# Concepts of Clinical Trials

Samiran Sinha
Texas A&M University
*sinha@stat.tamu.edu*

November 25, 2021

# What are clinical trials and why do people participate?

Clinical trials are methods (or studies) for investigating safety, side-effects, and effectiveness of a drug or a medical device. Clinical trials are used to compare several drugs.

Healthy and unhealthy people (ill) participate in clinical trials. Healthy people do it to help the medical research or to get some financial benefit. People with condition (disease) take part in clinical trials with the hope that they help the science, possibly get a newer medicine, to improve the quality of their life, or get some financial benefit[1].

[1] Motives for participating in a clinical research trial: a pilot study in Brazil by Nappo et al. appeared in BMC Public Health 2013

# Some concepts–

Before a clinical trial, usually the therapy (treatment) is tested in laboratories, and tested on animals. Using trials on human subjects we gain more understanding about the therapy, how it works, its side effect, the effectiveness of the therapy etc.

Protocol: A clinical trial is carried out according to a plan known as a protocol. The protocol is carefully designed to safeguard the participants' health and answer specific research questions. A protocol includes the following aspect of the trial design:
a) trial purpose,
b) who are eligible to participate in the trial and how many subjects to be included,
c) details about tests, procedures, medications, and dosages,
d) the length of the study and what information will be gathered,
e) how the subjects to be followed,
f) when to stop the experiment,
g) informed consent form

# Some concepts (continues) –

A clinical study is usually led by a principal investigator (PI), who is often a doctor. Members of the research team regularly monitor the participants' health to determine the study's safety and effectiveness.

IRB review: Before the study is conducted, this study protocol is reviewed by Institutional Review Board (IRB). This review board ensures, safety, welfare, and confidentiality of the human subjects. IRB also makes sure that ethical issues are maintained in a trial.

# Some concepts (continues) –

Sponsors: Clinical trials are sponsored by the government agencies, pharmaceutical companies, other corporate organizations.

Informed consent: Informed consent is a process of informing potential participants about the potential risk and potential benefit of the trials. This helps subjects to decide to take part in the study.

# Some classification of clinical trials

- Two main types: observational and interventional trial (Thiese, 2014)
- Observational clinical trial: progression of a disease and the potentially influential factors are studied. Usually cohort, case-control, and cross-sectional studies are used to investigate these issues.
- Interventional trial: prevention trial, diagnostic trial, screening trial, treatment trials, quality of life trials

# Prevention trials

These trials look for better ways to prevent a disease in people who have never had the disease or to prevent the disease from returning. Better approaches may include medicines, vaccines, or lifestyle changes, among other things. In this trial, subjects initially do not have the disease or the condition. One example of a prevention trial is the IBIS 2 breast cancer prevention trial. The goal of this trial was to investigate if the drug anastrozole prevent breast cancer development in post menopausal women who were at a high risk of getting it. The trial results showed that the drug reduce the risk of developing breast cancer in these women.

# Diagnostic trials

These trials test the presence of a disease among the subjects with some symptoms. Here we are more concerned about the high specificity of the test (high chance of diagnosing a true negative as negative). In terms of probability, we want a high value for pr(test result negative|subject does not have the disease).

# Screening trials

Screening trials test the best way to detect the presence of certain diseases or health conditions at the early stage. For controlling or for cure, for some diseases it is important to detect their presence before the symptoms show up. For these type of diseases, such as bowel cancer and pancreatic cancer, there have been several screening trials.

# Quality of life trials

Quality of life trials (or supportive care trials) explore and measure ways to improve the comfort and quality of life of people with a chronic illness.

# Main phases of a clinical trial (for interventional trials)

Phase I trials: Researchers test an experimental drug or treatment in a small group of people for the first time. The purpose is – to find out if a new treatment is safe,
– to find the best way to give the new treatment, such as by mouth or by vein,
–to see if there are signs that cancer responds to the new treatment,
– placebos (sham or inactive treatments) are not part of phase I trials.

Phase I clinical trials often last several months to a year. For a new drug trial for cancer treatment, doctors offer treatment in phase I clinical trials to people whose cancers won't respond to standard treatments. Although phase 1 design is not for testing how well a treatment or combination of treatments works, the treatment given during this phase may help to slow or stop the growth of cancer. One of the goals of this trial is to figure out the highest dose level of the drug that can be applied without any serious side effect.

# Phase II

The purpose is
– to test safety of the drug,
– know more about possible side effects,
– how effective the treatment is,
– no placebo is used in this trial.

Usually a moderate number volunteer patients join a Phase II trial. Usually, subjects receive the highest dose of the treatment that turned out to be safe in phase-I trial. If the new treatment works, doctors may go on to study it in a Phase III trial.

# Phase III

The treatments that show effectiveness at least under some condition in Phase II trial go to Phase III trial.

The experimental drug or treatment is administered to a large group of people to confirm its effectiveness, monitor side effects, compare it with standard or equivalent treatments, and collect information that will allow the experimental drug or treatment to be used safely. Placebos may be used in some phase III studies, but they are never used alone if there is a treatment available that works.

# Phase IV

Phase IV trials are conducted after a drug is approved by the FDA and made available to the public. In this trial researchers track its long-term safety and side effect, seek more information about a drug or treatment's risks, benefits, and optimal use. Usually many subjects with diverse background are involved in this trial. [2]

[2] www.cancer.org

# What is randomization?

Random assignment of treatments among subjects is called randomization. This is mainly used in Phase II or Phase III trials. Randomization is used to remove the selection bias, and to minimize the effect of uncontrollable confounding variables. There are different kinds of randomization, see Suresh (2011) for details.

# Concept of blinding

Blinding is a procedure to keep people unaware of which treatment a subject is assigned to. This is done to avoid unconscious or conscious bias. A study (trial) can fall in one of the following categories based on blinding.

| Category | Details |
|---|---|
| un-blinded | All parties know the assigned treatment to a patient |
| single blinded | Only the patient does not know which treatment he/she is receiving |
| double blinded | The patient and the clinician/data collector do not know which treatment the patient is receiving |

# Some terminologies

- Primary endpoint: The main variable that is measured at the end of a study to see if a given treatment worked. This variable could be the time-to-event for every subject in the trial, or the number of events in each of the treatment groups. The primary endpoint has to be decided before the study begins; actually the study sample size is calculated based on the chosen primary endpoint. The primary endpoint is used to address the main hypothesis of the study. The primary endpoint (outcome) should be easy to measure and should be clinically validated. (Reference: `https://www.cancer.gov/publications/dictionaries/cancer-terms/def/primary-endpoint`)
- The secondary endpoint is the variable that is needed to address the secondary hypothesis of the study. Sample size may not be calculated for the secondary endpoint.

# Some terminologies

- In order to better use of clinical study, several markers are collected, and the markers that are associated with the primary endpoints are referred to as surrogate markers. Often studying the primary end points requires a large sample size and or lengthy (consequently costly) follow-up. In those cases, one may use surrogate markers to find the efficacy of a therapy.

# Some terminologies

- Suppose that in a clinical study where each subject is supposed followed-up for two years, and we want to record time-to-event during this follow-up period. There will be some subjects who will not experience the event of interest during the follow-up period, and we call them administratively censored at the closeout. For some subjects we may not have any information after a few days/months/a year, for these subjects clinicians do not have any information beyond these days/months/a year. These subjects are called lost to follow-up.
- Loss to follow-up subjects are considered censored during the analysis, and the implicit assumption is that these subjects (lost to follow-up) do not carry any extra/new information about the time-to-event.

# Some terminologies

- Disease-free survival: In cancer, the length of time after primary treatment for a cancer ends that the patient survives without any signs or symptoms of that cancer. In a clinical trial, measuring the disease-free survival is one way to see how well a new treatment works. Also called DFS, relapse-free survival, and RFS. (Ref: NCI)
- Progression-free survival: The length of time during and after the treatment of a disease, such as cancer, that a patient lives with the disease but it does not get worse. In a clinical trial, measuring the progression-free survival is one way to see how well a new treatment works. Also called PFS. (Ref: NCI)
- One of the goals of the study described in the article *Progression-free survival as a surrogate endpoint of overall survival in patients with metastatic colorectal cancer* is to check if PFS can be used as a surrogate endpoint of the primary endpoint, the overall survival of the patients.

# Some terminologies

- Confounding variable: A variable(s) that influences both the outcome and the exposure independently. A confounder cannot be an effect of the treatment or an intermediate variable in the causal pathway between the exposure and the outcome. For instance, the relationship between diet and coronary heart disease may be explained by measuring serum cholesterol level. Cholesterol may not a confounder because it may well lie in the causal pathway between diet and coronary heart disease. [3]
- For estimating the treatment effect proper care must be taken to remove the possible confounding effect. Confounding happens when the treatment effect and the effect of other pre-treatment variables are mixed-up.

[3] Hernan MA, Robins JM. Causal Inference. I: Causal Inference Without Models 2013, `http://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/`.

# Some terminologies

- Prognostic factor: A prognostic factor is a variable (some scientists refer it to as parameter) that is associated with clinical outcome in the absence of therapy or with the application of a standard therapy that patients are likely to receive. [4]
- A control group from a randomized clinical trial is an ideal setting for evaluating the prognostic significance of a variable (most likely clinical).

