/Infection "Gold Standard" or "The Truth"		
New Test Result	**Yes**	**No**	**Total**
Positive	a	b	a + b
Negative	c	d	c + d
Total	**a + c**	**b + d**	**a + b + c + d**

Sensitivity = a/ a + c
Specificity = d/ b + d
Predictive value of a positive test = a/ a + b
Predictive value of a negative test = d/ c + d

Sensitivity indicates how well a test identifies if someone ***has*** a disease or infection:

Sensitivity

$$\frac{\text{\# of people } \textit{with} \text{ disease/infection who test } \textit{positive} \text{ for the disease/infection}}{\text{Total \# of people who truly } \textit{have} \text{ the disease/infection}} \quad \text{or} \quad \frac{a}{a+c}$$

If there are 100 people who are known to have a disease or infection (based on what is termed "the gold standard") and 90 of these 100 were identified as having this disease or infection using a new diagnostic test, then the new test is said to have 90/100 or 90% sensitivity.

Specificity indicates how well a test identifies if someone ***does not have*** a disease or infection.

Specificity

$$\frac{\text{\# of people } \textit{without} \text{ disease/infection who test } \textit{negative} \text{ for the disease/infection}}{\text{Total \# of people who truly } \textit{do not have} \text{ the disease/infection}} \quad \text{or} \quad \frac{d}{b+d}$$

Of 100 individuals who were known not to have a disease or infection, if 95 of these 100 were identified by the new test as not having the disease or infection, then the new test is said to have 95/100 or 95% specificity.

Sensitivity and **specificity** are values that are determined by using a test among people when it is known whether they actually have the disease or infection. Therefore, these measures are values that are determined in an "epidemiology laboratory."

Assuming that the "truth" can be known about any given individual, these measures can be calculated. In reality, the measurement that is called the "gold standard" is not perfect and there is some amount of error associated with it as well.

In order to know how well a screening or diagnostic test will perform in any population, the **positive predictive value** and the **negative predictive value** of the test result must be calculated.

The positive predictive value is a measure of the likelihood that a person who tests ***positive*** for a disease or infection actually has the disease or infection.

Positive Predictive Value

# of people who test positive who who actually *have* disease/infection
Total # of people who test positive for disease/infection

or

$$\frac{a}{a+b}$$

The **negative predictive value** is a measure of the likelihood that a person who tests ***negative*** for a disease or infection actually does not have the disease or infection.

Negative Predictive Value

# of people who test negative who actually *do not have* disease/infection
Total # of people who test negative for disease/infection

or

$$\frac{d}{c+d}$$

To summarize, **sensitivity** and **specificity** indicate how well a test performs in an ideal setting, while the **predictive values**, for any given patient or group of patients coming from a given high or low prevalence population, reveal how well the test predicts the presence of disease or infection. All 4 measures are expressed as percentages.

vi. Test Validity Example

Two examples of how to generate these values and how to interpret findings appear on the following pages.

For these examples, assume that the test result in the table is the TST result and the gold standard is the truth about whether someone is actually infected. The medical literature suggests that the TST performs quite well and has a sensitivity of approximately 99% and a specificity of approximately 95%.* These values are used in both examples.

1. First calculate the positive and negative predictive values of the TST.

Assume that the test is being conducted in a population of 1,000 with a TB prevalence of 1%. Since 1% of 1,000 people equals 10 people, 10 people of the population of 1,000 are infected and 990 people are not infected. These values are shown in the table below:

	Truly Infected		
TST Result	**Yes**	**No**	**Total**
Positive	a	b	a + b
Negative	c	d	c + d
Total	**10**	**990**	**1,000**

Since the sensitivity and specificity of the test are known, the values of **a**, **b**, **c**, and **d** can now be calculated:

With a sensitivity of 99%, this means that 99% of 10 infected people or 9.9 would replace the box where the "**a**" appears above. By subtraction, 0.1 person would appear in the box labeled "**c**."

With a specificity of 95%, this means that 95% of 990 infected people or 940.5 would be in the box previously labeled "**d**." By subtraction, 49.5 people would appear in the box previously labeled "**b**."

* Huebner E, Schein MF, Bass JB Jr. The Tuberculin Skin Test. Clin Infect Dis. 1993;17:968-75.

These values can then be used to fill in the remaining cells in the table. By adding the rows across, the table shows that 9.9 + 49.5 = 59.4 total TST-positive results and 0.1 + 940.5 = 940.6 total TST-negative results.

	Truly Infected		
TST Result	**Yes**	**No**	**Total**
Positive	9.9	49.5	59.4
Negative	0.1	940.5	940.6
Total	**10**	**990**	**1,000**

The predictive values for the TST may be calculated using the completed table above.

The **positive predictive value of a TST** will tell how likely it is that a patient who has a positive TST is really infected with TB.

$$\textbf{Positive Predictive Value of a TST} = \frac{\mathbf{a}}{\mathbf{a+b}} = \frac{9.9}{59.4} \times 100 = 17\%$$

The positive predictive value of a TST in this population is 17%. This means that approximately 17% of the time if a patient in this population has a positive TST, the patient is truly infected with TB.

The **negative predictive value of a TST** shows the likelihood that a patient with a negative TST is really NOT infected with TB.

$$\textbf{Negative Predictive Value of a TST} = \frac{\mathbf{d}}{\mathbf{c+d}} = \frac{940.5}{940.6} \times 100 = 99.9\%$$

The negative predictive value of the TST in this population is 99.9%. So, 99.9% of the time when a patient in this population has a negative TST, the patient truly is negative.

Interpretation: These data mean that in a population with a very low prevalence of TB infection (eg, 1%), even when the test has good sensitivity and specificity, the positive predictive value of the TST is not very good. Thus, there will likely be many results in which people who are not truly infected will receive a positive test result. This is known as a false-positive result. Since there is a low background prevalence of TB in the United States, testing is focused on high-risk groups, rather than the general population. In addition, for TB infection, the interpretation of a patient's positive TST result is based, in part, on the risk group to which the patient belongs.

2. If this same test were used in a population with a 20% prevalence of TB infection, 20% of 1,000 (or 200) cases would now appear in the (**a**+**c**) box. By subtraction, 800 people would appear in the (**b**+**d**) box.

Truly Infected			
TST Result	**Yes**	**No**	**Total**
Positive	a	b	a + b
Negative	c	d	c + d
Total	**200**	**800**	**1,000**

Using the original values of 99% sensitivity and 95% specificity from the previous example:

With a sensitivity of 99%, this means that 99% of 200 infected people or 198 would replace the box where "**a**" appears above. By subtraction, 2 people would appear in the box labeled "**c**."

With a specificity of 95%, this means that 95% of 800 infected people or 760 would be in the box previously labeled "**d**." By subtraction, 40 people would appear in the box previously labeled "**b**."

The completed table follows:

Truly Infected			
TST Result	**Yes**	**No**	**Total**
Positive	198	40	238
Negative	2	760	762
Total	**200**	**800**	**1,000**

Positive Predictive Value of the TST

$$= \frac{a}{a+b}$$

$$= \frac{198}{238} \times 100$$

$$= 83.2\%$$

Negative Predictive Value of a TST

$$= \frac{d}{c+d}$$

$$= \frac{760}{762} \times 100$$

$$= 99.7\%$$

Interpretation: In a population with a higher prevalence of infection (20% compared with 1%), the TST performs better. In a population with a TB infection rate of 20%, a patient with a positive TST will have an 83% likelihood of being truly infected, as compared to a 17% likelihood in a population with a TB infection rate of 1%.

3. Sample Problems: Sensitivity, Specificity, and Predictive Values

Suppose that a TB controller wanted to know how well an acid-fast bacilli (AFB) smear result predicts disease among patients who are suspected of having TB. These data are collected in a group of 630 suspects and are summarized in the table below:

Sputum Smear Result	Sputum Culture Result or "Gold Standard"		Total
	+	–	
+	185	45	230
–	95	305	400
Total	**280**	**350**	**630**

A. What is the prevalence of a positive sputum culture in this population?

B. What is the sensitivity of the sputum smear result?

C. What is the specificity of the sputum smear result?

D. What is the negative predictive value of the sputum smear result?

E. What is the positive predictive value of the sputum smear result?

Answers to these questions can be found in Appendix IV.

B. Mortality

i. Measures of Mortality

Mortality is easier to define than morbidity because death is a certain event. The main source of mortality data in the United States is the standard US death certificate. This information is collected by states and kept by the National Center for Health Statistics.

Taking the total number of people who died from all causes in 2003 in the United States and dividing that number by the total population during 2003, establishes the crude mortality rate, also known as the crude death rate. Population information is available through the US Census Bureau.

Crude Mortality Rate

$$\frac{\text{\# of deaths in 1 year}}{\text{Total midyear population}} \times 1{,}000$$

\# deaths → Vital Registration System
Total midyear population → Census Bureau

These rates are called crude rates because they do not account for other factors that might have an impact on the mortality rate, such as age, sex, and race of the population.

Age (or other factors) can be accounted for in several ways, first, by calculating the age-specific mortality rate using the formula in the next box. This calculation reports the death rate for a segment of the population within a specific age range.

Age-Specific Mortality Rate

$$\frac{\text{\# of deaths in 1 year in age group A}}{\text{Total midyear population of age group A}} \times 1{,}000$$

Note: "Specific" applies to both the numerator (the people who die) and the denominator (the people at risk). The death rate may be calculated per 100, 1,000, or 100,000.

