

PUTTING RESEARCH INTO CLINICAL PRACTICE

Lisa A. Jones-Jordan, PhD
Elizabeth S. Hoppe, OD, MPH, DrPH

The specific objectives of this chapter are after reading and completing the study questions the student should be able to:

- 1.) Identify different study designs;
- 2.) Recognize different statistical tests;
- 3.) Interpret the results of analyses presented; and
- 4.) Critique published literature to determine applicability

Introduction

The role of research in a clinical practice may be a subtle one, but it is a role that is vital to bringing the most contemporary methods of care and treatment to the patient. Research plays a key role in every step of the health care decision-making process. If we are looking to cure glaucoma, we need to research the mechanisms of disease development, and identify a treatment that is effective in curing glaucoma. Determining effective treatments requires clinical studies to establish what works. What about if we take a step back prior to the development of glaucoma and start looking for ways to prevent it to begin with? Identifying risk factors and determining where we can intervene in the development of a disease to prevent its occurrence begins in research as well. Research can ultimately help us save health care dollars as well, by making sure we spend money on the more effective and most appropriate treatments.

When you approach a patient with a particular diagnosis, you want to be armed with the most current information, but the degree to which information is current is not sufficient. Evidence should also be based on the best research available. To find and apply the best, you need to be able to assess the available literature to identify those studies that have followed appropriate research methodology. Because of the importance of the information that is generated in the research endeavor, understanding what is behind the results that we apply to patient care is the logical place to begin. The framework of research is the methodology: the process by which research questions (or hypotheses) are posed, studies are designed to answer the questions asked, and data are analyzed in a way to produce accurate and meaningful results.

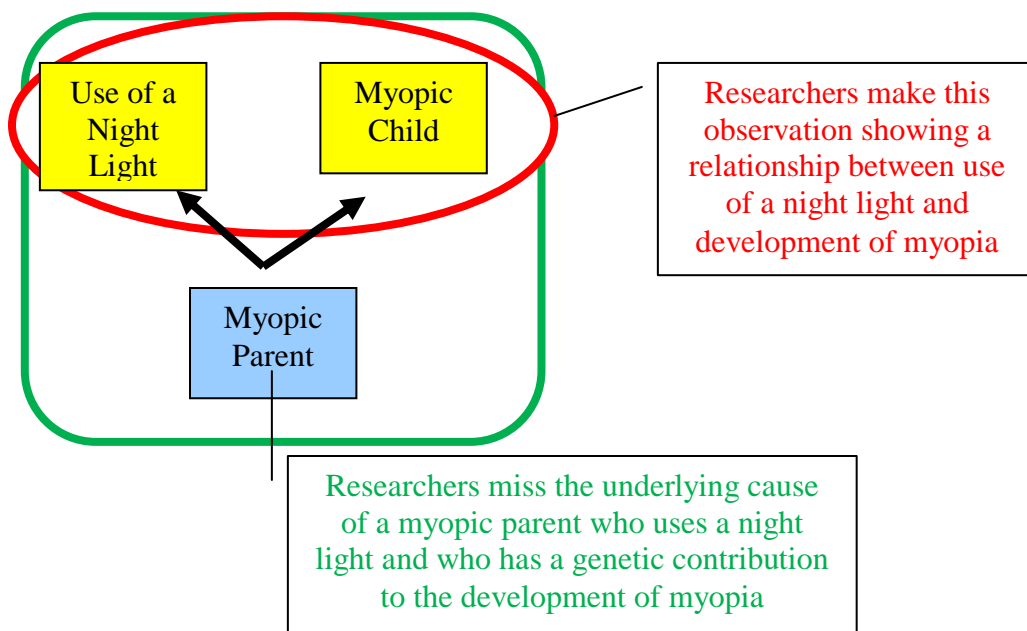
The primary objective of this chapter is to familiarize the reader with the essential concepts necessary to appreciate good research. Here we will outline the basics of study design and analysis to help you identify hallmarks of a well-conducted scientific study. With the tools to locate and critically appraise the results of the best quality research, you can then take the information and apply it to your practice.

As an example of why we want to be wary of accepting the results of a study without scrutiny, we will look an article entitled “Myopia and Ambient Lighting” which was published in the journal *Nature* in 1999.¹ The paper reported that there was a dose-

dependent increase in myopia found in children who slept in rooms with some kind of artificial light. The percentage of children with myopia was higher among children who slept with a night light before the age of 2 years old and even higher among children who slept in a room with full light. The article received wide-spread attention, and was cited in the popular press and in the media. Parents worried about causing life-long vision problems by behaviors that occurred when their children were still infants. However, on closer inspection, it became apparent that there were some significant problems with the research design, methodology, and the conclusions that were drawn.

First, critical consumers of the article found that the sample on whom these conclusions were based would not appear to be a representative sample of children in the 2 – 16 year old age group. The sample came from a tertiary referral facility, so one would expect that these children would differ from the general population of all children in that age range. In fact, the proportion of myopia in this group overall was around 29%, which is much higher than would be expected in this age group. Thus, selection bias or sampling bias may have been present. In addition, one of the major risk factors for myopia is parental history of myopia, which was not collected for this study, thereby not allowing the investigators to control for its effect. This raises the possibility of confounding. For instance, as a myopic parent, perhaps you are more likely to use a night light to see without correction to go into your child's room. This would make parental myopia related to the outcome, child's myopia, as well as being related to the risk factor, night light use. This type of variable is known as a *confounder*, and it has the ability to distort the results of a study. A figure illustrating a potentially confounding relationship is shown in Figure 1.

Figure 1. Illustration of Potentially Confounding Relationship



The Research Question

The research question is a statement of the goal of the study, detailing what the unanswered issue is and what type of information the researcher is collecting to answer the question. Good research questions are ones that are concise with a testable idea behind them. A frequent mistake is attempting to make the research question too complex for any one study to answer.

The research question is where study design begins. It helps define the group of people under study and should provide details about the outcome, or measure of interest, that is the focus of the study. This statement of outcome then has an impact on sample size for the chosen study design, as well. For studies that aim to compare groups or treatments, the research question is asked in a manner that allows for a statement as a hypothesis. A hypothesis is a specific question that allows for the results to be tested by statistical analysis.ⁱⁱ It outlines specific criteria that determine whether the hypothesis is accepted or rejected.

For example, the following would be considered a research question:

Is an increased level of low-density lipoprotein cholesterol (LDL) associated with an increased risk of glaucoma?

Once you have framed your research question, you then must develop a testable hypothesis designed to help answer the question. A hypothesis can be described as an assertion of a specific relationship between two or more variables. We typically call these variables either the independent variable and the dependent variable or the predictor variable and the outcome variable. The independent or predictor variable is the factor that you expect to influence the results of the study. The independent or predictor variable could be considered a risk factor. The dependent variable or the outcome variable is the variable whose values are the results or outcomes of the study. The dependent or outcome variable could be considered a disease or condition. Think about the relationship you wish to study. What relationship makes the most sense – which is the predictor and which is the outcome?

A testable hypothesis would be something like:

Does a low-density lipoprotein cholesterol level (LDL) above 150 mg/dL increase the risk of glaucoma by 50%?

One of the next steps is selecting the appropriate study design. Several key questions can help you narrow in on your best options. Should your study design be a descriptive study, an observational study, or an experimental study? Is the independent or predictor variable something that you can manipulate to determine the exposure (like a particular drug, therapy, treatment or educational program) or is it something inherent (like age, gender, or profession)? What is the optimal timing of the data collection (retrospective, prospective, at one point in time)? What is the best method also

consider the strengths and weaknesses of the various designs described in the following sections. Do you have resources available for the optimal study design? If you must compromise on study design can you implement additional controls to limit the potential for bias and confounding?

Once you have your research question, hypothesis, and study design selected it is important to get some advice and input from more experienced researchers to help you refine these key elements. These three things are critical to ensuring a successful project, so take time to work through several iterations. Some basic next steps to getting started in a research project are shown in Table 1.

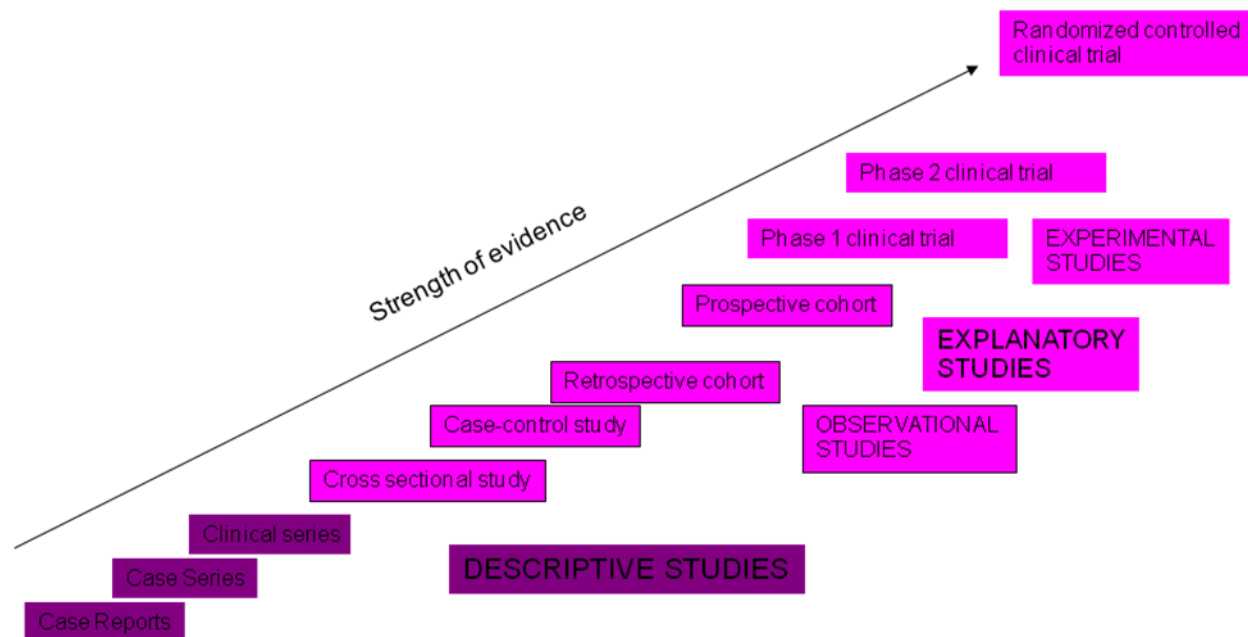
Table 1. Steps for Planning Research Studies

- Write a researchable question
- Critically review the existing literature
- State the purpose of the study
- Write your research hypothesis
- Define the variables of interest
- Select an appropriate study design
- Define the population to be studied
- Determine the method for selecting the sample
- Determine the ways for measuring the variables
- Review threats validity and reliability, identify potential bias and confounding
- Plan for analysis of data
- Ensure adherence to ethical guidelines

Study Design Overview

Why do we conduct health research? Simply stated, we do it to save lives, to ease suffering, and to advance knowledge. In its most basic form, epidemiology is making observations to describe normal from abnormal in populations. It is describing the characteristics of a disease or condition and the characteristics of the people effected and not effected. It is observing the characteristics over time. Research needs to define disease by showing repeating signs and symptoms and be able to predict the natural course of disease and its effects on quality of life. Then, results are interpreted to theorize the cause of disease. Finally, research is commissioned to treat the symptoms of disease and/or cure the disease based on the biology. Or, research investigates established treatments to confirm or refute current care. Thus, we conduct research to improve health care's efficacy and efficiency. The underlying purpose of the research study gives rise to classifications and organization of study designs, illustrated in Figure 2.

Figure 2. Clinical study designs.



In Figure 2 two basic study methods are shown: descriptive studies and explanatory studies. Descriptive studies are primarily used to describe events, observations, or activities. This type of study does not provide an explanation about a disease process or treatment but usually describes patient cases that include interesting or unusual observations for a small number of patients. Descriptive studies can sometimes provide preliminary evidence that lead to further research and ultimately yield exciting discoveries. Explanatory studies seek to make comparisons that allow us to answer questions about the effectiveness of treatments, to quantify relationships between risk factors and health conditions, to identify factors that predict health behaviors or health outcomes, or to offer insight into etiology of diseases. Explanatory studies usually include a research question that is framed as a hypothesis about the relationship between a risk factor and a health outcome.

The category of explanatory studies is further broken down into experimental studies and observational studies. In observational studies one or more groups of subjects are observed and characteristics about them are recorded and analyzed. In observational studies, the researchers observe and measure what happens naturally. Experimental studies include some type of intervention or manipulation, such as the use of a drug, procedure, or treatment. The researchers study the effect of the intervention on the study subjects.

Different studies are considered as stronger or weaker in terms of the value of the evidence produced and their potential to provide useful information that can impact health care delivery and clinical practice. The strength of the study is affected by potential sources of bias and limitations in the study design methodology. This chapter will highlight the explanatory studies show in the diagram in Figure 2.

Clinical Scenario:

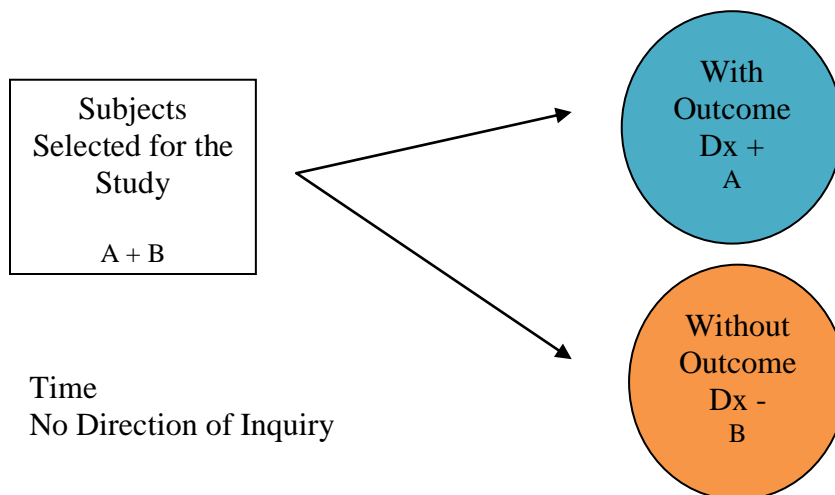
You are an active member of your community Lions Club and frequently participate in community-based screenings. The Club secretary has noticed that different screening locations have resulted in different rates of referrals, and different types of conditions are being referred for follow-up evaluations. In particular, she noticed that when the screening was done in a predominantly Latino community, more people had reduced visual acuity. She thinks this may be due to a problem with the eye chart projector. For example, perhaps the bulb needs to be replaced or the calibration distance was off. You double-check the equipment and everything seems to be fine. Can there be another possible explanation?

Cross-sectional Study Design

We are all familiar with the results of cross-sectional studies, but we may not have recognized them as a familiar study design. Every time that we hear the results of an opinion survey, an election poll, or an estimate of the magnitude of a social problem, those numbers likely came from a cross-sectional study. In health care, cross-sectional studies are often used to estimate the prevalence of a condition, to describe health behaviors, to compare two or more test procedures, to establish normative data, or to assess levels of patient satisfaction.

