Lucy Mastrandrea and Takara Stanley
& PES Fellow Education Committee, 10/19/12

Clinical Trials Study Design

Case Series: Case series are studies that describe in detail a group of subjects who share common characteristics and/or have received similar interventions.

Observational Studies: Observational studies provide a mechanism for clarifying the epidemiology and risk factors/exposures that place populations at risk for disease. Observational studies assist in identifying clinical outcomes that might aid in predicting future disease. This type of study allows researchers to draw conclusions about responses to variables; they allow for associations but not causation.

Cohort Studies: Longitudinal studies that are designed to follow a group of subjects over time. A cohort is a group of subjects that share a common characteristic, such as time of birth, exposure to an environmental agent, or risk for a specific disease. Cohort studies often test hypotheses aimed to identify associations between a given characteristic and an outcome. Cohort studies may be retrospective (looking back in time) or prospective (looking forward in time).

I. Retrospective Cohort Study: A study population is defined and the medical history is reviewed to identify associations between exposures and a given outcome. This type of study is also known as a historical cohort study. Retrospective cohort studies often serve as the background for development of an interventional trial.

   Example: Chart review of pediatric subjects with a known diagnosis of late-onset congenital adrenal hyperplasia to determine linear growth patterns prior to diagnosis.

II. Prospective Cohort Study: A defined study cohort is followed forward in time to determine how factors/exposures affect a given outcome. Prospective studies can help identify risk factors for disease because data is collected at set time intervals. Risk to these studies is loss to follow-up and length of time needed to determine associations.

   Example of a prospective cohort study is the Bogalusa Heart Study. This study began in 1973 to determine the natural history of cardiovascular disease in a group of 12,000 children from Bogalusa, LA.

III. Case/Control: Observational study in which likelihood of exposure is compared between representative groups with (case) and without (control) a disease. Data can be conducted either retrospectively or prospectively. These studies can identify associations between a disease and risk factors or disease outcomes. Case-control studies generally require fewer subjects and are less expensive than cohort studies, and they can be particularly useful in epidemiologic investigations to identify risk factors for disease. A case-control design may also be appropriate for studying very rare diseases/outcomes that would not be encountered in sufficient numbers using a prospective cohort design.

Cross-Sectional Studies: This type of cohort study takes a "snap-shot" of a given population at one point in time. In a cross-sectional study, a cohort of interest is identified and a set of factors related to an outcome is explored. In some cases, discrete biochemical data is obtained to strengthen associations. Cross-sectional studies can provide prevalence data for disease, as well as rates of risk exposures and associations between risk factors and an outcome of interest. Cross-sectional studies can be designed as case/control in order to determine whether a given condition is associated with risk factors or discrete data in cases vs. controls. Cross-sectional studies are powerful in that they can provide data for planning of prospective cohort studies or intervention trials.

Intervention Studies: Intervention studies test the impact of an intervention on a given outcome. In an interventional study, a variable is manipulated, and the effect of that manipulation on a defined outcome is measured.

Pilot Studies: Pilot studies are small, feasibility studies that are usually performed to test hypotheses in preparation for a larger interventional clinical trial. Pilot studies allow investigators to test experimental design, obtain preliminary data for power analysis (see below), and provide information about subject recruitment and study management before investing resources to a larger study.

Randomized Controlled Trials: Randomized controlled trials (RCTs) are intervention studies in which a group of subjects with similar characteristics are randomized to receive one of several defined interventions. The goal is to determine quantitatively the effect of an intervention on a defined outcome. RCTs are powerful tools to test a hypothesis. These studies are often the basis for evidence-based medicine practice, and are considered the gold-standard in clinical research. Pilot studies may provide the groundwork for the development of a RCT. There are several different components which must be defined when developing this type of trial.

i. **Randomization:** Randomization means that the intervention group into which a subject is placed is determined randomly. The goal of randomization is to minimize selection bias and to increase the likelihood that comparison groups are similar in all variables except the intervention of interest. The larger the clinical trial, the lower the risk that treatment groups will be different from each other. There are several strategies which can be used to randomize subjects.

   a. **Simple Randomization:** This method is similar to flipping a coin. A random number generator is used defining 1=placebo, 2=active (more numbers added if there are other interventions). This method provides an even distribution of subjects to each of the interventions.

