EPIDEMIOLOGY

Robert D. Newcomb, OD, MPH, FAAO

Whoever wishes to investigate medicine properly, should proceed thus: in the first place to consider the seasons of the year, and what effects each of them produces for they are not all alike, but differ much from themselves in regard to their changes. Then the winds, the hot and the cold, especially such as are common to all countries, and then such as are peculiar to each locality.

from Airs, Waters, and Places
Hippocrates (460-370 B.C.)

Chapter Overview

Epidemiology (*epi* = upon, *demos* = population, *ology* = word, thought, the study of) is the basic science of all public health activities. If diseases, disorders, and conditions affecting the human body were randomly distributed within a given population, thenthere would be no way for researchers to predict who might be affected by them at a later point in time; and therefore, certainly no way to prevent them from occurring. So the science of epidemiology helps us understand why some members of a population develop a clinical finding while other members of that same population do not. And having this type of information allows clinicians to prevent diseases, disorders and conditions in addition to treating them earlier and more effectively.

This chapter will briefly discuss the scientific principles, terminology, and applications of epidemiology. Examples from the medical, eye, and vision literature will be employed to emphasize the relevance of this basic science to the daily practice of optometry.

Objectives

On completion of this chapter, the reader should be able to
1. Calculate the Relative Risk that a given risk factor has on predicting a morbidity (or mortality) rate.
2. Categorize patients by age, sex, race, socioeconomic status, medical history, and other demographic variables and learn to use these categories as predictors of various diseases, disorders, or conditions.
3. Identify high-risk sub-groups of people as a way to begin addressing the causes of health disparities in large populations.
4. Design an epidemiological research project to identify pertinent risk factors that predict which members of a population are "at risk" for developing a specific disease, disorder, or condition.

Demography: The Study of Populations

While clinicians must think in terms of individual patients who are seated in their examination chairs, epidemiologists prefer to think of those same patients as members of a much larger group of comparable people. Austin and Werner[1] have defined the word "Epidemiology" as "…the study of how and why diseases and other conditions are distributed within the population the way they are." Populations can be defined as any small, medium, or large group of people. A population could be defined as all the students in one classroom, all the people working in an office building, all the citizens of a city or state, or even all of the people living on planet earth! Data from these studies would provide examples of descriptive epidemiology. But describing these populations of people solely on the basis of their common locations would be counterproductive unless one could analyze them in terms of other risk factors for certain diseases. So the next logical step of epidemiology research would be analytical. Only then would researchers be able to identify both inherent (ie, age, sex, race) and acquired (do you smoke? are you obese? have you been exposed to hazardous materials?) risk factors which may predispose a sub-set of that population to excessive disease rates.

For example, there are about 305 million people living in the United States; and about 6 billion people living on planet earth. If diseases were distributed randomly in these two large populations, everyone would need the same amount of health care at predictable stages in their lives. But, as shown later in this chapter, disease rates are not the same in the US as they are in the rest of the world; and they are not even the same between groups within the US borders.

Incidence and Prevalence

Incidence rates are measures of only the <u>new cases</u> of a disease that occur during a given time period such as months, years, or decades within a population susceptible to the disease of interest. Mortality, for example, is the incidence of deaths per thousand per year in a defined population. Prevalence rates are measures of <u>all cases (both new and old)</u> of a disease that are present (or prevail) at one point in time. Prevalence is not actually a rate but a proportion and is usually reported as a percentage. For example, in 2002, a meta-analysis by the National Eye Institute and Prevent Blindness America reported the prevalence of glaucoma was 1.9% of the total U.S. population age 40 and older.[2]

In acute diseases, such as conjunctivitis ("pink eye") in an elementary school, incidence measures are more useful because it is a limited but highly-contagious condition. Such conditions may be seasonal, so comparison of a current incidence rate with historical data may establish an epidemic (loosely defined as more cases of a disease than would normally be expected). But in chronic diseases that are not contagious, such as cancer or glaucoma, prevalence measures are more useful because public health officials can monitor whether or not these conditions are increasing, decreasing, or remaining constant in a community, state, or nation over time, and can then begin to research underlying causes of the variation in data.

Descriptive Epidemiology

At its most basic level, the science of epidemiology is used to simply count the number of people who have a certain disease, disorder, or condition. The Who? What? Where? and When? –type questions can be answered by descriptive epidemiology studies. The result of this counting is usually expressed as a fraction, where the number of people who are affected is shown in the numerator, and the denominator is the size of the population from which they came. This fraction can then be converted to a decimal or percentage, so different populations can be easily compared with each other.