Cross-sectional studies analyze information about a group of subjects that is collected at one point in time. Subjects are identified from the population of interest and information is gathered in a variety of ways such as a survey, a review of a medical record, analysis of imaging studies, or laboratory testing results. Subjects are not followed either forward in time, or retrospectively in past reviews. Cross-sectional studies offer a snap shot picture of what is happening at the time of data collection. The purpose of the study is usually descriptive, designed to identify the magnitude and details of the health status of the subjects. Cross-sectional studies are one of the easiest observational studies because the researchers observe the status quo. A schematic of the cross-sectional study design is shown in Figure 3.

Figure 3. Schematic Diagram: Cross Sectional Study Design



Cross-sectional studies can be useful for measuring the *prevalence* of a condition. Prevalence is defined as the proportion of persons in a given population who have a particular condition of interest at a point in time. Prevalence quantifies the number of *existing* cases of the condition in a population.

Cross-sectional studies are also useful for evaluating new clinical tests or new applications for existing tests, for example an evaluation of two different methods for doing the same thing. They can be used for establishing normal ranges of test results. When information about a variety of variables is collected, cross-sectional studies can be used to identify potential risk factors or associations. Cross-sectional studies can be helpful as pilot studies or to help generate hypotheses.

The results of cross-sectional studies may be presented using descriptive statistics or prevalence rates with 95% confidence intervals. If data are gathered for more than one variable, relationships may also be assessed using other appropriate techniques based on the type of data gathered (e.g. nominal, ordinal, numerical or categorical) and the appropriate statistical tests (see section on statistics). The formula, calculations and interpretation used for prevalence are shown in Table 2.

Table 2 Analysis: Cross Sectional Study

Calculate the Prevalence of the Condition (Outcome)

Prevalence is defined as the proportion of persons in a population of interest who have a particular condition or outcome at a given point in time.

$$\text{Prevalence (P)} = \frac{\text{Number with the Outcome}}{\text{Total population}} = \frac{A}{B}$$

Report Descriptive Statistics

- Central tendency
 - Mean, median, mode
- Variability
 - Range
 - Standard error, standard deviation
 - 95% confidence interval of the estimate

The main advantages for cross-sectional studies are efficiency and low cost. Investigators do not have to wait for outcomes to develop or for treatments to take effect. However, there are significant problems with this study design which make it a weaker design than other study options. The means of defining and sampling the population being studied and the response rate are the main pitfalls of the cross-sectional study.

When reading the published literature and when applying the results of cross-sectional studies it is important to question how the investigators defined the population

of interest. In many studies the population is drawn from a clinical base, which may not be reflective of the population as a whole. White *et al*, have developed a model depicting the selection process when a cross-sectional study is conducted using a tertiary hospital or referral center. Their example suggests that only a small fraction of possible subjects in a community become subjects for study in a referral center. They suggest that those who do are unlikely to be representative of the remaining population who are not included, thereby causing bias and misleading information.

Once the population has been defined, the next challenge for cross-sectional studies is to gather a representative sample. Ideally a random sample will be chosen, but this may be problematic (see section on sampling to follow). Cross-sectional studies may be more subject to volunteer bias by enrolling only those subjects who volunteer to participate, again limiting the ability to generalize the results.

For some studies, misclassification or measurement error can occur. In a survey, many individuals decline to provide information on income or personal health behaviors. Relying on information from medical records or laboratory testing can also result in incomplete and missing data. Testing may not have been standardized if different examiners and different sites collected data, especially in a retrospective study. If systematic differences between those who choose to participate and nonparticipants are related to the outcomes the investigators are attempting to measure, the results may be distorted.

Because cross-sectional studies are conducted at one point in time, they also have a problem with the temporal association between risk factors and outcomes. Which came first? Some authors describe this as a chicken-and-egg dilemma. Consider a study designed to evaluate potential associations between childhood obesity, physical activity levels, and myopia. Do early myopes who are uncorrected with reduced visual acuity feel uncomfortable playing sports because they can't see the baseball (soccer ball, basket, hockey puck, etc.) resulting in reduced physical activity and subsequent obesity? Do children who are obese struggle with physical activity and prefer more sedentary activities such as reading, watching television, or playing computer games? Does the added near point stress resulting from these activities induce myopia? Does excess caloric intake cause both obesity and ocular axial growth as an underlying cause of myopia? Cross-sectional studies do not allow the investigators to sort out the complexities of causative relationships, but only provide some evidence for possible associations. The strengths and weaknesses, potential sources of bias, and threats to validity for cross-sectional studies are summarized in Table 3.

Table 3. Characteristics of cross sectional studies.

Study Design: Cross-Sectional Study (observational, at one point in time)	
Brief Description: Subjects are selected and data are collected in a short period of time. Subjects are categorized based on current health status, or current presence of risk factors of interest.	
<p>Test Statistic and Interpretation: Prevalence, or estimated value using descriptive statistics (mean, median, proportion, standard error, standard deviation), 95% confidence interval</p> <ul style="list-style-type: none"> • Standard Error less than 10% of the estimate indicates a precise estimate • Standard Error 10% to 15% of the estimate indicates a fair estimate • Standard Error greater than 15% of the estimate indicates an imprecise estimate 	
<p>Design Strengths:</p> <ul style="list-style-type: none"> • Good for determining status quo of a disease or condition • Quick to complete • May be less expensive • Used to establish prevalence rates 	<p>Design Weaknesses:</p> <ul style="list-style-type: none"> • May have problem with response rate and measurement accuracy • Subject to sampling bias • Subject to volunteer bias
<p>Questions to Ask when Evaluating this Type of Study:</p> <p><i>Rule out Bias</i></p> <ol style="list-style-type: none"> 1. What are the characteristics of the population from which the sample was chosen? 2. Was an appropriate sampling method or selection of subjects used? 3. What is the response rate? 4. Are classifications, categories, and measurements accurate and valid? 5. If a survey was used are the questions unbiased? <p><i>Establish the Level of Precision</i></p> <ol style="list-style-type: none"> 1. Were confidence intervals reported for any estimated values? <p><i>Establish Clinical Significance</i></p> <ol style="list-style-type: none"> 1. Can the results of the study be generalized to <i>your</i> patients, e.g. are subjects similar to <i>your</i> patients with respect to age, gender, race/ethnicity, socio-economic status 	

APPLYING RESEARCH TO CLINICAL SCENARIO

CITATIONS

Lam BL, Lee DJ, Gomez-Marin O. Prevalence of usual-corrected binocular distance visual acuity impairment in Hispanic and non-Hispanic adults. *Ophthalmic Epidemiol.* 2000 Mar; 7(1):73-83.

Globe DR, Wu J, Varma R; Los Angeles Latino Eye Study Group. The impact of visual impairment on self-reported visual functioning in Latinos: The Los Angeles Latino Eye Study. *Ophthalmology.* 2004 Jun; 111(6):1141-9.

DESCRIPTION

In the study by Lam *et al* the investigators evaluate data from two large national cross-sectional studies, the Hispanic Health and Nutrition Examination Survey (HHANES) and the National Health and Nutrition Examination Augmentation Survey I (NHANES I-A). They compare the prevalence of visual acuity measurements that were found to be 20/50 or worse (with habitual correction) among various racial and ethnic groups.

Globe *et al* conducted a population-based cross-sectional study of 6,357 Latino adults aged 40 and over who reside within 6 specific census tracts in Los Angeles County. They measured the prevalence of presenting binocular visual acuity of 20/40 or worse and the prevalence of visual acuity worse than 20/40. They made comparisons based on age, gender, number of systemic comorbidities, depression, and self-reported visual function.

RELEVANT RESULTS

The results of several community-based screenings conducted by your Lions Club have raised some important questions. Are screening results expected to remain consistent across various populations? Do differences in the results obtained suggest a systematic measuring error, such as might occur with improper testing conditions? A brief review of the publications cited above provides the following relevant information:

- The analysis by Lam *et al* found the following prevalence of visual acuity levels less than 20/50 as stratified by race and ethnicity:
 - Cuban Americans 3.5%
 - Mexican Americans 4.6%
 - Puerto Ricans 6.6%
 - African Americans 7.7%
 - Non-Hispanic Whites 4.1%

- Data from the HHANES indicate that Puerto Rican adults had a significantly higher prevalence of visual impairment, even after adjusting for gender and age.

- Data from the NHANES I-A indicate that the difference in the difference in the prevalence rate of visual impairment was statistically significant when comparing African Americans and non-Hispanic Whites.

- Globe *et al* found that in the community studied the prevalence of visual acuity levels of 20/40 or worse was 6.3%.
- When the acuity level of 20/40 was excluded (reporting only acuity that was worse than 20/40) the prevalence of visual impairment was found to be 4.2%.

Clinical Scenario:

A 22 year-old Asian American female whom you fitted with contact lenses six months ago makes an emergency appointment with your office. She has just returned from an overseas vacation and is experiencing extremely red, painful eyes with an unusual discharge. Based on your examination you suspect a case of fungal keratitis, which is confirmed by lab culture. The patient and her mother are extremely concerned. How could the infection have been contracted? Is there any danger of passing on the infection to her younger siblings?

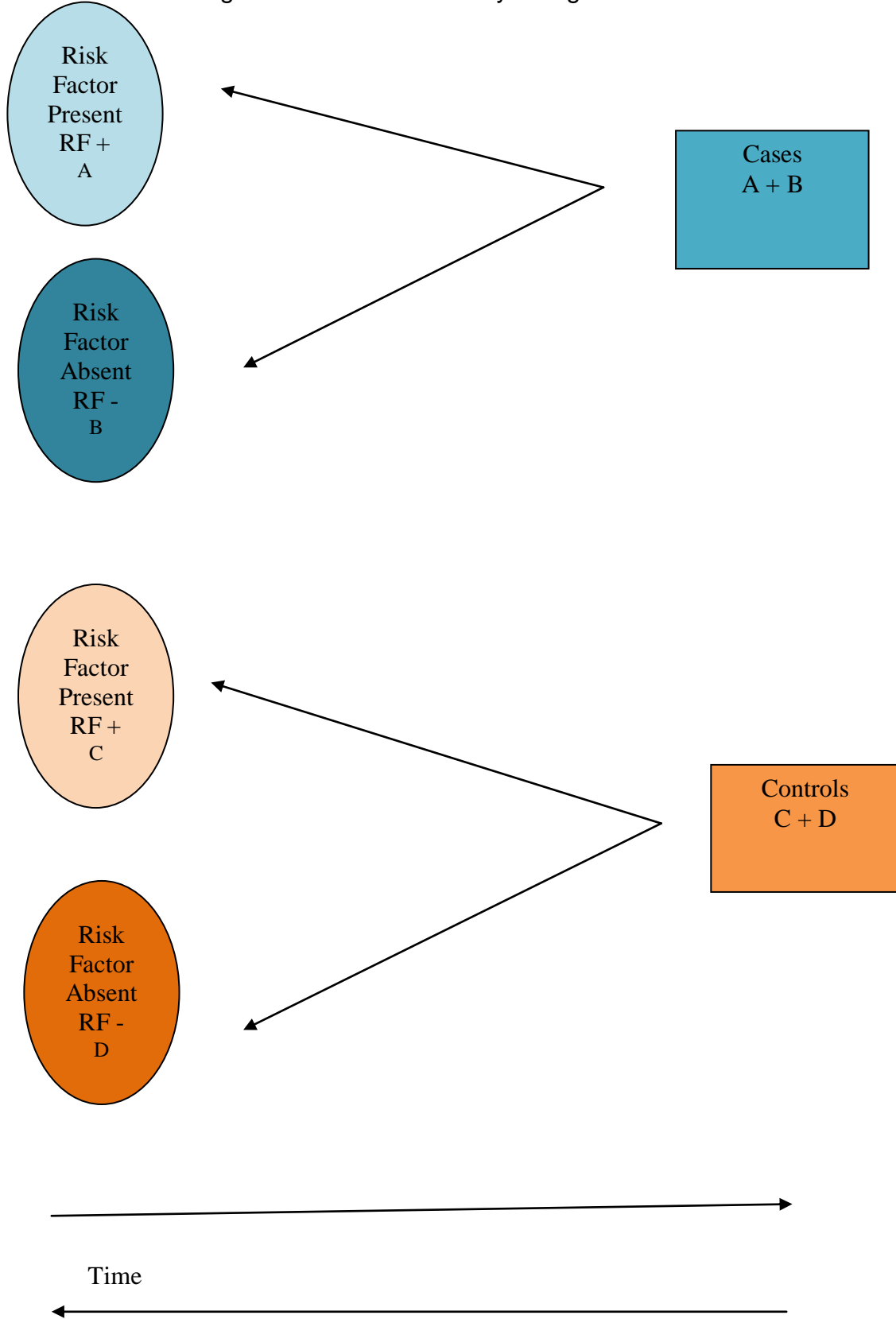
Case-control Study Design

With the advent of more effective contact lens disinfection systems and disposal lenses, contact lens-related eye infections have decreased dramatically in recent years. Fungal keratitis is an extremely rare condition, one which many practitioners can hope to never see in their office. How can you learn more about this unusual case? The case-control study design is ideally suited to studying rare outcomes.

To conduct a case-control study, individuals with the outcome of interest are first identified. Then, a comparison group is gathered, which forms the basis of the controls. There are a variety of methods for establishing the control group, but each method is designed to ensure that the cases and controls are comparable in many ways, with the exception of the outcome being studied. For example, if cases are all drawn from one hospital emergency room among patients who present with ocular injuries, controls might be drawn from the same emergency room among patients who have foot injuries. Cases and controls should be comparable on the basis of important demographic factors that may affect health status, such as age, gender, and ethnicity. It can also be important to match cases and controls on the basis of geographic location to avoid the introduction of possible biases related to environmental factors.

Once the cases and controls are identified, their past exposures to risk factors are then evaluated. This may be done using medical records, survey instruments, or sometimes even biological specimens. Case-control studies are always done retrospectively, because the outcome of interest is already present. Because the outcome is already present, the incidence rate (see description of incidence rate in the section on cohort studies) cannot be calculated. Case-control studies are also observational studies because the researchers do not manipulate an exposure or treatment, but rather observe naturally occurring disease states and past exposures. A schematic of the case-control study design is shown in Figure 4.

Figure 4. Schematic Diagram: Case-control Study Design



To compare the cases and controls, and to assess the strength of the association between the outcome of interest and the risk factors being studied, the proportion of subjects who have a history of a previous exposure to the risk factor is compared in each group, thereby generating the Odds Ratio. The Odds Ratio is the odds of having a past exposure *in cases* compared with the odds of having a past exposure *in controls*. The higher the calculated Odds Ratio, the stronger the relationship between the risk factor and the outcome. An Odds Ratio of 1.0 means there is exactly the same risk of outcome between those with and those without the exposure. An Odds Ratio of less than 1.0 indicates the risk factor may be protective. When assessing the strength of the relationship you should also consider not only the point estimate of Odds Ratio, but its 95% confidence interval as well. When expressing the difference between groups as a ratio of odds, we test whether the Odds Ratio is statistically different from a value of 1. If the 95% confidence interval does not include 1, we conclude that the Odds Ratio is significant at $p < 0.05$. The formula, calculations, and interpretation used for Odds Ratios are shown in Table 4.

