MedPage Tools

Guide to Biostatistics

Here is a compilation of important epidemiologic concepts and common biostatistical terms used in medical research. You can use it as a reference guide when reading articles published on *MedPage Today* or download it to keep near the reading stand where you keep your print journals. For more detailed information on these topics, use the reference list at the end of this presentation.

Study Designs

- How Research Is Classified
- Terminology
- Important epidemiologic concepts

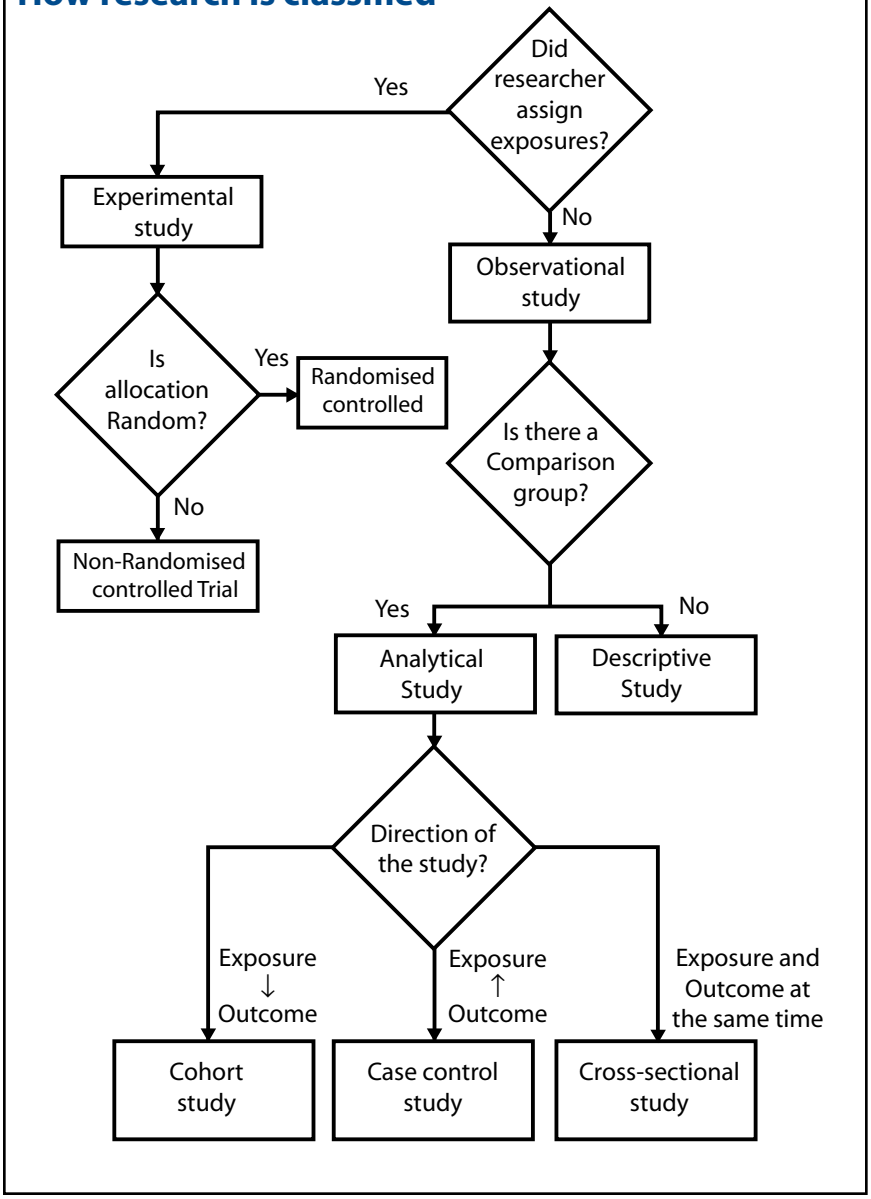
Descriptive Statistics

- Measures of central tendency
- Measures of spread
- Measures of frequency of events
- Measures of Association
- Terms used to describe the quality of measurements
- Measures of diagnostic test accuracy
- Expressions used when making inferences about data
- Multivariable Regression Methods