   b. **Block Randomization:** This method is used if there is concern that subjects may change over time. In block randomization, the number of subjects in a block is defined. Within the block, subjects are equally and randomly assigned to the interventions. Note that simple randomization is effectively randomization by blocks defined as the number of intervention variables. For example: The number of subjects in a block is defined as 4. There are two interventions –

1=placebo, 2=active. The distribution to interventions within the blocks could be as follows:

Block 1: 1,1,2,2
Block 2: 2,1,2,1
Block3: 1,2,1,2
Block 4: 1,1,2,2

c. **Stratified Randomization:** This method is used when there is concern that subjects may not be balanced based on baseline characteristics, such as age, length of time of disease, or disease control. When the strata are defined, subjects are divided along that stratification, and then randomization (either simple or block) occurs within that strata.

d. **Unequal Randomization:** This method does not allocate subjects equally to study interventions. Instead, subjects are randomized in greater proportions to a defined intervention. The randomization may be 2:1, in which twice the number of subjects is randomized to the active arm compared to those randomized to the control arm. This method may decrease statistical power.

ii. **Study Blinding:** Blinding in a study means that either subjects (single-blind) or subjects and investigators (double-blind) are not aware of the treatment into which subjects are randomized. Blinding decreases the likelihood that study outcome will be influenced by knowing the intervention to which a subject is randomized. Not every protocol lends itself to blinding. Subjects can be "unblinded" if there are any adverse events that occur during the trial.

iii. **Placebo-Controlled:** In placebo-controlled studies, subjects are randomized to receive either placebo (sugar-pill, inactive excipients) or the active intervention. Placebo-controlled trials may be challenging to perform, particularly if the intervention involves an injection or use of a device. In addition, in cases where an established therapy is already considered to be standard of care, placebo-controlled trials may be considered unethical if the usual treatment is withdrawn while the subject is enrolled in the study. In some cases, the placebo arm of the study is actually defined as a non-intervention arm.

iv. **Active Comparator:** In some RCTs, subjects are randomized to an active comparator arm which is compared to an experimental intervention arm. An active comparator is used when it would be unethical or impractical to randomize subjects to placebo or non-intervention. Active comparator trials are powerful methods to determine if a new therapy is non-inferior or superior to current standard of care treatments.

v. **Cross-Over Design**: In a RCT, there are situations when subjects cross-over into another intervention. For instance, all subjects randomized to active comparator may be crossed-over to receive the experimental agent after a set time period, while all subjects

receiving the experimental agent will cross-over to the active comparator. The advantages to this design are that each subject serves as his/her own control, and the number of subjects needed to answer a clinical question may be less. Thus, the investigator can analyze the data within and between subjects. A concern with this design is that it is important to make sure that subjects are the same at the time of the cross-over as they were at the initiation of the study. This design must also take into consideration the "order-effect" which states that the order of interventions may influence the outcome.

There are different factors that may influence the outcome of a study. These must be considered during study design as well as in the analysis phase.

Chance: Random variation can contribute to certain findings. Statistics help us minimize the risk that the results of a study are due to chance (see Type 1 error below). However, there is always the caveat that a given outcome is due to chance variation. Random variation results in imprecise data.

> **Example:** You could flip a coin 20 times and always have it land on tails. That does not mean that one could draw the conclusion that both sides of the coin have tails. Rather, it is random variation or chance.

Bias: Bias refers to systematic variations rather than random variation. Inaccurate results may be obtained if the study is biased with respect to subject recruitment, data collection, or data analysis. Bias results in inaccurate data.

i. Selection bias: Bias that results when treatment groups are inherently different or when individual allocation to interventions is not representative of the population. For example, people who volunteer for exercise studies may be more motivated to exercise and lifestyle changes than the general population. Randomization limits selection bias.

ii. Measurement bias: Bias that occurs when outcome measurement is inaccurate. This can be the result of improper tools for measuring outcome or preconceived notions of the outcome. Measurement bias is limited by careful study design, use of appropriate measurements, and study blinding.

iii. Analysis bias: Bias that occurs due to incomplete data collection. This type of bias is a particular problem when subjects withdraw from studies since non-completers may be different than those who complete a study. Methods for avoiding this form of bias include subject retention strategies as well as carrying data from last visit forward to the end of study.