Table 4. Analysis: Case-control Study

	Risk Factor Present RF+	Risk Factor Absent RF-	Totals
Cases	A	B	A+B
Controls	C	D	C+D
Totals	A+C	B+D	A+B+C+D

The probability that a case was previously exposed to the risk factor can be described using the following terms, all of which are synonymous

- Proportion exposed among those who have the condition (cases)
- Number who have the risk factor out of all those who have the condition
- Among persons who have the disease, the risk of the exposure to the risk factor

$$\text{Case exposure probability} = \frac{\text{Exposed Cases}}{\text{All Cases}} = \frac{A}{A+B}$$

The odds of exposure for cases represent the probability that a case was exposed divided by the probability that a case was not exposed.

$$\text{Odds of case exposure} = \frac{\text{Exposed Cases}}{\text{All Cases}} / \frac{\text{Unexposed Cases}}{\text{All Cases}}$$

$$\text{Odds of case exposure} = \frac{A}{A + B} / \frac{B}{A + B} = \frac{A}{B}$$

The probability that a control was previously exposed to the risk factor can be described using the following terms, all of which are synonymous

- Proportion exposed among those who *do not* have the condition (controls)
- Number who have the risk factor out of all those who *do not* have the condition
- Among persons who *do not* have the disease, the risk of the exposure to the risk factor

$$\text{Control exposure probability} = \frac{\text{Exposed Controls}}{\text{All Controls}} = \frac{C}{C+D}$$

The odds of exposure for cases represent the probability that a case was exposed divided by the probability that a case was not exposed.

$$\text{Odds of case exposure} = \frac{\text{Exposed Controls}}{\text{All Controls}} / \frac{\text{Unexposed Controls}}{\text{All Controls}}$$

$$\text{Odds of case exposure} = \frac{C}{C + D} / \frac{D}{C + D} = \frac{C}{D}$$

Compare the two groups using Odds Ratio

The odds of exposure for cases divided by the odds of exposure for controls represents the odds ratio.

$$\text{Odds Ratio} = \frac{\text{Odds of case exposure}}{\text{Odds of control exposure}}$$

$$\text{Odds Ratio} = \frac{A}{B} / \frac{C}{D} = \frac{A \times D}{B \times C}$$

Interpret the Value of the Odds Ratio

- If the cases and controls have the same rate of exposure to the risk factor, the Odds Ratio will equal 1.
- If the Odds Ratio is 1 the risk factor is not related to the outcome of disease.
- If cases have a *greater* rate of exposure to the risk factor, the Odds Ratio will be greater than 1.
- If the Odds Ratio is greater than 1 the risk factor is hazardous.

- If cases have a *lower* rate of exposure to the risk factor, the Odds Ratio will be less than 1.
- If the Odds Ratio is less than 1 the risk factor is beneficial or protective.
- The higher the Odds Ratio the greater the association between the risk factor and the disease outcome.

Calculate the 95% Confidence Interval for the Odds Ratio

$$95\% \text{ CI} = \text{OR} \pm [1.96 \sqrt{(1/A + 1/B + 1/C + 1/D)}]$$

Interpret the Range of the Confidence Interval

- If the range of the confidence interval contains a value of 1 the relationship is not statistically significant and may be due to chance.
- If the range of the confidence interval is narrow the estimate is more precise.
- If the range of the confidence interval is broad the estimate is less precise.

For example, a case-control study reports an Odds Ratio of 2.5. This means that the group exposed has a 2.5 times greater risk of developing the condition. The 95% confidence interval is calculated at 1.5 to 4.0. This means that if the study were repeated, we expect the Odds Ratio to fall between 1.5 and 4.0 in 95 out of 100 times the study is repeated. This is considered a statistically significant association. On the other hand, if the 95% confidence interval is 0.8 to 5.7, then there is not a statistically significant association because the Odds Ratio would be expected to be 1.0 or less in some number of repeated study conditions. Odds Ratios and 95% confidence intervals are commonly reported as OR = 2.5 [1.5,4.0].

As mentioned previously, case-control studies are well-suited for studying rare conditions. In comparison to a cohort study (see section following), investigators do not need to wait for the unlikely chance that a condition will develop. Studying a rare disease with a low incidence rate would require a very large sample size of non-diseased subjects to be followed forward in time to obtain a sufficient number of cases that would enable valid conclusions from the study. Case-control studies are also useful when there is a long lag time between the exposure to the risk factor and the development of the condition, requiring an extensive length of study when using a cohort design. Case-control studies can be quicker, more efficient, and less costly than cohort studies for these reasons.

The case-control design is also ideally suited for initial exploratory research or pilot studies. Many important discoveries might begin with an astute clinician's observation. A quick way to see if these clinical hunches lead to the discovery of more significant relationships is to make some quick comparisons with subjects who are readily available for investigation. When the risk factors are yet unknown, case-control studies can be used to quickly evaluate multiple risk factors in one study. The relative strengths of the risk factors can be compared individually and also in combination with each other. For example, diabetic patients with and without diabetic eye disease can be compared on factors such as diet, exercise, insulin usage, duration of disease,

frequency of blood glucose monitoring, and other potential risk factors all within the context of a single case-control study.

Case-control studies have some significant limitations, and are considered to be a weaker study design when compared with cohort studies. One of the weaknesses of the case-control design is the fact that cases selected for the study often do not represent the people with the disease in the general population. Cases are often selected from clinical settings, and can become so highly screened that they do not represent the larger population of those who have the condition. For example, if risk factors for fungal eye infections are being studied, cases might be selected from an emergency room and matched with controls who are the same age and live in the same community. One can argue that not everyone who has an eye infection will go to the emergency room; many will go to their own eye care provider. It may also be argued that those individuals who do seek care for their eye infections in the emergency room may be less likely to have a regular eye care provider and may therefore be less health conscious and more likely to engage in unhealthy or risky behaviors.

It is also important to ensure that cases are not selected based on their known risk factors or previous exposures. This can be difficult to do when the condition being studied is part of a disease outbreak or a highly publicized entity. If it is known that a study subject has a potential risk factor, and this knowledge prompts or suggests the diagnosis of the outcome being studied, then bias can be introduced into the study. For example, among patients who have a red eye, fungal keratitis might be more likely to be diagnosed if it is known that the patient wears contact lenses or was hit in the eye with a tree branch. Because of publicity implicating contact lens wear and use of certain solutions as risks for fungal keratitis, eye care providers might also be more likely to report an episode to the Centers for Disease Control and Prevention if the patient's history was positive for these two features.

Case-control studies are also subject to problems with recall bias. Because subjects are often interviewed to evaluate exposures to risk factors that happened in the past, their recollection is likely to be inaccurate. Recall bias occurs when there is a differential ability of subjects to remember their previous activities and past exposures. When an individual is diagnosed with a serious condition (case), she is more likely to search her memory in an effort to try to explain why she developed the condition. Control subjects without the condition may be less likely to remember an exposure because it has less meaning and less importance to them. Parents of children who are diagnosed with retinoblastoma may recall more types of pre-natal exposures or early risk factors than parents of healthy age-matched controls.

Reliance on existing records can also be a potential source of bias in case-control studies. Past medical records may be used to evaluate different risk factors, and records may come from a variety of examiners. Different standards for testing, measurement error, different interpretations, and misclassifications can all occur when previous records are used as sources of data. Records may be incomplete and not include all testing or risk factors of interest for the study, resulting in further loss of subjects in the study. Strengths and weaknesses, potential sources of bias, and threats to validity are summarized in Table 5.

Table 5. Characteristics of case-control studies.

Study Design: Case-control Study (observational, retrospective)	
Brief Description: Subjects are selected and categorized based on the outcome of interest (disease present /absent; condition mild / moderate / severe). Past history of risk factor status (exposed / unexposed; present / absent) is determined. Rates of exposure between those with and without the condition are compared.	
<p>Test Statistic and Interpretation: Odds Ratio, 95% confidence interval</p> <ul style="list-style-type: none"> • Odds Ratio less than 1 indicates a protective effect (or negative association) • Odds Ratio of 1 indicates no relationship • Odds Ratio of 2 indicates a weak relationship • Odds Ratio greater than 4 indicates a strong relationship • 95% Confidence Interval that contains a value of 1 indicates no statistically significant difference 	
<p>Design Strengths:</p> <ul style="list-style-type: none"> • Good for rare diseases • Good for chronic diseases • Useful for generating a hypothesis • Less expensive than other designs • May be completed more quickly than other designs • May need a smaller sample size 	<p>Design Weaknesses:</p> <ul style="list-style-type: none"> • Cannot directly estimate risk of disease • Can only study one outcome of interest • Rely on quality of existing records • Not good for rare risk factors • Problems with selection bias • Problems with recall bias • Information on risk factors may be less accurate • Cannot prove causality
<p>Questions to Ask when Evaluating this Type of Study:</p> <p><i>Rule out Bias</i></p> <ol style="list-style-type: none"> 1. Were all subjects taken from the same population? 2. Were subjects randomly selected from an eligible pool? 3. Were controls chosen from an appropriate group? 4. Were controls matched in some way to the cases, e.g. age, gender, location? 5. Were multiple controls chosen? 6. Were deceased cases included? 7. Were evaluators masked to the risk status of the subjects? 8. Were records reviewed by more than one person to ensure reliability of data? 9. Was information for risk factor status obtained in a similar manner for cases and controls? <p><i>Rule out Confounding</i></p> <ol style="list-style-type: none"> 1. Were the cases and controls similar at baseline on other important non-risk factors, e.g. by demographics or clinical characteristics? 	

2. Was the possibility of confounding controlled for in the study design, e.g. inclusion / exclusion criteria, stratification, matching, or in the analysis, e.g. logistic regression to control the effects of more than one variable?

Rule out Chance

1. Was a statistically significant difference found between the two groups, e.g. 95% confidence interval does not contain a value of 1; or statement that $p < 0.05$
2. Was the sample size sufficiently large to detect a difference, e.g. power of at least 80%?
3. Was the level of detectable difference between the two groups appropriate for clinical decision making?
4. Could bias or confounding make a real difference undetectable?

Establish a Causal Association

1. How strong is the association, e.g. the value of the Odds Ratio?
2. Are the results consistent with other published studies?
3. Do the results fit what is already known about the biological foundation or the clinical background?
4. Is the risk factor related to other outcomes that were not studied?
5. Is the outcome related to other risk factors that were not studied?

Establish Clinical Significance

1. What is the effect size as expressed as an absolute difference between the cases and controls?
2. Does the effect size of this magnitude influence your clinical decisions?
3. Are the subjects in the study similar to *your* patients, e.g. age, gender, race/ethnicity, socio-economic status, magnitude of exposure to risk factor, duration and level of disease process?

APPLYING RESEARCH TO CLINICAL SCENARIO

CITATIONS

Ma SK, So K, Chung PH, Tsang HF, Chuang SK. A multi-country outbreak of fungal keratitis associated with a brand of contact lens solution: the Hong Kong experience. *Int J Infect Dis.* 2008 Nov 17.

Saw SM, Ooi PL, Tan DT, *et al.* Risk factors for contact lens-related fusarium keratitis: a case-control study in Singapore. *Arch Ophthalmol.* 2007 May; 125(5):611-7.

DESCRIPTION

Both of these papers describe an analysis of outbreaks that took place in 2005 – 2006. Both papers utilized a case-control study design. Ma *et al.* defined cases as a disposable contact lens user with ophthalmologist-diagnosed keratitis and a positive culture of *Fusarium* spp during the specified time period. Controls were recruited from three outpatient clinics. Saw *et al.* defined cases as individuals diagnosed with *Fusarium* keratitis during the specified time period who were contact lens users. Two

groups of controls were gathered from the community and from a hospital-based sample, controlling for age and sex.

RELEVANT RESULTS

In the clinical scenario presented at the beginning of this section the patient and her mother had several questions: How could her eye infection have been contracted? Is there any danger of passing on the infection to her younger siblings? If the laboratory results confirm your suspected diagnosis, you may need ready access to information on this relatively rare condition. A brief review of the publications cited above provides the following relevant results:

- Saw *et al* found that after controlling for sex, age, and contact lens hygiene the use of ReNu with MoistureLoc significantly increased the risk of Fusarium keratitis; odds ratio 99.3; 95% confidence interval 18.4-535.4; $p < 0.001$
- The risk when using ReNu with MoistureLoc was 5 times higher compared with the risk when using ReNu MultiPlus; odds ratio 21.5; 95% confidence interval 4.0-115.5; $p < 0.001$
- Ma *et al* found that ReNu solution showed the strongest association with being a case; odds ratio 26.1; 95% confidence interval 3.0-225.3.

Clinical Scenario:

A 35 year-old African American male who is a new patient presents in your office and expresses a desire for refractive surgery correction. He states that he has been noticing deteriorating vision and increasing discomfort with his contact lenses. During the course of the examination you observe a scissors reflex during retinoscopy, high astigmatism, corneal thinning, and corneal irregularity. You make a tentative diagnosis of keratoconus. The patient is very concerned and would like to know more about this condition. What is the prognosis? What types of visual changes can be expected over time? Is he still a candidate for refractive surgery? If he has keratoconus, will continuing to wear his contact lenses increase or decrease the rate of progression?

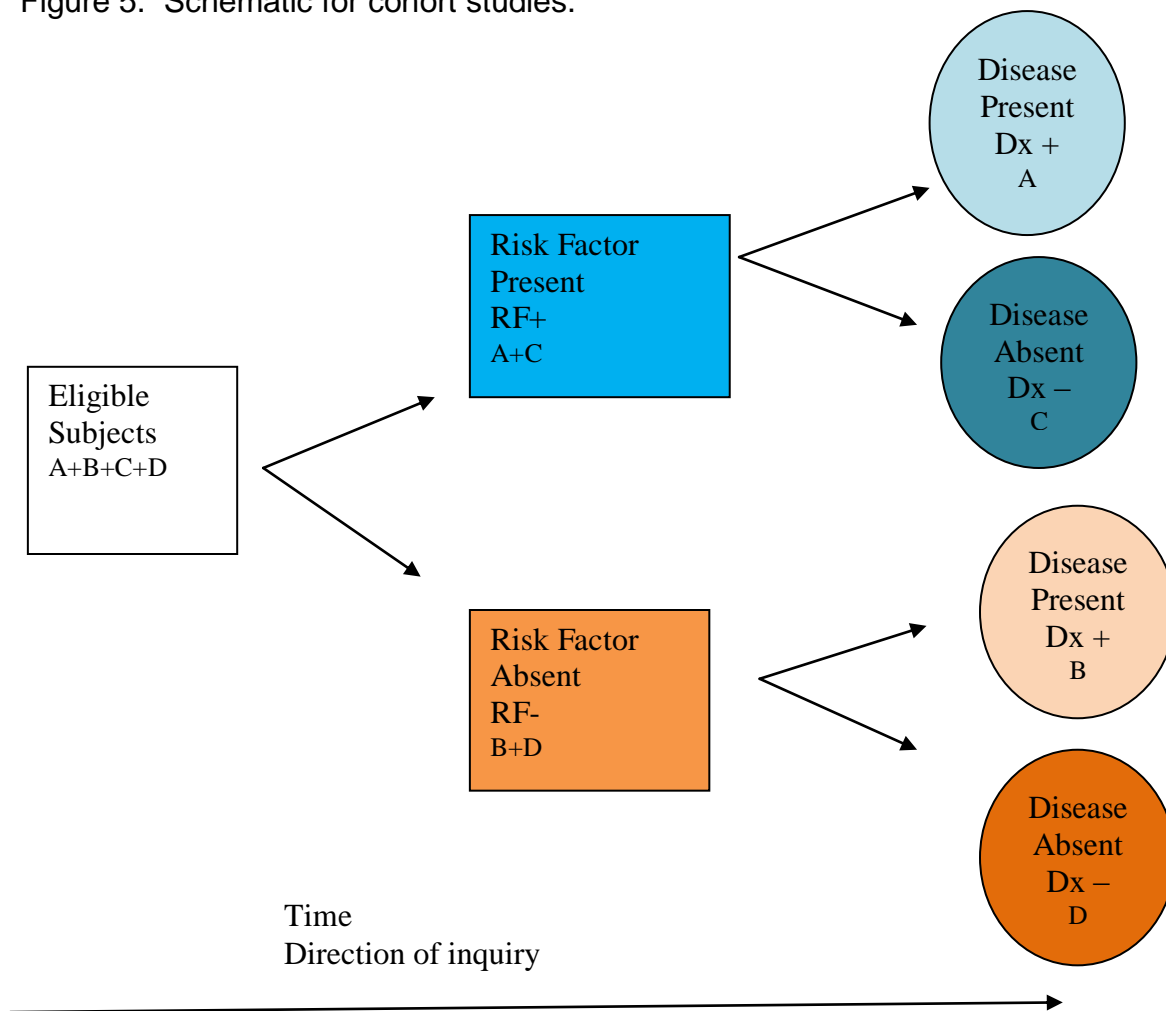
Cohort Study Design

Health care providers often find it helpful to understand the relationship between a risk factor and an outcome. In the clinical scenario described above, the outcome might be considered as decreased vision due to progression of keratoconus, and the risk factor could be described as use of contact lenses. In order to better understand the strength of the relationship and to consider the presence of a causal relationship, researchers make comparisons between two or more groups.