References

Study Designs in Clinical Research

How research is classified[1]



Study Designs

- How Research Is Classified
- Terminology
- Important epidemiologic concepts

Descriptive Statistics

- Measures of central tendency
- Measures of spread
- Measures of frequency of events
- Measures of Association
- Terms used to describe the quality of measurements
- Measures of diagnostic test accuracy
- Expressions used when making inferences about data
- Multivariable Regression Methods

References

Terminology

Clinical Trial Experimental study in which the exposure status (e.g. assigned to active drug versus placebo) is determined by the investigator.

Randomized Controlled Trial A special type of clinical trial in which assignment to an exposure is determined purely by chance.

Cohort Study Observational study in which subjects with an exposure of interest (e.g. hypertension) and subjects without the exposure are identified and then followed forward in time to determine outcomes (e.g. stroke).

Case-Control Study Observational study that first identifies a group of subjects with a certain disease and a control group without the disease, and then looks to back in time (e.g. chart review) to find exposure to risk factors for the disease. This type of study is well suited for rare diseases.

Cross-Sectional Study Observational study that is done to examine presence or absence of a disease or presence or absence of an exposure at a particular time. Since exposure and outcome are ascertained at the same time, it is often unclear if the exposure preceded the outcome.

Case Report or Case Series Descriptive study that reports on a single or a series of patients with a certain disease. This type of study usually generates a hypothesis but cannot test a hypothesis because it does not include an appropriate comparison group.

Important Epidemiologic Concepts

Bias Any systematic error in the design or conduct of a study that results in a mistaken estimate of an exposure's effect on risk of disease.

Selection Bias Bias introduced by the way in which participants are chosen for a study. For example, in a case-control study using different criteria to select cases (e.g. sick, hospitalized population) versus controls (young, healthy outpatients) other than the presence of disease can lead the investigator to a false conclusion about an exposure.

Confounding This occurs when an investigator falsely concludes that a particular exposure is causally related to a disease without adjusting for other factors that are known risk factors for the disease and are associated with the exposure.

Study Designs

- How Research Is Classified
- Terminology
- Important epidemiologic concepts

Descriptive Statistics

- Measures of central tendency
- Measures of spread
- Measures of frequency of events
- Measures of Association
- Terms used to describe the quality of measurements
- Measures of diagnostic test accuracy
- Expressions used when making inferences about data
- Multivariable Regression Methods

References

Descriptive Statistics

Measures of Central Tendency

Mean equals the sum of observations divided by the number of observations.

Median equals the observation in the center when all observations are ordered from smallest to largest; when there is an even number of observations the median is defined as the average of the middle two values.

Mode equals the most frequently occurring value among all observations.

Measures of Spread

Spread (or variability) describes the manner in which data are scattered around a specific value (such as the mean). The most commonly used measures of spread are:

Range is the difference between the largest observation and the smallest.

Standard Deviation measures the spread of data around the mean. One standard deviation includes 68% of the values in a sample population and two standard deviations include 95% of the values.

Standard Error of the Mean describes the amount of variability in the measurement of the population mean from several different samples. This is in contrast to the standard deviation which measures the variability of individual observations in a sample.

Percentile equals the percentage of a distribution that is below a specific value. As an example, a child is in the 80th percentile for height if only 20% of children of the same age are taller than he is.

Interquartile Range refers to the upper and lower boundary defining the middle 50 percent of observations. The upper boundary is the 75th percentile and the lower boundary is the 25th percentile.

Measures of Frequency of Events

Incidence The number of new events (e.g. death or a particular disease) that occur during a specified period of time in a population at risk for developing the events.

Incidence Rate A term related to incidence that reports the number of new events that occur over the sum of time individuals in the population were at risk for having the event (e.g.