Further discussion on crude vs age-adjusted mortality rates is found in the following sample calculation.

ii. Sample Calculations: Age-Specific Mortality Rate

Crude Mortality Rates

The **crude mortality rates** for Alaska and Florida in 2000 appear in the following table:

2000 Crude Mortality Rates: Alaska and Florida		
	Alaska	**Florida**
Number of deaths	2,922	162,804
Population	626,932	15,982,378
Crude mortality rate =	$\frac{2,922}{626,932} \times 100,000$	$\frac{162,804}{15,982,378} \times 100,000$
	466.1 per 100,000	1,018.6 per 100,000

Sources:
Alaska Bureau of Vital Statistics 2000 Annual Report, Table 15. Available at: http://health.hss.state.ak.us/dph/bvs/PDFs/2000/annual_report/Deaths.pdf

Florida Bureau of Vital Statistics 2000 Annual Report, Chart D-4. Available at: http://www.doh.state.fl.us/planning_eval/vital_statistics/00vitals/deaths.pdf

US Census Bureau State and County Quick Facts. Available at: http://quickfacts.census.gov/qfd/

Based on these crude death rates, a number of questions arise, as well as possible explanations, or hypotheses.

For example:

- Based on these crude rates, which population is healthier?
- Is Florida an unhealthy environment?
- Is the risk of dying in Florida more than double that of the risk of dying in Alaska?
- Is Florida an "older" population and, therefore, would more people be expected to die there than in "young" Alaska?

Some additional information can be found by looking at US Census information.

States Ranked by Percent of Population Age 65 or Older, 2000				
Rank	**State**	**Total resident population (thousands)**	**Population age 65+ (thousands)**	**Percentage of population age 65+**
1	Florida	15,982	2,808	17.6
51	Alaska	627	36	5.7

Source: US Census Bureau. *Demographic Profiles: Census 2000* (available at: www.census.gov/Press-Release/www/2001/demoprofile.htm). Christine L. Hilmes. PRB's Population Bulletin, *Elderly Americans.* Reprinted here with permission.

The US Census Bureau information reveals that Florida has the highest percentage of people 65 years of age or older, and Alaska has the lowest, suggesting that some of the difference in mortality could be explained by the different age distributions of these populations. One way to adjust or control for the difference in age distribution and to answer some of the previous questions is to calculate age-specific mortality rates.

Age-Specific Mortality Rate

The following table represents population and death statistics by age group for Alaska and Florida in 2000.

	Alaska (2000)		Florida (2000)	
Age group (years)	**Population**	**No. of Deaths**	**Population**	**No. of Deaths**
<5	47,591	88	945,823	1,716
5-14	110,432	27	2,088,742	380
15-24	89,986	130	1,942,377	1,734
25-44	203,522	359	4,569,347	8,280
45-64	139,702	747	3,628,492	25,066
>65	35,699	1,571	2,807,597	125,628
Total	**626,932**	**2,922**	**15,982,378**	**162,804**

Sources:
Alaska Bureau of Vital Statistics 2000 Annual Report, Table 16A.
Available at:
http://health.hss.state.ak.us/dph/bvs/PDFs/2000/annual_report/Deaths.pdf
Florida Bureau of Vital Statistics 2000 Annual Report, Chart D-4.
Available at:
http://www.doh.state.fl.us/planning_eval/vital_statistics/00vitals/deaths.pdf

A separate rate for each age grouping can be generated, using data from the previous tables and the formula for age-specific mortality rates:

$$\text{Age-Specific Mortality Rate} = \frac{\text{\# of deaths in 1 year to age group A}}{\text{Midyear population of age group A}} \times 100{,}000$$

For example, the age-specific mortality rate for children less than 5 years of age in Florida is:

$$\text{Age-Specific Mortality Rate} = \frac{1{,}716}{945{,}823} \times 100{,}000$$

$$= 181.4 \text{ per } 100{,}000$$

Finally, using the above calculation, the age-specific mortality rates for Alaska and Florida can be calculated. The table below shows the age-specific mortality rates for both states:

Age-Specific Mortality Rates (2000)		
Age group (years)	**Alaska**	**Florida**
<5	181.3	181.4
5-14	24.6	18.2
15-24	144.5	89.3
25-44	130.8	181.2
45-64	534.7	690.8
>65	4,400.7	4,474.6

A comparison of the age-specific mortality rates suggests that the mortality experience in Florida and Alaska is much more similar than suggested by the crude mortality rates. Although there are still differences in mortality rates between Florida and Alaska for each age group, the age-specific rates are clearly not twice as high in Florida as compared with Alaska.

iii. Age-Adjusted Rates

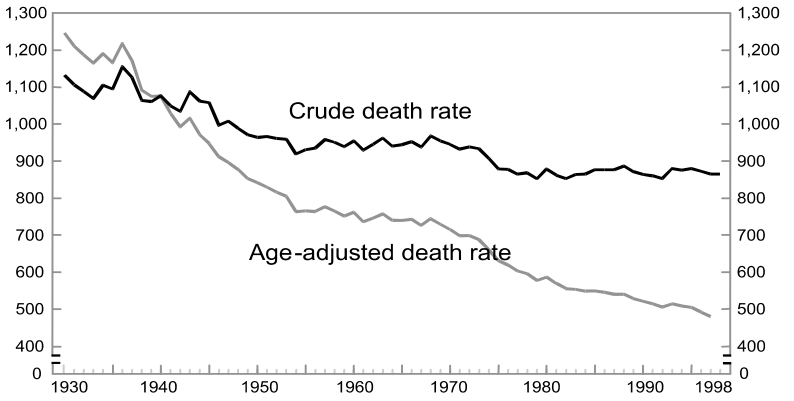
Another way to account for the age structure of a population is to calculate "age-adjusted" or "standardized" rates. This can be done using a few different methods, but the outcome is a summary measure in which age is no longer a factor.*

The figure on the following page presents the crude and age-adjusted death rates by year in the United States from 1930 through 1998. The crude death rate line suggests that mortality has been declining in the United States over time. The age-adjusted death rate line reveals an even more dramatic decline in mortality. Since the US population has been aging during this same time period, by 1998 the US population was an "older" population than it was in 1930, so more people would have been expected to die.

* Those interested in performing age adjustments may refer to the epidemiology textbooks listed at the end of this guide.

Crude and Age-Adjusted Death Rates: United States, 1930–1998

(Rate per 100,000 population)



NOTE: Crude death rates on an annual basis per 100,000 population; age-adjusted rates per 100,000 U.S. standard population

Source: Murphy SL. Deaths: Final Data for 1998. *National Vital Statistics Reports*. 2000; vol 48 no.11. Available at: http://www.cdc.gov/nchs/data/nvsr/nvsr48/nvs48_11.pdf

iv. Case-Fatality Rate

The case-fatality rate is a measure of the severity of a disease. The case-fatality rate presents the risk of dying during a defined period for those who have a particular disease. A disease in which everyone dies would have a case-fatality rate close to 100%. Case-fatality is often calculated when a disease outbreak occurs.

Case-Fatality Rate

$$\frac{\text{number of deaths during a specified time period after disease onset}}{\text{number of individuals with that disease during that time period}} \times 100$$

Using the data from the following articie excerpt, the TB case-fatality rate for Baltimore between January 1993 and June 1998 can be calculated.

> "Worldwide, the case-fatality rate of smear-positive pulmonary tuberculosis among patients on treatment is 3.8%. We assessed the case-fatality rate among such patients in Baltimore between January 1993 and June 1998. Tuberculosis incidence was less than 17/100,000 population and 99% of patients received DOT. Of the 174 study patients, 42 (24%) died on treatment. Patients who died were older (mean age: 62 vs. 47 years; $P<0.001$) and more likely to have underlying medical conditions. With effective control, tuberculosis may become concentrated in older persons with chronic diseases and be associated with high case-fatality rates. In such settings, acceptable treatment success rates may need to be revised."

Source: Fielder JF, Chaulk CP, Dalvi M, Gachuhi R, Comstock GW, Sterling TR. A high tuberculosis case-fatality rate in a setting of effective tuberculosis control: implications for acceptable treatment success rate. *Int J Tuberc Lung Dis*. 2002;6:1114-1117. Reprinted here with permission.

The authors of this article state that the case-fatality rate for Baltimore during this time period was 24%. They calculated this measure using the formula listed below:

$$\text{Case-Fatality Rate in Baltimore from 1/93 to 6/98} = \frac{\text{42 study participants who died}}{\text{174 study participants}} \times 100$$

$$= 24.1\ \%$$

In the next excerpt, the authors then compared this case-fatality rate with other populations and suggested that the difference in case-fatality rates may be due, in part, to the different age distributions of the populations being compared.

> "A study by the British Medical Research Council found a 15% fatality rate among patients from England and Wales, compared to 2% among patients from the Indian subcontinent; this difference was attributed in part to the older age of the patients from England and Wales."

Source: Fielder JF, Chaulk CP, Dalvi M, Gachuhi R, Comstock GW, Sterling TR. A high tuberculosis case-fatality rate in a setting of effective tuberculosis control: implications for acceptable treatment success rates. *Int J Lung Tuberc Dis*. 2002;6:1114-1117. Reprinted here with permission.

This is a good example of when age adjustment should be used to compare the case-fatality rates. An adjustment procedure would tell if the age distribution of these populations could account for the observed differences in case-fatality rates.

Sample Problems: Case-Fatality Rate

In the previous article, the authors stated that "A study by the British Medical Research Council found a 15% fatality rate among patients from England and Wales, compared with 2% among patients from the Indian subcontinent; this difference was attributed in part to the older age of the patients from England and Wales."

