****

**NAME:LOKWALE L.ROBERT**

**COURSE:DIPLOMA IN PUBLIC HEALTH**

**COURSE CODE:D012**

**REGISTRATION NO:** **AIPMS/237/2019.**

**DATE:31/05/2019**

**1. Distinguish between descriptive epidemiology and analytical epidemiology**

A descriptive study – as the name implies – **describes** the distributions of disease, injury or health in a population, outlining the burden of disease or the extent of exposure. While An analytic study tests a **hypothesis** about the causes of disease, injury or health, or the effectiveness of interventions, shedding some light on the determinants of these events.

**2. Write down and explain the mathematical expression of the following.**

**i. Incidence**

**ii. Prevalence**

**Incidence** refers to the occurrence of new cases of disease or injury in a population over a specified period of time.

Two types of incidence include: —

1. **incidence proportion** which is the proportion of an initially disease-free population that develops disease, becomes injured, or dies during a specified (usually limited) period of time.

2 **incidence rate or person-time rate** which is a measure of incidence that incorporates time directly into the denominator. A person-time rate is generally calculated from a long-term cohort follow-up study, wherein enrollees are followed over time and the occurrence of new cases of disease is documented.

**Method for calculating incidence rate**

EXAMPLES:

Calculating Incidence Rates

**Example A**: Investigators enrolled 2,100 women in a study and followed them annually for four years to determine the incidence rate of heart disease. After one year, none had a new diagnosis of heart disease, but 100 had been lost to follow-up. After two years, one had a new diagnosis of heart disease, and another 99 had been lost to follow-up. After three years, another seven had new diagnoses of heart disease, and 793 had been lost to follow-up. After four years, another 8 had new diagnoses with heart disease, and 392 more had been lost to follow-up. The study results could also be described as follows: No heart disease was diagnosed at the first year. Heart disease was diagnosed in one woman at the second year, in seven women at the third year, and in eight women at the fourth year of follow-up. One hundred women were lost to follow-up by the first year, another 99 were lost to follow-up after two years, another 793 were lost to follow-up after three years, and another 392 women were lost to follow-up after 4 years, leaving 700 women who were followed for four years and remained disease free. Calculate the incidence rate of heart disease among this cohort. Assume that persons with new diagnoses of heart disease and those lost to follow-up were disease-free for half the year, and thus contribute ½ year to the denominator.

Numerator = number of new cases of heart disease = 0 + 1 + 7 + 8 = 16

Denominator = person-years of observation = (2,000 + ½ x 100) + (1,900 + ½ x 1 + ½ x 99) + (1,100 + ½ x 7 + ½ x 793) + (700 + ½ x 8 + ½ x 392)

= 6,400 person-years of follow-up or

Denominator = person-years of observation = (1 x 1.5) + (7 x 2.5) + (8 x 3.5) + (100 x 0.5) + (99 x 1.5) + (793 x 2.5) + (392 x 3.5) + (700 x 4)

= 6,400 person-years of follow-up

Person-time rate = Number of new cases of disease or injury during specified period

–––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––––

Time each person was observed, totaled for all persons

= 16 / 6,400

= .0025 cases per person-year

= 2.5 cases per 1,000 person-years

Number of new cases of disease or injury during specified period

­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­­–––––––––––––––––––––––––––––––––––––––––––––––

Time each person was observed, totaled for all persons

**Method for calculating incidence proportion**

(risk) Equal to Number of new cases of disease or injury during specified period Size of population at start of period

EXAMPLES:

Calculating Incidence Proportion (Risk)

Example A: In the study of diabetics, 100 of the 189 diabetic men died during the 13-year follow-up period. Calculate the risk of death for these men.

Numerator = 100 deaths among the diabetic men

Denominator = 189 diabetic men 10n = 102 = 100

Risk = (100 / 189) x 100 = 52.9%

Example B:

In an outbreak of gastroenteritis among attendees of a corporate picnic, 99 persons ate potato salad, 30 of whom developed gastroenteritis. Calculate the risk of illness among persons who ate potato salad.