Consider the descriptive epidemiology data in Tables 1-4 from the World Health Organization.[3]

Table 1. Top Ten Causes of Death in the World

	Deaths in millions	% of deaths
Coronary heart disease	7.20	12.2
Stroke and other cerebrovascular diseases	5.71	9.7
Lower respiratory infections	4.18	7.1
Perinatal conditions	3.18	5.4
Chronic obstructive pulmonary disease	3.02	5.1
Diarrheal diseases	2.16	3.7
HIV/AIDS	2.04	3.5
Tuberculosis	1.46	2.5
Trachea, bronchus, lung cancers	1.32	2.3
Road traffic accidents	1.27	2.2

Source: World Health Organization, September, 2008

Table 2. Top Ten Causes of Death in low-income countries

	Deaths in millions	% of deaths
Lower respiratory infections	2.94	11.2
Coronary heart disease	2.47	9.4
Perinatal conditions	2.40	9.1
Diarrheal diseases	1.81	6.9
HIV/AIDS	1.51	5.7
Stroke and other cerebrovascular diseases	1.48	5.6
Chronic obstructive pulmonary disease	0.94	3.6
Tuberculosis	0.91	3.5
Malaria	0.86	3.3
Road traffic accidents	0.48	1.9

Source: World Health organization, September, 2008

Table 3. Top Ten Causes of death in middle-income countries

	Deaths in millions	% of deaths
Stroke and other cerebrovascular diseases	3.47	14.2
Coronary heart disease	3.40	13.9
Chronic obstructive pulmonary disease	1.80	7.4
Lower respiratory infection	0.92	3.8
Perinatal conditions	0.75	3.1
Trachea, bronchus, lung cancers	0.69	2.9
Road traffic accidents	0.67	2.8
Hypertensive heart disease	0.62	2.5

Stomach cancer	0.55	2.2
Tuberculosis	0.54	2.2

Source: World Health organization, September, 2008

Table 4. Top Ten causes of death in high-income countries

	Deaths in Millions	% of deaths
Coronary heart disease	1.33	16.3
Stroke and other cerebrovascular diseases	0.76	9.3
Trachea, bronchus, lung cancers	0.48	5.9
Lower respiratory infections	0.31	3.8
Chronic obstructive pulmonary disease	0.29	3.5
Alzheimer and other dementias	0.28	3.4
Colon and rectum cancers	0.27	3.3
Diabetes mellitus	0.22	2.8
Breast cancer	0.16	2.0
Stomach cancer	0.14	1.8

Source: World Health organization, September, 2008

They indicate different diseases - and different frequencies of the same diseases – occur depending upon where people live. Why is this? Is it because of Genetics? Economics? the Environment? Varying levels of Health Education and/or Health Literacy? Access to Health Care? Lack of basic Public Health Infrastructures such as surveillance systems, prevention strategies, quality assurance measures, and/or excessively low provider/population ratios? A descriptive epidemiology study cannot answer the Why? –type question; but a well-designed analytical epidemiologic research project (see below) in each country could begin to address all of these possible risk factors for its citizens.

Another example of descriptive epidemiology is shown in Figure 1, which shows how legal blindness increases with age among Americans who are 40 years old and older in four different race and ethnic groups.

Source: National Eye Institute, Prevent Blindness America. *Vision Problems in the U.S.: Prevalence of Adult Vision Impairment and Age-Related Eye Disease in America.* Bethesda, MD; 2002.