A. With a 15% case-fatality rate, if 100 people had TB, how many would die during the study period?

B. Why did the authors attribute the difference in case-fatality rate in England and Wales compared with the rate from the Indian subcontinent in part to the age distribution of these patients?

Answers to these questions can be found in Appendix IV.

v. Cause-Specific Mortality Rate

Another mortality measure that relates to cause of death is the cause-specific mortality rate, also known as the cause-specific death rate.

Cause-Specific Mortality Rate

$$\frac{\text{Deaths due to a cause during a specified time period}}{\text{Total population during that time period}} \times 100{,}000$$

Unlike the case-fatality rate in which the denominator is the number of people with the disease or infection during a specified time period, the denominator of a cause-specific mortality rate is the whole population. Since the numbers of people who die due to any one cause of death are quite small during a 1-year time period, the cause-specific death rate is expressed per 100,000 population. TB mortality in the United States is very low, so the cause-specific death rate is not a statistic that is usually reported in the United States. However, these rates may be calculated, and they are often reported in countries with higher mortality due to TB.

In the chart on the next page, Men and colleagues present the adjusted TB death rates for Russian men and women by year from 1991 to 2001.

Death Rate by Selected Causes at Age 35-69 Years per 100,000 (standardized to world population)

	Age 15-34 years							
	Men				Women			
Cause of death	**91**	**94**	**98**	**01**	**91**	**94**	**98**	**01**
All causes	298	457	392	454	82.1	117	109	124
Infectious diseases:								
All	6.5	11.2	16.9	21.6	2.1	3.1	4.3	5.6
Tuberculosis	5.2	9.2	13.2	17.5	1.1	1.7	2.9	3.6

	Age 35-69 years							
	Men				Women			
Cause of death	**91**	**94**	**98**	**01**	**91**	**94**	**98**	**01**
All causes	1,789	2,814	2,117	2,566	674	969	756	873
Infectious diseases:								
All	34	64.2	68	74.1	4.6	9	7.2	10.7
Tuberculosis	30.4	56.5	63.9	68	2.5	4.6	6	7.6

Source: Men T, Brennan P, Boffetta P, Zaridze D. Russian mortality trends for 1991-2001: analysis by cause and region. *BMJ.* 2003;327:964. Reprinted here with permission.

Sample Problem: Cause-Specific Mortality Rate

A. What type of TB rates are presented in the above table?

Answer to this question can be found in Appendix IV.

To summarize…

The quality of morbidity data is dependent upon whether a patient is willing or able to seek care, the severity of the illness, the type of public health surveillance required by law, whether the provider reports the illness, as well as the sensitivity, specificity, and predictive values of the tests used to identify the disease or infection.

When compared with morbidity data, mortality data are of much higher quality due to the certainty of the event and high level of death reporting in the United States. Although TB mortality rates are not often used to describe epidemiologic trends in the United States, in other areas of the world where TB is a major cause of death, TB mortality or death rates are often reported.

Finally, the adjustment procedures described in this section may be applied to morbidity as well as mortality rates, and can be used to adjust for factors other than age.

5.0 What Is Tuberculosis Genotyping?

Our understanding of TB epidemiology and transmission, which was traditionally based on findings of case and contact tracing, has been enhanced in recent years by TB genotyping. TB genotyping refers to several techniques used to analyze DNA from *Mycobacterium tuberculosis* colonies that have been cultured from specimens collected from TB patients. When TB bacteria reproduce, they create new genetically identical bacilli. However, in some cases, random mutations occur spontaneously, creating different strains of TB, which then reproduce. Because of this, there are now numerous diverse strains of *M tuberculosis* present around the world. TB genotyping techniques can identify the specific strain of bacteria with which a patient is infected.

Using TB genotyping to identify the strain of *M tuberculosis* can assist in:

- Identifying patients involved in recent transmission
- Confirming if 2 patients really share the same strain, or they acquired TB from different sources
- Tracing the chain of TB transmission
- Differentiating between reactivation and re-infection

For example, when a person has TB and is improving, but then becomes sick again, TB genotyping can identify whether the patient has the same strain as before. A different strain indicates that the patient was infected with a different strain of bacteria, as opposed to reactivation with the same strain. TB genotyping also allows TB programs to:

- Detect and control outbreaks earlier
- Identify false-positive culture results more easily
- Identify unknown relationships between cases and unrecognized places of transmission
- Detect transmission between patients in different jurisdictions
- Evaluate effectiveness of routine contact investigations

Further, since national and international databases and collections of clinical *M tuberculosis* strains were established in different TB centers worldwide, using these databases to compare strains isolated from individual TB patients might increase understanding of TB transmission pathways and serve as a tool for evaluation of TB program outcomes.

TB programs may take a variety of steps after analyzing TB genotyping results including: expanding contact investigations, conducting outbreak investigations, performing cluster investigations to locate epidemiologic links between patients, or assessing if a specific patient had a false-positive culture report.

Currently there are 3 main genotyping techniques; spoligotyping, mycobacterial interspersed repetitive units (MIRU) analysis, and IS6110-based restriction fragment length polymorphism (RFLP), also known as DNA fingerprinting. Spoligotyping and MIRU analysis are based on the polymerase chain reaction (PCR). All 3 methods require culture-positive samples from the patient. After the specimens are sent for culture evaluation and have grown out *M tuberculosis*, they can be sent for genotyping. PCR methods can be performed on nonviable cultures, and require much less isolate material than RFLP tests. Usually for RFLP, the genotyping laboratories must place the isolates in culture medium and wait until sufficient growth has taken place to perform the analysis.

TB programs in the United States can utilize the CDC Tuberculosis Genotyping Program. All isolates will be analyzed using spoligotyping and MIRU analysis and selected isolates will be analyzed using the RFLP method. Additional information on TB genotyping and its applicability for TB programs may be found in the CDC document: *Guide to the Application of Genotyping to Tuberculosis Prevention and Control: Handbook for TB Controllers, Epidemiologists, Laboratorians, and Other Program Staff* (available at http://www.cdc.gov/nchstp/tb/genotyping/toc.htm).

6.0 Study Design

There are 3 major types of epidemiologic studies that appear in the medical and public health literature:

- Cross-sectional studies
- Case-control studies
- Cohort studies

A. Cross-Sectional Studies

Cross-sectional studies provide information on possible risk factors and disease outcomes at the same point in time. They are sometimes called ***prevalence studies*** since they can provide prevalence ratios. The data collected presents a picture of what is occurring at a specific time. Cross-sectional studies cannot provide information on causes of diseases since it is unclear in these studies whether the disease or the supposed risk factor occurred first. Cross-sectional studies are usually descriptive, in that they describe the disease or condition in a population at a given time, in terms of person, place, and time. The following excerpt provides an example of a cross-sectional or prevalence study:

Study Design: Cross-Sectional Study

"**Objective:** To determine the prevalence of and risk factors for tuberculin skin test positivity and conversion among New York City Department of Health and Mental Hygiene employees.

Design: Point-prevalence survey. Sentinel surveillance was conducted from March 1, 1994 to December 31, 2001.

Participants: HCWs in high-risk and low-risk settings for occupational TB exposure.

Results: Baseline tuberculin positivity was 36.2% (600 of 1,658), 15.5% (143 of 922) among HCWs born in the United States, and 48.5% (182 of 375) among HCWs not born in the United States."

Source: Cook S, Maw KL, Munsiff SS, Fujiwara PI, Frieden TR. Prevalence of tuberculin skin test positivity and conversions among healthcare workers in New York City during 1994 to 2001. *Infect Control Hosp Epidemiol.* 2003;24:807-813. Reprinted here with permission.

Information from cross-sectional studies can help researchers formulate a hypothesis or theory, as discussed earlier. For example, in the above study, the prevalence survey suggests that health care workers born outside the United States are more likely to have a positive TST result. The reason for this high rate cannot be ascertained from the prevalence survey, however, since it contains only information from one point in time. Other study designs must be used (such as case-control and cohort studies) to more fully explore the relationship between the risk factor (in this example, place of birth) and the outcome (in this example, TST result).

B. Case-Control Studies

Case-control studies are a type of analytic epidemiologic study that allow the researcher to *estimate* the strength of the association between the disease and a particular risk factor. Cases are people with disease or infection, while controls do not have the disease or infection. Once the cases and controls are identified, they are then questioned about potential risk factors that occurred in their past. Case-control studies are especially useful when the disease outcome being studied is rare, since in an observational study of a rare event, only a few cases might ever be identified.

Study Design: Case-Control

An analytic study that compares a group of people with a certain disease, chronic condition, or type of injury (case-patients) with a group of people without the health problem (controls) to detect differences in characteristics such as exposure to an agent.

Available at: http://www.cdc.gov/excite/library/glossary.htm

The following excerpt from an article by Lobato and Hopewell describes a case-control study.

Study Design: Case-Control

"To assess whether there is increased risk of tuberculosis infection in children who traveled to or had a household visitor from a country having a high prevalence of tuberculosis, we conducted a case-control study. Children younger than 6 years of age who had a tuberculin skin test read at public health clinics in areas of California that have a high prevalence of tuberculosis were enrolled. Of the 953 children who had a skin test read, 72 (7.6%) had a positive reaction…"

Source: Lobato MN, Hopewell PC. *Mycobacterium tuberculosis* infection after travel to or contact with visitors from countries with a high prevalence of tuberculosis. *Am J Respir Crit Care Med*. 1998;158:1871-1875. Reprinted here with permission.

i. Odds Ratios

An odds ratio is the usual measurement that results from a case-control study.

