EPI-546: Fundamentals of Epidemiology and Biostatistics

Course Notes – Lecture 7 The RCT

Mat Reeves BVSc, PhD

Outline:

I. Introduction to the RCT
II. An Overview of the RCT Design
    1. Inclusion Criteria
    2. Exclusion Criteria
    3. Baseline Measurements
    4. Randomization and concealment
    5. Intervention
    6. Blinding
    7. Follow-up (FU), Non-Compliance and Contamination
    8. Measuring Outcomes, Sub-group analyses and Surrogate end points
    9. Statistical Analyses (ITT vs. PP vs. AS)
    10. RCTs and Meta-analyses - assessing trial quality, reporting, and trial registration
III. Advantages and disadvantages of RCT's

I. Introduction to the RCT

The randomized clinical trial is an *experimental study* conducted on clinical patients (with their consent of course!). The investigator seeks to completely control the exposure (in terms of its type, amount, and duration), and (most importantly) who receives it through the process of randomization.

RCT's are regarded as the most scientifically vigorous study design. Because:

- An unpredictable (i.e., concealed) random assignment eliminates (or at least greatly reduces) confounding from known and unknown prognostic factors (that is, it makes the groups equivalent in terms of their prognosis at baseline).

- Blinding eliminates biased measurement, so that outcomes are measured with the same degree of accuracy and completeness in every participant.

Because of these conditions, it is then possible to confidently attribute **cause and effect** – that is, because the only thing that differed between the groups (or arms) of the trial was the presence or absence of the intervention, any effect can be ascribed to it (assuming a well conducted, unbiased study). The RCT is therefore described as having high *internal validity* – the experimental design ensures that, within reason, strong cause and effect conclusions can be drawn from the results. While RCT's are the gold standard by which we determine the efficacy of treatments, trials are not always feasible, appropriate, or ethical.

There are two types of RCTs:

Prophylactic trials

- evaluate the efficacy of an intervention designed to prevent disease, e.g., vaccine, vitamin supplement, patient education, screening.

Treatment trials

- evaluate efficacy of a curative drug or intervention or a drug or intervention designed to manage or mitigate signs and symptoms of a disease (e.g., arthritis, hypertension).

Also, RCTs can be done at the individual level, where highly select groups of individuals are randomized under tightly controlled conditions, or they can be done at the community level, where large groups are randomized (e.g., cities, regions) under less rigidly controlled conditions. Community trials are usually conducted to test interventions for primary prevention purposes – so they are prophylactic trials done on a wide scale.

II An Overview of the RCT Design and Process

1. Inclusion Criteria

Specific inclusion criteria are used to optimize the following:

- The rate of the primary outcome
- The expected efficacy of treatment
- The generalizability of the results
- The recruitment, follow-up, and compliance of patients

So, the goal is to identify the sub-population of patients in whom the intervention is feasible, and that will produce the desired effect. To this end, the choice of inclusion criteria represents a balance between picking the people who are most likely to benefit without sacrificing the generalizability of the study. If you make the inclusion criteria too restrictive you run the risk of having the study population so unique that no one else will be able to apply your findings to their population.

2. Exclusion Criteria

Specific exclusion criteria are used to exclude subjects who would "mess up" the study. Valid reasons for excluding patients might include the following:

- When the risk of treatment (or placebo) is unacceptable.
- When the treatment is unlikely to be effective because the disease is too severe, or too mild, or perhaps the patient has already received (and failed) the treatment.
- When the patient has other conditions (co-morbidities) that would either interfere with the intervention, or the measurement of the outcome, or the expected length of follow-up e.g., terminal cancer.
- When the patient is unlikely to complete follow-up or adhere to the protocol.
- Other practical reasons e.g., language/cognitive barriers/no phone at home etc.

Again, one needs to be careful to avoid using excessive exclusions even when they appear perfectly rational. Excessive number of exclusions can add to the complexity to the screening process (remember, every exclusion criteria needs to be assessed in every patient), and ultimately

to decreased recruitment. Once again, the choice of exclusion criteria represents a balance between picking subjects who are more likely to make your study a success without sacrificing *generalizability*. The real danger of setting very restrictive inclusion and exclusion criteria is that the final study population becomes so highly selective that nobody is interested in the results because they don't apply to real-world patients. In other words, while *internal validity* may have been maximized, the study's generalizability or *external validity* is badly compromised.