Confounders: Variables that may create errors in interpretation of data. For example, an association between low BMI and mortality may be confounded by smoking, which prevents weight gain, but also increases cancer risk. Confounders may increase or decrease the strength of an association between

intervention and outcomes. Confounders can be controlled for in statistical analysis as long as they are recognized.

Null Hypothesis (H_0): States that there is no relationship between two variables. In research, one cannot "prove" **H_0**, one can only "reject" or "fail to reject" the **H_0**.

Alternate Hypothesis (H_A): States that there is a relationship between two variables. The research hypothesis may be "one-tailed"/"one-sided" (e.g., going to college *improves* income) or "two-tailed"/"two-sided" (e.g., going to college *changes* income).

> **Example**: For a study to test the effects of continuous glucose monitoring on HbA1c, the null hypothesis would be that CGM has no effect on HbA1c. In the study design, the researchers could specify either a one-sided hypothesis, e.g., CGM decreases HbA1c, or a two-sided hypothesis, e.g., CGM affects (either increases or decreases) HbA1c.

P-value: Indicates the likelihood of observing the same data or more extreme data given that the null hypothesis is true. The null hypothesis states that there is no difference or no relationship between variables being tested. When this is the case, the p-value refers to the probability of getting the results "by chance alone," meaning "if no true relationship exists." **For example**, you are investigating the rates of lung cancer in smokers vs. non-smokers, and you observe higher rates in smokers with a p-value for your data of 0.03. This indicates that there is a 3% chance of observing the same or more extreme results by chance, if there were no relationship between smoking and lung cancer.

Type 1 error: A type 1 error in statistical analysis occurs when the **H_0** is rejected when **H_0** is actually true. A type 1 error is present when analysis of the data shows that there is a difference between two study groups/outcomes when in fact there is not. The probability of a Type 1 error is the false positive rate, defined as α. For most studies $\alpha = 0.05$, indicating that there is a 5% chance that a Type 1 error will be committed. When the p-value is $< \alpha$ (e.g., less than 0.05), the null hypothesis can be rejected, whereas H_0 cannot be rejected if p-value $> \alpha$.

Type 2 error: A type 2 error in statistical analysis occurs when the **H_A** is rejected when **H_A** is actually true. A type 2 error is present when analysis of the data shows that there is no difference between two study groups/outcomes, when in fact there is. The probability of a Type 2 error is the false negative rate, defined as β.

	H_0 =True	**H_0** = False
Reject **H_0**	False positive = Type 1 error (α)	True positive =Correct outcome
Fail to reject **H_0**	Correct outcome = True negative	False negative = Type 2 error (β)

Statistical Power: Power is the likelihood that a study will find a statistical difference between groups. Power is the probability that the $\mathbf{H_0}$ will be rejected. Power = 1-β which is the sensitivity of the test.

Power calculations: Power calculations are performed in order to plan statistical studies (***a priori)*** or to analyze the statistical power of data that has already been collected **(post hoc**). A variety of statistical software tools are available to perform power calculations (see below).

i. ***A priori* power calculations**: Used to determine the number of subjects that need to be recruited in order to detect a statistical difference between groups. Data from pilot studies, animal studies, or equivalent clinical studies can provide the tools for performing this calculation. The following information is defined in order to do this calculation 1) Power, 2) α, 3) mean, 4) standard deviation, and 5) type of study design (paired or independent).

ii. **Post-hoc power calculations:** Used to determine the power of a study based on the data that has already been collected. The following information is defined in order to do this calculation 1) Number of subjects in each group, 2) α, 3) mean, 4) standard deviation, and 5) type of study design (paired or independent). This determination is helpful for designing future experiments if statistical power is not reached.

Web-based tools are available for performing power calculations. Some helpful sites are:
1) http://www.dssresearch.com/KnowledgeCenter/toolkitcalculators/statisticalpowercalculators.aspx
2) http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize
Alternatively, you can seek assistance from a biostatistician.