In the case of a cohort study, a sample of subjects is selected from a larger population and then assessed and assigned into two groups: those individuals who have a risk factor present (or sometimes described as having had a previous exposure to a risk factor) and those individuals who do not have the risk factor or exposure. The entire group is followed forward over time to determine each individual's health outcome. That is, whether the individual develops the disease or condition of interest, or whether he or she remains healthy and disease-free. To quickly summarize the

study design, the researchers select or define a sample of subjects. The research team measures characteristics in each subject (such as habits, behaviors, diet, environmental exposures, etc.) that might predict the subsequent outcomes (such as development of a disease or condition). The subjects are then followed with periodic measurements of the outcomes of interest. A cohort study is an observational study, meaning that the researchers do not manipulate an exposure or treatment, but rather observe naturally occurring risk factors and their outcomes. A schematic of the cohort study design is shown in Figure 5.

Figure 5. Schematic for cohort studies.



Cohort studies are usually prospective, when the risk factor groups are identified at the start of the study and followed forward over time to determine the outcome status of each subject. Occasionally a cohort study is retrospective (also known as historical). Retrospective cohort studies utilize information on prior exposure to a risk factor and subsequent disease status. The exposure to the risk factors and the subsequent development of the health outcome of interest have already occurred prior to beginning the retrospective cohort study design. The main advantage of this type of study design is that all of the events being evaluated have already taken place and the study may be completed more quickly than following a group forward, as is the case with prospective cohort studies. Because the retrospective cohort study relies on data gathered from past medical records or from subject recall, it is subject to bias from measurement error and inaccurate recollections.

It is important to note that in a cohort study information about the risk factor (exposure) is determined prior to the observation of disease. In both the prospective and retrospective cohort study design, all subjects must be free of the disease or health outcome that is to be studied at the start of the study period. This criterion is very important to allow the determination of the incidence rate of disease, which is defined as the rate at which new cases of a disease or condition develop within a given population at risk.

To compare the two groups, and to assess the strength of the association between the risk factor and the disease, the proportion of subjects who develop the disease (incidence rate) is compared in each group, thereby generating the Relative Risk. The Relative Risk is the incidence of disease in people who *have* the risk factor compared with the incidence in people who *do not have* the risk factor. The higher the calculated Relative Risk, the stronger the relationship between the risk factor and the outcome. When assessing the strength of the relationship you should also consider not only the point estimate of Relative Risk, but its 95% confidence interval as well. When expressing the difference between groups as a ratio of incidence rates (Relative Risk), we test whether the Relative Risk is different from a value of 1. If the 95% confidence interval does not include 1, we conclude that the Relative Risk is significant at $p < 0.05$. The formula, calculations and interpretation used for Relative Risk are shown in Table 6.

Table 6. Analysis: Cohort Study.

	Risk Factor Present RF+	Risk Factor Absent RF-	Totals
Disease Present Dx+	A	B	A+B
Disease Absent Dx-	C	D	C+D
Totals	A+C	B+D	A+B+C+D

The risk of disease, *with* exposure can be described using the following terms, all of which are synonymous

- Incidence rate of disease among those who have the risk factor
- Number who have the disease out of all those who have the risk factor
- Among persons who have the risk factor, the risk of the disease

$$R_{(RF+)} = A / (A+C)$$

The risk of disease, *without* exposure can be described using the following terms, all of which are synonymous

- Incidence rate of disease among those who do not have the risk factor
- Number who have the disease out of all those who do not have the risk factor
- Among persons who do not have the risk factor, the risk of the disease

$$R_{(RF-)} = B / (B+D)$$

Compare the two groups using Relative Risk

$$RR = R_{(RF+)} \div R_{(RF-)} = A / (A+C) \div B / (B+D)$$

Interpret the Value of the Relative Risk

- If the subjects *with* and *without* the risk factor have the same rate of disease, the Relative Risk will equal 1.
- If the Relative Risk is 1 the risk factor is not related to the outcome of disease.
- If subjects *with* the risk factor have a *greater* rate of disease, the Relative Risk will be greater than 1.
- If the Relative Risk is greater than 1 the risk factor is hazardous.
- If the subjects *with* the risk factor have a *lower* rate of disease, the Relative Risk will be less than 1.
- If the Relative Risk is less than 1 the risk factor is beneficial or protective.
- The higher the Relative Risk the greater the association between the risk factor and the disease outcome.

Calculate the 95% Confidence Interval for the Relative Risk

$$95\% \text{ CI} = RR \pm [1.96 \sqrt{ \frac{1-A/(A+C)}{A} + \frac{1-B/(B+D)}{B} }]$$

Interpret the Range of the Confidence Interval

- If the range of the confidence interval contains a value of 1 the relationship is not statistically significant and may be due to chance.
- If the range of the confidence interval is narrow the estimate is more precise.

- If the range of the confidence interval is broad the estimate is less precise.

Cohort studies are considered to be the best of the observational methodologies and they have several key strengths. Prospective cohort studies in particular allow researchers to control all of the testing and measurements. The methods for evaluating the risk factors and assessing health outcomes will be more consistent and reliable when they can be carefully defined in advance by the investigator. Measurements and assessments will be most accurate when carried out by a limited number of carefully trained study personnel. Cohort studies are one of the only ways available to estimate the incidence rate of disease in a population. Because multiple measurements can be made within one study to assess a number of different risk factors and a variety of outcomes, cohort studies are also useful when the investigator wants to study many outcomes and many possible risk factors. Recall bias (see section on case control studies) is not a problem, because subjects are not being asked to recall events of the past, but are instead followed forward in time to see which health outcomes develop. The time-order relationship between risk factor and disease is also made clear, because subjects are classified by risk factor before the disease or outcome develops.

Cohort studies also have their own unique weaknesses. Disadvantages to the study design include a longer time needed to conduct the study, a requirement for more financial and personnel resources, changes in subject characteristics, selection problems, surveillance bias, and loss to follow up.

Because subjects are enrolled prior to the development of the disease outcome being studied, researchers must watch and wait until the condition becomes manifest and clinically measurable. It may take years for the outcome to develop, and even for relatively common conditions, few cases may eventually occur. This can result in the need to enroll greater numbers of subjects, thereby increasing the costs of the study and the need for more resources, such as more personnel and more study equipment.

Because of the duration of the study required to assess newly developed conditions, subjects who are enrolled may change their habits or behaviors during the course of the study. For example, if researchers want to know how diet affects the development of macular degeneration, individuals in the study may change their dietary habits over the years due to changes in the culture, advertising, or awareness campaigns. For instance, campaigns to increase the consumption of green leafy vegetables to reduce the risk of cancer may also affect the risk of developing macular degeneration.

Cohort studies are also particularly susceptible to sample selection problems. A cohort may be drawn from clinic-based patients, from those who receive care at specialty referral centers, or from population-based residents. Each of these groups may result in studies with very different subject characteristics and therefore potentially varying rates of disease.

Surveillance bias can also be a problem for cohort studies, particularly when measurements are not done using masked examiners. At the outset of the study, the subjects who are enrolled are assessed to determine their different risk levels. The subjects who are considered to be at the highest risk levels may unintentionally be subject to more scrutiny and additional evaluations. The closer you look, the more you

find. Low risk subjects may also have some behaviors or exposures that could contribute to the development of the condition, but perhaps are not as closely observed. For example, if researchers were studying the development of hypertensive retinopathy in a group of healthy subjects, individuals who are at greater risk of hypertension (males, older, African American, obese, history of high cholesterol) may be more closely monitored.

The most significant weakness of cohort studies is the loss of valuable information due to subject drop out. Subjects move, fail to respond, miss appointments, decide they no longer wish to participate, or cannot be found. Dropouts reduce the numbers of subjects observed, but also because the reasons that they drop out may be related to the outcomes under study, dropouts thereby add another source of bias. Consider this hypothetical example of a cohort study evaluating the relationship between loss of vision, visual impairment, and depression. If more individuals who develop visual impairment are lost to follow up compared with the normally sighted group, it may be an indication that they are actually too depressed to continue engaging in the study process. The differential rate of depression between the two groups may remain hidden due to subject loss. Strengths and weaknesses, potential sources of bias, and threats to validity are summarized in Table 7.

Table 7. Characteristics of cohort studies.

Study Design: Cohort Study (observational, may be prospective or retrospective)	
Brief Description: Subjects are selected and categorized based on risk factor status (exposed / unexposed; present / absent) then followed to determine whether the outcome of interest occurs (disease present /absent; condition mild / moderate / severe). Rates of disease between the exposed and unexposed are compared.	
Test Statistic and Interpretation: Incidence Rate, Relative Risk, 95% confidence interval	
<ul style="list-style-type: none"> • Relative Risk less than 1 indicates a protective effect (or negative association) • Relative Risk of 1 indicates no relationship • Relative Risk of 2 indicates a weak relationship • Relative Risk greater than 4 indicates a strong relationship • 95% Confidence Interval that contains a value of 1 indicates no statistically significant difference 	
Design Strengths: <ul style="list-style-type: none"> • Direct calculation of risk and relative risk ratios • Able to measure incidence rates • Clear timing of exposure and outcomes • Good for rare exposures • Good for studying multiple exposures • Good for studying multiple outcomes • Consistent measurements and 	Design Weaknesses: <ul style="list-style-type: none"> • Time consuming to wait for outcomes to develop • Requires large staff and financial resources • Often requires a large sample size • Not good for rare diseases • Loss of subjects over time • Changes in diagnostic criteria, measurements, and treatments over time introduce variability and

classifications minimize bias <ul style="list-style-type: none"> • Considered to be the strongest observational design for establishing case and effect relationships 	bias
<p>Questions to Ask when Evaluating this Type of Study:</p> <p><i>Rule out Bias</i></p> <ol style="list-style-type: none"> 1. Were all subjects free of the condition at the beginning of the study period? 2. Were all subjects taken from the same pool? 3. Were evaluators masked to the risk factor status of the subjects? 4. Was the outcome assessed in a similar manner for all groups? 5. Was there differential attrition from the study groups? <p><i>Rule out Confounding</i></p> <ol style="list-style-type: none"> 1. Were the groups similar at baseline, e.g. by demographics or clinical characteristics? 2. Was the possibility of confounding controlled for in the study design, e.g. inclusion / exclusion criteria, stratification, matching, or in the analysis, e.g. regression analysis to control for interactions between risk factors? <p><i>Rule out Chance</i></p> <ol style="list-style-type: none"> 1. Was a statistically significant difference found between the two groups, e.g. 95% confidence interval does not contain a value of 1; or statement that $p < 0.05$ 2. Was the sample size sufficiently large to detect a difference, e.g. power of at least 80%? 3. Was the level of detectable difference between the two groups appropriate for clinical decision making? 4. Could bias or confounding make a real difference undetectable? <p><i>Establish a Causal Association</i></p> <ol style="list-style-type: none"> 1. How strong is the association, e.g. the value of the relative risk? 2. Are the results consistent with other published studies? 3. Do the results fit what is already known about the biological foundation or the clinical background? 4. Did exposure to the risk factor precede the development of the outcome? 5. Is the risk factor related to other outcomes that were not studied? 6. Is the outcome related to other risk factors that were not studied? <p><i>Establish Clinical Significance</i></p> <ol style="list-style-type: none"> 1. What is the effect size as expressed as an absolute difference between the comparison groups? 2. Does the effect size of this magnitude influence your clinical decisions? 3. Are the subjects in the study similar to <i>your</i> patients, e.g. age, gender, race/ethnicity, socio-economic status, magnitude of exposure to risk factor, duration and level of disease process? 	

APPLYING RESEARCH TO CLINICAL SCENARIO

CITATIONS

Wagner H, Barr JT, Zadnik K. Collaborative Longitudinal Evaluation of Keratoconus (CLEK) Study: methods and findings to date. *Cont Lens Anterior Eye*. 2007 Sep; 30(4):223-32.

Zadnik K, Barr JT, Steger-May K, Edrington TB, McMahon TT, Gordon MO; The Collaborative Longitudinal Evaluation of Keratoconus (CLEK) Study Group. Comparison of flat and steep rigid contact lens fitting methods in keratoconus. *Optom Vis Sci*. 2005 Dec;82(12):1041-21.

Barr JT, Wilson BS, Gordon MO, Rah MJ, Riley C, Kollbaum PS, Zadnik K; CLEK Study Group. Estimation of the incidence and factors predictive of corneal scarring in the Collaborative Longitudinal Evaluation of Keratoconus (CLEK) Study. *Cornea*. 2006 Jan;25(1):16-25.

DESCRIPTION

The Collaborative Longitudinal Evaluation of Keratoconus (CLEK) Study was an 8-year, multi-center prospective cohort study that enrolled 1,209 patients with keratoconus and followed them forward in time to observe the natural history of the disease process.

RELEVANT RESULTS

Several questions were raised in the clinical scenario presented at the beginning of this section: What is this patient's prognosis? What types of visual changes can be expected over time? Is he still a candidate for refractive surgery? If he has keratoconus, will continuing to wear his contact lenses slow the rate of progression?

The CLEK Study added a great deal to our knowledge about keratoconus and provided us with many excellent peer-reviewed publications. A brief review of the publications cited above provides the following relevant results:

- Subjects exhibited a 7-year decrease in both high and low contrast, best-corrected visual acuity.
- 19% demonstrated a decrease of 10 or more letters in high-contrast, best-corrected acuity and 31% demonstrated decreases of 10 or more letters in low-contrast best-corrected acuity in at least one eye.
- The 8-year incidence of corneal scarring was 20%, with younger age (and several other factors) predictive.
- Initial results suggested that contact lens wear increased the risk of corneal scarring by more than 2-fold.
- Further analysis found that keratoconic eyes fitted with a rigid contact lens using a method that resulted in an apical touch fluorescein pattern did not have an increased risk of scarring.