[4] Clark, et al. Clinical utility of epidermal growth factor receptor expression for selecting patients with advanced non-small cell lung cancer for treatment with Erlotinib. *Journal of Thoracic Oncology*, **1**, 837–846, 2006.

# Screening trials

- Screening tests are used to detect a health condition before clinical symptoms appear. For example, mammography test is now used for screening breast cancer for women of age 40 and above. The purpose introducing these tests is to reduce the burden of the disease on the society.
- Screening trials are used to evaluate a new screening test for detecting cancer and other health conditions in people before symptoms are present. The goal of such trials is to study whether or not the screening test saves lives and at what cost.

# Treatment trials

- Treatment trials: As the name indicates it involves one or two treatments. Phases I, II, and III are part of treatment trials. In the early phases, side-effects and safe dose-levels are studied.
- Suppose $\Delta$ denotes the difference in efficacy between the two treatments, treatment of interest (A) and active control or placebo (B). Specifically, $\Delta$ =Efficacy of treatment A−Efficacy of treatment B.
- Superiority trial: As the name indicates the purpose of a superiority trial to check if a drug (therapy) is superior than another drug. In this case $H_0 : \Delta \leq 0$ versus $H_a : \Delta > 0$. A significant result ($p$-value smaller than a desired level) implies that A is superior than B. However, a non-significant result does not indicate that A and B are equivalent.

# Superiority trial:

- The superiority of the treatment can be measured via a 1) numeric measurement (e.g., reduction in systolic blood pressure), 2) binary response (e.g., improved versus no-improvement), or 3) time-to-event (e.g., time to become symptom free).
- The analysis strategy depends on the type of outcome. First consider the case where the main outcome of interest is a numeric variable. Suppose that
  - $X_1, X_2, \ldots, X_m$ is a random sample from a population with mean $\mu_1$ and variance $\sigma_1^2$.
  - $Y_1, Y_2, \ldots, Y_n$ is a random sample from a population with mean $\mu_2$ and variance $\sigma_2^2$.
  - The $X$ and $Y$ samples are independent of one another.
  - The interest is in testing $H_0 : \mu_1 = \mu_2$ against ones of these alternatives: $H_a : \mu_1 > \mu_2$ or $H_a : \mu_1 < \mu_2$

We will use $\bar{X} - \bar{Y}$ to estimate $\mu_1 - \mu_2$.

- Note that:

$$\begin{aligned} E(\bar{X} - \bar{Y}) &= \frac{1}{m} E[\sum_{i=1}^{m} X] - \frac{1}{n} E[\sum_{j=1}^{n} Y] \\ &= \frac{1}{m} m\mu_1 - \frac{1}{n} n\mu_2 = \mu_1 - \mu_2. \end{aligned}$$

  So, $\bar{X} - \bar{Y}$ is an unbiased estimator of $\mu_1 - \mu_2$,

- and the variance of $(\bar{X} - \bar{Y})$ is:

$$\begin{aligned} \text{Var}(\bar{X} - \bar{Y}) &= \frac{1}{m^2} \text{Var}(\sum_{i=1}^{m} X_i) + \frac{1}{n^2} \text{Var}(\sum_{i=1}^{n} Y_i) \\ &= \frac{1}{m^2} m\sigma_1^2 + \frac{1}{n^2} n\sigma_2^2 = \frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n} \end{aligned}$$

  when $X$ and $Y$ are independent of one another.

In general for testing $H_0 : \mu_1 - \mu_2 = \Delta_0$, the test statistic is

$$TS = \frac{\bar{X} - \bar{Y} - \Delta_0}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}}$$

| Alternative hypothesis | Rejection region for level $\alpha$ test |
|---|---|
| $H_a : \mu_1 - \mu_2 > \Delta_0$ | $TS \geq Z_\alpha$ |
| $H_a : \mu_1 - \mu_2 < \Delta_0$ | $TS \leq -Z_\alpha$ |
| $H_a : \mu_1 - \mu_2 \neq \Delta_0$ | $TS \geq Z_{\alpha/2}$ or $TS \leq -Z_{\alpha/2}$ |

# Probability of type-II error

Let $\alpha = 0.05$, so $Z_{\alpha/2} = 1.96$.

$$
\begin{aligned}
\beta(\Delta') &= P(\text{Do not reject } H_0 \text{ when } \mu_1 - \mu_2 = \Delta') \\
&= P(-1.96 < TS < 1.96 \text{ when } \mu_1 - \mu_2 = \Delta') \\
&= P(-1.96 < \frac{\bar{X} - \bar{Y}}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}} < 1.96 \text{ when } \mu_1 - \mu_2 = \Delta') \\
&= P(-1.96\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}} < (\bar{X} - \bar{Y}) < 1.96\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}} \\
&\qquad \text{when } \mu_1 - \mu_2 = \Delta') \\
&= P(-1.96 - \frac{\Delta'}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} < \frac{(\bar{X} - \bar{Y}) - \Delta'}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} < 1.96 - \frac{\Delta'}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}) \\
&= \Phi(1.96 - \frac{\Delta'}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}) - \Phi(-1.96 - \frac{\Delta'}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}}),
\end{aligned}
$$

where $\Phi$ denotes the CDF of the Normal$(0, 1)$ distribution.

# Sample size calculation

Now, suppose that for given effect size $\Delta^{'}$, $\alpha = 0.05$, and power 80% we want to determine the sample size. Recall $\beta(\Delta^{'}) = 1-$power. So, for 80% power, set

$$0.20 = \beta(\Delta^{'}) \quad = \Phi(1.96 - \frac{\Delta^{'}}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}) - \Phi(-1.96 - \frac{\Delta^{'}}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}).$$

Next set $\Phi(-1.96 - \Delta^{'}/\sqrt{\sigma_1^2/m + \sigma_2^2/n}) = 0$, then

$$0.20 = \Phi(1.96 - \frac{\Delta^{'}}{\sqrt{\frac{\sigma_1^2}{m}+\frac{\sigma_2^2}{n}}}).$$

Since $\text{pr}(Z < -0.84) = 0.2$, setting $m = n$, we obtain

$$-0.84 = 1.96 - \frac{\Delta^{'}}{\sqrt{\frac{\sigma_1^2}{n}+\frac{\sigma_2^2}{n}}}$$

so

$$n = \frac{(0.84 + 1.96)^2(\sigma_1^2 + \sigma_2^2)}{(\Delta^{'})^2}.$$

# Sample size calculation

- The general rule is for the level $\alpha$ test with power $1-\beta$, the required sample size in each arm for an equal allocation trial is

$$n = \frac{(Z_\beta + Z_{\alpha/2})^2(\sigma_1^2 + \sigma_2^2)}{(\Delta')^2},$$

  where $\text{pr}(Z > Z_{\alpha/2}) = \alpha/2$ and $\text{pr}(Z > Z_\beta) = \beta$. This formula is for the two-sided test.

- If assumed that $\sigma_1^2 = \sigma_2^2 = \sigma^2$, and allocation ratio is $r = n_b/n_a$, then

$$n_a = \frac{(1+r)(Z_\beta + Z_{\alpha/2})^2\sigma^2}{r(\Delta')^2}.$$

- If the interest is in the one-sided test, which maybe the case for some superiority trials, then we should replace $Z_{\alpha/2}$ by $Z_\alpha$ in the above formula.
- If we know $100p^*\%$ of the subjects will be lost to follow-up, then the sample sizes must be $n_a/(1-p^*)$ and $n_a/(1-p^*)$.

# Sample size calculation, example

In the CACTUS clinical trial[5], participants suffering from long-standing aphasia post stroke are to be randomized to one of three arms: (1) usual care, (2) self-managed computerized speech and language therapy in addition to usual care and (3) attention control in addition to usual care.

The primary objective of the trial is to compare usual care with self-managed computerized therapy. The primary outcome of interest is the change in the number of words named correctly at 6 months. The minimal clinically meaningful difference is improvement in word retrieval of 10%. Based on a pilot study, the population standard deviation is assumed to be 17.38%. The dropout rate applied is 15% (95% confidence interval (CI), 5–32%). [5]

[5] Palmer R, Enderby P, Cooper C, Latimer N, Julious S, Paterson G, Dimairo M, Dixon S, Mortley J, Hilton R, Delaney A, Hughes H. Computer therapy compared with usual care for people with long-standing aphasia poststroke: a pilot randomized controlled trial. Stroke 2012; 43:1904–1911.