Study Designs

- How Research Is Classified
- Terminology
- Important epidemiologic concepts

Descriptive Statistics

- Measures of central tendency
- Measures of spread
- Measures of frequency of events
- Measures of Association
- Terms used to describe the quality of measurements
- Measures of diagnostic test accuracy
- Expressions used when making inferences about data
- Multivariable Regression Methods

References

events/person-years).

Prevalence The number of persons in the population affected by a disease at a specific time divided by the number of persons in the population at the time.

Measures of Association

The types of measures used to define the association between exposures and outcome depends upon the type of data. For categorical variables, the relative risk and odds ratio are commonly used to describe the relationship between exposures and outcome.

Relative risk and cohort studies The relative risk (or risk ratio) is defined as the ratio of the incidence of disease in the exposed group divided by the corresponding incidence of disease in the unexposed group (Figure 2). Relative risk can be calculated in cohort studies such as the Framingham Heart Study where subjects with certain exposures (e.g. hypertension, hyperlipidemia) were followed prospectively for cardiovascular outcomes. The incidence of cardiac events in subjects with and without exposures was then used to calculate relative risk and determine whether exposures were cardiac risk factors.

Odds ratio and case-control studies The odds ratio is defined as the odds of exposure in the group with disease divided by the odds of exposure in the control group (Figure 1). As described above, subjects are selected on the basis of disease status in case-control studies, therefore it is not possible to calculate the rate of development of disease given presence or absence of exposure. So, the odds ratio is often used to approximate the

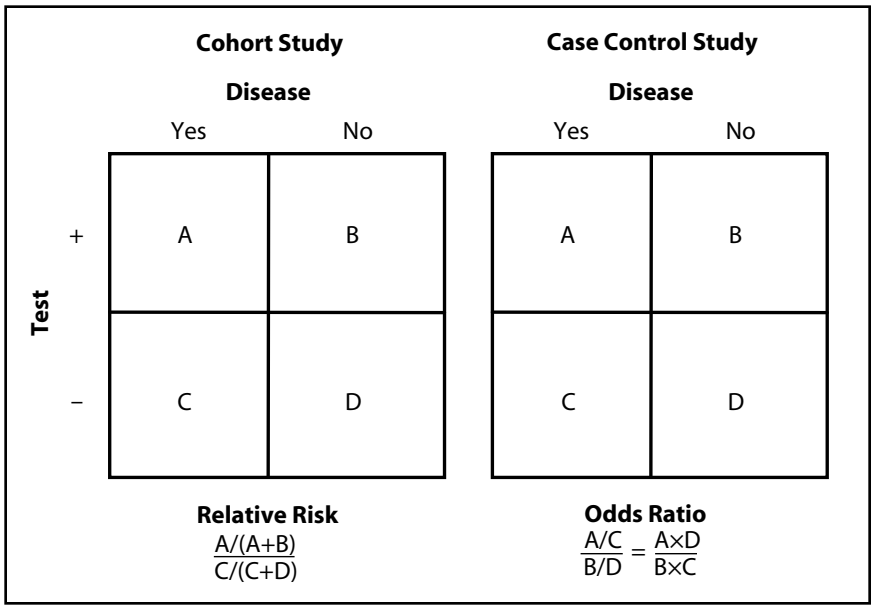


Figure 1: In a case-control study, the odds ratio can be used to approximate the relative risk under the assumption that the disease is rare.

Study Designs

- How Research Is Classified
- Terminology
- Important epidemiologic concepts

Descriptive Statistics

- Measures of central tendency
- Measures of spread
- Measures of frequency of events
- Measures of Association
- Terms used to describe the quality of measurements
- Measures of diagnostic test accuracy
- Expressions used when making inferences about data
- Multivariable Regression Methods

References

relative risk in case-control studies. For example, a case-control study was done to evaluate the relationship between artificial sweeteners and bladder cancer. The odds of artificial sweetener use in the cases and controls were used to calculate an odds ratio and determine whether sweeteners were associated with bladder cancer. Under the assumption that the disease under consideration is rare (e.g. bladder cancer), the odds ratio gives a stable, unbiased estimate of the relative risk (Figure 1). The odds ratio from a case-control study nested within a defined cohort also approximates the relative risk even when the rare disease assumption is not held.