Clinical Scenario:

You have been providing eye care for a 75 year-old Caucasian female for the past 10 years. During her annual dilated fundus examination you notice an increase in the number of drusen and some mottling of the macular pigment. When you share your findings your patient asks whether anything can be done to improve her eye health. She remembers that her mother went blind later in life and she tells you that her sister has been “bugging” her to take vitamins. Can vitamins really help your eyes? If so, which ones are best to take?

Randomized Controlled Trial Study Design

Randomized controlled trials are considered to be the gold standard in health care research. Randomized controlled trials are experimental studies and are considered the only way to truly establish a causal relationship. To conduct a randomized controlled trial, investigators gather a sample of the population then assign subjects to treatment groups rather than simply making observations about the risk factor status of subjects. The researchers randomly assign the subjects to two (or more) treatment groups. Subjects are then followed in each group over time to determine outcome (or disease) status. Treatment groups may make comparisons between currently accepted therapies and new therapies, or in some cases one group serves as a control through the use of a placebo or sham procedure. A schematic of the randomized controlled trial study design is shown in Figure 6.

The results of the randomized controlled trial are assessed using appropriate techniques based on the type of data gathered (e.g. nominal, ordinal, numerical or categorical) and the appropriate statistical tests (see section on statistics). In general, the investigators are charged with determining whether any differences in the outcomes or health status of the two groups are solely due to chance or whether a statistically significant difference is demonstrated. A brief description is provided in Table 8.

Randomized controlled trials are considered to be the study design with the least number of problems or biases. Randomized controlled trials are considered to provide the strongest evidence and are the best type of study to use to establish the efficacy of a treatment or procedure. They allow the investigator to control how the treatment is administered to ensure consistency and maximum treatment effectiveness and also allow for the uniform collection of data by a skilled research team.

The randomization used to allocate subjects into treatment groups is a crucial step because it avoids confounding bias that might occur if treatment were decided by the patient or his health care provider. It also maximizes the likelihood that the treatment groups will have similar baseline characteristics on other nontreatment factors such as age, gender, and the level of disease present, as well as unknown associated factors. This makes sense because the group is selected first, then divided randomly, so the only differences have to occur by chance. Randomization also satisfies the sampling assumptions of statistical testing and estimation concerning random selection of subjects thereby strengthening the statistical analysis of the study results.

Figure 6. Schematic Diagram: Randomized Controlled Trial Study Design

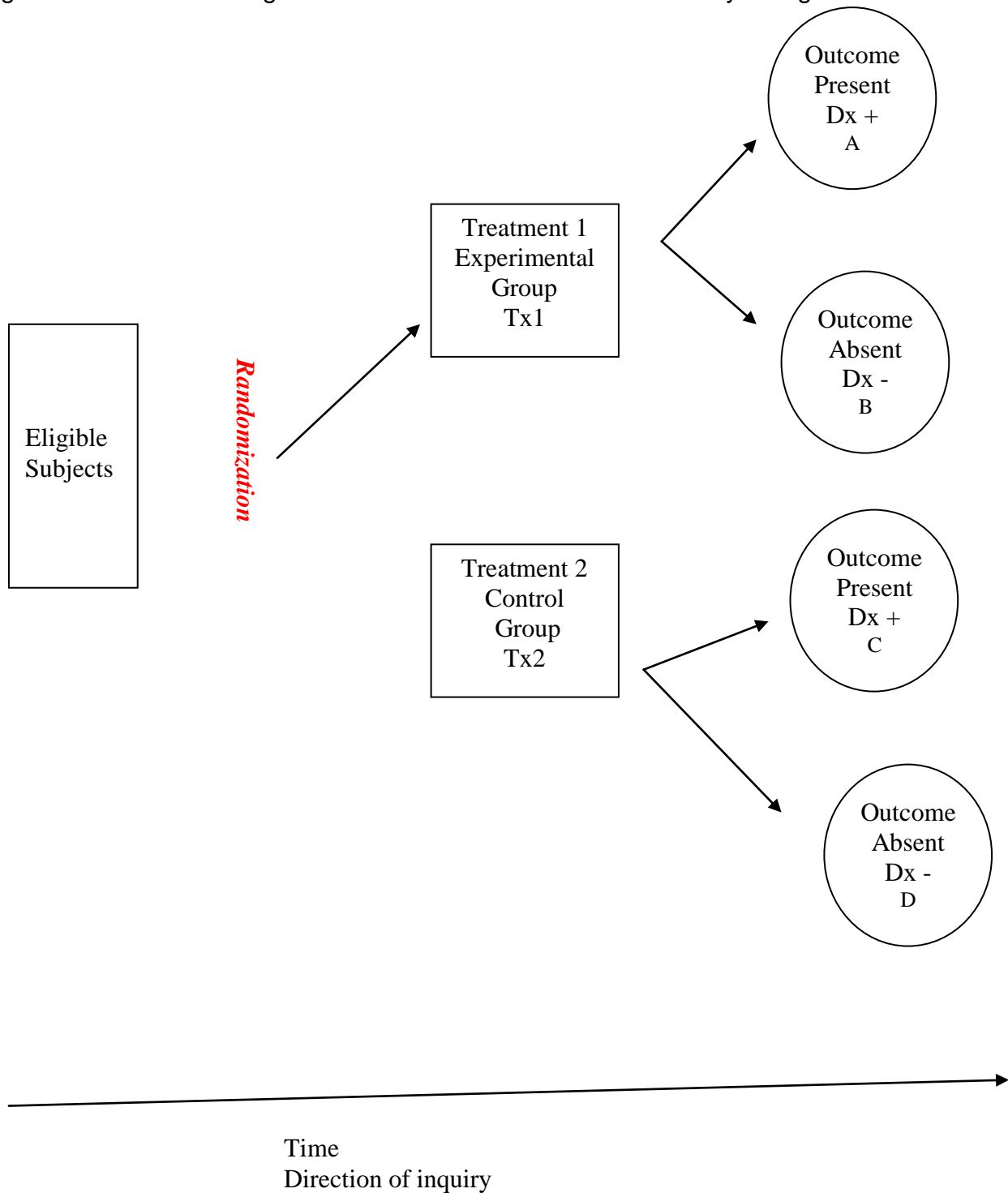


Table 8. Analysis: Randomized Controlled Trial Study

A variety of statistical tests can be used to evaluate the results of a Randomized Controlled Trial. The selection of the test is dependent upon the variables being studied. Some examples include, but are not limited to, those shown in the table below.

Variable of Interest	Statistical Test	Interpretation
Incidence rate of disease in the experimental group versus incidence of disease in the control group	Relative Risk	See section describing cohort studies
Mean value of a variable comparing the experimental and control groups	t-test, ANOVA, or Confidence Intervals	Evaluate the absolute difference in means; check the p-value attained; check the width and overlap of confidence intervals
Difference in proportions (percentage with and without the variable of interest) comparing the experimental and control groups	Chi-square test or Confidence Intervals	Evaluate the absolute difference in proportions; check the p-value attained; check the width and overlap of confidence intervals

Randomized controlled trials are often very expensive and take a long time to complete. Expenses can increase due to the large numbers of subjects that are usually required and due to the tight controls that must be maintained to ensure good quality data. Because time is required for outcomes of interest to develop, subjects must be followed forward in time for a sufficient period, thereby delaying the results of the study.

Because subjects are being asked to undergo treatments or procedures, randomized controlled trials rely heavily on volunteers and recruitment of subjects and do not typically utilize any type of population sampling. Randomized controlled trials often have very strict inclusion and exclusion criteria for their subjects. This is done to further limit the potential for bias and confounding and to ensure more uniform responses to the treatment being evaluated. Because of these subject exclusions, the results of the study may not apply to the general population or to your patients.

Two other potential problems are related to the treatment aspect of the trials. First, it is often difficult for randomized controlled trials to evaluate standard and accepted therapies, even if the therapy was never fully evaluated since its inception. Accepted therapies become the “usual treatment” by endorsement and it may be considered unethical to withhold the therapy or to substitute an alternative. Secondly, subject behaviors may influence the treatment. Even with rigorous controls in place, subjects may either reject their treatment group assignment or they may not actually comply with the prescribed regimen or therapy. Sophisticated trials use methods to limit these potential problems by masking subjects to their assigned group so they do not know whether they are taking the new drug, the standard drug, or a placebo and by

utilizing techniques such as directly observed therapies and dosage meters. The strengths and weaknesses, potential sources of bias, and threats to validity for randomized controlled trials are summarized in Table 9.

Table 9. Characteristics of randomized controlled trial.

Study Design: Randomized Controlled Trial (experimental, prospective)	
Brief Description: Subjects are enrolled and randomly assigned to treatment groups then followed to determine whether the outcome of interest occurs (disease present /absent; condition mild / moderate / severe). Rates (of disease, side effects, outcomes, etc.) between the treatment groups are compared.	
Test Statistic and Interpretation: Test statistics are dependent upon the variables measured and may include ANOVA, t-test, correlations, regression analysis and others.	
Design Strengths: <ul style="list-style-type: none"> • Considered the gold standard of study designs • Greatest justification for establishing causality • Good for studying efficacy of treatment or procedure • Measurements can be controlled for high validity and reliability • Assumptions of statistical tests tend to be met 	Design Weaknesses: <ul style="list-style-type: none"> • Expensive • Potentially long duration • Difficulty evaluating accepted therapies • Subject exclusions may limit ability to generalize the findings • Large number of subjects usually required • Subjects may not comply with treatment assignments
Questions to Ask when Evaluating this Type of Study: <i>Rule out Bias</i> <ol style="list-style-type: none"> 1. Were all subjects taken from the same pool? 2. How were subjects recruited? 3. Was appropriate randomization used to assign subjects to treatment groups? 4. Were evaluators masked to the group assignment? 5. Were subjects masked to the group assignment? 6. Was information on the outcomes obtained in a similar manner for all groups? 7. Was there a differential attrition from the study groups? 8. Are all study subjects who were assigned to the treatment groups accounted for? <i>Rule out Confounding</i> <ol style="list-style-type: none"> 1. Are appropriate therapies included? 2. If indicated, is a placebo (or sham treatment) used? Were examiners masked? 3. How is compliance with the treatment assured and evaluated? 4. Were treatment groups similar at baseline, e.g. by demographics or clinical characteristics? 5. Was the possibility of confounding controlled for in the study design, e.g. 	

inclusion / exclusion criteria, stratification, matching, or in the analysis, e.g. regression analysis to control for interactions between risk factors?

Rule out Chance

1. Was a statistically significant difference found between the two groups, e.g. statement that $p < 0.05$
2. Was the sample size sufficiently large to detect a difference, e.g. power of at least 80%?
3. Was the level of detectable difference between the two groups appropriate for clinical decision making?
4. Could bias or confounding make a real difference undetectable?

Establish a Causal Association

1. How strong is the association, e.g. the strength of the correlation?
2. Are the results consistent with other published studies?
3. Do the results fit what is already known about the biological foundation or the clinical background?
4. Is the treatment related to other outcomes that were not studied?
5. Is the outcome related to other treatments that were not studied?

Establish Clinical Significance

1. What is the effect size as expressed as an absolute difference between the treatment groups?
2. Does the effect size of this magnitude influence your clinical decisions?
3. Are the subjects in the study similar to *your* patients, e.g. age, gender, race/ethnicity, socio-economic status, magnitude of exposure to risk factor, duration and level of disease process?

APPLYING RESEARCH TO CLINICAL SCENARIO

CITATION

Richer S, Stiles W, Statkute L, Pulido J, *et al.* Double-masked, placebo-controlled, randomized trial of lutein and antioxidant supplementation in the intervention of atrophic age-related macular degeneration: the Veterans LAST study (Lutein Antioxidant Supplementation Trial). *Optometry* 2004 Apr; 75(4):216-30.

DESCRIPTION

The investigators conducted a 12-month prospective randomized double-masked, placebo-controlled trial in a Department of Veterans Affairs hospital in the Midwestern region of the United States. A total of 90 patients with atrophic macular degeneration were randomly assigned into three groups. Subjects in group 1 received 10 mg of lutein; subjects in group 2 received lutein, antioxidants, vitamins, and minerals combined in a broad-spectrum supplement; subjects in group 3 received a placebo.

The researchers evaluated visual acuity, contrast sensitivity, and macular pigment and compared the three groups.

RELEVANT RESULTS

Television, newspapers, magazines, and the popular literature have been an abundant source of information about vitamins in recent years. But how can you separate the hype from reality? Your patient wants to know if vitamins would be helpful to her, and which vitamins might be best for her. A brief review of the publication cited above provides the following relevant information:

- Subjects enrolled in group 1 showed an increase in the mean eye macular pigment optical density, an average improvement in Snellen visual acuity of 5.4 letters, an improvement in contrast sensitivity, and a subjective improvement in the Amsler grid findings.
- Subjects enrolled in group 2 showed an increase in the mean eye macular pigment optical density, an average improvement in Snellen visual acuity of 3.5 letters, and an improvement in contrast sensitivity. There was a trend towards improvement in self-reported visual function, but the change was not statistically significant.
- Subjects enrolled in group 3 showed no difference in any of the measured findings.

Getting Started in Research

Anyone, in any setting, can engage in research if they wish to. At a minimum, practitioners must be regular consumers of current research by virtue of their own self-study through reading quality publications or attending high level continuing education courses. Practitioners owe it to their patients to maintain their knowledge base and to expand their knowledge to include the latest research developments. Some individuals may also choose to participate in research as data collectors, research subjects, or even the originators of research ideas.

While this chapter may not fully prepare you to engage in your first research project, it can definitely help you become a more informed consumer, and it can give you some ideas to get you started. Suppose you decide that you want to conduct a study. Where should you start? Often the first step is to define a specific research question.

Where do research questions come from? Sometimes they can come from a clinical observation. Good questions can also come from other published studies. For example, most papers will identify weaknesses of the study and areas for further research. Ideas can come from an understanding of the biological underpinnings. Ideas can come from novel approaches to using existing instrumentation. Ideas for research can come from a healthy skepticism about what “everybody knows”.

As noted at the beginning of the chapter, a researchable question is not just any question. In order to be a question worthy of research, it must have the potential to produce hard facts and evidence, to add to a theory or understanding, or to improve the clinical practice of optometry. A good research question will provide answers that explain, describe, define, identify, substantiate, predict, contradict, or qualify.

Many experts suggest applying what is called the “FINER” test to potential research questions. This is an acronym for the attributes Feasible, Interesting, Novel, Ethical, and Relevant. Feasibility relates to your capabilities to recruit and enroll an adequate number of subjects; whether you and your team have the required technical expertise; whether the project is affordable in terms of time and money; and whether the question is manageable in scope and not too broad. The research question not only needs to be of interest to the investigator in order to maintain a high level of enthusiasm, it must also be of interest (and thereby beneficial to) the larger potential audience, for example practicing optometrists, optometric students, optometric educators, patients of optometrists, etc. A good research question should be novel in its approach and should have the potential to either confirm or refute previous findings, enhance and extend previous findings with more information, or to provide new findings. Duplicating previous research questions does little to enhance our body of knowledge. Research must always be ethical and adhere to ethical standards and guidelines. And finally, relevance is of the utmost importance. To be of value, a research question must have relevance to scientific knowledge, clinical practice, health policy, educational programs, health outcomes, or future research directions. Research questions usually start with “what” or “why”. A research question must be clear. Lack of clarity yields unusable results. The research question sets the foundation for the study design and guides you through the whole research process.