# Sample size calculation, example

With $Z_{.1} = 1.28$, $Z_{0.025} = 1.96$,

$$n_a = \frac{(1+1) \times (1.28 + 1.96)^2 17.38^2}{1 \times 10^2} = 63.41.$$

Considering 15% drop-out, the needed sample size for each arm is $75 (\approx 63.41/(1 - 0.15))$

# Pooled Variance Estimator

The estimator of our common variance $\sigma^2$ is called the pooled variance estimator and is denoted by $s_p^2$.
Remember that we require that $\sigma_1^2 = \sigma_2^2$. Therefore the two sample standard deviations $s_1$ and $s_2$ are really two estimates of the same quantity $\sigma$. Our estimate for $\hat{\sigma}^2 = s_p^2$ is a weighted average of the sample variances $s_1$ and $s_2$:

$$\begin{aligned} s_p^2 &= \frac{m-1}{m+n-2} s_1^2 + \frac{n-1}{m+n-2} s_2^2 \\ &= \frac{(m-1)s_1^2 + (n-1)s_2^2}{m+n-2} \end{aligned}$$

# Pooled $t$-test

Given the random variable T, we can construct a test and confidence interval for two Normal population means given $\sigma_1^2 \approx \sigma_2^2$: Null hypothesis: $H_0 : \mu_1 - \mu_2 = \Delta_0$
Test statistic:

$$TS = \frac{\bar{x} - \bar{y} - \Delta_0}{s_p\sqrt{1/m + 1/n}}$$

| Alternative hypothesis | Rejection region for level $\alpha$ test |
|---|---|
| $H_a : \mu_1 - \mu_2 > \Delta_0$ | $t \geq t_{\alpha,m+n-2}$ |
| $H_a : \mu_1 - \mu_2 < \Delta_0$ | $t \leq -t_{\alpha,m+n-2}$ |
| $H_a : \mu_1 - \mu_2 \neq \Delta_0$ | $t \geq t_{\alpha/2,m+n-2}$ or $t \leq -t_{\alpha/2,m+n-2}$ |

A $100(1-\alpha)\%$ confidence interval for $\mu_1 - \mu_2$ is:

$$\bar{x} - \bar{y} \pm t_{\alpha/2,m+n-2} s_p \sqrt{\frac{1}{m} + \frac{1}{n}}$$

# Inferences for the Difference of Two Population Proportions

Two compare two population proportions $\pi_1$ and $\pi_2$ we will use the statistic: $\hat{p}_1 - \hat{p}_2 = X/m - Y/n$. As always, to construct a test we need to know the 1) the mean, 2) the variance of our statistic and 3) the distribution,

- The mean:

$$E(\hat{p}_1 - \hat{p}_2) = E\left(\frac{X}{m} - \frac{Y}{n}\right) \quad = \frac{1}{m}E(X) - \frac{1}{n}E(Y)$$
$$= \frac{1}{m}mp_1 - \frac{1}{n}np_2 = p_1 - p_2$$

Therefore the estimator $\hat{p}_1 - \hat{p}_2$ is unbiased for $p_1 - p_2$.

- The variance:

$$V(\hat{p}_1 - \hat{p}_2) = V\left(\frac{X}{m} - \frac{Y}{n}\right) \quad = \frac{1}{m^2}V(X) + \frac{1}{n^2}V(Y)$$
$$= \frac{1}{m^2}mp_1q_1 + \frac{1}{n^2}np_2q_2 = \frac{p_1q_1}{n} + \frac{p_2q_2}{m}$$

- We know that if $n$ and $m$ are large and

$$X \sim \text{binomial}(m, p_1)$$

$$Y \sim \text{binomial}(n, p_2)$$

then $\hat{p}_1$ and $\hat{p}_2$ have an approximate normal distribution.
Given the mean, variance and distribution we know that the random variable:

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{m} + \frac{p_2 q_2}{n}}}$$

has an approximate standard normal distribution. We use this variable as the basis of our large sample test concerning the difference between two population proportions.

Under the null hypothesis we assume $p_1 = p_2$, therefore the estimates $\hat{p}_1$ and $\hat{p}_2$ are really two separate estimate of the same quantity $p = p_1 = p_2$. We combine the two estimates $\hat{p}_1$ and $\hat{p}_2$ into one:

$$\hat{p} = \frac{X + Y}{m + n} = \left(\frac{m}{m + n}\right)\hat{p}_1 + \left(\frac{n}{m + n}\right)\hat{p}_2$$

The revised random variable:

$$Z = \frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sqrt{\hat{p}\hat{q}(\frac{1}{n} + \frac{1}{m})}}.$$

Null hypothesis:

$$H_0 : p_1 - p_2 = 0$$

Test statistic:

$$TS = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}\hat{q}(\frac{1}{m} + \frac{1}{n})}}$$

where

$$\hat{p} = \frac{X + Y}{m + n} = \frac{m}{m + n}\hat{p}_1 + \frac{n}{m + n}\hat{p}_2$$

| Alternative hypothesis | Rejection region for level $\alpha$ test |
|---|---|
| $H_a : p_1 - p_2 > 0$ | $TS \geq Z_\alpha$ |
| $H_a : p_1 - p_2 < 0$ | $TS \leq -Z_\alpha$ |
| $H_a : p_1 - p_2 \neq 0$ | $TS \geq Z_{\alpha/2}$ or $TS \leq -Z_{\alpha/2}$ |

$p$-values are calculated the same way as for all Z-tests.

A $100(1-\alpha)\%$ confidence interval for $p_1 - p_2$:

$$\hat{p}_1 - \hat{p}_2 \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}_1\hat{q}_1}{m} + \frac{\hat{p}_2\hat{q}_2}{n}}$$

# Type II Error Probabilities

The derivation of type II error probabilities follows the same steps as before. For an upper tailed test $H_a : p_1 - p_2 > 0$:

$$\beta(p1, p2) = P(H_0 \quad \text{is not rejected when } p_1 - p_2 > 0)$$

For this test the rejection region is $Z \geq Z_\alpha$, therefore the type II error is:

$$\begin{aligned} \beta(p_1, p_2) &= P(Z < Z_\alpha) \\ &= P\left(\hat{p}_1 - \hat{p}_2 < Z_\alpha \sqrt{\hat{p}\hat{q}(1/m + 1/n)}\right) \\ &= P\left(\frac{\hat{p}_1 - \hat{p}_2 - (p_1 - p_2)}{\sigma} < \frac{Z_\alpha \sqrt{\hat{p}\hat{q}(1/m + 1/n)} - (p_1 - p_2)}{\sigma}\right) \end{aligned}$$

where $\sigma$ is the standard deviation under the alternative hypothesis:

$$\sigma = \sqrt{\frac{p_1 q_1}{n} + \frac{p_2 q_2}{m}}$$

# Type II Error Formulas for $\beta(p_1, p_2)$

We can do similar derivations for all three case of our alternative hypothesis:

| Alternative hypothesis | Type II error for size $\alpha$ test |
|---|---|
| $H_a : p_1 - p_2 > 0$ | $\Phi\left(\frac{Z_\alpha\sqrt{\bar{p}\bar{q}(1/m+1/n)}-(p_1-p_2)}{\sigma}\right)$ |
| $H_a : p_1 - p_2 < 0$ | $1-\Phi\left(\frac{-Z_\alpha\sqrt{\bar{p}\bar{q}(1/m+1/n)}-(p_1-p_2)}{\sigma}\right)$ |
| $H_a : p_1 - p_2 \neq 0$ | $\Phi\left(\frac{Z_{\alpha/2}\sqrt{\bar{p}\bar{q}(1/m+1/n)}-(p_1-p_2)}{\sigma}\right)$ $-\Phi\left(\frac{-Z_{\alpha/2}\sqrt{\bar{p}\bar{q}(1/m+1/n)}-(p_1-p_2)}{\sigma}\right)$ |

where

$$\bar{p} = \frac{mp_1 + np_2}{m+n} \qquad \bar{q} = 1 - \bar{p} = \frac{mq_1 + nq_2}{m+n}$$

and

$$\sigma = \sqrt{\frac{p_1 q_1}{m} + \frac{p_2 q_2}{n}}$$

# Sample Sizes and Type II Errors for $\beta(p_1, p_2)$

Given a known alternative $\beta(p_1, p_2)$, we can calculate the required sample size necessary to have a type II error of $\beta$:
For an upper tailed test when we solve

$$\beta = \Phi\left(\frac{Z_\alpha\sqrt{\bar{p}\bar{q}(1/m + 1/n)} - (p_1 - p_2)}{\sigma}\right)$$

we find

$$m = n = \frac{\left[Z_\alpha\sqrt{(p_1 + p_2)(q_1 + q_2)/2} + Z_\beta\sqrt{p_1q_1 + p_2q_2}\right]^2}{d^2}$$

where $d = p_1 - p_2$. The same results holds for a lower tailed test.

For a two tailed test we replace $\alpha$ by $\alpha/2$.

# Sample size calculation example

Suppose that 60% of subjects respond to a standard treatment. If the response rate for the new treatment is at least 20% better than the current success rate then the new treatment would be considered to be superior. What would be the required sample size to establish superiority of the new treatment with 80% power and level 0.05.

Set $p_1 = 0.72$, $p_2 = 0.60$, $Z_{.2} = 0.84$ and $Z_{0.025} = 1.96$, then

$$\begin{aligned} m = n &= \frac{\left[1.96\sqrt{(0.72+0.6)(0.28+0.4)/2} + 0.84\sqrt{0.72\times 0.28 + 0.6\times 0.4}\right]^2}{0.12^2} \\ &= 243.12. \end{aligned}$$

Thus, needed sample size for every arm is 244.