If the disease is rare, A<<B and C<<D. So, A/(A + B) is approximated by A/B and C/(C + D) approximated by C/D. In this situation, the relative risk equals (A/B)/(C/D) which, rearranged, equals the odds ratio A×D/B×C

Absolute risk The relative risk and odds ratio provide a measure of risk compared with a standard. However, it is sometimes desirable to know the absolute risk. For example, a 40% increase in risk of heart disease because of a particular exposure does not provide insight into the likelihood that exposure in an individual patient will lead to heart disease.

The **Attributable risk** or **Risk difference** is a measure of absolute risk. It represents the excess risk of disease in those exposed taking into account the background rate of disease. The attributable risk is defined as the difference between the incidence rates in the exposed and non-exposed groups.

A related term, the **Population Attributable Risk** is used to describe the excess rate of disease in the total study population of exposed and non-exposed individuals that is attributable to the exposure. This measure is calculated by multiplying the Attributable risk by the proportion of exposed individuals in the population.

Number needed to treat (NNT) The number of patients who would need to be treated to prevent one adverse outcome is often used to present the results of randomized trials. NNT is the reciprocal of the absolute risk reduction (the absolute adverse event rate for placebo minus the absolute adverse event rate for treated patients). This approach can be used in studies of various interventions including both treatment and prevention. The estimate for NNT is subject to considerable error and is generally presented with 95% confidence intervals so that it can be properly interpreted.

Terms Used To Describe The Quality Of Measurements

Reliability The concept of reliability or reproducibility is related to the amount of error in any measurement (e.g. blood pressure

Study Designs

- How Research Is Classified
- Terminology
- Important epidemiologic concepts

Descriptive Statistics

- Measures of central tendency
- Measures of spread
- Measures of frequency of events
- Measures of Association
- Terms used to describe the quality of measurements
- Measures of diagnostic test accuracy
- Expressions used when making inferences about data
- Multivariable Regression Methods

References

measurement). A more formal definition of reliability is variability between subjects divided by inter-subject variability plus measurement error. Thus, reliability is greater when measurement error is minimal. There are several types of reliability including: inter and intra-observer reliability and test-retest reliability.

Percent agreement and the kappa statistic are often used to report reliability. The kappa statistic takes into account agreement that would be seen by chance alone while percent agreement does not. Generally, a kappa greater than 0.75 represents excellent agreement beyond chance, a kappa below 0.40 represents poor agreement and a kappa of 0.40-0.75 represents intermediate to good agreement.

Validity refers to the extent to which a test or surrogate is measuring what we think it is measuring. There are several types of validity that can be measured including content validity (the extent to which the measure reflects the dimensions of a particular problem), construct validity (the extent to which a measure conforms to an external established phenomenon), and criterion validity (the extent to which a measure correlates with a gold standard or can predict an observable phenomenon). These types of validity are often applied to questionnaires in which the truth is not physically verifiable.

Measures Of Diagnostic Test Accuracy

Sensitivity is defined as the ability of the test to identify correctly those who have the disease. It is the number of subjects with a positive test who have disease divided by all subjects who have the disease. A test with high sensitivity has few false negative results.

Specificity is defined as the ability of the test to identify correctly those who do not have the disease. It is the number of subjects who have a negative test and do not have the disease divided by the number of subjects who do not have the disease. A test with high specificity has few false positive results.

Sensitivity and specificity are test characteristics that are most useful when assessing a test used to screen a free-living population. These test characteristics are also interdependent: an increase in sensitivity is accompanied by a decrease in specificity and visa versa. This is illustrated best by continuous tests where the cut-off for a positive test result can be varied. For example, consider the use of the white blood cell (WBC) count as a test to diagnose bacterial infection. If one sets a high cut-off for a positive test (e.g. WBC> 25,000) then the test will have a low sensitivity and high specificity compared to the test characteristics if the cut-off is lower (e.g. WBC>10,000).