Statistics

Quantifying the Research Question

To set the stage for the following discussion, a review of biostatistics vocabulary is appropriate. The basis of testing arises from the null hypothesis. The null hypothesis is what the statistical test we are using is actually testing in order to determine if we accept the results of the study. It assumes that there is no relationship between the independent variable(s) (or predictors) under study and the dependent variable (or outcome) in question. If we accept the null hypothesis, we have determined that no association exists in our sample between the independent and dependent variable. If we reject the null hypothesis, then we have accepted the alternative hypothesis. Like the null, we specify the alternative hypothesis before we begin the study. It is the hypothesis we believe to be true of the data when our statistical testing shows that a difference exists. We base the alternative hypothesis on physiologically plausible expectations. Determining which of the hypotheses are true requires evaluating the statistical significance of the test that has been performed on the data. Statistical significance relates to the probability that the result obtained in our study or a more extreme result would be expected given the data at hand. This is identified by the p-value.

When specifying the alternative hypothesis, we are indicating what we believe the difference will be. This can be a non-directional, or two-tailed hypothesis, or a directional or one-tailed hypothesis. Two-tailed means that the parameter estimate may lie on either side of what you specified in your null hypothesis, while a one-sided

hypothesis specifies whether the parameter is larger or smaller than your null value. The use of the one-sided hypothesis garners a lot of attention because it makes the null hypothesis easier to reject. Use of the two-sided hypothesis is preferred as it makes no assumptions about what the outcome might be. One-tailed hypotheses are found in instances where a certain set of values may be impossible or where the investigator has no interest in the conclusion if the result falls on the other side of the expected parameter.

Let's use the example of cholesterol and glaucoma from earlier in the chapter to illustrate these concepts.

Research Question: Is an increased level of low-density lipoprotein cholesterol (LDL) associated with an increased risk of glaucoma?

Testable Hypothesis: Does a low-density lipoprotein cholesterol level (LDL) above 150 mg/dL increase the risk of glaucoma by 50%?

Null Hypothesis: There is no relationship between LDL and the incidence rate of glaucoma.

One-tailed Alternative Hypothesis: An increased level of LDL above 150 mg/dL is associated with an increased incidence rate of glaucoma.

Two-tailed Alternative Hypothesis: Variation in the level of LDL is associated with variation in the incidence rate of glaucoma.

Deciding whether to use a one or two-tailed alternative hypothesis depends on whether we think increased LDL increases glaucoma risk or whether we really have no idea if it increases or decreases risk of glaucoma. Unless there is previous evidence or a very strong biologic plausibility that increased LDL increases the likelihood of glaucoma, then the two-tailed alternative hypothesis is to be favored.

When trying to determine whether we think the results of a study are meaningful, there are some key concepts to consider when evaluating the research results that the authors may present as the "answer." No result is fool proof, and it is not possible for an author to conclude they have definitively proven a hypothesis true or false. Any answer that comes from a study is subject to error. Here we will discuss, briefly, what those errors are.

Type I error results when there is a difference that is declared to be statistically significant when, in fact, no difference actually exists. Said in terms of hypothesis testing, it is rejecting a null hypothesis that should have been accepted. This is also known as an alpha error (α), or the significant level of a test.

Type II error accepts the null hypothesis when it should be rejected. That is, the research finds no difference when in reality there is actually a true difference. Also called a beta error (β), it is directly related to the power of a test as we will see shortly. As with many concepts in epidemiology, study errors can be summarized neatly in a two by two table as shown below.

Table 10. Types of study errors.

	Truth A difference exists	Truth No difference exists
Study finds a difference	Correct conclusion	Type I error
Study finds no difference	Type II error	Correct conclusion

The type I and type II errors work in opposition to one another. As we attempt to decrease the type I error, decreasing the chance of a false positive, we correspondingly increase the type II error. This works the other way as well, as we try to minimize the frequency of reporting that no differences exists when in fact it does, we increase the type I error.

The power of a study is calculated as $1 - \beta$. Power represents the ability of a study to find a difference, should one actually exist. It is determined based on a specific set of assumptions, which includes the desired type I error level, the sample size of the study, and the size of effect expected. The size of the effect requires deciding what constitutes a clinically meaningful difference between the groups under question. That is, how much difference can there be between two groups before we say that we consider the difference important. These things working together contribute to the power of the study.

CLINICAL PEARL – Non-significance ≠ no difference

It is important to remember that just because a research study finds no statistically significant difference, that does not mean that no difference exists. The difference may be too small to detect, there may be too few subjects enrolled in the study to find a difference, or the lack of a difference may be due to measurement error or bias.

Who?

The group of people with the characteristic or disease of interest is the population you seek to study. Logistically, however, it is impossible to study every individual in a population. This requires selecting a smaller group of people, a sample, in order to estimate what the values would be in the population of interest.

Investigators who take steps in the design and conduct of a study to reduce the sources of bias end up with study results that are a better estimate of the measure in question. These biases, or errors, that move the observed value away from the true value in the sample affect the internal validity of a study. External validity implies that the measurements of the study represent the actual target population values, after accounting for random error. The list of biases is quite longⁱⁱⁱ, and definitions can overlap, but there are a few more common ones to look for. These sources of bias were also described as potential problems based on the study design selected (see previous section).

Selection bias results from the way in which subjects entered the sample, which may make them different from the target population (see also the sections on study design). This can result in a biased estimate, where the sample differs from the target

population and affects the conclusions drawn. A volunteer sample is one example of this. There are factors that may influence whether a person is a participant, including personality, logistics (distance, transportation), the use of a particular source to recruit (i.e. a hospital, which requires a certain level of sickness for entrance and an ability to pay). If the sample is selected in a non-random way, then sampling bias exists. Conclusions resulting can vary from the true value due to certain subjects being more likely to be included than others. An example of this may be a study that looked at intraocular pressure (IOP) solely in the practice of a glaucoma specialist as a measure of what IOP is in those subjects with ocular hypertension. It is likely that the patients in the specialist's office have particularly elevated IOP levels and thus were referred to the specialist for follow-up. As a result, these subjects may have a higher IOP than a random sample of all individuals with ocular hypertension might suffer.

Information bias (resulting from misclassification or measurement bias) affects internal validity. Internal validity is when the measurements do not accurately reflect the sample. One example, recall bias, is particularly common when people are asked to provide information from their past. The accuracy of the information as well as the fullness of the details provided can be affected by time or intervening events. Someone with a disease may attempt to remember every detail in order to try and explain what has happened to them. As described previously, recall bias can be particularly problematic in case-control studies (see section on case-control strengths and weaknesses).

In situations where study personnel are involved in collecting information from subjects, interviewer bias is a possibility. The way questions are posed may be based upon interviewer assumptions, and can affect the way in which subjects respond to the question. This is particularly germane to studies that use survey instruments that are verbally administered.

Confounding bias, often referred to simply as confounding, is when groups being tested for differences have some underlying difference which is not controlled. In the example at the beginning of the chapter, infants who slept with nightlights on had higher incidence of myopia, but this may have been due to confounding variables such as parental myopia, socioeconomic status, or environmental differences between the groups with and without night lights. Randomization controls for confounding because known and unknown confounding variables are randomly distributed between the groups.

Bias in clinical studies can be hard to detect, in part because accounting for all potential sources is difficult to achieve. Being able to identify possible sources in studies can help provide an assessment about how much caution should be applied to interpreting the results from the sample.

Testing Overview

When approaching statistical testing, you need to consider what your outcome(s) of interest looks like in order to determine the appropriate analysis. Here we will briefly discuss the types of variables that occur in scientific research and the basics of the test statistics that are derived from the data.

Statistical analysis, and selection of the appropriate statistical test, are based on the form of the measure of interest. Two examples of types of variables are “continuous variables” and “categorical variables”. The most flexible variable, i.e. the one suitable for the widest range of analyses, is the continuous variable (also called interval-scaled variable). These are the outcomes whose responses can fall anywhere along a continuum and which have equidistant spacing between units. Blood pressure is a frequently used example of a continuous variable, where the range is, in theory, unlimited and the difference between units on any part of the scale is the same. Categorical variables cover a variety of variables which are grouped based according to some characteristic. The categories can be nominal, where the categories are just group names (like gender), or ordinal, where there is some ordered progression from one category to the next, but the differences between the categories are not quantitative or evenly spaced (for example, a clinical scale to describe corneal staining ranging from 0 to 4).

When the data are analyzed, the calculations produce a test statistic which, while specific to the analysis, has the same function across analyses. The test statistic has a distribution that is associated with the probability that this test statistic, and therefore the hypothesis being tested by the data, could have occurred if the null hypothesis were true. This probability is known as a p-value. The p-value is what is used to assess the statistical significance associated with a test. Historically, researchers have looked at a p-value of 0.05 as a rejection of the null hypothesis, i.e. a statistically significant result. Practically speaking, this means that when you have a p-value that is 0.05, you would expect to see a result more extreme than this result only 5 times out of 100 samples.

When assessing the significance of a result, judging only by a p-value can be misleading. P-values are affected by the sample size of a study, so that a large study is able to detect smaller differences and therefore may have statistically significant results due to this fact. Looking at the clinical relevance of the reported results is an important companion to assessing statistical significance. It is possible, even in samples that are not extremely large, to have statistical significance, but a result that is not clinically meaningful. If you look at the results and see that the difference between groups is not one that you find important, then that is a gauge of how relevant the finding is. Well-designed studies should indicate what kind of effect sizes are considered important. If not, determine for yourself what difference would be clinically meaningful for your patients. For example, a study finds a new anti-hypertensive drug lowers IOP by 1 mmHg. This result is statistically significant. Do you consider this effect size of 1 mmHg meaningful enough that you would consider administering this as a treatment for your patients with ocular hypertension?

The effect size, or parameter obtained from a study, is an estimate of what the true value in the sample is. Associated with this estimate is the variability. Variability is measurement of the amount of spread, or the variation, in the data. In order to express the interval where the parameter estimate will most likely fall, a 95% confidence interval is constructed based on the variability and the sample size. The 95% confidence interval means that in repeated samples of the same size, the parameter value will fall in the interval for 95% of the samples. Because the variability is included in the calculation, the width of the confidence interval is an indication of how variable the sample was. If the sample was drawn from a population that was homogeneous for the

characteristics under study, one would expect the range of values to be narrow, indicating little variability. Therefore the confidence interval around the mean drawn from this population will be tighter.

We can think about two different scenarios. The first would be estimating refractive error in a group of first grade children, all of the same racial/ethnic background. We would expect a tight range spherical equivalent refractive error, so the variability of the estimate for the mean should be small. The accompanying 95% confidence interval will be narrow as a result. Now let's consider refractive error from a different sample. Now we draw our sample from a population whose ages ranged from 6 years old to 70 years old and that had a greater racial/ethnic diversity. We would expect to get a wide range of refractive errors from this group, leading to greater variability. When we calculate the 95% confidence interval, the bounds would be wider because it is more difficult to give a precise estimate of where the mean will actually fall.

One can also conclude significance from looking at the interval, though the interpretation depends on the analysis. In an analysis of means, we are looking to see if there is a difference between the means, i.e. is the difference equal to zero or not. Therefore, we look to see if the 95% CI includes a value of 0. If it does not, then the difference is statistically significant. If we have an analysis of odds ratios, for example, we look for 1 to be in the interval (no difference if ratio is one) to conclude no significant difference (see examples described in Tables 4 and 6).

Measures of Central Tendency

One of the basic descriptive statistics used to present information about the sample is the mean, as shown in Table 11. The mean, or average, is used as a way to summarize continuous data and is simply the sum of the observations divided by the number of observations. Most often you will see the mean presented with the standard deviation (sd), which is the square root of the variance (Table 11). This represents the distribution of data around the mean. The larger the standard deviation/variance, the more widely spread the observations are (i.e. the wider the distribution of the numbers is) around the mean.

The normal distribution curve, more commonly identified as the bell curve, is a probability curve whose x-axis is made of values of the variable under question, and whose y-axis represents the probability of a given value occurring, shown in Figure 7. The center of this distribution is the mean (identified in Figure 7 by 0 on the x-axis at the peak of the curve), and the curve is symmetric on each side of the mean. Each standard deviation can be identified from each side of the mean. For example -1 and +1 on the x-axis is one standard deviation from the mean in each direction. In the normal distribution, the mean is also the mode and median (as can be seen from the symmetry) (Table 11).

Table 11. Useful terms for describing data

Term	Definition	Calculation*
Mean (arithmetic)	The average of a group of data points	$\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$
Median	The value where 50% of the data points fall above and 50% of the data points fall below	In an odd number of data points, the median is the middle observation. In an even number of data points, the middle two points are averaged.
Mode	The values that occur most frequently	Count the frequency of the different values and the most frequently occurring is the mode. There can be multiple modes.
Variance	How much variability exists in the data.	$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$
Standard deviation	Another measure of how much variability exists in the data.	$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$

* $\sum_{i=1}^n$ = sum of all observations

\bar{x} = mean

n = number of observations

The normal distribution is a frequently used distribution in statistics, shown in Figure 7. An assumption of many statistical tests is that the variable under consideration is normally distributed. There are other distributions that exist (for example, Poisson and binomial), many of which have normal approximations that make them easier to work with.

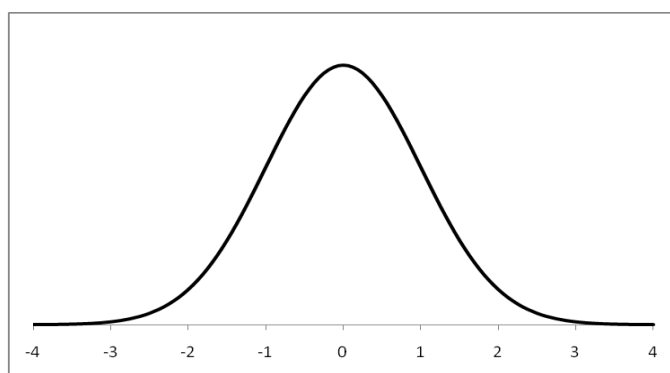


Figure 7.

An understanding of the normal distribution is very helpful when trying to determine whether a given clinical result should be considered “abnormal”. When it is known that a given clinical measurement is normally distributed, and the mean and standard deviation are known, it is possible to predict the likelihood of a given finding. For normally distributed data, 68.2% of the values will fall within the range of +/- 1 s.d. from the mean value; 95.4% of the values will fall within the range of +/- 2 s.d. from the mean; and 98.8% of the values will fall within +/- 3 s.d. from the mean.

This expected distribution can translate to a quick interpretation of clinical data. It has been reported that in the “normal” (non-glaucomatous) population, the mean value of IOP is 16 mmHg and the standard deviation is 2.5 mmHg.^{iv} If you measure an IOP of 21 mmHg (equivalent to 2 s.d. above the mean value) you know that this measurement exceeds 97.7% of the IOP values in the normal population (50% fall below the mean, 13.2% fall between the mean and +1 s.d. and 13.6% fall within 1 s.d. and 2 s.d. totaling 97.7%).