Numerator = 30 persons who ate potato salad and developed gastroenteritis

Denominator = 99 persons who ate potato salad 10n = 102 = 100

Risk = ―Food-specific attack rate‖ = (30 / 99) x 100 = 0.303 x 100 = 30.3%

***Prevalence*** refers to the total number of individuals in a population who have a disease or health condition at a specific period of time, usually expressed as a percentage of the population.

**EXAMPLE: Calculating Prevalence**

In a survey of 1,150 women who gave birth in Maine in 2000, a total of 468 reported taking a multivitamin at least 4 times a week during the month before becoming pregnant. 7 Calculate the prevalence of frequent multivitamin use in this group.

Numerator = 468 multivitamin users

Denominator = 1,150 women Prevalence

= (468 / 1,150) x 100 = 0.407 x 100 = 40.7%

**3. Apart from Randomized trials, describe four (4) other epidemiological research designs**

**The four basic study designs**

It should first be emphasized that all epidemiological studies are (or should be) based on a particular population (the ‘source population’) followed over a particular period of time (the ‘risk period’). Within this framework, the most fundamental distinction is between studies of disease ‘incidence’ and studies of disease ‘prevalence’. Once this distinction has been drawn, then the different epidemiological study designs differ primarily in the manner in which information is drawn from the source population and risk period.

**Incidence studies**

Incidence studies ideally measure exposures, confounders and outcome times of all population members. . When the source population has been formally defined and enumerated (e.g. a group of workers exposed to a particular chemical), then the study may be termed a ‘cohort study’ or ‘follow-up study’ and the former terminology will be used here. Incidence studies also include studies where the source population has been defined but a cohort has not been formally enumerated by the investigator, e.g. ‘descriptive’ studies of national death rates. Furthermore, there is no fundamental distinction between incidence studies based on a broad population (e.g. all workers at a particular factory or all persons living in a particular geographical area) and incidence studies involving sampling on the basis of exposure, since the latter procedure merely redefines the study population (cohort).4

**Incidence case–control studies**

Incidence studies are usually the preferred approach to studying the causes of disease, because they use all of the available information on the source population over the risk period. However, they are often very expensive in terms of time and resources, and the equivalent results may be achieved more efficiently by using an incidence case–control study design.

When the outcome under study is rare, an even more remarkable gain in efficiency can be achieved with only a minimal reduction in the precision of the effect estimate.

**Prevalence studies**

Incidence studies are usually the preferred approach to studying the causes of disease, but they often involve lengthy periods of follow-up and large resources.Also, for some diseases (e.g. asthma and diabetes), incidence may be difficult to measure without very intensive follow-up. Thus, it is often more practical to study the ‘prevalence’ of disease at a particular point in time. This approach has one major potential shortcoming, since disease prevalence may differ between two groups because of differences in age-specific disease incidence, disease duration or other population parameters; thus, it is much more difficult to assess causation (i.e. whether an exposure increases disease incidence) in prevalence studies. Nevertheless, for many common diseases, studying prevalence is often the only practical option and may be an important first step in the research process; furthermore, prevalence may be of interest in itself, e.g. because it measures the population burden of disease. For example, motor neurone disease and multiple sclerosis have similar incidence and mortality rates, but multiple sclerosis represents a greater burden of morbidity for the health services, because survival for motor neurone disease is so short.

Note that this definition of prevalence studies does not involve any specification of the timing of the measurement of exposure. In many prevalence studies, information on exposure will be physically collected by the investigator and at the same time information on disease prevalence is collected. Nonetheless, exposure information may include factors that do not change over time (e.g. gender) or change in a predictable manner (e.g. age), as well as factors that do change over time. The latter may have been measured at the time of data collection [e.g. current levels of airborne asbestos exposure, body mass index (BMI)] or at a previous time (e.g. historical records on past asbestos exposure levels, birthweight recorded in hospital records), or integrated over time (e.g. using a job–exposure matrix and work history records). The sole defining feature of prevalence studies is that they involve studying disease prevalence. There is no restriction on when the exposure information is collected or whether it relates to current and/or historical exposures.