Study Designs

- How Research Is Classified
- Terminology
- Important epidemiologic concepts

Descriptive Statistics

- Measures of central tendency
- Measures of spread
- Measures of frequency of events
- Measures of Association
- Terms used to describe the quality of measurements
- Measures of diagnostic test accuracy
- Expressions used when making inferences about data
- Multivariable Regression Methods

References

Predictive values are important for assessing how useful a test will be in the clinical setting at the individual patient level. The **positive predictive value** is the probability of disease in a patient with a positive test. Conversely, the **negative predictive value** is the probability that the patient does not have disease if he has a negative test result.

Predictive values depend on the prevalence of a disease in a population. A test with a given sensitivity and specificity can have different predictive values in different patient populations. If the test is used in a population with a high prevalence, it will have a high positive predictive value and the same test will have a low positive predictive value when used in a population with low disease prevalence. For example, a positive stool test for occult blood is much more likely to be predictive of colon cancer in a population of elderly people compared with twenty year olds.

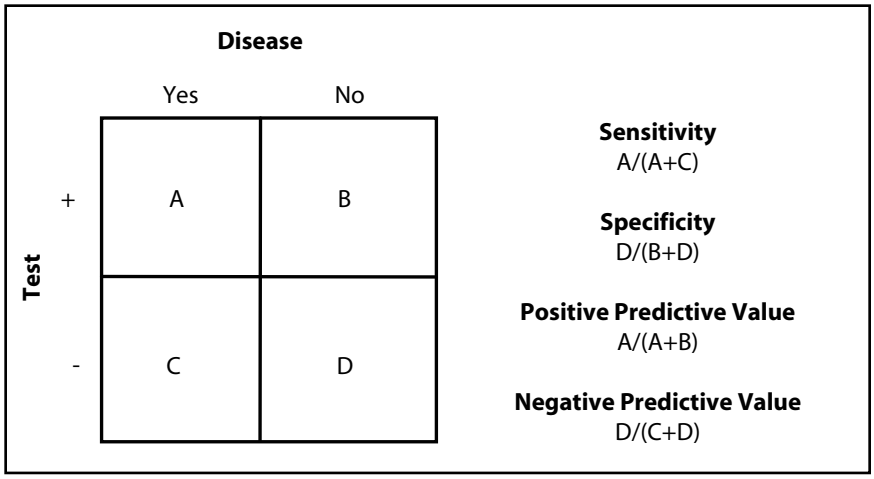


Figure 2: Calculating sensitivity, specificity, and predictive values

Likelihood ratios Calculating likelihood ratios is another method of assessing the accuracy of a test in the clinical setting. Likelihood ratios also offer the advantage of being independent of disease prevalence.

The likelihood ratio indicates how much a given diagnostic test result will raise or lower the odds of having a disease relative to the prior probability of disease. Each diagnostic test is characterized by two likelihood ratios: a positive likelihood ratio that tells us the odds of disease if the test result is positive and a negative likelihood ratio that tells us the odds of disease if the test result is negative:

LR+ = Sensitivity / (1- Specificity)

LR- = (1- Sensitivity) / Specificity

A likelihood ratio greater than 1 increases the odds that the per-

Study Designs

- How Research Is Classified
- Terminology
- Important epidemiologic concepts

Descriptive Statistics

- Measures of central tendency
- Measures of spread
- Measures of frequency of events
- Measures of Association
- Terms used to describe the quality of measurements
- Measures of diagnostic test accuracy
- Expressions used when making inferences about data
- Multivariable Regression Methods

References

son has the target disease, and the higher the LR the greater this increase in odds. Conversely, a likelihood ratio less than 1 diminishes the odds that the patient has the target disease.

Expressions Used When Making Inferences About Data

Confidence Intervals The results of any study sample are an estimate of the true value in the entire population. The true value may actually be greater or less than what is observed. A confidence interval gives a range of values within which there is a high probability (95% by convention) that the true population value can be found. The confidence interval takes into consideration the number of observations and the standard deviation in the sample population. The confidence interval narrows as the number of observations increases or standard deviation decreases.