Odds Ratio

A measure of association used in comparative studies to quantify the relationship between an ***exposure*** and a health ***outcome***; also known as the cross-product ratio.

Available at: http://www.cdc.gov/excite/library/glossary.htm

The odds ratio is the ratio of the odds that cases were exposed to a particular risk factor as compared with the odds that the controls were exposed to that same risk factor. The odds ratio can be calculated using a simple 2-by-2 table similar to the one used to calculate measures of test validity. This table includes information on the suspected risk factor (travel, in this case) and the outcome (TST result, in this case). The odds ratio is calculated by generating a cross-products ratio (see examples and interpretation in the following example).

The standard 2-by-2 table used to calculate odds ratios is outlined below:

	Cases	Controls	Total
Exposed	a	b	a + b
Not Exposed	c	d	c + d
Total	**a + c**	**b + d**	**a + b + c + d**

Using this table, the odds ratio can be calculated as follows:

$$\textbf{Odds ratio} = \frac{\mathbf{a \times d}}{\mathbf{b \times c}}$$

ii. Sample Calculation: Odds Ratio

In a hypothetical study similar to the one conducted by Lobato and Hopewell, the odds ratio can be calculated to assess whether children who travel to, or have household visitors from, countries with a high prevalence of TB are more likely to have LTBI than children who do not travel to or have household visitors from high prevalence countries.

A 2-by-2 table can be created as follows:

	Skin Test Result		
Risk Factor	**Positive**	**Negative**	**Total**
Travel	a	b	a + b
No Travel	c	d	c + d
Total	**a + c**	**b + d**	**a + b + c + d**

By inserting the hypothetical data, the table would provide the following information:

	Skin Test Result		
Risk Factor	**Positive**	**Negative**	**Total**
Travel	20	100	120
No Travel	45	900	945
Total	**65**	**1,000**	**1,065**

In this hypothetical study, 1,065 children were included. Of the 1,065 children, 65 had positive TST results; of these 65, 20 reported travel to or household visitors from a high prevalence country.

Using the basic calculation:

$$\text{odds ratio} = \frac{\mathbf{a} \times \mathbf{d}}{\mathbf{b} \times \mathbf{c}}$$

$$\text{odds ratio} = \frac{(20 \times 900)}{(100 \times 45)}$$

$$= 4$$

The interpretation of this hypothetical odds ratio is: Children who are TST-positive have 4 times the odds than those who are TST-negative to have traveled to or had visitors from a high prevalence country.

While case-control studies are quite useful, cases and controls are asked to recall events that occurred in the past. For example, in the Lobato and Hopewell study, the parents were asked to recall events that occurred up to 1 year ago. In contrast, cohort studies, described in the next section, do not require participants to recall past events.

C. Cohort Studies

In a cohort study, researchers collect information on a group of exposed and unexposed individuals over time and then calculate incidence rates. These incidence rates allow for the direct calculation of a measure of association between a risk factor and an outcome, called the relative risk.

Study Design: Cohort Study

- An observational analytic study in which enrollment is based on status of exposure to a certain factor or membership in a certain group. Populations are followed and disease, death, or other health-related outcomes are determined and compared.
- Rate ratio. A comparison of two groups in terms of incidence rates, person-time rates, or mortality rates.
- Relative risk. A comparison of the risk of a health problem in 2 groups.

Available: http://www.cdc.gov/excite/library/glossary.htm

Cohort studies, as compared with cross-sectional and case-control studies, provide the most useful epidemiologic measures (incidence rates), but in general they take the longest to complete and are more costly and labor intensive. In addition, some participants will fail to complete the study and this loss (known as loss to follow-up) could bias the results of the study.

i. Relative Risk

The relative risk (RR) is sometimes called a rate ratio.

Relative Risk

$$\frac{\text{Incidence rate in the group exposed to the risk factor}}{\text{Incidence rate in the unexposed group}}$$

A relative risk of 2 means that the risk of developing a particular outcome or disease is twice as high among those with the risk factor as among those without the risk factor. To calculate the relative risk, the incidence of the disease in both unexposed and exposed groups must be known.

The following abstract describes a cohort study.

Study Design: Cohort Study

"**Objectives:** 1) Demonstrate the importance of maintaining a tuberculosis (TB) control program even in low-incidence areas by studying a TB-contact investigation of a highly infectious high school student in rural Missouri, and 2) discuss factors that perpetuated or contained this school-based outbreak.

Methods: A case review of the index patient, a 15-year-old high school student, established estimates of his level and duration of infectiousness. Contact investigations of his household ($n = 5$), high school ($n = 781$), and school bus ($n = 67$) were administered according to guidelines established by the Centers for Disease Control and Prevention. High school students were stratified further based on classroom exposure, and relative risks were calculated for each risk group.

Results: The case review revealed that the index patient had evidence of a pulmonary cavity on chest radiograph 6 months before his TB diagnosis. Of the 5 household contacts, all were infected and 3 (60%) had developed active TB disease. Of the 781 high school students sought for TB screening, 559 (72%) completed testing, and 58 (10%) were PPD-positive. Sixty-seven bus riders were sought for testing and 7 (19%) were purified protein derivative (PPD)-positive, with 1 bus rider subsequently diagnosed with active disease.

Risks were calculated based on classroom and bus exposure to the patient. The relative risks for a positive PPD were 3.2 for attending any class with the patient ($n = 25$), 4.2 for classes with less ventilation ($n = 21$), and 5.7 for ≥ 3 classes ($n = 7$) with the patient. A total of 62 students started treatment for latent TB infection, and 49 have completed it. Forty-two of these students received directly observed therapy through the local public health agency and the high school.

Conclusion: This investigation demonstrated widespread adult-type transmission from a pediatric TB case with a 6-month delay in diagnosis. Several actions contributed to the success of this investigation, including rapidly mobilizing the public health system, centralizing follow-up, and on-site testing and treatment with directly observed therapy. Pediatricians need to maintain awareness of TB and risk factors in children, even in low-incidence areas. Prompt diagnosis would have reduced the severity of illness in the patient and potentially prevented widespread school-based transmission. Public health authorities must maintain an infrastructure to respond to large TB outbreaks."

Source: Phillips L, Carlile J, Smith D. Epidemiology of a tuberculosis outbreak in a rural Missouri high school. *Pediatrics*. 2004;113:e514-519. Reprinted here with permission.

According to the authors, information was collected on all school and bus contacts. Relative risks of TB infection were calculated according to estimated exposure to the index case. The high school had a population of 781 students. Of these 781 students, 559 completed skin testing. The following table presents TST results for students who were in at least 1 class with the index case, compared with those who were not in class with the index case.

Exposure Group	TST Results		
	Positive	Negative	Total
In class with index case	25	81	106
Not in class with index case	33	420	453
Total tested	**58**	**501**	**559**

This table reveals that overall 559 students were tested and 58 were TST positive. Assuming that none of these students had a prior positive TST result and, therefore, they were all "new" infections, then incidence rates for each group can be calculated.

Incidence rate of TST positivity among those who **attended** class with the index case

$$= \frac{\text{25 students attending class with index case who are TST positive}}{\text{106 total students attending class with index case}} \times 100$$

$$= 23.6\%$$

Incidence rate of TST positivity among those who **did not attend** class with the index case

$$= \frac{\text{33 students NOT attending class with index case who are TST positive}}{\text{453 total students NOT attending class with index case}} \times 100$$

$$= 7.3\%$$

Therefore, the relative risk for TB infection (which is calculated as the incidence among the exposed divided by the incidence among the unexposed) would be:

$$\frac{23.6}{7.3} = 3.2$$

This means that students who attended at least 1 class with the index case were slightly more than 3 times as likely to have a positive TST result compared with those who did not attend class with the index case.

The following table is excerpted from the article.

Relative Risks and Confidence Intervals for Results of High School Students and Bus Riders

Persons Tested	No. Sought for Testing	No. Completing Testing	Induration >5 mm	RR	95% Confidence Interval
All students*	781	559	58	NA	NA
All bus riders	67	27	7	2.5†	1.26–4.93
Students in ≥ 1 class	137	106	25	3.2	2.0–5.18
Students in periods 1, 2, and 5–7	80	66	21	4.2	2.6–6.75
Students in ≥ 3 classes	14	13	7	5.7	3.26–10.13

NA indicates not applicable.

*Two students rode the bus and attended school but were not in class with the index patient. The 2 were counted as bus riders when calculating RRs.

†RR compares risk of riding the bus with the index patient versus attending school with index patient.

Source: Phillips L, Carlile J, Smith D. Epidemiology of a tuberculosis outbreak in a rural Missouri high school. *Pediatrics*. 2004;113:e514-519. Reprinted here with permission.

Note that the relative risk with a value of 3.2 that was calculated in the previous example appears in the third row of the RR column in this table. The 95% confidence intervals, which appear in the final column of the table, provide an estimate of how much variation might be expected for this estimate of risk. The 95% confidence interval for the relative risk of 3.2 is 2.0-5.18, meaning that the estimate of increased risk for those attending class with the index case could reasonably vary from 2 times as high to 5.18 times as high as among those who were not exposed to the index case at all. The endpoints which define the confidence interval, in this case 2.01 and 5.18, are also called confidence limits.