Figure 1. Estimated specific prevalence rates for blindness among Americans 40 years and older, by age group and race/ethnicity.
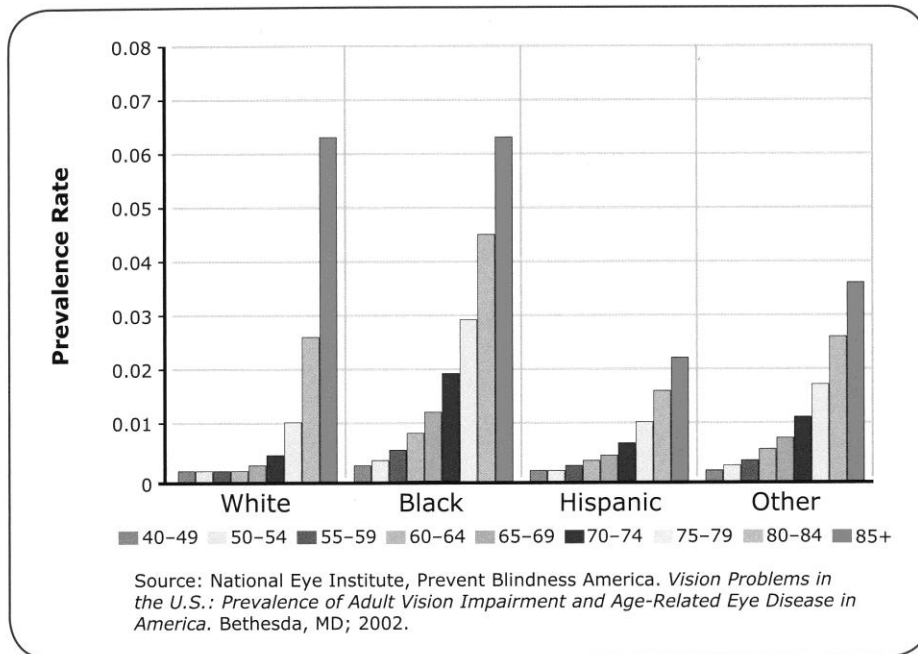
Multi-factorial diseases...likely have more than one risk factor contributing to the onset of the

The first observation from these data would be that the frequency of blindness increases dramatically as people age, regardless of their race or ethnicity. This fact can be explained by knowing that the number of cataract, age-related macular degeneration, glaucoma, and diabetic retinopathy cases also increase as people age. But a more detailed examination of the data show that black patients in the 75-79 year-old age category have 3 times the amount of blindness as similarly-aged white and Hispanic patients. And Hispanic patients in the 85+ age category have only one-third the amount of blindness as similarly-age white and black patients. What variables (factors) might explain these marked differences in blindness? Again, a descriptive epidemiology study such as this cannot answer the Why? –type question; but a well-designed analytical epidemiologic research project can.

Analytical Epidemiology

To better understand and test theories for why certain diseases are distributed differently in different populations, epidemiologists must perform analytical studies. These studies attempt to explain the differences based upon certain variables, or risk factors, that are present in some people but not present in others. Analytical

epidemiologists then attempt to explain, or attribute, these differences by using mathematical concepts such as rates, ratios, and risk factors. Often, these analytical studies establish a relationship, also known as an association, between a specific risk factor and a disease; these associations may be weak or strong, and even direct or inverse The ultimate goal for analytical epidemiologists is to discover a cause-and-effect association between a risk factor and a disease. This direct cause-and-effect association is more commonly found in infectious diseases, where the cause is a specific microorganism such as the *tubercule bacillus* in tuberculosis. But in chronic diseases such as heart disease, hypertension, and diabetes, there is rarely a single cause that can be identified. So these conditions may be termed multi-factorial diseases, which means they likely have more than one risk factor contributing to the onset of the disease.

A very good recent example of analytical epidemiology in optometry was published by Charlotte Joslin, OD, et al, in 2006 entitled *Epidemiological Characteristics of a Chicago-area Acanthamoeba Keratitis Outbreak*[4]. In her study of 40 patients diagnosed with *Acanthamoeba* keratitis (AK) over a thirty-month time period, she found a statistically and clinically significant increase in the incidence of AK cases around Chicago; and hypothesized that the U.S. Environmental Protection Agency's 1998 regulations that reduced the allowable amount of carcinogenic disinfection by-products in public water treatment systems may be a likely explanation for her findings.

Experimental Epidemiology

Since any given patient can have any specific diagnosis at any time ("it ain't rare if it's in your chair"), the science of epidemiology requires researchers to study and compare two similar groups of people which, ideally, differ only in having one isolated risk factor.

This is known as experimental epidemiology. Both descriptive and analytical epidemiology studies are known as observational, or "natural experience" studies, where the investigator observes but does not manipulate either disease frequencies or risk factor exposures in the population. In contrast, experimental epidemiology studies are "investigator controlled," so that research studies called Randomized Clinical Trials (RCT's) can be designed to statistically measure the effect of a given risk factor on the presence or absence of a certain disease.