# Some terminologies

- Equivalence trial: As the name indicates the purpose of an equivalence trial is to verify if two treatment groups have equal efficacy. Equivalence trials are usually used in making generic drugs.
- Suppose that clinicians agree that when $\Delta$, a measure of the difference of efficacy, lies in a small interval $(-\Delta_*, \Delta_*)$, where $\Delta_* > 0$, then the two treatments can be considered to be equivalent.
- Consequently, we can express the statistical hypotheses as $H_0 : \Delta \leq -\Delta_*$ or $\Delta \geq \Delta_*$ versus $H_a : -\Delta_* < \Delta < \Delta_*$.
- If a $(1-\alpha)100\%$ CI for $\Delta$ is completely within $(-\Delta_*, \Delta_*)$, then the test is statistically significant at the $100\alpha\%$ level of significance. A significant result ($p$-value smaller than a desired level $\alpha$) implies that A and B are equivalent. However, a non-significant result indicates that A and B are not-equivalent.

# Sample size calculation

- Suppose that for two treatments, $A$ and $B$, the success probabilities are $\pi_a$ and $\pi_b$, respectively.
- Define $\Delta = \pi_a - \pi_b$. If $\Delta \in (-\Delta^*, \Delta^*)$, then $A$ and $B$ are equivalent, otherwise not.
- So, the required sample size, for the type-I error rate $\alpha$, type-II error rate $\beta$, a given equivalence margin $\Delta^*$, and the success probabilities $\pi_a$ and $\pi_b$ is

$$m = n = \frac{(Z_{\alpha/2} + Z_\beta)^2 \{\pi_a(1-\pi_a) + \pi_b(1-\pi_b)\}}{\Delta^{*,2}}.$$

- The choice of $\Delta^*$ is one of the most critical issues of equivalence trials.

# Non-inferiority trial (NIT):

Define $\psi$ =Efficacy of treatment B (active control)− Efficacy of treatment A. Suppose that $\psi_*$ $(> 0)$is the maximum acceptable extent of clinical noninferiority of an experimental treatment. In this trial we test $H_0 : \psi \geq \psi_*$(A is inferior to B) versus $H_a : \psi < \psi_*$ (A is non-inferior to B).
The alternative hypothesis ($H_a$) states that the new treatment may have a negative effect compared to the active control, but by no more than $\psi_*$. For a new treatment to at least have better efficacy than the active control while it can be less effective than the active control within the extent of $\psi_*$, the size of the margin allowed to the maximum limit would be the entire effect size of the control treatment.

- Suppose that new treatment has been developed that is easier to access and use, has fewer side effects, and less costly that that of the standard treatment. If the new treatment offers these more appealing characteristics, then the research goal may be to show that its effectiveness is not much less than that of the standard treatment.
- A critical requirement for this trial is the existence of standard treatment. Also, there must be a genuine need/question for the equivalence or noninferiority of the new treatment. If the new treatment is expected to perform better than the standard treatment, a superiority trial must be used.
- In a NI trial, the interest is in checking if the efficacy is not worse than a set margin. On the other hand, in an EQ trial, the interest is in checking if the efficacy is not less than certain limit and not more than a certain limit.

# Sample size, dichotomous outcome

Let $p_s$ and $p_d$ be the proportion of *success* in the standard and treatment arms, respectively. The level of significance/type-I error rate is $\alpha$ and the type-II error rate is $\beta$. Assume we have equal allocation (equal number of subjects in both arms), and the margin (this margin could be equivalence margin or the margin for the non-inferiority) is $\Delta^*$. The required sample size for any arm for the non-inferiority trial is

$$n = \frac{2(Z_\alpha + Z_\beta)^2 p_s(1-p_s)}{\Delta^{*,2}}.$$

By changing the variance formula $2p_s(1-p_s)$, we may get the alternative form

$$n = \frac{(Z_\alpha + Z_\beta)^2 \{p_s(1-p_s) + p_d(1-p_d)\}}{\Delta^{*,2}}.$$

Another alternative is

$$n = \frac{2(Z_\alpha + Z_\beta)^2 \overline{p}(1-\overline{p})}{\Delta^{*,2}},$$

where $\overline{p} = (p_d + p_s)/2$. Although the first one is commonly used, one may use the other two formulea. However, one should clearly state which formula is used to calculate the sample size. Note that in all these formulae, one-sided test is used and use $Z_\alpha$ not $Z_{\alpha/2}$.

# Sample size, dichotomous outcome

The required sample size for any arm for an equivalence trial is

$$n = \frac{2(Z_{\alpha/2} + Z_{\beta})^2 \overline{p}(1-\overline{p})}{\Delta^{*,2}}.$$

By changing the variance formula $2\overline{p}(1-\overline{p})$, we may get the alternative form

$$n = \frac{(Z_{\alpha/2} + Z_{\beta})^2 \{p_s(1-p_s) + p_d(1-p_d)\}}{\Delta^{*,2}}.$$

One may use either of the formulea. However, one should clearly state which formula is used to calculate the sample size. Note that in all these formulae, two-sided test is used and use $Z_{\alpha}/2$ not $Z_{\alpha}$.

# Sample size, dichotomous outcome

The required sample size for any arm for a superiority trial is

$$n = \frac{\{Z_{\alpha/2}\sqrt{(p_d + p_s)(\overline{p}_d + \overline{p}_s)/2} + Z_{\beta}\sqrt{p_d\overline{p}_d + p_s\overline{p}_s}\}^2}{(p_d - p_s)^2}.$$

Although there is a longstanding debt whether one-sided or two sided test to be used in the superiority trial, the regulatory agencies nowadays recommend to use two-sided test, so $Z_\alpha/2$ should be used in the formula. Here, $\overline{p}_d = 1 - p_d$, $\overline{p}_s = 1 - p_s$.
For clinical superiority trial use the formula

$$n = \frac{\{Z_{\alpha/2}\sqrt{(p_d + p_s)(\overline{p}_d + \overline{p}_s)/2} + Z_{\beta}\sqrt{p_d\overline{p}_d + p_s\overline{p}_s}\}^2}{(|p_d - p_s| - \Delta^*)^2}.$$

# Sample size, numeric outcome

Let $\mu_s$ and $\mu_d$ be the mean value of the numeric response in the standard and treatment arms, respectively. The level of significance/type-I error rate is $\alpha$ and the type-II error rate is $\beta$. Assume we have equal allocation (equal number of subjects in both arms), and the margin (this margin could be equivalence margin or the margin for the non-inferiority) is $\Delta^*$. Let $s$ be the standard deviation of the underlying population.

| Design (trial) | Required sample size |
|---|---|
| Non-inferiority | $n = 2(Z_\alpha + Z_\beta)^2 s^2/\Delta^{*,2}$ |
| Equivalence | $n = 2(Z_{\alpha/2} + Z_\beta)^2 s^2/\Delta^{*,2}$ |
| Superiority | $n = 2(Z_{\alpha/2} + Z_\beta)^2 s^2/(\mu_d - \mu_s)^2$ |
| Clinical superiority | $n = 2(Z_{\alpha/2} + Z_\beta)^2 s^2/(\lvert\mu_d - \mu_s\rvert - \Delta^*)^2$ |

# Protocol deviation

A protocol deviation occurs when a patient departs from the defined experimental procedure (either drops out from the study, does not follow the instructions of taking the medicine etc)

Here is one such example. A randomized double-blind trial compared a low dose of new antidepressant with a high dose of the drug and with a control treatment. The data are

| Effectiveness measure | low | high | control | |
|---|---|---|---|---|
| very effective | 2 | 8 | 6 | |
| effective | 4 | 2 | 8 | |
| ineffective | 3 | 2 | 0 | |
| total assessed | 9 | 12 | 14 | 35 |
| withdrawn | 6 | 8 | 1 | 15 |

# Protocol deviation– continues

So the withdrawn patients do not follow the protocol– a violation. Here are the options for analyzing such data: per protocol analysis and intent-to-treat analysis.

per protocol analysis: ignore the withdrawn subjects in the analysis

intention-to-treat analysis: include the withdrawn subjects and consider their responses as 'ineffective'

The results of these two analyses could be very different– see the next page

# Protocol deviation– continues

□: proportions based on per-protocol analysis
□: proportions based on intent-to-treat analysis

| Effectiveness measure | low | high | control | |
|---|---|---|---|---|
| very effective | 0.22(0.13) | 0.66(0.4) | 0.43(0.40) | |
| effective | 0.44(0.27) | 0.17(0.1) | 0.57(0.53) | |
| ineffective | 0.33(0.60) | 0.17(0.5) | 0.00(0.07) | |
| total assessed | 9 | 12 | 14 | 35 |
| withdrawn | 6 | 8 | 1 | 15 |

- Intent to treat (ITT) analysis is generally recommended for the superiority trial. ITT generally reduces the treatment effect, minimizing the difference between groups– in turn favoring the null hypothesis of no difference. Should we reduce the level of significance to make ST more conservative?
- ICH E10 (2001)[6] and CONSORT[7] also recommend to use ITT for the analysis of noninferiority trials.
- However, in the NI trial, because the null and alternative hypotheses are reversed, a dilution difference of the two groups actually favors the alternative hypothesis, making it more likely that true inferiority is masked– leading a higher Type-II error. Does increasing the Type-I error rate help to reduce the Type-II error rate?