In addition to being an example of a cohort study, this is also a good example of how epidemiology and statistics can be used in an outbreak investigation. The relative risks and confidence intervals provide TB controllers with a good estimate of where exposure occurred, information that they then used to concentrate their efforts on testing additional at-risk students. According to the authors, "The students at highest risk for infection, identified through the risk gradient...received a letter from the MO DHSS [Missouri Department of Health and Senior Services]. After these efforts, an additional 87 students were tested; none were positive."

ii. Clinical Trials

A special type of cohort study, which is often used to assess the effectiveness of clinical therapies (eg, a new TB drug regimen), is called a clinical trial. In a clinical trial, individuals are assigned to different therapies and then followed over time to measure the outcome of the therapy.

The most valuable clinical trials are those in which patients are **randomly assigned** to the treatment options, so that high and low-risk patients have an equal chance of receiving each treatment. In addition to random assignment, it is important that clinical trials be "**blinded"** or **"masked**" so that the person receiving the treatment *and* the study evaluators are both unaware of the assigned treatment group.

This blinding or masking avoids a situation whereby a patient or a physician feels so strongly that a new treatment is better than an old one that he or she might unintentionally bias the study outcome. When both patient and evaluator are unaware of the treatment assignment, the study is **"double-blinded."**

When possible, researchers use a **placebo**, or inert substance, in the comparison group, so that patients really do not know which treatment is being used. For ethical reasons, a placebo group may not be used when a standard proven therapy is available. An example of ***a randomized, placebo-controlled, double-blinded trial*** appears in the next abstract.

Study Design: Randomized, Placebo-Controlled, Double-Blinded Trial

"Interleukin (IL)-2 has a central role in regulating T cell responses to *Mycobacterium tuberculosis*. Adjunctive immunotherapy with recombinant human IL-2 was studied in a **randomized, placebo-controlled, double-blinded trial** in 110 human immunodeficiency virus-seronegative adults in whom smear-positive, drug-susceptible pulmonary tuberculosis was newly diagnosed. Patients were randomly assigned to receive twice-daily injections of 225,000 IU of IL-2 or placebo for the first 30 days of treatment in addition to standard chemotherapy. Subjects were followed for 1 year. The **primary endpoint was the proportion of patients with sputum culture conversion** after 1 and 2 months of treatment."

Note that patients are receiving the new treatment or the placebo *in addition* to the standard therapy.

Source: Johnson JL, Ssekasanvu E, Okwera A, et al, Uganda-Case Western Reserve University Research Collaboration. Randomized trial of adjunctive interleukin-2 in adults with pulmonary tuberculosis. *Am J Respir Crit Care Med*. 2003;168:185-191. Epub 2003 Apr 17. Reprinted here with permission.

To summarize:

Cross-Sectional Studies provide a snap shot of the health status of a group at one particular time. They are usually quicker and less expensive than other study designs. They can be used to generate hypotheses regarding risk factors and disease outcome, but they cannot be used to support a causal association. The measurement most often produced is a prevalence ratio.

Case-Control Studies take less time and are less expensive than cohort studies. They are particularly good when the outcome being studied is rare. The study design requires that participants recall exposure to particular risk factors. Therefore, the measure of association may be affected by faulty or biased recall. The measurement most often produced is an odds ratio, which is an estimate of relative risk, when the disease being studied is rare.

Cohort Studies are the most expensive and time consuming of all epidemiologic studies, but they produce incidence rates and relative risks. Cohort studies may be observational, whereby a researcher observes a group over time, or they may be clinical trials used to test new therapies. Since individuals are studied over longer periods of time, compared with case-control or cross-sectional studies, some people may drop out of the study which may bias the incidence rates and relative risk.

Odds Ratio versus Relative Risk: In case-control studies, the incidence of disease in the exposed and unexposed groups is unknown, since some preset number of people with disease and without disease (cases and controls) is specifically selected. Therefore, the relative risk in a case-control study cannot be calculated. Instead, researchers calculate the odds ratio as the measure of association between a risk factor and a disease in a case-control study. In a cohort study, relative risk is calculated directly using available incidence rates. The odds ratio is a good approximation of the relative risk, when the disease being studied is rare.

7.0 Assessing Epidemiologic Studies

When reading articles and assessing epidemiologic studies in journals, it is important to understand how these results are evaluated as statistically significant or not significant.

A. *P*-Values

When testing a hypothesis or research question, the researcher must decide how sure he or she wishes to be about the study results, prior to conducting the study. This is done by choosing a significance or risk level (called the alpha level). The alpha level represents the risk that the researcher is willing to accept that any differences found are due to chance alone. If a test is conducted at the alpha = 0.05 level, it is accepted that 5 out of 100 times or 5% of the time something might be found to be statistically significant when the result actually occurred by chance. If a test is conducted at the alpha = 0.01 level, then the researcher is being more averse to the risk of falsely reporting a significant finding, so that 1 out of 100 times or only 1% of the time, this result will be due to chance alone. Once the statistical test is completed, the *p*-value, generated by a statistical package, is compared to the preset alpha level. If the *p*-value is smaller than the alpha level then the result is statistically significant and unlikely to be due to chance alone.

B. Confidence Intervals

Incidence rates, prevalence ratios, odds ratios, and relative risks are often presented with confidence intervals (CI). *Usually 95% and 99% confidence intervals are reported in the medical literature.* A confidence interval tells the reader how much variability there is associated with a prevalence or incidence estimate or with an odds ratio or relative risk.

For example, with a hypothetical relative risk of 2.0, a hypothetical 95% confidence interval might appear as (CI: 1.5, 2.5), establishing that with 95% certainty the true level of risk due to the risk factor could be as low as 50% higher than those

without the risk factor and as high as 2.5 times higher than those without the risk factor.

If the confidence interval includes the number 1.0, this means that the increased risk is not statistically significant, since a relative risk of 1 means that the incidence of disease or infection among those with the risk factor is the same as the incidence of disease or infection among those without the risk factor. Examining the 95% confidence interval to see if it includes the number 1, is equivalent to conducting a significance test at the alpha = 0.05 level to see if the risk factor is significantly associated with a particular outcome.

C. Confounding Factors

In epidemiologic studies, there may be other factors in addition to the risk factor being tested that will affect or "confound" the results. For example, if researchers using a cohort study to investigate whether men are more likely than women to develop TB disease when infected, the researchers might pick a population of 100 women and 100 men who have LTBI and follow them for 10 years. However, clearly there are other factors that will affect whether a patient progresses to TB disease, such as HIV infection, age, presence of diabetes, etc.

Statistical techniques exist to adjust for other factors that have been identified as confounders. When a statistic has been "adjusted" for race, age, or some other factor, the affect of this factor has been removed. When reviewing articles, it is important to note which other variables have been adjusted for, or if the researchers neglected to adjust for other important confounding factors.

D. Types of Data

In traditional epidemiologic studies, data are collected on study subjects using 3 basic measurement scales: **nominal, ordinal, and numerical**. A nominal scale is used to record categorical data. Race, sex, or place of residence are examples of nominal data. An ordinal scale is used to collect information, which has some order, but the distance between each point on the scale is not necessarily the same. For example, patients are often described as having Stage I, II, III, or IV cancer. Stage IV is a more advanced stage of the disease than Stage II, but Stage IV is not necessarily twice as severe as Stage II.

Finally, data are often collected on a numerical scale**.** Numerical data include discrete variables like the number of prior pregnancies or continuous variables such as blood pressure or body weight. All 3 types of data are often described as quantitative, although some researchers refer to data collected on a nominal or ordinal scale as qualitative.

In addition to data collected on nominal, ordinal, or numerical scales, respondents may be asked to describe their feelings about a particular treatment or about their health using open-ended questions. These open-ended questions allow the researchers to collect qualitative information through an analysis of the language the respondents use. An example of such a question is: "Please describe anything which you believe made it difficult for you to complete your treatment for latent tuberculosis infection." Once these responses are transcribed, they are analyzed using a qualitative data analysis software package.

Combined quantitative and qualitative techniques can provide a rich source of information and can be used to validate responses.

Notes:

Appendix I

Common Statistical Terms Used in Epidemiology

Adapted from:
CDC EXCITE Resource Library: Glossary of Epidemiologic Terms.
Available at: http://www.cdc.gov/excite/library/glossary.htm

Agent. A factor that is essential for a disease, chronic conditions, or injury to occur. Examples of agents include microorganisms, chemical substances, forms of radiation, and, in the case of injury, physical force. Agents can cause a health problem by either being introduced, being present in excess, or being present at deficient levels.

Association. The statistical relationship between 2 or more events, characteristics, or other variables.

Case. An instance of a particular disease, chronic condition, or type of injury. A variety of criteria may be used to identify cases (See Case definition), and the epidemiological definition of a case is not necessarily the same as the ordinary clinical definition. (See also Case-patient.)

Case definition. A set of standard criteria for determining whether a person has a particular disease or health condition. A case definition specifies clinical criteria and details of time, place, and person.

Case-fatality rate. The proportion of people with a particular condition (case-patients) who die from that condition. In calculating case-fatality rates, the numerator is the number of people who die from the condition, and the denominator is the total number of people with the condition.

Case-patient. A person in a case-control study who has the disease or health condition under investigation.

Cause of disease. A factor (characteristic, behavior, event, etc) that directly influences the occurrence of a disease. Reducing such a factor in a population should reduce occurrence of the disease.

Census. The enumeration of an entire population, usually including details on residence, age, sex, occupation, ethnic group, marital status, birth history, and relationship to head of household.

Cohort. A well-defined group of people who have had a common experience or exposure and are then followed up, as in a cohort study or prospective study, to determine the incidence of new diseases or health conditions.

Confidence interval. A range of values for a variable (eg, a rate).