**Prevalence case–control studies**

Just as an incidence case–control study can be used to obtain the same findings as a full cohort study, a prevalence case–control study can be used to obtain the same findings as a full prevalence study in a more efficient manner. In particular, if obtaining exposure information is difficult or costly, then it may be more efficient to conduct a prevalence case–control study by obtaining exposure information on some or all of the prevalent cases and a sample of controls selected from the non-cases.

Suppose that a prevalence case–control study is conducted using the source population in Table 4, involving all the 1385 prevalent cases and a group of 1385 controls In this instance, there is one main option for selecting controls, namely to select them from the non-cases. This will enable us to estimate the exposure odds of the non-cases, and the OR obtained in the prevalence case–control study will therefore estimate the POR in the source population  Alternatively, if the PR is the effect measure of interest, controls can be sampled from the entire source population (i.e. in a manner analogous to case–cohort sampling) and the resulting prevalence case–control ‘OR’ will estimate the PR in the source population.

**Cross-sectional studies**

In the presentation of prevalence studies above, the health outcome under study was a ‘state’ (e.g. having or not having hypertension). Studies could involve observing the incidence of the ‘event’ of acquiring the disease state (e.g. the incidence of being diagnosed with hypertension), or the prevalence of the disease state (e.g. the prevalence of hypertension). More generally, the health state under study may have multiple categories (e.g. non-hypertensive, mild hypertension, moderate hypertension and severe hypertension) or may be represented by a continuous measurement (e.g. blood pressure). Since these measurements are taken at a particular point in time, such studies are often referred to as ‘cross-sectional studies’. Prevalence studies are a subgroup of cross-sectional studies in which the disease outcome is dichotomous.

**Longitudinal studies**

Longitudinal studies (cohort studies) involve repeated observation of study participants over time. They represent the most comprehensive approach since they use all of the available information on the source population over the risk period. Incidence studies are a subgroup of longitudinal study in which the outcome measure is dichotomous. More generally, longitudinal studies may involve repeated assessment of categorical or continuous outcome measures over time (e.g. a series of linked cross-sectional studies in the same population). A simple longitudinal study may involve comparing the disease outcome measure or more usually changes in the measure, over time, between exposed and non-exposed groups. For example, rather than comparing the incidence of hypertension (as in an incidence study) or the prevalence at a particular time (as in a prevalence study), or the mean blood pressure at a particular point in time (as in a cross-sectional study), a longitudinal study might involve measuring baseline blood pressure in exposed and non-exposed persons and then comparing changes in blood pressure (i.e. the change from the baseline measure) over time in the two groups. One special type of longitudinal study is that of ‘time series’ comparisons in which variations in exposure levels and symptom levels are assessed over time with each individual serving as their own comparison.

**4. Data from hospital records are one of the most important sources of information in**

**epidemiologic studies.**

**a) Outline the limitations of using hospital data.**

The benefits of big data are indisputable, but there are also some limitations that need to be discussed as well.

Claims data captures the services provided to a patient. This information can be grouped into different cohorts—those getting preventive exams, those seeing specific physicians or hospitals for conditions, etc. The data can be grouped by diagnosis. However, all claims data is just a collection of medical bills. Medical bills do not contain a complete look at the patient, such as important information about a patient’s prognosis. That’s a gap. Thus, it is important to set appropriate expectations on the use of the data.

Here are some limitations that should be placed on the expectations:

**Number 1 (one of the most important): Avoid the averages**  
Most claims data sets are not normally distributed, so the averages do not provide relevant information. In most discussions today, employers evaluate the average cost of employees with specific conditions, e.g., diabetes or high blood pressure. This is a flawed approach because spending by employees with various chronic conditions is skewed, thus not really “averageable.” For example, assume 90% of an employee population with diabetes is spending $10,000/year and 10% is spending $250,000/year; the average will be a meaningless $34,000/year. All too often, a wild goose chase ensues, when in fact the focus should be on the $250,000 cohort to understand why they were so much more expensive.