CLINICAL PEARL – outside normal range ≠ disease

A clinical test result that falls outside of the normal range does not mean that the patient definitely has a disease. A “normal” finding in statistical terms is simply a value that falls within a defined range around the mean, given normally distributed data.

When evaluating a random sample where the variable of interest is normally distributed in the population, the sample mean will be normally distributed. Often we can perform statistical analysis assuming the underlying distribution is approximately normal by relying on the central limit theorem. The basic idea of the central limit theorem is that as the sample size increases, non-normal distributions will increase in their symmetry, thereby allowing the use of normal statistical methods.

Comparing Groups (Parametric or Normally Distributed Data)

t- tests

A hypothesis about a mean can be tested in a one-sample setting or a two-sample setting, shown in Table 11. In the one sample scenario, you test whether the data from your sample (sample mean identified by \bar{x}) are similar to a known population mean (denoted in terms of μ). In other words, the population mean and standard deviation are known. Under these circumstances, the underlying distribution is normal, and the z-test is the appropriate statistical test for comparison. The probability determining whether a statistical difference exists comes from normal distribution probability table.

More frequently, however, we want to compare two independent groups where the parameter of interest, specifically the standard deviation, is not known and must be estimated from the sample. For comparison, we use a different distribution, the t-distribution, and compare means using the t-test (also known as the Student’s t-test). The hypothesis being tested is that the mean in group 1 is equal to the mean in group 2. The regular two-sample t-test compares the means from each sample to indicate a difference between the two. The information required to perform a t-test is similar to that

required for a z-test, the sample mean and the estimated standard deviation. The t-distribution is the source of the probability to determine whether there is a statistical difference between the two means when using this test.

When the two samples are not independent, for example when measurements are made on the same person under two different conditions, (e.g., wearing glasses and then wearing contact lenses) or at two different time points, then a paired t-test is used. This is a test that compares person differences for the group to see if there is a difference between conditions (glasses or contact lenses) or time points (before and after).

Table 12. Different scenarios using the t-test

Test	Variables for test	Hypothesis
One-sample t-test	Mean (\bar{x}_1) for the sample and mean for the comparison population (μ_p)	Null: $\mu_1 = \mu_p$ Alternative: $\mu_1 \neq \mu_p$
Two-sample t-test	Mean for sample 1 (\bar{x}_1) and mean for sample 2 (\bar{x}_2)	Null: $\mu_1 = \mu_2$ Alternative: $\mu_1 \neq \mu_2$
Paired t-test	Difference between mean at time 1 (\bar{x}_1) and mean at time 2 (\bar{x}_2)	Null: The change ($\mu_1 - \mu_2$) between two time points is not different than 0. Alternative: $(\mu_1 - \mu_2) \neq 0$

We assume that for performing calculations, the reader will use a statistical software program. Thus, we are not providing formulas and arithmetic. For those preferring to do calculations in their own research, the reader is referred to various statistical texts listed in the references.

Clinical Scenario:

A prospective study drew consecutive pediatric patients from a clinical setting. The purpose of the study was to compare the central corneal thickness, as measured by ultrasound pachymetry, among three different ethnic groups. Shall we proceed with a t-test or is there some other analysis?

What are the basics of this scenario?

Research Question: Does central corneal thickness vary based on an individual's ethnic background?

Null Hypothesis: There is no difference in corneal thickness based on ethnicity.

$$\mu_1 = \mu_2 = \mu_3$$

Alternative Hypothesis (two-tailed): Corneal thickness is statistically significantly different between various ethnic groups.

$$\mu_1 \neq \mu_2 \neq \mu_3$$

Outcome Variable: central corneal thickness (continuous variable measurement)

Predictor Variable: ethnicity (categorical variable measurement)

Often there will be more than two means that need to be compared. Sometimes, you will see an investigator that has compared three or more groups using all the possible pairs of groups and reports p-values for each of these comparisons. By doing this, the risk of making a type I error has increased significantly because of the multiple tests run to answer the question. Remember, the more times a test is repeated the more likely one result will be outside the 95% confidence interval just by chance. Instead we use a test that will consider all of the groups at one time – an analysis of variance. An analysis of variance (or ANOVA) looks at the means in all the groups at one time and determines whether a difference exists among any of the groups (i.e. looks at the hypothesis that all the means are equal). Only if there is a significant difference among all the groups do we then move on to analyzing pairs, and this is done in such a way, using post-hoc comparisons, that we minimize the chances of making a type I error.

The point of the ANOVA is to evaluate different sources of variation in the means to determine if there are differences between the two groups beyond chance. The basic process is that the variability within the groups is calculated, the variability between groups is calculated, and then the two sources of variability are compared between the two groups. If the between group variability is found to be statistically different from the within group variability, then a group difference exists.

As with most statistical tests, there are some assumptions for the ANOVA that you should be aware of. Assumptions are criteria that are necessary to be able to use a statistical test for a given set of data. This test assumes that data are from a random sample and that the data are on an interval scale. The variables should also be normally distributed with a common variance. Violation of the assumptions of a statistical test can violate the accuracy of the test, but it is not uncommon that manuscripts do not actually address this issue. Additional statistical analysis is necessary to confirm normal distribution.

In addition to using an analysis of variance to compare multiple groups, there are other uses. This analysis allows for the addition of covariates into a model, even when there are only two groups to compare. A covariate is another independent variable (or variables) or possible risk factors outside of the outcome and the group variable that may affect the outcome. It is added into the model to determine if it has an association with the outcome (in other words, we look at whether this covariate is able to explain any of the variability in the outcome variable).

Another use of the ANOVA is an extension of the paired t-test where we have measures on the same subjects at more than one time point. This analysis is a repeated measures ANOVA, and it is frequently used in the analysis of longitudinal studies.

APPLYING RESEARCH TO CLINICAL SCENARIO

CITATION

Dai, E, Gunderson CA. Pediatric central corneal thickness variation among major ethnic populations. *Journal of AAPOS*. 2006;10:22-25.

DESCRIPTION

Revisiting the scenario from the beginning of the section, we have three means that the investigator needed to compare. The analysis used was an analysis of variance to compare the mean corneal thickness among African-American, White, and Hispanic children.

There were 106 left eyes that were analyzed, with 21% African-American children, 33% Hispanic subjects and the remaining 46% were White. The mean central corneal thickness values (\pm standard deviation (sd)) were as follows:

African-Americans = $523 \pm 40 \mu\text{m}$

Hispanics = $568 \pm 44 \mu\text{m}$

Whites = $563 \pm 36 \mu\text{m}$

RELEVANT RESULTS

The analysis showed a significant p-value for race, $p = 0.00002$. This meant that there were differences among the three groups, but it does not tell which groups were different from each other. Post-hoc t-tests, statistical comparisons done after the original analysis to determine where differences lie, were done compare each race group to the other. The Hispanics and the Whites were not significantly different. African-Americans were statistically different from the Whites ($p = 0.0001$), with their corneal thickness being on average $39.2 \mu\text{m}$ thinner than the Whites. They were also significantly different from the Hispanics ($p = 0.0003$). Their corneal thicknesses were, on average, $44.4 \mu\text{m}$ thinner than the Hispanics.

Clinical Scenario:

A randomized controlled trial was designed to compare two treatments used for neovascular age-related macular degeneration (ARMD): photodynamic therapy, a commonly used treatment, and transpupillary thermotherapy. The main outcome was measured 12 months after treatment, and the null hypothesis tested was that the two treatments were equally effective at maintaining best-corrected visual acuity, while the alternative hypothesis was that the transpupillary thermotherapy would result in less loss of visual acuity. The outcome was the proportion of patients in each treatment arm who lost less than 15 ETDRS letters of best-corrected visual acuity. What kind of analysis should we use?

What are the basics of this scenario?

Research Question: Are photodynamic therapy and transpupillary thermotherapy equally effective for maintaining best-corrected visual acuity in neovascular age-related macular degeneration patients?

Null hypothesis: The proportion of subjects who received photodynamic therapy and lost less than 15 letters of best-corrected visual acuity = the proportion of subjects who received transpupillary thermotherapy and lost less than 15 letters of best-corrected visual acuity.

Alternative hypothesis (one-tailed): The proportion of subjects who received photodynamic therapy and lost less than 15 letters of best-corrected visual acuity < the proportion of subjects who received transpupillary thermotherapy and lost less than 15 letters of best-corrected visual acuity.

Outcome Variable: Proportion of subjects who lost less than 15 ETDRS letters of best-corrected visual acuity (continuous variable measurement).

Predictor Variable: Treatment group: photodynamic therapy or transpupillary thermotherapy (categorical variable measurement).

When the variable under study is categorical (regardless of the number of categories), a basic descriptor is the proportion (for a two category variable this is $X/(X+Y)$, where X and Y are the counts for each variable). A simple comparison of proportions, where the hypothesis is that proportion 1 is equal to proportion 2, can be accomplished using the z-test to find a difference between the proportions. When we want to expand beyond a comparison of two proportions for a single variable, and looking at more than one categorical variable, we create what is called a contingency table. In this table, all possible combinations of the variable are represented with the sample size in each cell. The proportions can then be calculated for each cell. The classic contingency table is the 2X2 table (see Tables 4 and 6 for examples of 2X2 tables). To determine whether the proportions across two variables in a contingency table are similar, a chi-squared (χ^2) test is used. This test determines whether the observed proportions in the table are different than what the expected values of the proportions for each cell are. If the observed data differ from the expected, then an association exists.

APPLYING RESEARCH TO CLINICAL SCENARIO

CITATION

Odergren A, Algvere PV, Seregard S, Kuanda A. A prospective randomized study on low-dose transpupillary thermotherapy versus photodynamic therapy for neovascular age-related macular degeneration. *British Journal of Ophthalmology*. 2008;92:757-761.

DESCRIPTION

Taking the proportions from the randomized trial example in the scenario, we need to compare the proportion of success in each group. In order to test the difference between the success in the two groups the χ^2 statistic was used. The sample size for this study was based upon the aim of showing that the transpupillary thermotherapy would result in 30% (size of the effect) more of the subjects retaining their visual acuity (with a power of 0.90 and an alpha error of 0.05).

RELEVANT RESULTS

At twelve months, 73.9% of those in the photodynamic therapy arm had lost fewer than 15 letters on the ETDRS visual acuity chart. Among those in the transpupillary thermotherapy, 75.0% had lost fewer than 15 letters on the visual acuity chart. The p-value from the χ^2 statistic comparing these two proportions was greater than 0.05, therefore the study was unable to reject the null hypothesis.

Correlation**Clinical Scenario:**

Let's say we've taken data on a sample of myopic subjects before LASIK, immediately after LASIK, and ten years after they received LASIK. We want to see whether or not their refractive error has regressed back into myopia after ten years, and determine whether the amount of myopia they had before LASIK predicts this regression. We've looked at the data points plotted on a graph, and they look to be fairly linear. What analysis do we use for the data?

What are the basics of this scenario?

Research Question: Is regression of myopia after LASIK related to the amount of myopia at the time of LASIK surgery?

Null hypothesis: The change in myopia after LASIK is not related to the amount of myopia at the time of surgery.

Alternative hypothesis (two-tailed): The change in post-LASIK myopia is related to the amount of myopia at the time of surgery.

Outcome Variable: Amount of refractive error change from LASIK surgery to 10 years post-LASIK (continuous variable measurement).

Predictor Variable: Amount of myopia at baseline (continuous variable measurement).

As previously discussed in this chapter, many of the types of measurements that we are interested in studying in health research are continuous variables. This is common for measurements of biological factors. Intraocular pressure and amount of refractive error are two commonly seen variables in optometry. When both our dependent and independent variables are continuous, there are a couple of analysis options. The analysis that is selected depends on what question is asked. If we want to see whether one model can predict the other, and we have reason to suspect a linear relationship between the two variables, a linear regression model will show if one model can predict the other. The regression model is defined by the intercept (α), the slope (or steepness) of the line (β), as well as the error term which allows for some variability around the line as we don't really expect a perfect linear fit. The general form of the equation is

$$\text{Outcome} = \alpha + \beta \cdot \text{independent variable} + \text{error}.$$

The hypothesis tested here is whether the slope of the line is equal to zero, which means there is no linear association between the two variables (Figure 8a). If the null

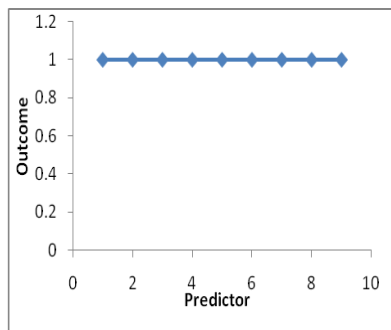


Figure 8 a. slope = 0

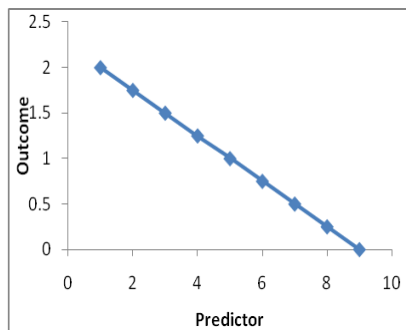


Figure 8 b. slope is (-)

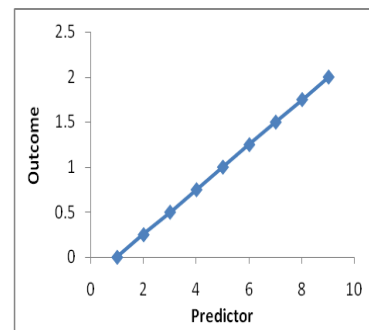


Figure 8 c. slope is (+)

hypothesis that the slope is zero is rejected, and the slope is positive, then as the value of our independent variable increases, we predict that the value of our outcome variable increases as well (Figure 8b). Conversely, if our slope is negative, then as the independent variable increases, we expect the value of the outcome variable to decrease (Figure 8c).

Should there be more than one independent variable to consider, as there most frequently is, a multiple regression analysis can be used to control for these other variables. That way, when we look at the slope for one independent variable, it has been adjusted for the effects that the other independent variables might have on the outcome. Thus, the effect of each independent variable can be estimated and accounted for in a cause and effect model. It is also one way confounding bias might be controlled for after a study is completed.

Statistical interactions (also known as effect modification) are often assessed in linear regression models. An interaction takes two independent variables and determines if the outcome under study differs for one variable depending on the amount or presence of the other variable, that is that effect of one variable varies depending on the other variable(s) under study. Let's say, for instance, we find that LDL cholesterol is related to glaucoma as is number of cigarettes smoked per day. We can look for an interaction between cholesterol level and number of cigarettes. We may find a significant interaction between the two variables. Upon looking we find no association between level of cholesterol and number of cigarettes smoked when less than 20 cigarettes are smoked in those with LDL cholesterol less than 150 mg/dL. However, in those that smoke 20 or more cigarettes and also had an LDL cholesterol level of 150 mg/dL or more, there was an increased risk of glaucoma. In the presence of an interaction, we do not focus on what the individual variables are presenting. We must focus on the interaction because it tells us that without knowing what is occurring with both variables, we know nothing about either one.

A key assumption underlying the linear regression model is that for a given value of the independent variable, there is an outcome response that is a linear function of the independent variable. Additionally the outcome variable is from a normal distribution, which has the same variance as the independent variable, and the errors between any pairs of points are independent (for example, the pairs of points are from different people).