Confidence limits. The endpoints (ie, the minimum and maximum values) of a confidence interval.

Contingency table. A table of cross-tabulated data that allows for calculating associations. The 2-by-2 table, with cases tabulated by exposure and outcome, is the contingency table most commonly used in epidemiology.

Control. The group of people without the health problem under study in a case-control study; a person in that group. For controls, investigators choose people who are as similar as possible to the cases, but without the health problem under study. In a case-control study, the control group is compared with the case group to determine associations between exposures and outcomes and to test hypotheses. (See also Study, case-control.)

Demographic information. The personal characteristics of age, sex, race, residence, and occupation. Demographic information is used in descriptive epidemiology to define the population at risk.

Denominator. The lower portion of a fraction. Epidemiologists use fractions to calculate rates or ratios. The denominator is usually the population at risk, although it may also be a measure, such as person-time, that quantifies the population's exposure.

Determinant. Any factor that brings about change in a health condition or in other defined characteristics.

Distribution. The complete summary of the frequency and pattern of the values or categories of a measurement. In epidemiology, distribution is the frequency and pattern of health-related characteristics and events in a population.

Endemic. The constant presence of a disease, chronic condition, or type of injury in a given geographic area or population group; may also refer to the usual prevalence of a disease or condition.

Epidemic. (Syn: outbreak) The occurrence of more cases of a particular type of disease, chronic condition, or injury than expected in a given area, or among a specific group of people, over a particular period of time. (See also Outbreak.)

Epidemiology. The study of the distribution and determinants of health conditions or events in populations, and the application of this study to control health problems.

Epidemiology, analytic. The aspect of epidemiology concerned with why and how a health problem occurs. Analytic epidemiology uses comparison groups to provide baseline data so that associations between exposures and outcomes can be quantified and hypotheses about the cause of the problem can be tested. Examples include cohort studies and case-control studies.

Epidemiology, descriptive. The aspect of epidemiology concerned with gathering, organizing, and summarizing data on "person" (Who is ill?), "time" (When did they become ill?), and "place" (Where could they have been exposed to the illness?). This information is then used to conduct analytic epidemiology.

Exposed group. A group whose members have had contact with a cause of, or possess a characteristic that is a determinant of, a particular health problem.

Exposure. Coming into contact with a cause of, or possessing a characteristic that is a determinant of, a particular health problem.

Frequency. The amount, or number of occurrences, of a disease, chronic condition, injury, or other attribute or event in a population.

High-risk group. A group of people whose risk for a particular disease, health condition, or type of injury is higher than that of the rest of their community or population.

Hyperendemic health problem. A disease, chronic condition, or type of injury that is constantly present at a high incidence and/or prevalence.

Hypothesis. A supposition, arrived at from observation or reflection, that leads to refutable predictions; any conjecture cast in a form that will allow it to be tested and refuted.

Incidence. A rate that measures the frequency with which a health problem, such as a new injury or case of illness, occurs in a population. In calculating incidence, the numerator is the number of new cases occurring in the population during a given period of time, and the denominator is the total population at risk during that time.

Mean, arithmetic. The measure of central location commonly called the average. The arithmetic mean is calculated by adding all the values in a group of measurements and dividing by the number of values in the group.

Measure of association. A quantified relationship between exposure and a particular health problem. Commonly used measures of association include relative risk, rate ratio, and odds ratio.

Median. The middle value in a set of numbers (or the average of two middle numbers) above and below which lie an equal number of values.

Morbidity. Disease; any departure, subjective or objective, from a state of physiological or psychological health and well-being.

Mortality rate. A measure of the frequency of occurrence of death in a defined population during a specified time interval.

Mortality rate, age-adjusted. A mortality rate that has been statistically modified to account for the effect of different age distributions in different populations in a study.

Mortality rate, age-specific. A mortality rate limited to a particular age group. In calculating age-specific mortality rates, the numerator is the number of deaths in the age group, and the denominator is the number of people in that age group.

Mortality rate, cause-specific. The mortality rate from a specified cause. In calculating cause-specific mortality rates, the numerator is the number of deaths attributed to a specific cause during a specified time interval in a population, and the denominator is the size of the population at the midpoint of the time interval.

Mortality rate, crude. A population's mortality rate from all causes of death.

Numerator. The upper portion of a fraction. Epidemiologists use fractions to calculate rates or ratios.

Observational study. An epidemiologic study in which there is no intervention and nature is allowed to take its course. Changes or differences in one characteristic are studied in relation to changes or differences in others.

Odds ratio. A measure of association used in comparative studies to quantify the relationship between an exposure and a health outcome; also known as the cross-product ratio.

Outbreak. (Syn: epidemic) Because the public sometimes perceives "outbreak" as less sensational than "epidemic," it is sometimes the preferred word. Sometimes the two words are sometimes differentiated, with "outbreak" referring to a localized health problem, and "epidemic," to one that takes in a more general area. (See also Epidemic.)

Outcome(s). Any or all of the possible results that may stem from exposure to a causal factor or from preventive or therapeutic interventions; all identified changes in health status that result from the handling of a health problem.

Pandemic. An epidemic occurring over a very wide area (several countries or continents) and usually affecting a large proportion of the population.

Person-time. The total of the units of time, whether weeks, months, or years, that people were exposed to a condition or were actively involved in a study. One person-year can represent a single person who was exposed for one year or an accumulation, such as two people who were each exposed for half a year.

Person-time rate. A measure of the incidence rate of an event (eg, disease, injury, or death) in a population over an observed period. The person-time rate directly incorporates time into the denominator.

Population. The total number of inhabitants of a given area or country. In sampling, the population may refer to the units from which the sample is drawn, not necessarily the total population of people. A population can also be a particular group at risk, such as everyone who is engaged in a certain occupation.

Predictive value, positive. A measure of the likelihood that a person who tests positive for a disease or infection acually has the disease or infection.

Predictive value, negative. A measure of the likelihood that a person who tests negative for a disease or infection acually does not have the disease or infection.

Prevalence. The number or proportion of cases or events or conditions in a given population.

Prevalence, period. The amount of a particular disease, chronic condition, or type of injury present in a population over a period of time. (See also Prevalence, point.)

Prevalence, point. The amount of a particular disease, chronic condition, or type of injury present in a population at a single point in time. (See also Prevalence, period.)

Prevalence ratio. The proportion of people in a population who have a particular disease, chronic condition, injury, or attribute at a specified point in time or over a specified period of time.

Proportion. A ratio in which the numerator is included in the denominator; the ratio of a part to the whole, expressed as a "decimal fraction" (e.g., 0.2), a fraction (1/5), or a percentage (20%).

Public health surveillance. The systematic, ongoing collection, analysis, interpretation, and dissemination of health data. The purpose of public health surveillance is to gain knowledge of the patterns of disease, injury, and other health problems in a community so that we can work toward controlling and preventing them.

Rate. An expression of the relative frequency with which an event occurs in a defined population.

Rate ratio. A comparison of two groups in terms of incidence rates, person-time rates, or mortality rates.

Relative risk. A comparison of the risk of a health problem in two groups.

Risk. The probability that an individual will be affected by, or die from, an illness or injury within a stated time or age span.

Risk factor. An aspect of personal behavior or lifestyle, an environmental exposure, or a hereditary characteristic that is associated with an increase in the occurrence of a particular disease, chronic condition, or injury.

Risk ratio. A comparison of the risk of a particular health problem in two groups.

Sensitivity. The ability of a system to detect epidemics and other changes in the occurrence of health problems; the proportion of people with a health problem who are correctly identified by a screening test or case definition. (See also Specificity.)

Specificity. The proportion of people without a particular disease, chronic condition, or type of injury who are correctly identified by a screening test or case definition. (See also Sensitivity.)

Statistical significance. The measure of how likely it is that a set of study results could have occurred by chance alone. Statistical significance is based on an estimate of the probability of the observed or a greater degree of association between independent and dependent variables occurring under the null hypothesis. The level of statistical significance is usually expressed by the *p*-value.

Study, analytic. A study in which groups are compared to identify and quantify associations, test hypotheses, and identify causes. Two common types are cohort studies and case-control studies.

Study, case-control. An analytic study that compares a group of people with a certain disease, chronic condition, or type of injury (case-patients) with a group of people without the health problem (controls) to detect differences in characteristics such as exposure to an agent.

Study, cohort. (Syn: follow-up, longitudinal, and prospective study) An observational analytic study in which enrollment is based on status of exposure to a certain factor or membership in a certain group. Populations are followed and disease, death, or other health-related outcomes are determined and compared.

Trial, clinical. An experimental study using data from individuals. Investigators identify the type of exposure that each individual has had and then follow the individuals' health status to determine the effects of the exposure.

Trial, randomized clinical. A clinical trial in which individuals are randomly assigned to exposure or treatment groups. (See also Trial, clinical)

Validity. The degree of accuracy of a measurement. For survey instruments, validity refers to what the questions actually measure in practice, as compared with what they are intended to measure.

Variable. Any characteristic or attribute that can be measured and can have different values.

Variable, continuous. A variable that has the potential for having an infinite number of values along a continuum. Common examples are height and weight. (See also Variable, discrete.)

Variable (or data), discrete. A variable that is limited to a finite number of values; data for such a variable. One example would be the number of cigarettes smoked per day by people in a study of smoking and lung cancer. (See also Variable, continuous.)

Appendix II

RVCT Form: Report of Verified Case of Tuberculosis

Patient's Name: _____ (Last) (First) (M.I.)