[6] International Conference on Harmonisation *ICH Topic E10: Choice of control group and related issues in clinical trials*. 2001.

[7] Piaggio G, Elbourne DR, Altman DG, Pocock SJ, Evans SJ. Reporting of noninferiority and equivalence randomized trials: An extension of the CONSORT statement. *Journal of the American Medical Association*. 2006;295:1152D1160.

# Interim analysis

- This is the analysis of the trial data before the clinical trial is completed. The analysis is based on the partial data from the trial. The plan for the interim analysis including the statistical approaches should be disclosed in the protocol.
- Any adverse (unexpected) result of interim analysis may lead to early stopping of a trial, appropriate modification in the sample size, study design. Alternatively, expected result of the interim analysis may call for an early declaration of success. This approach allows clinical researchers to employ the same basic management principles as typical modern businesses, using real-time data and analysis to inform decisions that continually optimize operations. [8]

[8] Kumar A, Chakraborty BS. Interim analysis: A rational approach of decision making in clinical trial. *J Adv Pharm Technol Res.* 2016; 7(4):118–122.

# Crossover trials

- Suppose that there are two treatments $A$ and $B$.
- Subjects are randomly divided into two groups.
- One group receives treatment sequence A and then B, while the other group receives treatment sequence B and then A. It is usually referred to as AB/BA study.
- Thus in the AB arm subjects receive treatment A first, followed by treatment B, and vice versa in the BA arm.
- Crossover trials allow the response of a subject to treatment A to be compared with the same subject's response to treatment B. Removing patient variation due to several confounding factors in this way makes crossover trials potentially more efficient than similar sized, parallel group trials in which each subject is exposed to only one treatment. In theory treatment effects can be estimated with greater precision given the same number of subjects.

# Crossover trials

- Crossover trials are generally used for chronic diseases or processes because the disease or process needs to persist long enough for the investigator to expose the subject to each of the experimental treatments and measure the response.
- Treatments must be the one that do not permanently alter the disease or process under study.

# Crossover trials

- Although the design can remove the potential confounding effect and aims to estimate the effect of the treatment accurately, the effects of one treatment may be carried over and alter the response to subsequent treatments.
- The usual approach to preventing this is to introduce a washout period between consecutive treatments that is long enough to allow the effects of a treatment to wear off. [9]

[9] Sibbald, B and Roberts, C. (1998). Understanding controlled trials Crossover trials, *BMJ*, 316, 1719–1720.

# Crossover trials with continuous endpoint

- Suppose that there are two arms AB and BA, and there are $n_1$ and $n_2$ subjects in arms AB and BA, respectively.
- Denote the continuous end point as $Y$.
- Define $Y_{AB,i,j}$ as the response of the $j$th subject in period $i$, where $j = 1, \ldots, n_1$, $i = 1, 2$, in the AB arm. Similarly, $Y_{BA,i,j}$, $i = 1, 2$, $j = 1, \ldots, n_2$ is defined.

| Sequence | Period 1 | Period 2 |
|---|---|---|
| AB | $Y_{AB,1,1}, \ldots, Y_{AB,1,n_1}$ | $Y_{AB,2,1}, \ldots, Y_{AB,2,n_1}$ |
| BA | $Y_{BA,1,1}, \ldots, Y_{BA,1,n_2}$ | $Y_{BA,2,1}, \ldots, Y_{BA,2,n_2}$ |

# Crossover trials with continuous endpoint

- Here is the mean model for each cell.
- $\mu_A$ ($\mu_B$): mean effect of treatment A (B)
- $\gamma$ ($-\gamma$): mean effect of period 1 (2)
- $\beta$ ($-\beta$): mean effect of sequence AB (BA)
- $\lambda_A(\lambda_B)$ : carry over effect of A (B)
- If A (B) is a placebo, then the corresponding carry-over effect is zero, $\lambda_A = 0$ ($\lambda_B = 0$).

| Sequence | Period 1 | Period 2 |
|---|---|---|
| AB | $\mu_A + \gamma + \beta$ | $\mu_B - \gamma + \beta + \lambda_A$ |
| BA | $\mu_B + \gamma - \beta$ | $\mu_A - \gamma - \beta + \lambda_B$ |

# Crossover trials with continuous endpoint

- $\overline{Y}_{AB,1} = \sum_{j=1}^{n_1} Y_{AB,1,j}/n_1$ estimates $\mu_A + \gamma + \beta$.
- $\overline{Y}_{AB,2} = \sum_{j=1}^{n_1} Y_{AB,2,j}/n_1$ estimates $\mu_B - \gamma + \beta + \lambda_A$.
- $\overline{Y}_{BA,1} = \sum_{j=1}^{n_2} Y_{BA,1,j}/n_2$ estimates $\mu_B + \gamma - \beta$.
- $\overline{Y}_{BA,2} = \sum_{j=1}^{n_2} Y_{BA,2,j}/n_2$ estimates $\mu_A - \gamma - \beta + \lambda_B$.
- So, $(\overline{Y}_{AB,1} + \overline{Y}_{BA,2})/2$ estimates $(\mu_A + \gamma + \beta + \mu_A - \gamma - \beta + \lambda_B)/2 = \mu_A + \lambda_B/2$.
- So, $(\overline{Y}_{AB,2} + \overline{Y}_{BA,1})/2$ estimates $(\mu_B - \gamma + \beta + \lambda_A + \mu_B + \gamma - \beta)/2 = \mu_B + \lambda_A/2$.

| Sequence | Period 1 | Period 2 |
|---|---|---|
| AB | $\overline{Y}_{AB,1}$ | $\overline{Y}_{AB,2}$ |
| BA | $\overline{Y}_{BA,1}$ | $\overline{Y}_{BA,2}$ |

# Crossover trials with continuous endpoint

- If there is no carry-over effect $\lambda_A = \lambda_B = 0$, then $(\overline{Y}_{AB,1} + \overline{Y}_{BA,2})/2$ will estimate the effect of treatment A, while $(\overline{Y}_{AB,2} + \overline{Y}_{BA,1})/2$ will estimate the effect of treatment B.
- Most importantly, if the two carry-over effects are equal, $\lambda_A = \lambda_B$, the difference of treatment efect $\mu_A - \mu_B$ will be estimated by

$$\begin{aligned}
\widehat{\mu_A - \mu_B} &= \frac{\overline{Y}_{AB,1} + \overline{Y}_{BA,2}}{2} - \frac{\overline{Y}_{AB,2} + \overline{Y}_{BA,1}}{2} \\
&= \frac{\overline{Y}_{AB,1} - \overline{Y}_{AB,2}}{2} - \frac{\overline{Y}_{BA,1} - \overline{Y}_{BA,2}}{2} \\
&= \frac{1}{2}\left\{ \frac{1}{n_1}\sum_{j=1}^{n_1}(Y_{AB,1,j} - Y_{AB,2,j}) - \frac{1}{n_2}\sum_{j=1}^{n_2}(Y_{BA,1,j} - Y_{BA,2,j}) \right\}
\end{aligned}$$

# Crossover trials with continuous endpoint

- Define the $Y_{AB,d,j} = Y_{AB,1,j} - Y_{AB,2,j}$, and $s^2_{AB}$ as the sample variance of $Y_{AB,d,1}, \ldots, Y_{AB,d,n_1}$.
- Define the $Y_{BA,d,j} = Y_{BA,1,j} - Y_{BA,2,j}$, and $s^2_{BA}$ as the sample variance of $Y_{BA,d,1}, \ldots, Y_{BA,d,n_2}$.
- The standard error of the estimator $\widehat{\mu_A - \mu_B}$ is $0.5\sqrt{s^2_{AB}/n_1 + s^2_{BA}/n_2}$. So, the 95% two-sided CI for $\mu_A - \mu_B$ is $\widehat{\mu_A - \mu_B} \pm 1.96 \times 0.5 \times \sqrt{s^2_{AB}/n_1 + s^2_{BA}/n_2}$.