Statistical Measures
Two-by-two table

	Disease	**No Disease**
Positive Test	**A**	**B**
Negative Test	**C**	**D**

Sensitivity: Likelihood that a person with the disease will have a true positive test (low false negative rate).

A/A+C

Specificity: Likelihood that a person without the disease will have a true negative test (low false positive rate).

D/C+D

Positive Predictive Value: Likelihood that a person with a positive test will have a disease.

A/A+B

Negative Predictive Value: Likelihood that a person with a negative disease will not have a disease.
D/C+D

Disease Prevalence: The number of individuals with the disease in a given population.
A+C/A+B+C+D

Disease Incidence: The number of new cases of disease during a set time period in a given population. Incidence rates are expressed in terms of unit of time.

Relative Risk: Also called the risk ratio. The relative risk is the ratio of incidence of disease in individuals with a risk factor/exposure to the incidence in individuals without a risk factor/exposure.

Disease/Outcome

(+)	(-)		
18	32	**Exposed**	**RR= (18/18+32) / (6/6+44) = 3**
6	44	**Control**	

Odds Ratio: The ratio of the probability that an event will occur in an exposed/at risk group vs. the probability that an event will occur in the non-exposed group. If an exposure does not increase the odds of the event, then the OR = 1.

Disease/Outcome

(+)	(-)		
18	32	**Exposed**	**OR= (18/32) / (6/44) = 4.12**
6	44	**Control**	

Relative risk and odds ratio cannot be used interchangeably. This link gives an excellent explanation of the difference between relative risk and odds ratio:
http://stats.org/stories/2008/odds_ratios_april4_2008.html

Number Needed to Treat: Number of subjects who would need to be treated in order to "prevent" an outcome. The optimal NNT is 1 – every person who is treated, benefits from the treatment. The higher the NNT, the less effective is the therapy. An example using the table above:

NNT = 1/(exposed event rate – control event rate)
NNT = 1/ ((18/18+32) – (6/6+44))
NNT = 4.17

Categorical Data: Data that fits into categories. Examples of categorical data include gender, race, age group (ie <20 years or >20 years), and presence or absence of a given disease. In data analysis, each category is assigned a numerical value.

Lucy Mastrandrea and Takara Stanley
& PES Fellow Education Committee, 10/19/12

Discrete Data: Data that has values equal to an integer. Examples of discrete data include number of family members, number of correct answers on a test, and number of medications an individual takes.

Continuous Data: Data that has values on a continuum. Examples of continuous data include age, height/weight, and insulin levels.

Mean: The average of all the data. The mean is calculated by adding all the observations and dividing the sum by the number of observations.

$$\bar{x} = \Sigma X/n$$

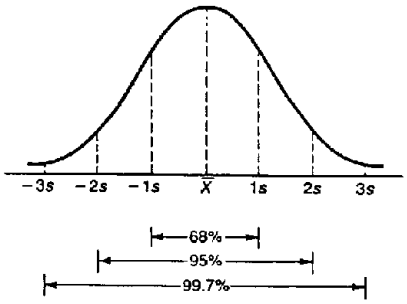
Median: The middle value of the data. To determine the median, arrange all the data from least to greatest value and then determine the value in the middle. If the number of data points is even, then the median is the mean of the two middle data points.

Mode: The data value that occurs most frequently. If there are two observations that occur most frequently, the data is described as bimodal.

Range: Lowest to highest data point.

Standard Deviation: Standard deviation is the spread of the data around the mean.
S=standard deviation; X=discrete data value; $\bar{x}$ = mean; n = number of observations

$$S = \sqrt{\frac{\Sigma(X-\bar{X})^2}{n-1}}$$



If the data is normally distributed then the standard deviation follows a bell-shaped curve. The distribution of the data about the mean plays a role in choosing tools for statistical analysis.