Street Address: _____ (Number, Street, City, State) (Zip Code)

REPORT OF VERIFIED CASE OF TUBERCULOSIS

CDC Centers for Disease Control and Prevention

DEPARTMENT OF HEALTH & HUMAN SERVICES
PUBLIC HEALTH SERVICE
CENTERS FOR DISEASE CONTROL AND PREVENTION (CDC)
ATLANTA, GEORGIA 30333

FORM APPROVED OMB NO. 0920-0026 Exp. Date 09/30/2005

REPORT OF VERIFIED CASE OF TUBERCULOSIS

SOUNDEX ☐☐☐☐

1. State Reporting:
Specify: _____
Alpha State Code ☐☐

2.
State Case Number: ☐☐☐☐☐☐☐☐☐☐
City/County Case Number: ☐☐☐☐☐☐☐☐☐☐

3. Date Submitted: Mo. ☐☐ Day ☐☐ Yr. ☐☐
By: _____

4. Address for Case Counting:
City ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐
Within City Limits 1 ☐ Yes 2 ☐ No
County ☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐☐
Zip Code ☐☐☐☐☐ – ☐☐☐☐

5. Month-Year Reported: Mo. ☐☐ Yr. ☐☐☐☐

6. Month-Year Counted: Mo. ☐☐ Yr. ☐☐☐☐

7. Date of Birth: Mo. ☐☐ Day ☐☐ Yr. ☐☐☐☐

8. Sex:
1 ☐ Male
2 ☐ Female

9. Ethnicity: (Select one)
1 ☐ Hispanic or Latino
2 ☐ Not Hispanic or Latino

10. Race: (Select one or more)
1 ☐ American Indian or Alaska Native
2 ☐ Asian Specify (Optional): _____
3 ☐ Black or African American
4 ☐ Native Hawaiian or Other Pacific Islander Specify (Optional): _____
5 ☐ White

11. Country of Origin:
If U.S., check here ☐ If not U.S., enter country code (see list) ☐☐

12. Month-Year Arrived in U.S.: Mo. ☐☐ Yr. ☐☐☐☐

13. Status at Diagnosis of TB:
1 ☐ Alive 2 ☐ Dead

14. Previous Diagnosis of Tuberculosis:
1 ☐ Yes
2 ☐ No
Yr. ☐☐☐☐ If yes, list year of previous diagnosis
1 ☐ If more than one previous episode, check here

15. Major Site of Disease:
00 ☐ Pulmonary
10 ☐ Pleural
21 ☐ Lymphatic: Cervical
22 ☐ Lymphatic: Intrathoracic
23 ☐ Lymphatic: Other
29 ☐ Lymphatic: Unknown
30 ☐ Bone and/or Joint
40 ☐ Genitourinary
50 ☐ Miliary
60 ☐ Meningeal
70 ☐ Peritoneal
80 ☐ Other*
90 ☐ Site not Stated
*If site is "Other", enter anatomic code (see list) ☐☐

16. Additional Site of Disease:
00 ☐ Pulmonary
10 ☐ Pleural
21 ☐ Lymphatic: Cervical
22 ☐ Lymphatic: Intrathoracic
23 ☐ Lymphatic: Other
29 ☐ Lymphatic: Unknown
30 ☐ Bone and/or Joint
40 ☐ Genitourinary
50 ☐ Miliary
60 ☐ Meningeal
70 ☐ Peritoneal
80 ☐ Other*
*If site is "Other", enter anatomic code (see list) ☐☐
If more than one additional site, check here ☐ 88

17. Sputum Smear:
1 ☐ Positive 3 ☐ Not Done
2 ☐ Negative 9 ☐ Unknown

18. Sputum Culture:
1 ☐ Positive 3 ☐ Not Done
2 ☐ Negative 9 ☐ Unknown

19. Microscopic Exam of Tissue and Other Body Fluids:
1 ☐ Positive 3 ☐ Not Done
2 ☐ Negative 9 ☐ Unknown
If positive, enter anatomic code(s) (see list) ☐☐ ☐☐

20. Culture of Tissue and Other Body Fluids:
1 ☐ Positive 3 ☐ Not Done
2 ☐ Negative 9 ☐ Unknown
If positive, enter anatomic code(s) (see list) ☐☐ ☐☐

21. Chest X-Ray:
1 ☐ Normal 2 ☐ Abnormal 3 ☐ Not Done 9 ☐ Unknown
If Abnormal (check one) 1 ☐ Cavitary 2 ☐ Noncavitary Consistent with TB 3 ☐ Noncavitary Not Consistent with TB
If Abnormal (check one) 1 ☐ Stable 3 ☐ Improving 2 ☐ Worsening 9 ☐ Unknown

22. Tuberculin (Mantoux) Skin Test at Diagnosis:
1 ☐ Positive 3 ☐ Not Done
2 ☐ Negative 9 ☐ Unknown
Millimeters (mm) of Induration ☐☐
If Negative, was patient anergic? 1 ☐ Yes 2 ☐ No 9 ☐ Unknown

REPORT OF VERIFIED CASE OF TUBERCULOSIS

23. HIV Status: 0 ☐ Negative 3 ☐ Refused 9 ☐ Unknown
1 ☐ Positive 4 ☐ Not Offered
2 ☐ Indeterminate 5 ☐ Test Done, Results Unknown

If Positive, Based on: 1 ☐ Medical Documentation 2 ☐ Patient History 9 ☐ Unknown

If Positive, List: CDC AIDS Patient Number ☐☐☐☐☐☐☐ (If AIDS Reported before 1993)
State HIV/AIDS Patient Number ☐☐☐☐☐☐☐☐☐☐ (If AIDS Reported 1993 or Later)
City/County HIV/AIDS Patient Number ☐☐☐☐☐☐☐☐☐☐ (If AIDS Reported 1993 or Later)

24. Homeless Within Past Year: 0 ☐ No 1 ☐ Yes 9 ☐ Unknown

25. Resident of Correctional Facility at Time of Diagnosis: 0 ☐ No 1 ☐ Yes 9 ☐ Unknown

If Yes, 1 ☐ Federal Prison 3 ☐ Local Jail 5 ☐ Other Correctional Facility
2 ☐ State Prison 4 ☐ Juvenile Correctional Facility 9 ☐ Unknown

26. Resident of Long-Term Care Facility at Time of Diagnosis: 0 ☐ No 1 ☐ Yes 9 ☐ Unknown

If Yes, 1 ☐ Nursing Home 4 ☐ Mental Health Residential Facility 6 ☐ Other Long-Term Care Facility
2 ☐ Hospital-Based Facility 5 ☐ Alcohol or Drug Treatment Facility 9 ☐ Unknown
3 ☐ Residential Facility

27. Initial Drug Regimen:

	NO	YES	UNK.		NO	YES	UNK.		NO	YES	UNK.
Isoniazid	0 ☐	1 ☐	9 ☐	Ethionamide	0 ☐	1 ☐	9 ☐	Amikacin	0 ☐	1 ☐	9 ☐
Rifampin	0 ☐	1 ☐	9 ☐	Kanamycin	0 ☐	1 ☐	9 ☐	Rifabutine	0 ☐	1 ☐	9 ☐
Pyrazinamide	0 ☐	1 ☐	9 ☐	Cycloserine	0 ☐	1 ☐	9 ☐	Ciprofloxacin	0 ☐	1 ☐	9 ☐
Ethambutol	0 ☐	1 ☐	9 ☐	Capreomycin	0 ☐	1 ☐	9 ☐	Ofloxacin	0 ☐	1 ☐	9 ☐
Streptomycin	0 ☐	1 ☐	9 ☐	Para-Amino Salicylic Acid	0 ☐	1 ☐	9 ☐	Other	0 ☐	1 ☐	9 ☐

28. Date Therapy Started: Mo. ☐☐ Day ☐☐ Yr. ☐☐

29. Injecting Drug Use Within Past Year: 0 ☐ No 1 ☐ Yes 9 ☐ Unknown

30. Non-Injecting Drug Use Within Past Year: 0 ☐ No 1 ☐ Yes 9 ☐ Unknown

31. Excess Alcohol Use Within Past Year: 0 ☐ No 1 ☐ Yes 9 ☐ Unknown

32. Occupation (Check all that apply within the past 24 months):
1 ☐ Health Care Worker 3 ☐ Migratory Agricultural Worker 5 ☐ Not Employed within Past 24 Months
2 ☐ Correctional Employee 4 ☐ Other Occupation 9 ☐ Unknown

Comments:

Patient's Name: _____ (Last) (First) (M.I.)

Street Address: _____ (Number, Street, City, State) _____ (Zip Code)

REPORT OF VERIFIED CASE OF TUBERCULOSIS

DEPARTMENT OF HEALTH & HUMAN SERVICES
PUBLIC HEALTH SERVICE
CENTERS FOR DISEASE CONTROL AND PREVENTION (CDC)
ATLANTA, GEORGIA 30333

CDC
CENTERS FOR DISEASE CONTROL AND PREVENTION

REPORT OF VERIFIED CASE OF TUBERCULOSIS

FORM APPROVED OMB NO. 0920-0026 Exp. Date 09/30/2005

Initial Drug Susceptibility Report (Follow Up Report – 1)

SOUNDEX ☐☐☐☐

State Reporting:
Specify: _____
Alpha State Code ☐☐

Year Counted: ☐☐

State Case Number: ☐☐☐☐☐☐☐☐☐☐
City/County Case Number: ☐☐☐☐☐☐☐☐☐☐

Submit this report for all culture-positive cases.