# Testing the period effect in crossover trials

- $\overline{Y}_{AB,1} + \overline{Y}_{BA,1}$ estimates $\mu_A + \mu_B + 2\gamma$.
- $\overline{Y}_{AB,2} + \overline{Y}_{BA,2}$ estimates $\mu_A + \mu_B - 2\gamma + \lambda_A + \lambda_B$.
- $(\overline{Y}_{AB,1} + \overline{Y}_{BA,1}) - (\overline{Y}_{AB,2} + \overline{Y}_{BA,2})$ estimates
  $\mu_A + \mu_B + 2\gamma - (\mu_A + \mu_B - 2\gamma + \lambda_A + \lambda_B) = 4\gamma - \lambda_A - \lambda_B$.
- Suppose that there is no carry-over effect, $\lambda_A = \lambda_B = 0$, then we can test the presence of period effect by testing $H_0 : \gamma = 0$ versus $H_a : \gamma \neq 0$. The estimator of $4\gamma$ is

$$\begin{aligned} 4\widehat{\gamma} &= (\overline{Y}_{AB,1} + \overline{Y}_{BA,1}) - (\overline{Y}_{AB,2} + \overline{Y}_{BA,2}) \\ &= (\overline{Y}_{AB,1} - \overline{Y}_{AB,2}) + (\overline{Y}_{BA,1} - \overline{Y}_{BA,2}) \\ &= \frac{1}{n_1}\sum_{j=1}^{n_1} Y_{AB,d,j} + \frac{1}{n_2}\sum_{j=1}^{n_2} Y_{BA,d,j} \end{aligned}$$

# Testing the period effect in crossover trials

- The variance of the estimator of $4\gamma$ is

$$\text{Var}\left(\frac{1}{n_1}\sum_{j=1}^{n_1} Y_{AB,d,j} + \frac{1}{n_2}\sum_{j=1}^{n_2} Y_{BA,d,j}\right) = \frac{s_{AB}^2}{n_1} + \frac{s_{BA}^2}{n_2}$$

- The test statistic is

$$T = \frac{(\overline{Y}_{AB,1} + \overline{Y}_{BA,1}) - (\overline{Y}_{AB,2} + \overline{Y}_{BA,2})}{\sqrt{s_{AB}^2/n_1 + s_{BA}^2/n_2}}$$

- Under a large sample $T$ follows approximate Normal$(0,1)$ distribution. So, the $p$-value is $2\text{pr}(Z > |\text{observed } T|)$. We reject $H_0$ if the $p$-value is smaller than $\alpha$.
- The two-sided $(1-\alpha)100\%$ CI is $(\overline{Y}_{AB,1} + \overline{Y}_{BA,1}) - (\overline{Y}_{AB,2} + \overline{Y}_{BA,2}) \pm Z_{\alpha/2}\sqrt{s_{AB}^2/n_1 + s_{BA}^2/n_2}$. If the CI includes (does not include) 0, then we fail to reject (reject) $H_0$ at the $100\alpha\%$ level of significance.
- The presence of the period effect may have important medical significance.

# Testing the presence of carry over effect

- Carryover effects likely to cause the difference between the two treatments to be different in the two time periods, resulting in a significant treatment$\times$period interaction. Hence, the presence of carryover effect is equivalent to the presence of significant treatment$\times$period interactions.
- When the carry-over effect is statistically significant, a usual practice is to set aside the data from the second time period and analyze the data from the first period only, because the first period is free of any carryover effects. If the preliminary test for carryover is not significant, then the data from both periods are analyzed in the usual manner.

# Testing the presence of carry over effect/equal carry over effect

- We want to test $H_0 : \lambda_A = \lambda_B$ versus $H_a : \lambda_A \neq \lambda_B$. The null and alternative hypotheses can be equivalently written as $H_0 : \lambda_A - \lambda_B = 0$ versus $H_a : \lambda_A - \lambda_B \neq 0$.
- $\overline{Y}_{AB,1} + \overline{Y}_{AB,2}$ estimates $\mu_A + \gamma + \beta + \mu_B - \gamma + \beta + \lambda_A = 2\beta + \lambda_A$.
- $\overline{Y}_{BA,1} + \overline{Y}_{BA,2}$ estimates $\mu_B + \gamma - \beta + \mu_A - \gamma - \beta + \lambda_B = -2\beta + \lambda_B$.
- Although $\beta$ is referred as the sequence effect, in reality it represents subject and sequence effects together. We assume that $\beta$ is a random effect with mean zero and variance $\sigma^2$. As a consequence, $(\overline{Y}_{AB,1} + \overline{Y}_{AB,2}) - (\overline{Y}_{BA,1} + \overline{Y}_{BA,2})$ estimates $\lambda_A - \lambda_B$.
- The variance of $(\overline{Y}_{AB,1} + \overline{Y}_{AB,2}) - (\overline{Y}_{BA,1} + \overline{Y}_{BA,2})$ is $\tau_1^2/n_1 + \tau_2^2/n_2$, where $\tau_1^2$ is the sample variance of $(Y_{AB,1,1} + Y_{AB,2,1}), \ldots, (Y_{AB,1,n_1} + Y_{AB,2,n_1})$, and $\tau_2^2$ is the sample variance of $(Y_{BA,1,1} + Y_{BA,2,1}), \ldots, (Y_{BA,1,n_2} + Y_{BA,2,n_2})$.

# Testing the presence of carry over effect/equal carry over effect

- The test statistic is

$$T = \frac{(\overline{Y}_{AB,1} + \overline{Y}_{AB,2}) - (\overline{Y}_{BA,1} + \overline{Y}_{BA,2})}{\sqrt{\tau_1^2/n_1 + \tau_2^2/n_2}}$$

- For a large sample, $T$ approximately follows $N(0,1)$ under $H_0$.
- So, the $p$-value is $2\text{pr}(Z > |\text{observed } T|)$. We reject $H_0$ if the $p$-value is smaller than $\alpha$.
- The two-sided $(1-\alpha)100\%$ CI is $(\overline{Y}_{AB,1} + \overline{Y}_{AB,2}) - (\overline{Y}_{BA,1} + \overline{Y}_{BA,2}) \pm Z_{\alpha/2}\sqrt{\tau_1^2/n_1 + \tau_2^2/n_2}$. If the CI includes (does not include) 0, then we fail to reject (reject) $H_0$ at the $100\alpha\%$ level of significance.

# A more formal way of analyzing this numeric response from a cross-over trial

- The model is $Y_{AB,i,j} = \mu + I(\text{Treat} = B)\eta + \zeta I(i = 2) + \tau_j + e_{i,j}$ and $Y_{BA,i,j} = \mu + I(\text{Treat} = B)\eta + \zeta I(i = 2) + \tau_j + e_{i,j}$
  - $\mu$ : grand mean
  - $\eta$ : effect of treatment B compared to treatment A $(\mu_B - \mu_A)$
  - $\zeta$ : effect of period 2 compared to period 1
  - $\tau_j$ : random subject effect and assume that $\tau_j \sim \text{Normal}(0, \sigma^2)$, where $\sigma^2$ is unknown. Here $\tau_j$ can be seen as a subject+ group (AB or BA) effect.
  - $e_{i,j}$ : random noise, and assumed to follow Normal$(0, \sigma_e^2)$

# A more formal way of analyzing this numeric response from a cross-over trial

- Note that $Y_{AB,1,j}$ and $Y_{AB,2,j}$ are not independent because these two responses came from the same $j$th subject, for every $j = 1, \ldots, n_1$.
- For the same reason $Y_{BA,1,j}$ and $Y_{BA,2,j}$ are not independent.
- We bring the dependence in the model through $\tau_j$.
- For analyzing this dependent data we shall use the linear mixed model method. We need to use the lme4 package.

# Cross-over clinical trial continues

Here we have measurements of peak expiratory flow (PEF), a measure of lung function, made on 13 children aged 7 to 14 with moderate or severe asthma in a two-period cross-over trial comparing the effects of a single inhaled dose of 200 $\mu g$ Salbutamol and 12 $\mu g$ Formeterol. Here is the dataset.
First we shall fit a model with interaction between period and treatment.

# Cross-over clinical trial continues

| | Group 1 | | | Group 2 | |
|---|---|---|---|---|---|
| Patient id | Formoterol | Salbutamol | Patient id | Salbutamol | Formoterol |
| 1 | 310 | 270 | 8 | 370 | 385 |
| 2 | 310 | 260 | 9 | 310 | 400 |
| 3 | 370 | 300 | 10 | 380 | 410 |
| 4 | 410 | 390 | 11 | 290 | 320 |
| 5 | 250 | 210 | 12 | 260 | 340 |
| 6 | 380 | 350 | 13 | 90 | 220 |
| 7 | 330 | 365 | | | |

# Example

## Example

```
subject.id=rep(1:13, each=2);
period=c(rep(c(1:2), 7), rep(c(1, 2), 6))
treatment=c(rep(c("Formoterol", "Salbutamol"), 7),
            rep(c("Salbutamol", "Formoterol"), 6))
response=c(
c(310, 270, 310, 260, 370, 300, 410, 390, 250, 210, 380, 350, 330, 365),
c(370, 385, 310, 400, 380, 410, 290, 320, 260, 340, 90, 220)
  )
period=as.factor(period)
 mydata=data.frame(subject.id, period, treatment, response)
  head(mydata)
  subject.id period  treatment response
1          1      1 Formoterol      310
2          1      2 Salbutamol      270
3          2      1 Formoterol      310
4          2      2 Salbutamol      260
5          3      1 Formoterol      370
6          3      2 Salbutamol      300
```

# Example

## Example

```
tail(mydata)
   subject.id period  treatment response
21         11      1 Salbutamol      290
22         11      2 Formoterol      320
23         12      1 Salbutamol      260
24         12      2 Formoterol      340
25         13      1 Salbutamol       90
26         13      2 Formoterol      220
```