An example of a recent and well-designed Randomized Clinical Trial in optometry is the Convergence Insufficiency Treatment Trial (CITT).[5] This National Eye Institute (NEI) sponsored multi-center clinical trial enrolled 221 children, aged 9 to 17, who had symptomatic convergence insufficiencies. These children were then randomly-assigned by the investigators to one of four treatment groups. And after 12 weeks of treatment, the office-based vergence/accommodation therapy with home reinforcement group had statistically-significant improvements in their symptom survey scores, and also increases in their near points of convergence and increased positive fusional vergence ranges compared to the other three treatment groups.

This study achieved one of the ultimate goals that all experimental epidemiologists have: to present such compelling research data that it establishes a

new standard of care for treating patients. Dr. Paul A. Sieving, Director of the National Eye Institute, complimented the CITT investigators by stating "The CITT will provide eye care professionals with the research they need to assist children with this condition."[6]

Experimental epidemiology can be either prospective (where two groups of subjects, one with the risk factor and the other without, are followed over time to see who eventually get a disease) or retrospective (where the histories of affected and unaffected subjects are studied to determine which risk factors may or may not have been responsible for a disease which occurred at a later time. After collecting the pertinent data, comparisons can then be made to determine which one of the two groups has developed more "disease" (or, in the case of the CITT study, which group has better clinical outcomes from different therapeutic interventions).

The Natural History of Disease

To appropriately apply scientific epidemiological principles and methods to the everyday practice of optometry, it is important to understand a concept known as the Natural History of Disease. As stated previously in the chapter, if diseases, disorders, and conditions affecting the human body were randomly distributed within a given population, then there would be no way for researchers to predict who might be affected by them at a later point in time.

Weiss[7] and many others have stated that, with respect to disease prevention and control, a good clinician's knowledge of the natural history of a disease is as important as his or her understanding its pathophysiological cause.

> **...a good clinician's knowledge of the natural history of a disease is as important as his or her understanding its pathophysiological cause.**

Consider the following timeline for a chronic disease like diabetes:

$\longrightarrow$ $\longrightarrow$ $\longrightarrow$ $\longrightarrow$ $\longrightarrow$ $\longrightarrow$ $\longrightarrow$ $\longrightarrow$ $\longrightarrow$

| Pre-natal Treatment (genetic) Factors | Birth Outcomes Factors | Risk Onset | Pathological Symptoms | Clinical Signs | Clinical | Diagnosis |

In this construct, if a patient has a mother or father with diabetes, that patient inherits an increased risk for developing the disease later in life from prenatal, or genetic, factors. Sometime after birth, that same patient may be exposed to risk factors such as obesity, sedentary lifestyle, hypertension, hypercholesterolemia, cigarette smoking, etc, which may lead to cellular changes within the body (pathological onset) which then lead to clinical symptoms (ie, polydipsia, polyphagia, polyuria) followed by clinical signs such as an abnormal glucose tolerance test. After a diagnosis of diabetes is confirmed, the patient's physician will develop a treatment plan and monitor the patient for clinical outcomes such good or poor glucose control and the presence of any complications such as diabetic retinopathy.

So when does the patient's optometrist begin intervention to improve visual acuity and prevent blindness? Ideally, it should be early in the natural history of the disease and not after hemorrhages and exudates are observed in the late stages of the disease. If an optometrist can examine a patient in the susceptible "risk factors" stage of this terrible disease and provide health education to interrupt its natural progression on to the pathological onset stage, blindness from diabetic retinopathy can be eliminated because the precedent diabetic disease will have been eliminated!

Comparing Two Groups

The world-famous ophthalmic epidemiologist, Dr. Alfred Sommer, once said

"Epidemiology has two overriding characteristics: a preference for rates rather than absolute numbers, and a peculiarly thoughtful approach to studies amounting to applied common sense.[8]"

Comparisons between two groups can actually be made by using either Rates (a fraction in which those affected are divided by everyone in a specified population) during a specified period of time or Ratios (a fraction in which those affected are divided by those not-affected in a specified population). Fractions describing diseases or conditions are called morbidity data, whereas fractions describing death are called mortality data. And when using rates or ratios to compare two groups of people, the quintessential research tool is a 2x2 contingency table, as shown below:

	Factor	No Factor	Total
Case	A	C	A+C
Control	B	D	B+D
Total	A+B	C+D	A+B+C+D

Relative Risk (RR) is defined as the risk of a disease (also known as a "case") occurring in the presence of a certain factor as compared to its occurrence in the absence of that same factor. Mathematically, it can be calculated as follows:

RR   = $\dfrac{\text{prevalence of a case occurring in a population exposed to the factor}}{\text{prevalence of a case occurring in a population } not \text{ exposed to the factor}}$

Or, using data from the 2x2 contingency table above,

RR   = $\dfrac{A/\ A+B}{C/C+D}$

And the Ratio of people affected with the "case" compared with those unaffected would be

$$\text{Ratio} = \frac{A+C}{B+D}$$

To demonstrate how these concepts can be useful in an optometric practice, consider the following scenario:

An optometrist has noticed a lot of "red eye" patients in her practice recently, and wonders if they are related to wearing a new brand of soft contact lenses (SCLs). She reviews her records for the past six months and discovers 20 of 50 patients fit with the same new brand of SCLs have developed red eyes. Her records of 30 other patients in comparable age, sex, and racial categories were also reviewed for comparison, and she found 15 of those patients had also developed red eyes.

So 20 cases of red eyes developed out of 50 patients exposed to the new SCL and 15 cases out of 30 non exposed controls also developed red eyes. What is the Relative Risk of getting red eyes if one wears the new brand of SCLs? And what is the Ratio of getting red eyes compared to not getting red eyes in the two patient sub-populations?

Step 1: Construct a 2x2 Contingency Table, as follows:

	Wore SCL's	Did not wear SCL's	Total
Developed red eyes	20	15	35
Did not develop red eyes	30	15	45
Total	50	30	80

Step 2: Calculate Relative Risk (RR), as follows

$$RR = \frac{\text{Rate of red eyes in a population wearing SCLs}}{\text{Rate of red eyes in a comparable population not wearing SCL's}}$$

$$RR = \frac{20/(20+30)}{15/(15+15)} = \frac{0.4}{0.5} = 0.80$$

Step 2 Interpretation:

If RR is greater than 1.00, then the isolated factor increases the risk;
If RR is less than 1.00, then the isolated factor does not increase the risk (and may even be protective!).

If RR equals 1.00, then the isolated factor has no contribution to either getting or not getting the disorder; the isolated factor is therefore an independent variable.

In this example, since the RR is calculated to be less than 1.00, wearing SCLs does not increase the risk (also known as the chance or probability) of having red eyes; and the underlying cause of the red eyes has not yet been definitively determined.

Step 3: In this small clinical study, what is the Ratio of red eye patients among SCL wearers? What is the Ratio of red eye patients among non-SCL wearers?

$$\text{Ratio (in SCL)} = \frac{\text{All Affected}}{\text{All Non-affected}} = \frac{20}{50} = 0.4$$

$$\text{Ratio (in non-SCL)} = \frac{\text{All Affected}}{\text{All Non-Affected}} = \frac{15}{30} = 0.5$$

$$\text{Ratio (in entire study)} = \frac{\text{All Affected}}{\text{All-Non-Affected}} = \frac{35}{45} = 0.78$$

Step 3 Interpretation:

In this small study, 40% of SCL wearers developed red eyes while 50% of non SCL wearers developed red eyes.

Odds Ratio (OR)

Another useful concept in epidemiology research is Odds Ratio (OR), which is used when researchers want to compare the chance of an event happening in two different groups of people (ie, cases and controls).[9] Statistically, the odds of an event occurring is the probability of the event [p/(1-p)] divided by the probability of the same event *not* occurring [q/(1-q)] or

$$OR = \frac{p/(1-p)}{q/(1-q)}$$

where p is the probability for the first group and q is the probability of the second group.

Like RR, an odds ratio of 1.00 indicates that the event being studied is equally likely in both cases and controls. But unlike Relative Risk (RR), which is expressed as a decimal or percentage, the Odds Ratio (OR) is actually

expressed as a ratio. Using data from the table above, p = A/(A+C) and q = D/(B+D), so

OR = Odds of exposure in cases
         Odds of exposure in controls

$$OR = \dfrac{\dfrac{A/(A+C)}{1-[A/(A+C)]}}{\dfrac{B/(B+D)}{1-[B/(B+D)]}} = \dfrac{\dfrac{0.4}{0.6}}{\dfrac{0.5}{0.5}} = \dfrac{0.67}{1.0}$$

OR = 0.67

A popular short cut formula is OR = (A*D)/(B*C) = (20)*(15)/(15)*(30) = 0.67

In this example, since the OR is less than 1.00, the interpretation is that patients wearing SCL are actually less likely to have red eyes than a comparable group of non-SCL wearers. RR and OR values are rarely equal to each other; but in large, well-designed research studies of uncommon events, the values of RR and OR may be almost the same.