Errors In hypothesis testing, there are two types of errors:

Type I error (alpha) is the probability of incorrectly concluding there is a statistically significant difference in the population when none exists. This type of error is also called alpha and is the number after a P-value. A $P<0.05$ means that there is a less than 5% chance that the difference could have occurred by chance.

Type II error (beta) is the probability of incorrectly concluding that there is no statistically significant difference in a population when one exists.

Power is a measure of the ability of a study to detect a true difference. It is measured as 1- type II error rate or 1-beta. Every researcher should perform a power calculation prior to carrying out a study to determine the number of observations needed to detect a desired degree of difference. Ideally this difference should equal the smallest difference that would still be considered to be clinically important. However, the smaller the difference, the greater the number of observations needed. For example, it takes fewer patients to observe a 50% reduction in mortality from a new therapy than a 5% reduction.

Multivariable Regression Methods

In medical research, one is often interested in studying the independent effect of multiple risk factors on outcome. For example, we may want to know the independent effect of age, gender and smoking status on the risk of having a myocardial infarction. Furthermore, we may want to know if smoking raises the risk equally in men and women. Multivariable regression methods allow us to answer these types of questions by simultaneously accounting for multiple variables. The type of regression model used depends on the type of outcome data being evaluated.

Study Designs

- How Research Is Classified
- Terminology
- Important epidemiologic concepts

Descriptive Statistics

- Measures of central tendency
- Measures of spread
- Measures of frequency of events
- Measures of Association
- Terms used to describe the quality of measurements
- Measures of diagnostic test accuracy
- Expressions used when making inferences about data
- Multivariable Regression Methods

References

Multiple linear regression is used when the outcome data is a continuous variable such as weight. For example, one could estimate the effect of a diet on weight after adjusting for the effect of confounders such as smoking status. Another use of this method is to predict a linear variable based on known variables.

Logistic regression is used when the outcome data is binary such as cure or no cure. Logistic regression can be used to estimate the effect of an exposure on a binary outcome after adjusting for confounders. Logistic regression can also be used to find factors that discriminate two groups or to find prognostic indicators for a binary outcome. This method can also be applied to case-control studies.

Survival Analysis

In survival analysis, one is commonly interested in the time until some event such as the time from treatment of disease to death. In the study population, only some subjects will have the event of interest (e.g. death, stroke), others will have alternate events or no events. The duration of follow-up will also vary among subjects and it is important to account for the different follow-up times.

The Kaplan-Meier analysis and a regression method, the Cox proportional hazards analysis are two methods of survival analysis that account for inter-subject variation in events and follow-up time.

Kaplan-Meier analysis measures the ratio of surviving subjects (or those without an event) divided by the total number of subjects at risk for the event. Every time a subject has an event, the ratio is recalculated. These ratios are then used to generate a curve to graphically depict the probability of survival (Figure 3).