# Example

## Example

```
library(lme4)
out=lmer(response~period+treatment+treatment*(period)+(1|subject.id))
```

## Example

```
summary(out)
Linear mixed model fit by REML ['lmerMod']
Formula:
response ~ period + treatment + treatment * (period) + (1 | subject.id)

REML criterion at convergence: 244.5

Scaled residuals:
     Min       1Q   Median       3Q      Max
-1.65063 -0.40667  0.04785  0.52623  1.26690

Random effects:
 Groups     Name        Variance Std.Dev.
 subject.id (Intercept) 4846.5   69.62
 Residual                750.4   27.39
Number of obs: 26, groups:  subject.id, 13

Fixed effects:
                            Estimate Std. Error t value
(Intercept)                   337.14      28.28  11.923
period2                         8.69      41.62   0.209
treatmentSalbutamol           -53.81      41.62  -1.293
period2:treatmentSalbutamol    14.40      80.41   0.179

Correlation of Fixed Effects:
            (Intr) period2 trtmnS
period2     -0.679
trtmntSlbtm -0.679  0.928
prd2:trtmnS  0.656 -0.966 -0.966
```

# Testing the interaction effect

## Example

```
out=lmer(response~period+treatment+treatment*(period)+(1|subject.id))
out0=lmer(response~period+treatment+(1|subject.id))
 anova(out0, out)
refitting model(s) with ML (instead of REML)
Data: NULL
Models:
out0: response ~ period + treatment + (1 | subject.id)
out: response ~ period + treatment + treatment * (period) + (1 | subject.id)
     Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
out0  5 285.85 292.14 -137.92   275.85
out   6 287.81 295.36 -137.90   275.81 0.0379      1     0.8457
```

# Results

- We obtain $\widehat{\sigma}^2 = 4846.5$ that is much larger than $\widehat{\sigma}_e^2 = 750.4$.
- The interaction effect turns out to be statistically non-significant with a $p$-value of 0.85. Also, period seems to have no effect on the mean of the response variable. Therefore, we shall fit the model without any interaction.

# Example

## Example

```
library(lme4)
out1=lmer(response~period+treatment+(1|subject.id))
```

## Example

```
 summary(out1)
Linear mixed model fit by REML ['lmerMod']
Formula: response ~ period + treatment + (1 | subject.id)

REML criterion at convergence: 255.1

Scaled residuals:
     Min       1Q   Median       3Q      Max
-1.69842 -0.41583  0.06268  0.52315  1.28230

Random effects:
 Groups     Name        Variance Std.Dev.
 subject.id (Intercept) 4425.4   66.52
 Residual                750.4   27.39
Number of obs: 26, groups:  subject.id, 13

Fixed effects:
                   Estimate Std. Error t value
(Intercept)          333.82      20.56  16.233
period2               15.89      10.78   1.475
treatmentSalbutamol  -46.61      10.78  -4.325

Correlation of Fixed Effects:
            (Intr) perid2
period2     -0.242
trtmntSlbtm -0.242 -0.077
```

# Results

- The results indicate that there is a statistically significant effect of the treatment.
- Compared to Salbutamol, the mean peak expiratory flow is 46.61 unit higher in the Formoterol group.
- This is a perfect scenario of random effect ANOVA.

# Example: ANOVA

## Example

```
 anova(out1)
Analysis of Variance Table
          Df  Sum Sq Mean Sq F value
period     1   984.6   984.6  1.3121
treatment  1 14035.9 14035.9 18.7044
```

# Cross-over clinical trial with a binary response

- There[10] are two treatments, standard treatment $A$ and experimental treatment $B$, and two groups are group 1: $A \to B$, and group 2: $B \to A$. The binary response $Y$ takes on 1 for success and 0 for failure.
- $Y_{AB,i,j}$ : Outcome of patient $j$ ($j = 1, \ldots, n_1$) in period $i$ ($i = 1, 2$), in group 1 (AB group)
- $Y_{BA,i,j}$ : Outcome of patient $j$ ($j = 1, \ldots, n_2$) in period $i$ ($i = 1, 2$), in group 2 (BA group)
- Our (conditional) model is
  $\text{logit}\{\text{pr}(Y_{AB,i,j} = 1)\} = \mu_0 + \tau_j + \eta I(\text{Treat} = B) + \zeta I(i = 2)$, where
  - $\mu_0$: intercept (grand mean)
  - $\eta$ : treatment effect (effect of $B$ with respect to treatment $A$) (according to our previous notations, it is $\mu_B - \mu_A$)
  - $\zeta$: period effect (effect of period 2 with respect to period 1) (according to our previous notations it is $-2\gamma$)
  - $\tau_j$: random subject effect (we can say subject+ group effect) and assume that $\tau_j \sim \text{Normal}(0, \sigma^2)$, where $\sigma^2$ is unknown.

[10] CSDA, 2012, Vol 56, p 522–530

# Cross-over clinical trial with a binary response

- More explicitly our model is

$$\text{pr}(Y_{AB,1,j}=1)=\frac{\exp\{\mu_0+\tau_j+\eta\times 0+\zeta\times 0\}}{1+\exp\{\mu_0+\tau_j+\eta\times 0+\zeta\times 0\}}=\frac{\exp(\mu_0+\tau_j)}{1+\exp(\mu_0+\tau_j)},$$

and

$$\text{pr}(Y_{AB,2,j}=1)=\frac{\exp\{\mu_0+\tau_j+\eta\times 1+\zeta\times 1\}}{1+\exp\{\mu_0+\tau_j+\eta\times 1+\zeta\times 1\}}=\frac{\exp(\mu_0+\tau_j+\eta+\zeta)}{1+\exp(\mu_0+\tau_j+\eta+\zeta)}.$$

- Similarly,

$$\text{pr}(Y_{BA,1,j}=1)=\frac{\exp\{\mu_0+\tau_j+\eta\times 1+\zeta\times 0\}}{1+\exp\{\mu_0+\tau_j+\eta\times 1+\zeta\times 0\}}=\frac{\exp(\mu_0+\tau_j+\eta)}{1+\exp(\mu_0+\tau_j+\eta)},$$

and

$$\text{pr}(Y_{BA,2,j}=1)=\frac{\exp\{\mu_0+\tau_j+\eta\times 0+\zeta\times 1\}}{1+\exp\{\mu_0+\tau_j+\eta\times 0+\zeta\times 1\}}=\frac{\exp(\mu_0+\tau_j+\zeta)}{1+\exp(\mu_0+\tau_j+\zeta)}.$$

# Cross-over clinical trial continues

- Conditional treatment effect is $Odds_{AB,2}/Odds_{BA,2}$,

$$\begin{aligned} Odds_{AB,2} &= \frac{\text{pr}(Y_{AB,2,j}=1)}{\text{pr}(Y_{AB,2,j}=0)} = \exp(\mu_0+\tau_j+\eta+\zeta), \\ Odds_{BA,2} &= \frac{\text{pr}(Y_{BA,2,j}=1)}{\text{pr}(Y_{BA,2,j}=0)} = \exp(\mu_0+\tau_j+\zeta), \end{aligned}$$

  so $Odds_{AB,2}/Odds_{BA,2} = \exp(\eta)$.

- Conditional period effect is $Odds_{BA,2}/Odds_{AB,1}$

$$\begin{aligned} Odds_{BA,2} &= \frac{\text{pr}(Y_{BA,2,j}=1)}{\text{pr}(Y_{BA,2,j}=0)} = \exp(\mu_0+\tau_j+\zeta), \\ Odds_{AB,1} &= \frac{\text{pr}(Y_{AB,1,j}=1)}{\text{pr}(Y_{AB,1,j}=0)} = \exp(\mu_0+\tau_j), \end{aligned}$$

  so $Odds_{BA,2}/Odds_{AB,1} = \exp(\zeta)$.

- Our goal is estimation of $\eta$ (or $exp(\eta)$) and $\zeta$ (or $\exp(\zeta)$).

# Cross-over clinical trial continues

- The observed data in each group can be classified according to the following tables

| | | Group 1 (AB) | | Group 2 (BA) | |
|---|---|---|---|---|---|
| | | Response in period 2 | | Response in period 2 | |
| | | 0 | 1 | 0 | 1 |
| Response in period 1 | 0 | $n_{100}$ | $n_{101}$ | $n_{200}$ | $n_{201}$ |
| | 1 | $n_{110}$ | $n_{101}$ | $n_{200}$ | $n_{201}$ |

# Cross-over clinical trial continues

- We shall use the maximum likelihood method to estimate the model parameters.
- Important issue is that $Y_{AB,1,j}$ is not independent of $Y_{AB,2,j}$, for $j = 1, \ldots, n_1$. Similarly, $Y_{BA,1,j}$ and $Y_{BA,2,j}$ are not independent, for $j = 1, \ldots, n_2$.
- This is case of dependent (correlated) data. The model involving the random subject effect is known as Generalized Linear Mixed Model.