You can take the same two variables and ask a different question. If you are interested in how strongly associated two variables are, you can use a correlation coefficient to assess this. The correlation coefficient has a value between -1 and +1. A

correlation near zero indicates no association. As the magnitude of the correlation increases, the association is stronger. Positive correlations mean as one variable increases the other does as well, while negative correlations indicate a decrease in one variable as the other one increases.

APPLYING RESEARCH TO CLINICAL SCENARIO

CITATION

Gwiazda J, *et al.* A randomized clinical trial of progressive addition lenses versus single vision lenses on the progression of myopia in children. *IOVS*. 2003;44:1492-1500.

DESCRIPTION

This paper presents the main results of a randomized controlled trial that evaluated the effect of progressive addition lenses (PAL) compared to single vision lenses (SVL) in preventing the progression of myopia in children (known as the Correction of Myopia Evaluation Trial, or COMET). The primary outcome we will discuss here is the comparison of the myopia progression in spherical equivalent (SE) between the PAL and SVL groups after three years. Analyses including assessing the effectiveness of randomization at baseline using a t-test for continuous variables that were normally distributed, and the Wilcoxon test for those that were not continuous. Categorical variables were tested using a χ^2 test or a Fisher's Exact test if the assumptions for the χ^2 were not met. The analysis of the main outcome of myopia progression was assessed using linear regression. Multiple linear regression models were completed by general linear modeling, and were adjusted for covariates.

RELEVANT RESULTS

The first set of comparisons checked to make sure that randomization had worked correctly and that the known covariates were balanced between the two groups, which had a total of 469 subjects. For example, a χ^2 test was used to compare the gender assignment between the two lens groups. There were 52% of those in the PAL group who were female and 53% of the SVL group who were female. The χ^2 showed no statistically significant difference ($p = 0.85$), so the gender representation was balanced within the two groups. The distribution of spherical equivalent was another variable that was important to assess at baseline for comparability. The mean SE (\pm sd) at baseline was -2.40 ± 0.75 D in the PAL group and -2.37 ± 0.84 D in the group randomized to SVLs, which was not significantly different by t-test ($p = 0.38$).

The primary outcome of three-year progression by treatment group was analyzed adjusting for covariates that may be related to myopia progression (age, gender, ethnicity, baseline refractive error, axial length, accommodative response, and phoria). The analysis included an interaction between level of myopia at baseline (more myopia < -2.25 D, versus less myopia ≥ -2.25 D) and treatment group. An interaction looks to see if the outcome is different within subgroups of the variables being looked at for an interaction. The treatment group p-value from the model was statistically significant ($p < 0.01$), and the interaction between level of myopia and treatment group was significant as well ($p < 0.05$). Where there is a significant interaction, the main effect, in this case treatment group, cannot be interpreted by itself, but must be interpreted as an interaction. The interaction was that the two types of lenses had different effects on the

progression of myopia depending upon the amount of baseline myopia. The effect of treatment was larger for those who had less myopia at baseline. The adjusted means and standard deviations by treatment group, the difference between the two treatment groups and the 95% confidence interval associated with the differences are shown below.

Baseline myopia	PALs	SVLs	Difference (95% CI)
≥ -2.25 D	-1.17 ± 0.08	-1.47 ± 0.08	0.30 ± 0.11 (0.04 to 0.55)
< -2.25 D	-1.38 ± 0.07	-1.48 ± 0.08	0.10 ± 0.11 (-0.15 to 0.36)

This shows that the significant treatment effect occurred in those subjects who had less myopia at baseline, and indicates that those in SVLs progressed more over three years (0.30 D) than those in PALs. In those with higher levels of myopia at baseline, progression over three years did not differ between the two treatment groups. We see that there was a statistically significant difference found here, but we should also look at the results and ask ourselves about the clinical relevance. We see that over three years, there was a difference of 0.30 D between the progressive addition lenses and the single vision lenses among those subjects that had higher myopia at baseline. This is, on average, a change of 0.10 D a year. What would you do if you were treating one of these subjects? Prescriptions are given in 0.25 D increments, so it may be year three before a prescription change is warranted. Is this clinically relevant, under those circumstances?

Clinical Scenario:

So if we look at the example described at the beginning of the section describing linear regression, we can take data on LASIK subjects and ask a research question that can be answered using a logistic regression analysis. Let's say we expect some increase in myopia over time after the LASIK procedure, but we are only interested in the amount of myopia when it reaches a clinically significant level. For this example, let's use a 1.00 D change in the myopic direction. The outcome, therefore, becomes dichotomous, a 1.00 D change occurred, yes or no. We can then look at the odds ratio of a 1.00 D change to see if the odds differ based upon the amount of baseline refractive error. If we want to make this a multiple logistic regression, we can add in age at LASIK surgery to identify if age has an effect on the odds of seeing a 1.00 D myopic change, after controlling for the effects of baseline refractive error.

What are the basics of this scenario?

Research Question: Is a 1.00 D or more myopic change 10 years post-LASIK surgery related to the amount of baseline myopia?

Null hypothesis: The odds of at least a 1.00 D myopic change 10 years post-LASIK surgery is not associated with the amount of myopia at the time of LASIK surgery.

Alternative hypothesis (two-tailed): The odds of a 1.00 D or more myopic change 10 years post-LASIK surgery is associated with the amount of myopia at the time of LASIK surgery.

Outcome Variable: Progression of myopia change of at least 1.00 D in the myopic direction 10 years after LASIK surgery (dichotomous variable measurement).

Predictor Variable: Amount of myopia at the time of LASIK surgery (continuous variable measurement).

When your outcome of interest is a dichotomous variable (e.g. yes/no; progression/no progression; glaucoma/no glaucoma) and you want to be able to evaluate one or more independent continuous or categorical variables, a logistic regression analysis will allow you to do this. This situation is common in health research if you think about it.

Examples include things like disease or no disease, death or survival, and progression or no progression. A logistic regression analysis allows you to calculate the probability of disease (which has values between 0 and 1). From this you can calculate the odds ratio (see Table 4 for odds ratio calculation example), which gives the odds of disease to no disease based upon our independent variable of interest.

As with linear regression, we commonly want to be able to consider more than one independent variable at a time. Multiple logistic regression allows you to estimate the odds ratio for disease based on a given risk factor while controlling for other factors of interest or potential confounders.

APPLYING RESEARCH TO CLINICAL SCENARIO

CITATION

Ramamoorthy P, Sinnott LT, Nichols JJ. Treatment, material, care, and patient-related factors in contact-lens related dry eye. *OVS*. 2008;85:764-772.

DESCRIPTION

These data originate from a two-phase cross-sectional survey and examination visit on dry eye. The design for this analysis was a nested case-control study to examine the factors associated with contact-lens related dry eye. These factors included contact lens materials and characteristics, patient-specific factors (age, gender, etc.), lens care regime, and dry eye treatments used. Descriptive statistics were provided as means \pm sd for the continuous variables for both the dry eye group and the non-dry eye group. Groups were presented using contingency tables for those variables that were categorical in nature. Analyses consisted of univariate and multivariate logistic regression analysis which model the odds of self-reported dry eye due to the factors of interest. Univariate analyses were reported as odds ratios, 95% confidence intervals (95% CI), and p-values showing the association of each individual variable with the odds of dry eye. Using the significant variables, in this case those variables with p-values <0.25 , multivariate logistic regression models were created in a stepwise fashion, where each variable was added to determine the most significant variable. This variable was added to the model permanently, and then the process was started again to see if any other variables would add to the model significantly. The final model was controlled by age, gender, and current dry eye treatments.

RELEVANT RESULTS

The paper examined the different factors from above by topic. Here are presented a few representative examples. There were 306 subjects. Dry eye subjects represented 65% of the sample (199) and the remaining 35% did not report dry eye symptoms. As an example, we will look at the use of artificial tears/rewetting drops (AT), which was analyzed both as a dichotomous variable (use or no use) and as a continuous variable of average number of times used during a week. Of the dry eye subjects, 61.6% of them reported using AT, while 38.4% of the non-dry eye subjects reported AT use. The odds ratio was 1.57 with a 95% CI of 1.02 to 2.41 (p -value = 0.04). The interpretation of this result shows the odds that a dry eye subject used artificial tears was 1.57 higher than a non-dry eye subject. The 95% confidence interval does not include 1.0, indicating that this is a result that is statistically significant. For the average number of uses of AT per week, the dry eye subjects reported 2.96 ± 5.05 drops while the non-dry eye subjects reported 1.46 ± 3.21 drops per week. The odds ratio was 1.10 (95% CI = 1.03, 1.17, p -value = 0.003). Again, the confidence interval does not include 1, so the odds show the dry eye subject using more drops per week than a non-dry eye subject.

The final model was controlled for age, gender, recent contact lens refitting, and number of weekly applications of artificial tears). Four independent variables were significantly associated with an increased odds of dry eye: decreased overall satisfaction with contact lens wear; dry eye when not wearing contact lenses; a decrease in daily lens wearing time (hours/day); and a decreased ability to wear lenses as long as desired.

Decreased overall satisfaction with contact lens wear: OR = 3.57 (95% CI = 2.08, 5.88), $p < 0.0001$

Dry eye without contact lenses: OR = 6.54 (95% CI = 2.57, 16.62), $p < 0.0001$

Decreased lens wearing time: OR = 1.16 (95% CI = 1.06, 1.26, $p = 0.0007$)

Decreased ability to wear lenses as long as desired: OR = 2.44 (95% CI = 1.30, 4.54), $p = 0.005$

So the interpretation of each of these independent variables is as follows. Using decreased contact lens satisfaction as an example, the odds of dry eye is increased by 3.57 if the subject is dissatisfied with lens wearing, after controlling for the effect of age, gender, recent contact lens fitting, artificial tears, dry eye without contact lenses, decreased wearing time and decreased ability to wear lenses as desired. The 95% confidence interval indicates that if the odds ratio would fall within 2.08 and 5.88 in 95 out of a 100 samples drawn.

Non-parametric Methods (For Data Not Normally Distributed)

All the statistical analyses we've discussed have been parametric tests. These tests rely on assumptions about the distribution from which the data come from (i.e. normal distribution). Since these assumptions do not always apply, there needs to be an alternative way to analyze data. Non-parametric tests do not depend on the distribution or require assumptions be met, so these tests are useful when assumptions are not met, or when the sample size too small to rely on the central limit theorem. While these tests will not be addressed in detail, some of the most frequently seen include the Sign test, the Wilcoxon Signed Rank test, the Wilcoxon Rank Sum test (also known as the

Mann-Whitney U test), and the Kruskal-Wallis one-way analysis of variance. Most have corresponding parametric tests, so developing the research question and test hypothesis is the same as described above. The arithmetic is different because the variable formats are different. Interpretation, however, is very similar. The study reports a statistic and interprets whether it meets statistical significance. Then, it should be determined what the clinical significance is. The reader is referred to the various statistical texts in the reference list for more information.

Putting It All Together

When determining whether the analysis of a dataset is appropriate, it is first necessary to go back to the hypothesis and revisit the underlying research question and how the hypotheses were framed in context. For instance, does the null hypothesis address mean differences or odds ratios? How many groups are being compared? Then look at the independent and dependent variables. What form do they take, the dependent (outcome) variable in particular? Are the data categorical or dichotomous? These are the questions that lead to the determination of whether the analysis of the data is the correct one. Finally, look at the summary data, the measures of central tendency and variability. Do they make sense? Perhaps there was a statistically significant difference between group A's 75.9% success rate and group B's 79.2% success rate, but does that mean treatment A is inferior to treatment B? The statistics should be easily understood. More complicated calculations for rarely used statistical tests do not mean the results are more valid. If the summary data look equivalent, but complicated statistics find a statistically significant difference, ask yourself if the finding is important or meaningful in a clinical context.

The purpose of conducting research is to be able to affect the care of our patients. By maintaining a current literacy of available research, you will be able to assess patterns among the studies. Replication of results strengthens our belief in them, making the application of the results more natural. When these studies have been well conducted, this adds additional confidence to the appropriateness of incorporating the findings of studies into practice. Determining where the results of research you read might be useful in your practice is the first step towards implementing them.

Study Questions

1. You read a head to head study comparing two antihistamine eye drops. One drop was used in a study site in Vermont. The other was used in a study site in Florida. What concerns do you have about the study design?
2. A mutation of an adenoviral strain known to cause conjunctivitis is found in several subjects in different areas of the western United States. In order to quickly determine if the new strain could start an epidemic, what study design(s) would be appropriate and why?
3. Why is randomization important? What type of bias does it control for?
4. Design a study to determine if a vasoendothelial growth factor inhibitor can successfully treat proliferative diabetic retinopathy.

5. What statistical test can you use to determine if three means are significantly different from each other?
6. A statistically significant result can be not clinically meaningful. How can this be?
7. What is a statistical interaction?
8. A study comparing the incidence of microbial keratitis between two groups of contact lens wearers found a difference in the incidences of 10% which was not statistically significant different. The study had a power of 55% to find a difference of 10%. The authors concluded that the incidence in the two groups of lens wearers was the same. Is this true? Why or why not?
9. A study is comparing the effectiveness of drug X in preventing vision loss in subjects with ARMD. The outcome is vision loss – yes/no. The investigators need to control for covariates such as age and gender to determine whether drug X was effective. What statistical analysis should they use?

Take Home Conclusions

- Research is journey to find truth in the natural world. Without research, optometric practice cannot provide care based on evidence.
- To discover truth, research must use studies which control, quantitate or eliminate bias.
- Correlation does not equal causation.
- Research results can only be generalized to populations similar to the sample.
- Statistics quantify data samples and are used to infer how likely similar results will occur for a population group.
- Statistical results are only as valid as the sample and the accuracy of the data.

References

- Dawson B, Trapp RG. Basic and clinical biostatistics, third edition. New York: Lange, 2001.
- Garb JL. Understanding Medical Research: A Practitioner's Guide. Boston: Little, Brown and Company, 1996.
- Gehlbach SH. Interpreting the Medical Literature, third edition. New York: McGraw-Hill, Inc., 1993.
- Greenberg RS, Danies SR, Flanders WD, Eley JW, Boring JR. Medical epidemiology, third edition. New York: Lange, 2001.
- Hatch SW. Ophthalmic epidemiology and research. Boston: Butterworth-Heinemann, 1998: 265-78.
- Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB. Designing Clinical Research, second edition. Philadelphia: Lippincott Williams & Wilkins, 2001.
- White KL, Williams TF, Greenberg BG: The ecology of medical care. N Engl J Med 265:885, 1961.
- Wood MJ, Ross-Kerr JC. Basic Steps in Planning Nursing Research: From Questions to Proposal, sixth edition. Boston: Jones and Bartlett Publishing, 2006.

Endnotes

-
- ⁱ Quinn GE, Shin CH, Maguire MG, Stone RA. Myopia and ambient lighting. Nature. 1999;399(113-114).
- ⁱⁱ Friedman, Furberg, and DeMets, Fundamental of Clinical Trials. New York: Springer-Verlag, New York. 3rd edition.
- ⁱⁱⁱ Sackett, Bias in analytic research. J Chr Dis. 1979, 32:51-63.
- ^{iv} Fingeret M, Lewis T. Primary Care of Glaucoma, second edition. Appleton & Lange, p 116.