**Number 2: Follow the money**  
A superior use of claims data is to look at distributions of spending. In most plans today, roughly 8% of enrollees are consuming 80% of plan dollars, and these 8% typically change every 12 to 18 months. (We still run into benefit managers who were unaware of that turnover.) The future belongs to micro-managing these “outliers,” rather than the 92% who spend only 20% of the dollars. If you study those outliers carefully, you will find that only about 7% of their spending possibly would have been preventable, and then only if they faithfully did what their doctors told them to do decades earlier. A cardiologist recently told me that, of the patients he has seen with a significant acute blockage, about 25% had no known health risks of any kind…no high blood pressure, cholesterol, diabetes, obesity, smoking, genetic predisposition, etc. As such, there is a component of randomness in terms of who gets blocked arteries. The same holds true for cancer. For the other 75%, their physicians have usually counseled them on the importance of exercise and nutrition and the dangers of tobacco use, but to no avail.

**Number 3: Realize the limitations for quality designations**  
Yet another big error is trying to use claims data to determine the best-quality doctors. You had better be really, really talented to try that one. Why? We are in an era in which many doctors are making their “quality” and “outcomes” look better by referring their most complex and riskiest patients to someone else. (Much has been written about this.) On the other hand, there are highly effective doctors who take responsibility for their riskiest patients, but as a consequence score poorly on so-called “quality measures.” The real travesty is that the low-scoring doctors may be the most cost-effective and provide the best care.

**Number 4: Misdiagnoses are a real cost driver**  
Another huge shortcoming of claims data is one that readers of [Cracking Health Costs](http://www.amazon.com/Cracking-Health-Costs-Companys-Employees/dp/1118636481/ref=sr_1_1?ie=UTF8&qid=1458598533&sr=8-1&keywords=cracking+health+costs)know about. Namely, a large number of patients with complex health problems are simply misdiagnosed – today, that’s about 20% of the outliers in benefit plans, accounting for 18% of claim dollars. Thus, you cannot rely on diagnoses in claims data, and you cannot tell who is getting diagnoses right or wrong – this takes detective work beyond claims data by the Mayo Clinic on rates of misdiagnoses. We have sent hundreds of people to the Mayo Clinic for second opinions and can verify by personal experience the truth in that article…same for other clinics we have used for employers. Our first rule in selecting a Center of Excellence is its success in correctly diagnosing patients with complex health problems. Huge amounts of claim dollars are spent on treatments or surgeries that are either completely erroneous or clearly suboptimal. An executive at a Fortune 100 company once said to me that the biggest quality failure in healthcare is to misdiagnose a patient…everything that follows harms the patient.

**Number 5: Coding can affect the data analysis**  
During a data analysis for a very larger employer, with more than 250,000 covered lives, executives told me they had not paid for a solid organ transplant in a number of years. Based on their size, they should have been paying for about 25 a year. After further detective work, we discovered their consultant was using a DRG grouper that coded all transplants as ventilator cases…who knows why…but a huge error. The benefit team had no idea they were really paying for about 25 a year at an average cost over five years of about $1.5 million each.

**Number 6: Reversion to the mean**One thing we’ve learned from years of claims analysis of big companies’ benefit programs is that if you have enough life years of data, it all looks about the same, i.e., it reverts to the mean. If the workforce is comparatively older, they will have somewhat more high-cost claims.

**Number 7:Data Availability and Reliability**

Big data healthcare models require reliable and detailed data sets. This means healthcare providers need access to as much data on their patients as possible. They also need to vet it carefully, because inaccuracies can destabilize their entire healthcare models. Social media, in particular is often unreliable, because patients are less likely to double check what they post on their profiles. They may even intentionally post inaccurate information to embellish their resumes or look more appealing to their friends.

**b) Describe the possible sources of error in interview surveys**

errors are sources of uncertainty, both in the estimates in the data and the inferences about the results.

 Survey errors reduce, but don’t necessarily eliminate, our ability to accurately make inference to the larger population. Consequently, understanding survey errors is key to understanding survey data quality. Increasing error typically results in larger confidence intervals (reduced certainty) around the estimates in the data and inferences made about the population of interest. If these confidence intervals grow too large, the quality of the data and inferences can be degraded to the point of making them uninformative.

Consider the below sources of errors;

1. [**NONRESPONSE ERROR**](https://www.nap.edu/read/18605/chapter/1#a34)

Nonresponse error in surveys arises from the inability to obtain a useful response to all survey items from the entire sample. A critical concern is when that nonresponse leads to biased estimates. Nonresponse bias is a product of the difference between respondents and nonrespondents on a particular measure and the size of the nonresponse population. A lower response rate increases the potential for greater nonresponse bias, but when the data are missing at random, a lower response rate will neither create nor increase nonresponse error.