# Cross-over clinical trial continues

$$
\begin{aligned}
\mathcal{L} \quad = \quad & \left\{ \int \text{pr}(Y_{AB,1,j} = 0, Y_{AB,2,j} = 0) \frac{\exp(-\tau_j^2/2\sigma^2)}{\sqrt{2\pi}\sigma} d\tau_j \right\}^{n_{100}} \\
& \times \left\{ \int \text{pr}(Y_{AB,1,j} = 1, Y_{AB,2,j} = 0) \frac{\exp(-\tau_j^2/2\sigma^2)}{\sqrt{2\pi}\sigma} d\tau_j \right\}^{n_{110}} \\
& \times \left\{ \int \text{pr}(Y_{AB,1,j} = 0, Y_{AB,2,j} = 1) \frac{\exp(-\tau_j^2/2\sigma^2)}{\sqrt{2\pi}\sigma} d\tau_j \right\}^{n_{101}} \\
& \times \left\{ \int \text{pr}(Y_{AB,1,j} = 1, Y_{AB,2,j} = 1) \frac{\exp(-\tau_j^2/2\sigma^2)}{\sqrt{2\pi}\sigma} d\tau_j \right\}^{n_{111}} \\
& \times \left\{ \int \text{pr}(Y_{BA,1,j} = 0, Y_{BA,2,j} = 0) \frac{\exp(-\tau_j^2/2\sigma^2)}{\sqrt{2\pi}\sigma} d\tau_j \right\}^{n_{200}} \\
& \times \left\{ \int \text{pr}(Y_{BA,1,j} = 1, Y_{BA,2,j} = 0) \frac{\exp(-\tau_j^2/2\sigma^2)}{\sqrt{2\pi}\sigma} d\tau_j \right\}^{n_{210}} \\
& \times \left\{ \int \text{pr}(Y_{BA,1,j} = 0, Y_{BA,2,j} = 1) \frac{\exp(-\tau_j^2/2\sigma^2)}{\sqrt{2\pi}\sigma} d\tau_j \right\}^{n_{201}} \\
& \times \left\{ \int \text{pr}(Y_{BA,1,j} = 1, Y_{BA,2,j} = 1) \frac{\exp(-\tau_j^2/2\sigma^2)}{\sqrt{2\pi}\sigma} d\tau_j \right\}^{n_{211}}
\end{aligned}
$$

# Cross-over clinical trial continues

[11] A two-period double blind crossover trial of 12 $\mu g$ formoterol solution compared with 200 $\mu g$ salbutamol solution administered to 24 children with exercise induced athsma. Response is coded as S and F corresponding to 'good' and 'not good' based upon the investigators overall assessment. Subjects were randomized to one of two groups: group 1 received the treatments in the order formoterol$\rightarrow$ salbutamol; group 2 in the order salbutamol$\rightarrow$formoterol. The results are given below:

[11] Cross-over Trials in Clinical Research by Stephen Senn, 2002

# Cross-over clinical trial continues

| | Group 1 | | | Group 2 | |
|---|---|---|---|---|---|
| Subject | Formoterol | Salbutamol | Subject | Salbutamol | Formoterol |
| 1 | S | S | 13 | F | S |
| 2 | F | F | 14 | F | S |
| 3 | S | F | 15 | S | S |
| 4 | S | F | 16 | S | S |
| 5 | S | S | 17 | S | S |
| 6 | S | F | 18 | S | S |
| 7 | S | F | 19 | S | S |
| 8 | S | F | 20 | S | F |
| 9 | S | F | 21 | F | S |
| 10 | S | F | 22 | F | S |
| 11 | S | F | 23 | F | S |
| 12 | S | F | 24 | F | S |

# Cross-over clinical trial with a binary response

- Before estimating the treatment effect, we shall first check if there is any carry-over effect.
- The presence of carry over effect means the presence of interaction between treatment and period.
- To check the presence of carry-over effect we shall first fit the following model
- Our (conditional) model is $\text{logit}\{\text{pr}(Y_{AB,i,j} = 1)\} = \mu_0 + \tau_j + \eta I(\text{Treat} = B) + \zeta I(i = 2) + \psi I(\text{Treat} = B) I(i = 2)$, where $\psi$ is the interaction effect between treatment and period on the success probability. As before $\tau_j$ is assumed to be random and follow Normal$(0, \sigma^2)$ distribution.

# Example

## Example

```
subject.id=rep(1:24, each=2);
period=rep(rep(c(1:2), 12), 2)
treatment=c(rep(c("Formoterol", "Salbutamol"), 12),
            rep(c("Salbutamol", "Formoterol"), 12))
response=c(c(1, 1, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 0,
        1, 0, 1, 0, 1, 0, 1, 0, 1, 0),
        c(0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
          1, 0, 0, 1, 0, 1, 0, 1, 0, 1))
 mydata=data.frame(subject.id, period, treatment, response)
  head(mydata)
  subject.id period  treatment response
1          1      1 Formoterol        1
2          1      2 Salbutamol        1
3          2      1 Formoterol        0
4          2      2 Salbutamol        0
5          3      1 Formoterol        1
6          3      2 Salbutamol        0
```

# Example

## Example

```
library(lme4)
out=glmer(response~as.factor(period)+
            treatment+treatment*as.factor(period)+
            (1|subject.id),
          data=mydata, family=binomial)
```

## Example

```
 summary(out)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) [
glmerMod]
 Family: binomial  ( logit )
Formula:
response ~ as.factor(period) + treatment + treatment * as.factor(period) +
    (1 | subject.id)
   Data: mydata
     AIC      BIC   logLik deviance df.resid
    51.2     60.6    -20.6     41.2       43
Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.3166 -0.4472  0.3015  0.3015  2.2361
Random effects:
 Groups     Name        Variance  Std.Dev.
 subject.id (Intercept) 1.772e-16 1.331e-08
Number of obs: 48, groups:  subject.id, 24

Fixed effects:
                                    Estimate Std. Error z value Pr(>|z|)
(Intercept)                        2.398e+00  1.044e+00   2.296   0.0217 *
as.factor(period)2                 4.513e-13  1.477e+00   0.000   1.0000
treatmentSalbutamol               -2.398e+00  1.193e+00  -2.009   0.0445 *
as.factor(period)2:treatmentSalbutamol -1.609e+00  1.765e+00  -0.912   0.3618
---

Correlation of Fixed Effects:
            (Intr) as.()2 trtmnS
as.fctr(p)2 -0.707
trtmntSlbtm -0.875  0.619
```

# Cross-over clinical trial with a binary response

- The estimate of $\psi$ is $-1.69$ with a 95% CI $(-1.69 \pm 1.96 \times 1.76)$ that includes the null value 0. Note that $\psi = 0$ signifies no interaction (no carry-over effect). Based on the analysis we fail to reject $H_0 : \psi = 0$ at the 5% level.
- In this model the estimate of $\sigma^2$ was $1.77 \times 10^{-16}$, a small value.
- Next, we shall fit a model without the interaction term.

# Example

## Example

```
out1=glmer(response~as.factor(period)+
             treatment+
             (1|subject.id),
           data=mydata, family=binomial)
```

## Example

```
 summary(out1)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) [
glmerMod]
 Family: binomial  ( logit )
Formula: response ~ as.factor(period) + treatment + (1 | subject.id)
   Data: mydata

     AIC      BIC   logLik deviance df.resid
    50.0     57.5    -21.0     42.0       44

Scaled residuals:
    Min      1Q  Median      3Q     Max
-4.7667 -0.5122  0.2098  0.3771  1.9523

Random effects:
 Groups     Name        Variance Std.Dev.
 subject.id (Intercept) 0        0
Number of obs: 48, groups:  subject.id, 24

Fixed effects:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)           3.1233     0.9504   3.286 0.001015 **
as.factor(period)2   -1.1729     0.8054  -1.456 0.145318
treatmentSalbutamol  -3.2884     0.9114  -3.608 0.000309 ***
---

Correlation of Fixed Effects:
            (Intr) as.()2
as.fctr(p)2 -0.618
trtmntSlbtm -0.825  0.292
```

# Cross-over clinical trial with a binary response

- This model has less AIC (BIC) than the model with the interaction term.
- In this model the estimate of $\sigma^2$ was 0. That means the subject specific random effect not needed.
- The treatment effect is $-3.288$ for Salbutamol compared to Formoterol. In other words, the odds of success for Salbutamol is $\exp(-3.29) = 0.037$ times that of Formoterol. Likely we can say, the odds of success is reduced by 96% in Salbutamol compared to Formoterol.
- The effect of period came out to be statistically non-significant.
- Since $\widehat{\sigma^2} = 0$, we can fit a simple logistic model without the random effect taking only period and treatment effect, and compare the results. A quick comparison between the existing results and the following results shows that the parameter estimates remain unchanged, and the AIC value is further reduced in the model without the subject specific random effect.

# Example

## Example

```
out2=glm(response~as.factor(period)+
           treatment,
         data=mydata, family=binomial)
```

## Example

```
 summary(out2)
Call:
glm(formula = response ~ as.factor(period) + treatment, family = binomial,
    data = mydata)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.5165  -0.6826   0.2935   0.5157   1.7726

Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)           3.1233     0.9504   3.286 0.001015 **
as.factor(period)2   -1.1729     0.8054  -1.456 0.145318
treatmentSalbutamol  -3.2884     0.9114  -3.608 0.000309 ***
---
(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 63.510  on 47  degrees of freedom
Residual deviance: 42.033  on 45  degrees of freedom
AIC: 48.033

Number of Fisher Scoring iterations: 5
```