33. Initial Drug Susceptibility Results:

Was Drug Susceptibility Testing Done: 0 ☐ No 1 ☐ Yes 9 ☐ Unknown

If answer is No or Unknown, do not complete rest of report.

If Yes, Enter Date First Isolate Collected for Which Drug Susceptibility Was Done? Mo. ☐☐ Day ☐☐ Yr. ☐☐

34. Susceptibility Results:

	Resistant	Susceptible	Not Done	Unknown
Isoniazid	1 ☐	2 ☐	3 ☐	9 ☐
Rifampin	1 ☐	2 ☐	3 ☐	9 ☐
Pyrazinamide	1 ☐	2 ☐	3 ☐	9 ☐
Ethambutol	1 ☐	2 ☐	3 ☐	9 ☐
Streptomycin	1 ☐	2 ☐	3 ☐	9 ☐
Ethionamide	1 ☐	2 ☐	3 ☐	9 ☐
Kanamycin	1 ☐	2 ☐	3 ☐	9 ☐
Cycloserine	1 ☐	2 ☐	3 ☐	9 ☐
Capreomycin	1 ☐	2 ☐	3 ☐	9 ☐
Para-Amino Salicylic Acid	1 ☐	2 ☐	3 ☐	9 ☐
Amikacin	1 ☐	2 ☐	3 ☐	9 ☐
Rifabutine	1 ☐	2 ☐	3 ☐	9 ☐
Ciprofloxacin	1 ☐	2 ☐	3 ☐	9 ☐
Ofloxacin	1 ☐	2 ☐	3 ☐	9 ☐
Other	1 ☐	2 ☐	3 ☐	9 ☐

Comments:

Patient's Name: _____ (Last) (First) (M.I.)

Street Address: _____ (Number, Street, City, State) (Zip Code)

REPORT OF VERIFIED CASE OF TUBERCULOSIS

CDC Centers for Disease Control and Prevention

REPORT OF VERIFIED CASE OF TUBERCULOSIS

DEPARTMENT OF HEALTH & HUMAN SERVICES
PUBLIC HEALTH SERVICE
CENTERS FOR DISEASE CONTROL AND PREVENTION (CDC)
ATLANTA, GEORGIA 30333

FORM APPROVED OMB NO. 0920-0026 Exp. Date 09/30/2005

Case Completion Report (Follow Up Report – 2)

SOUNDEX

State Reporting:
Specify: _____
Alpha State Code

Year Counted:

State Case Number:

City/County Case Number:

Submit this report for all cases in which the patient was alive at diagnosis.

35. Sputum Culture Conversion Documented:
0 ☐ No 1 ☐ Yes 9 ☐ Unknown

If Yes, Date Specimen Collected on Initial Positive Sputum Culture: Mo. Day Yr.

If Yes, Date Specimen Collected on First Consistently Negative Culture: Mo. Day Yr.

36. Date Therapy Stopped: Mo. Day Yr.

37. Reason Therapy Stopped:
1 ☐ Completed Therapy 3 ☐ Lost 5 ☐ Not TB 7 ☐ Other
2 ☐ Moved 4 ☐ Uncooperative or Refused 6 ☐ Died 9 ☐ Unknown

38. Type of Health Care Provider:
1 ☐ Health Department
2 ☐ Private/Other
3 ☐ Both Health Department and Private/Other
9 ☐ Unknown

39. Directly Observed Therapy:
0 ☐ No, Totally Self-Administered
1 ☐ Yes, Totally Directly Observed
2 ☐ Yes, Both Directly Observed and Self-Administered
9 ☐ Unknown

If Yes, Give Site(s) of Directly Observed Therapy:
1 ☐ In Clinic or Other Facility
2 ☐ In the Field
3 ☐ Both in Facility and in the Field
9 ☐ Unknown

Number of Weeks of Directly Observed Therapy: Weeks

40. Final Drug Susceptibility Results:
Was Follow-up Drug Susceptibility Testing Done? 0 ☐ No 1 ☐ Yes 9 ☐ Unk.

If answer is No or Unknown, do not complete rest of report.

If Yes, Enter Date Final Isolate Collected for Which Drug Susceptibility Was Done: Mo. Day Yr.

41. Final Susceptibility Results:

	Resistant	Susceptible	Not Done	Unknown
Isoniazid	1 ☐	2 ☐	3 ☐	9 ☐
Rifampin	1 ☐	2 ☐	3 ☐	9 ☐
Pyrazinamide	1 ☐	2 ☐	3 ☐	9 ☐
Ethambutol	1 ☐	2 ☐	3 ☐	9 ☐
Streptomycin	1 ☐	2 ☐	3 ☐	9 ☐
Ethionamide	1 ☐	2 ☐	3 ☐	9 ☐
Kanamycin	1 ☐	2 ☐	3 ☐	9 ☐
Cycloserine	1 ☐	2 ☐	3 ☐	9 ☐
Capreomycin	1 ☐	2 ☐	3 ☐	9 ☐
Para-Amino Salicylic Acid	1 ☐	2 ☐	3 ☐	9 ☐
Amikacin	1 ☐	2 ☐	3 ☐	9 ☐
Rifabutine	1 ☐	2 ☐	3 ☐	9 ☐
Ciprofloxacin	1 ☐	2 ☐	3 ☐	9 ☐
Ofloxacin	1 ☐	2 ☐	3 ☐	9 ☐
Other	1 ☐	2 ☐	3 ☐	9 ☐

Comments:

Appendix III

Program Objectives Using TIMS

The following program objectives are examples of performance measures that are currently being used by the New Jersey Department of Health and Senior Services to assess program effectiveness.

Objective #1: At least 90% of all TB cases initiating therapy for pulmonary disease and with no documented Rifampin resistance complete therapy within 12 months.

Basis: National Objective

Objective #2: At least 90% of all TB cases initiating therapy for extra-pulmonary disease and with no documented Rifampin resistance complete therapy within 12 months.

Basis: National Objective

Objective #3: At least 90% of all initial *Mycobacterium tuberculosis* isolates are tested for drug susceptibility.

Basis: National Objective

Objective #4: HIV status is known for at least 75% of all individuals with TB disease aged 25 to 44 years, regardless of site of disease.

Basis: National Objective

Objective #5: At least 75% of all individuals with sputum culture confirmed pulmonary TB convert their sputum to negative within 60 days.

Basis: New Jersey Objective

Objective #6: Rule out or confirm disease within 90 days for at least 75% of all individuals suspected of having TB.

Basis: New Jersey Objective

Objective #7: At least 90% of all individuals with pulmonary TB receive at least 2 weeks of DOT upon initiation of therapy.

Basis: New Jersey State Statute

Appendix IV

Solutions for Sample Problems

I. Sample Problems: Incidence and Prevalence

A. Baseline prevalence of TB infection = 40/100 or 4 per 1,000 residents

B. Incidence Rate = 20/60 or 33.3/1,000 residents

II. Sample Problems: Sensitivity, Specificity and Predictive Values

	Sputum Smear Result	Sputum Culture Result or "Gold Standard"		Total
		+	–	
Count	+	185	45	230
Total %		29.37	7.14	36.51
Column %		66.07	12.86	
Row %		80.43	19.57	
Count	–	95	305	400
Total %		15.08	48.41	63.49
Column %		33.93	87.14	
Row %		23.75	76.25	
Total Count		280	350	630
Total %		44.44%	55.56%	100%

A. What is the prevalence of a positive sputum culture in this population?

280/630 × 100 = 44%

B. What is the sensitivity of the sputum smear result?

185/280 × 100 = 66%

C. What is the specificity of the sputum smear result?

305/350 × 100 = 87%

D. What is the negative predictive value of the sputum smear result?

305/400 × 100 = 76%

E. What is the positive predictive value of the sputum smear result?

185/230 × 100 = 80%

III. Sample Problems: Case-Fatality Rate

A. 15

B. The population of England and Wales is "older" than the population of India, and older patients may have other conditions that would make them more likely to die.

IV. Sample Problem: Cause-Specific Mortality Rate

A. These are age, sex, and cause-specific death rates.

Suggested Epidemiology Reading List

1. Centers for Disease Control, Division of Tuberculosis Elimination *Outbreak Response Plan (working draft).* Available at http://www.doh.wa.gov/cfh/tb/2004_guidelines/case_management/CDC%20Outbreak%20Plan.pdf

2. Centers for Disease Control, *Excellence in Curriculum Integration through Teaching Epidemiology* (EXCITE) Website. Available at http://www.cdc.gov/excite/index.htm

3. Haupt A. *Population Reference Bureau's Population Handbook*. 5th ed, PRB, US; 2005.

4. Lillienfeld D, Stolley, PD. *Foundations of Epidemiology*. 3rd ed, Oxford University Press; 1994.

5. Gordis L. *Epidemiology*. 2nd ed. New York:W. B. Saunders Company; 2000.

6. Greenberg R. *Medical Epidemiology*. 3rd ed. New York: Lange Medical Books. McGraw-Hill; 2001.

7. National TB Controllers Association/CDC Advisory Group on TB Genotyping. *Guide to the Application of Genotyping to Tuberculosis Prevention and Control.* Atlanta, GA: United States Department of Health and Human Services, CDC; June 2004. Available at http://www.cdc.gov/nchstp/tb/genotyping/toc.htm

New Jersey Medical School
National Tuberculosis Center
A Founding Component of the International Center for Public Health
225 Warren Street,1st Floor, West Wing • PO Box 1709
Newark, NJ 07101-1709
(973) 972-0979 • www.umdnj.edu/ntbcweb