> **...in large, well-designed research studies of uncommon events, the values of RR and OR may be almost the same.**

Summary thoughts

In his book entitled *The Ghost Map*, author Steven Johnson describes the terrible cholera epidemic in London during 1850's. At the time, most people thought cholera was spread person-to-person through contaminated air. But John Snow, who was a respected anesthesiologist in London, observed that the gastrointestinal tract – and not the lung - was the first organ affected in cholera patients. Because of this astute clinical observation, he hypothesized that the mysterious agent causing the cholera epidemic must be contained in food and/or water, not the air. And when plotted the locations of death from cholera as recorded in the meticulous vital statistics that his medical colleague, William Farr, recorded in his *Weekly Returns* (a forerunner of the U.S. Centers for Disease Control and Preventions's *WMMR: Weekly Morbidity and Mortality Report*), Snow concluded the "risk factor" for cholera was the water from the Broad Street pump. When Snow showed his analytical epidemiologic data to the local authorities, they removed the Broad Street pump handle and the cholera epidemic was over. This story is significant in the history of epidemiology, because the noted German microbiologist, Robert Koch, would not be able to identify the true causative agent of cholera (which is the bacterium *Vibrio cholera*) for another thirty years.

Johnson brings this historic and premiere epidemiological success story to bear on the major public health problems of the 21st century by stating

The influence of the Broad Street maps extends beyond the realm of disease. The Web is teeming with new forms of amateur cartography, thanks to services like Google Earth and Yahoo! Maps. Where Snow inscribed the location of pumps and cholera fatalities over the street grid, today's mapmakers record a different kind of data: good public schools, Chinese takeout places, playgrounds, gay-friendly bars, open houses….Anyone can create a map that shows you where streets intersect and where hotels are….The maps now appearing are of a different breed altogether: maps of local knowledge created by actual locals.…[10]

Using sound, time-honored epidemiological principles, and having accurate, contemporary data readily available now via the internet, allows today's epidemiologists an unparalleled opportunity to study the distribution and determinants of health problems in populations at risk. And with this knowledge will come better strategies to reduce the morbidity and mortality rates for a wide-variety of diseases throughout the world.

References

1. Austin, DF and Werner, SB: Epidemiology for the health sciences. Charles C Thomas, Springfield, IL, 1974, p. 4.

2. National Eye Institute and Prevent Blindness America. Vision problems in the U.S: Prevalence of adult vision impairment and age-related eye disease in America. Bethesda, MD, 2002. Available from: www.usvisionproblems.org

3. The 10 Leading Causes of Death by Broad Income Group, 2004. World Health Organization, Geneva, Switzerland, Fact Sheet No 310/November, 2008. Available from: http://www.who.int/mediacentre/factsheets/fs310_2008.pdf

4. Joslin, CE, Tu, EY, McMahon, TT, et al: Epidemiological characteristics of a Chicago-area Acanthamoeba keratitis outbreak. Am J Ophthal, 142(2): 212-217, 2006.

5. Convergence Insufficiency Treatment Trial Study Group. Randomized clinical trial trial of treatments for symptomatic convergence insufficiency in children. Arch Ophthal, 126(10): 1336-1349, 2008.

6. NEI Press Release, Monday, October 13, 2008: More Effective Treatment Identified for Common Childhood Vision Disorder. National Eye Institute, National Institutes of Health, Bethesda, MD. Available from: http://www.nei.nih.gov/news/pressreleases/101308.asp

7. Weiss, NS: Clinical epidemiology: the study of the outcome of illness, third ed. Oxford Univ press, New York, 2006, pp. 143-156.

8.   J Sommer, A: Epidemiology and statistics for the ophthalmologist. Oxford University Press, New York, 1980, p. 3.

9.   Marshall, EC: Research methodologies and statistics. Chapter 5 in Public health and Community optometry, second edition, by RD Newcomb and EC Marshall. Butterworths, Boston, 1990, pp. 52-53.

10. Johnson, S: The Ghost Map: the story of London's most terrifying epidemic- and how it changed science, cities, and the modern world. Riverhead books, New York, 2006, pp 219-220.