The NCVS, like most federal household surveys, is voluntary and not required by law. The challenges facing today’s federal household surveys were recently summarized by the National Research Council (2013a, p. 68):

[They] include maintaining adequate response from increasingly busy and reluctant respondents. More and more households are non-English speaking, and a growing number of higher income households have controlled-access residences….Today’s household surveys face confidentiality and privacy concerns, a public growing more suspicious of its government, and competition from an increasing number of private as well as government surveys vying for the public’s attention.

2.[**SPECIFICATION ERROR**](https://www.nap.edu/read/18605/chapter/1#a35)

For any survey, its intended purpose and concepts must be clearly defined in order for survey instruments and procedures to accurately translate those concepts into the collection of data. In surveys, specification error may occur when there is a mismatch between what the survey is measuring and what it is intended to measure.[5](https://www.nap.edu/read/18605/chapter/10#ch8_fn5) As defined by Biemer (2010, p. 31): “specification error pertains specifically to the problem of measuring the wrong concept in a survey, rather than measuring the right concept poorly.” This section examines a key concept associated with the NCVS to see if it is clearly defined and consistent between the survey’s purposes and processes.

This key concept is to identify if and when a respondent has been the victim of a rape or sexual assault. BJS has developed a clear definition of what the survey is intended to measure In the omnibus screener that is currently used in the NVCS, the deliberate approach is to soften the link between the screening cues and any particular type of criminal victimization. In particular, for rape and sexual assault, as BJS translates

This definition is different from that used by economists and other mathematical modelers, for whom “specification error” refers to an incorrect statement of an empirical model. We use the term differently in the report.

3.[**MEASUREMENT ERROR**](https://www.nap.edu/read/18605/chapter/1#a36)

Measurement error includes a large family of errors that may occur when response on a survey results in the collection of inaccurate or incomplete information. In this section, the report discusses potential measurement errors on the NCVS associated with the respondent, the questionnaire, the mode of collection, and with the interviewer/respondent interaction. These issues are interrelated, and each has the potential to result in measurement error on the NCVS.

**Respondents**

Survey research has mapped a respondent’s cognitive process in answering survey questions (Schwarz, 1996; Strack and Martin, 1987; Tourangeau, 1984; Tourangeau, Rips, and Rasinski, 2000). In particular, Tourangeau, describes four steps a respondent goes through in responding to a survey question:

**5. Explain the main determinants of health**

### Introduction

Many factors combine together to affect the health of individuals and communities. Whether people are healthy or not, is determined by their circumstances and environment. To a large extent, factors such as where we live, the state of our environment, genetics, our income and education level, and our relationships with friends and family all have considerable impacts on health, whereas the more commonly considered factors such as access and use of health care services often have less of an impact.

#### The determinants of health include:

The social and economic environment,

The physical environment, and

The person’s individual characteristics and behaviours.

## Social Determinants of Health

The complex, integrated, and overlapping social structures and economic systems that are responsible for most health inequities. These social structures and economic systems include the social environment, physical environment, health services, and structural and societal factors. Social determinants of health are shaped by the distribution of money, power, and resources throughout local communities, nations, and the world.

The context of people’s lives determine their health, and so blaming individuals for having poor health or crediting them for good health is inappropriate. Individuals are unlikely to be able to directly control many of the determinants of health. These determinants—or things that make people healthy or not—include the above factors, and many others:

Income and social status - higher income and social status are linked to better health. The greater the gap between the richest and poorest people, the greater the differences in health.

Education – low education levels are linked with poor health, more stress and lower self-confidence.

Physical environment – safe water and clean air, healthy workplaces, safe houses, communities and roads all contribute to good health. Employment and working conditions – people in employment are healthier, particularly those who have more control over their working conditions.

Social support networks – greater support from families, friends and communities is linked to better health. Culture - customs and traditions, and the beliefs of the family and community all affect health.