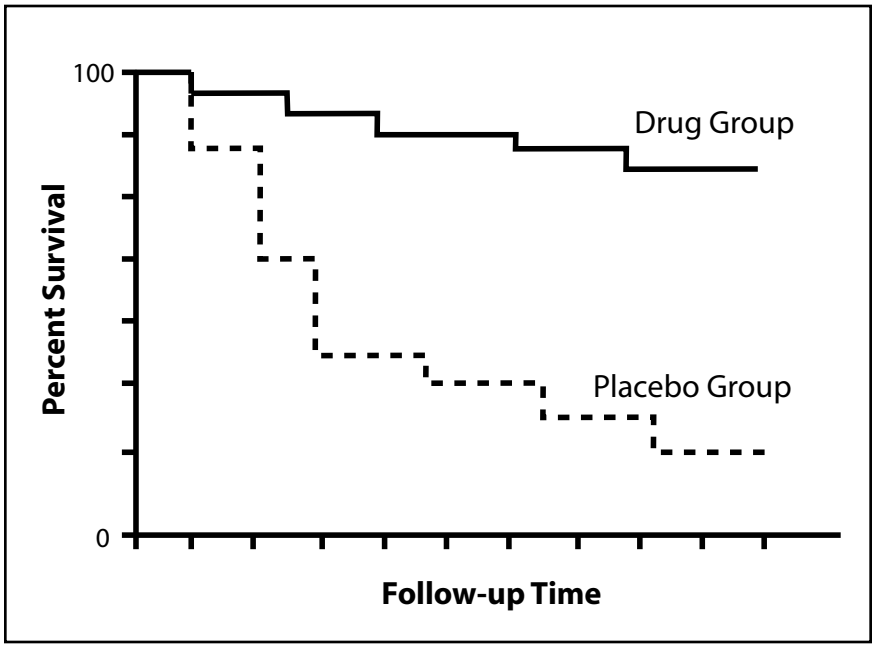


Figure 3: Kaplan-Meier Survival Curves

Study Designs

- How Research Is Classified
- Terminology
- Important epidemiologic concepts

Descriptive Statistics

- Measures of central tendency
- Measures of spread
- Measures of frequency of events
- Measures of Association
- Terms used to describe the quality of measurements
- Measures of diagnostic test accuracy
- Expressions used when making inferences about data
- Multivariable Regression Methods

References

In studies with an intervention arm and a control arm, one can generate two Kaplan-Meier curves. If the curves are close together or cross, a statistically significant difference is unlikely to exist. Statistical tests such as the log-rank test can be used to confirm the presence of a significant difference.

Cox proportional hazards analysis is similar to the logistic regression method described above with the added advantage that it accounts for time to a binary event in the outcome variable. Thus, one can account for variation in follow-up time among subjects. Like the other regression methods described above, it can be used to study the effect of an exposure on outcome after adjusting for confounders. Cox analysis can also be used to find prognostic indicators for survival in a given disease.

The hazard ratio that results from this analysis can be interpreted as a relative risk (risk ratio). For example, a hazard ratio of 5 means that the exposed group has five times the risk of having the event compared to the unexposed group.

Rubeen K. Israni, M.D.
Fellow, Renal-Electrolyte and Hypertension Division, University of Pennsylvania School of Medicine.

Study Designs

- How Research Is Classified
- Terminology
- Important epidemiologic concepts

Descriptive Statistics

- Measures of central tendency
- Measures of spread
- Measures of frequency of events
- Measures of Association
- Terms used to describe the quality of measurements
- Measures of diagnostic test accuracy
- Expressions used when making inferences about data
- Multivariable Regression Methods

References

References

1. Grimes DA, Schultz KF: An overview of clinical research: the lay of the land. *Lancet* 359:57-61, 2002
2. Grimes DA, Schultz KF: Bias and causal associations in observational research. *Lancet* 359:248-252, 2002.
3. Gordis L: *Epidemiology*, 3rd Edition, Philadelphia, Elsevier Saunders, 2004.
4. Rosner B: *Fundamentals of Biostatistics*, 4th Edition, Daxbury Press, 1995.
5. Grimes DA, Schultz KF: Cohort studies: marching towards outcomes. *Lancet* 359: 341-345, 2002.
6. Schultz KF, Grimes DA: Case-control studies: research in reverse. *Lancet* 359:431-434, 2002.
7. Streiner DL, Norman GR: *Health Measurement Scales: A Practical Guide to their Development and Use,* 2nd Edition, New York, Oxford University Press, 2000.
8. Jaeschke R, Guyatt GH, Sackett DL: Users' guides to the medical literature. III. How to use an article about a diagnostic test. B. What are the results and will they help me in caring for my patients? The Evidence-Based Medicine Working Group. *Jama* 271:703-707, 1994.
9. Guyatt G, Jaeshke R, Heddle N, et al. Basic statistics for clinicians: 2. Interpreting study results: confidence intervals. *Cmaj* 152:169-173, 1995.
10. Katz MH: Multivariable analysis: a primer for readers of medical research. *Ann Intern Med* 138:644-650, 2003.
11. Campbell MJ: *Statistics at Square Two*, 4th Edition, London, BMJ Publishing Group, 2004.