Z-score: Z-score is the number of standard deviations above or below the mean a data value is. For example, Z-score = +1 is 1 SD above the mean. Z-scores are particularly useful for continuous data that changes with age such as Body Mass Index. For populations, Z-score between -1 and +1 SD is representative of 68% of the population. For data that follow a normal distribution, a Z-score for a particular measurement X_i can be calculated as follows (SD = standard deviation and $\bar{x}$ = mean):

$$\frac{Xi - \overline{X}}{SD}$$

Standard Error of the Mean: The amount that a sample mean will vary from the true mean. SEM represents the variability in sampling means.
σ_m = standard error of mean; σ = standard deviation; N= number of observations

$$\sigma_m = \frac{\sigma}{\sqrt{N}}$$

Difference between SD and SEM: SD and SEM describe very different things. The standard deviation describes the distribution of a set of sample data, whereas the SEM describes how much the true population mean may vary from the mean obtained in the sample. ***For example,*** a large diabetes clinic wants to investigate the HbA1c levels of its patients, so its director randomly samples the most recent HbA1c levels from 225 patients and obtains data approximating a normal distribution with a mean of 8% and an SD of 1%. As shown in the bell curve above, the SD indicates that about 68% of patients have HbA1c between 7-9% (8±1SD), and about 95% of the patients have HbA1c between 6-10% (8±2SD). The SEM of this data is 1%/√225, or 1/15 = 0.07%. This SEM indicates that the mean for the sample, 8%, may vary from the true mean HbA1c of all the clinic patients by about 0.07%.

Confidence Intervals: A range of values that is expected to include the true mean of the population. Generally set as 95% confidence interval, i.e. 95% likelihood that the range will include the mean. Data with confidence intervals that cross zero are not statistically significant. Confidence intervals are also calculated for risk ratios and odds ratios.

t-test: A statistical hypothesis test that allows you to compare the means between two groups to determine whether or not they differ from each other. P-value is generally set at 0.05 – 95% likelihood that the difference between the means is not due to chance.

i. Independent (unpaired) t-test: Comparison of two populations when sample sizes and variances are equally distributed. An example of an unpaired t-test would be when you want to test the effect of a treatment and an equal number of subjects are randomized either to placebo or intervention.

ii. Dependent (paired) t-test: Comparison of matched pairs or one group that is tested twice. An example of a paired t-test would be when you test HbA1c in a group of subjects with Type 1 diabetes and then retest the HbA1c after an intervention in the same group of subjects.

ANOVA: Analysis of Variance. If multiple groups are compared by t-test, the likelihood of a Type 1 error is increased. ANOVA is a statistical hypothesis test used to compare the means of two or more groups. ANOVA assumes that variances between the groups are normal. If ANOVA testing tells you that there is a difference between the means, post-hoc analysis must be performed in order to determine which of the groups contributed to the differences.

Confounding variable: A variable that correlates both with the exposure variable and the outcome. Confounders may lead to establishment of associations between variables which do not exist. Mechanisms to prevent the effect of confounding variables include careful selection of subject populations and post-hoc multi-variable regression analysis to control for confounders.

Association: A relationship between two variables that is statistically dependent. Association does not imply causality.

Correlation: Gives the relationship between two variables. A relationship can be described as positive, negative, and can define the strength of the relationship. Correlations do not imply causality.

Correlation coefficient (r): A measure of the strength of the linear correlation between two variables. The range of values for correlation coefficient (r) is between -1 and 1 with -1 representing a 1:1 negative correlation between variables, 0 representing no correlation between variables, and 1 representing a 1:1 positive correlation between variables. The distribution of the variables determines which tool is used to calculate the correlation coefficient. Pearson's correlation coefficient is determined for variables that are normally distributed. When there are outlier variables, the Spearman rank correlation is used.

Coefficient of determination (r^2): Measures the strength of the relationship between two variables. It is the square of the correlation coefficient and has a range of values between 0-1, with 0 representing no relationship between the variables, and 1 representing a perfect relationship between the variables. **For example**, if one looked at the relationship between height (independent variable) and weight (dependent variable) and obtained an r^2 of 0.67, that would suggest that 67% of the variance in weight can be explained by variance in height.

Reference/Resourses:

(1) http://www.collemergencymed.ac.uk/CEM/Research/technical_guide/
(2) http://www.statsoft.com/textbook/
(3) http://stats.org/stories/2008/odds_ratios_april4_2008.html
(4) http://www.dssresearch.com/KnowledgeCenter/toolkitcalculators/statisticalpowercalculators.aspx
(5) http://biostat.mc.vanderbilt.edu/wiki/Main/PowerSampleSize