Genetics - inheritance plays a part in determining lifespan, healthiness and the likelihood of developing certain illnesses. Personal behaviour and coping skills – balanced eating, keeping active, smoking, drinking, and how we deal with life’s stresses and challenges all affect health.

Health services - access and use of services that prevent and treat disease influences health

Gender - Men and women suffer from different types of diseases at different ages.

**References**

**Suggested Citation:**"8 Potential Sources of Error: Nonresponse, Specification, and Measurement." National Research Council. 2014. *Estimating the Incidence of Rape and Sexual Assault*. Washington, DC: The National Academies Press. doi: 10.17226/18605.

(Schwarz, 1996; Strack and Martin, 1987; Tourangeau, 1984; Tourangeau, Rips, and Rasinski, 2000). In particular, Tourangeau, describes four steps a respondent goes through in responding to a survey question:

<https://cursos.campusvirtualsp.org/mod/tab/view.php?id=34133&forceview=1>

SOURCE: NORC at the University of Chicago (2009, p. 19, Table 2.5).

." National Research Council. 2014. *Estimating the Incidence of Rape and Sexual Assault*. Washington, DC: The National Academies Press. doi: 10.17226/18605

SOURCE: U.S. Census Bureau and Bureau of Justice Statistics (2011a, pp. 2-33-2-34).

*Journal of Marketing* © 1955 [Sage Publications, Inc.](https://www.jstor.org/publisher/sage)

* [HIA for regeneration projects volume II: Selected evidence base](http://www.geog.qmul.ac.uk/health/Vol2.pdf)  
  London: East London and the City Health Action Zone (2001)
* [Transport and health study group](http://www.phel.gov.uk/hiadocs/hia_transport_policy_and_guidance.pdf)  
  [PDF 81KB]

[Informing Transport Health Impact Assessment in London](http://www.phel.gov.uk/hiadocs/informing_transport_hia_in_london.pdf)

Tom Emerick is president of [Emerick Consulting](http://www.emerickconsulting.com/) and cofounder of EdisonHealth and Thera Advisors.  Emerick’s years with Wal-Mart Stores, Burger King, British Petroleum and American Fidelity Assurance have provided him with an excellent blend of experience and contacts.

**More articles by Tom Emerick:**

* [Ethics of Workplace Wellness Industry](http://insurancethoughtleadership.com/ethics-of-workplace-wellness-industry/)
* [What Caused the ACA Rate Surge?](http://insurancethoughtleadership.com/huge-aca-rate-hikes-explained-in-100-words-or-less/)
* [Walmart Shows Way on Health Benefits](http://insurancethoughtleadership.com/walmart-shows-way-on-health-benefits/)

follow [@CrackingHC](http://twitter.com/#!/CrackingHC) or[@ITLUpdates](https://twitter.com/itlupdates) for more stories like this

David Toomey is a senior healthcare executive with 30 years of extensive healthcare expertise in addressing the fragmentation within the healthcare system. He spent 25 years with two national insurance companies, before he spent close to 5 years in an early stage healthcare company.

**More articles by David Toomey:**

* [New, Troubling Healthcare Model](http://insurancethoughtleadership.com/new-troubling-healthcare-model/)
* [Is Transparency the Answer in Healthcare?](http://insurancethoughtleadership.com/is-transparency-the-answer-in-healthcare/)
* [Obamacare: Where Do We Stand Today?](http://insurancethoughtleadership.com/obamacare-where-do-we-stand-today/)

follow [@ITLUpdates](https://twitter.com/itlupdates) for more stories like this

Greenland and Morgenstern

.

Classification schemes for epidemiologic research designs

,

*J Clin Epidemiol*

 ,

1988

2

.

Principles of study design in environmental epidemiology

,

*Environ Health Perspect*

 ,

1993

 Miettinen

 OS

*Theoretical Epidemiology*

 ,

1985

New York

John Wiley & Sons, Inc.

4

Rothman

KJ

*Modern Epidemiology*

 ,

1986

Boston

Little, Brown

5

Pearce

 N

Crane

 J

Balmes

 J

.

Epidemiologic methods

,

*Occupational and Environmental Respiratory Disease*

 ,

1995

St Louis, MI