3. Baseline Measurements

At baseline you want to collect information that will:

- Describe the characteristics of the subjects in your study i.e., demographics. This is especially important in demonstrating that the randomization process worked (i.e., that you achieved balance between the intervention and control groups).
- Assist in the tracking the subject during the study (to prevent loss-to-follow-up). This includes names, address, tel/fax #'s, e-mail, SSN, plus contact information of friends, neighbours, and family (you can never collect too much information here).
- Identify major clinical characteristics and prognostic factors for the primary outcome that can be evaluated in pre-specified subgroup analyses. For example, if you thought that the treatment effect could be different in women compared to men, then you would collect information on gender, so that the effect could be evaluated within each of these two sub-groups.

The amount of baseline data that needs to be collected does not have to be excessive, because the randomization process should result in identical groups. However, it is always necessary to collect sufficient baseline information regarding demographics and clinical characteristics to prove this point.

4. Randomization and concealment

Randomization results in balance of known and unknown confounders. The randomization process should be reproducible, unpredictable, and tamper proof. It is critical that the randomization scheme itself should be **unpredictable** – so that it is not possible ahead of time to predict which group a given subject would be randomized to. Unpredictability is assured through the process of concealment which is critical in preventing selection bias – that is, the potential for investigators to manipulate who gets what treatment. Such "manipulation" in clinical trials has been well documented (for an example, see Schulz KF Subverting randomization in clinical trials. JAMA 1995;274:1456-8). Finally, note that the issues of randomization and concealment should be kept separate from blinding – they are completely different![1]

The actual process of generating the randomization scheme and the steps taken to ensure concealment should be described in detail - whether it is a simple coin toss (the least preferable

[1] Note that the Chapter 8 of the FF text refers to concealment as allocation concealment and includes it under the description of blinding. This is unfortunate in my opinion because concealment has a fundamentally different purpose from other aspects of blinding. Concealment is designed to prevent selection bias, whereas all other forms of blinding are undertaken to reduce measurement bias. So in my opinion these steps should be kept separate. Also note that the two conditions are mutually independent – so it's possible to have a concealed but not-blinded study or an un-concealed but blinded study (further justification for keeping the concepts separate).

method), or the use of sealed opaque envelopes, or a sophisticated off-site centralized randomization centre.

There are several different modifications to simple (individual level) randomization schemes such as a coin flip. Blocked randomization refers to the randomization done within blocks of 4 to 8 subjects. It is used to ensure that there is equal balance in the number of treatment and control subjects throughout the study.

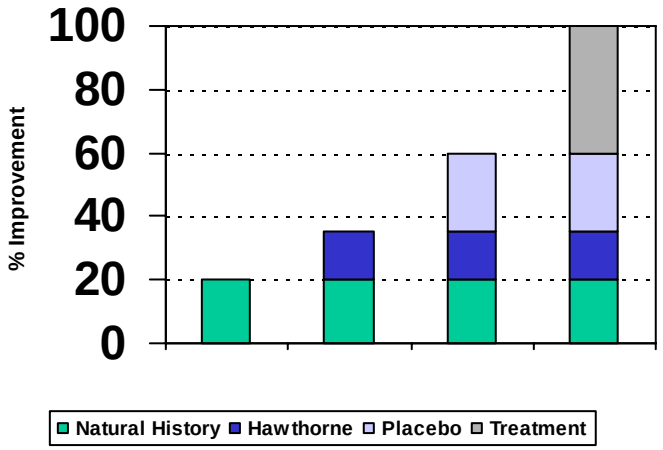
Stratified blocked randomization refers to the process whereby strata are defined according to a critically important factor or subgroup (e.g., gender, disease severity, or study center), and the randomization process conducted within each strata. Again, this design ensures that there is balance in the number of treatment and control subjects within each of the sub-groups (so that at the end of the trial the factor in question is distributed equally among the treatment and control groups).

5. Intervention

It is important that the balance between the potential benefits and risks of an intervention be considered carefully before a trial is begun. Remember that everyone is exposed to potential side effects of an intervention, whereas not everyone can benefit from the intervention itself (because not everyone will have or develop the outcome you are trying to prevent and no intervention is ever 100% effective). So caution dictates using the "lowest effective dose". Since RCTs are designed under the premise that serious side effects are expected to occur much less frequently that the outcome, it is not surprising that RCT's are invariably under-powered to detect side effects. This is the reason for Phase IV post-marketing surveillance studies. It is only after the drug has reached the market, and has been used on many, many people, that evidence for rare but serious side effects is found.

All RCT's require a control group. The purpose of the control group is to measure the cumulative effects of all the other factors that can influence the outcome over time – other than the active treatment itself. As shown in the figure below, this includes spontaneous improvements (due to the natural history and the Hawthorne effect), as well as the well documented placebo effect (see below).

Figure 1. Cumulative effects of spontaneous improvements, non-specific responses, and specific treatment effects (from Fletcher)



6. Blinding (a.k.a. masking)

The use of blinding is another cardinal feature of RCT's. Blinding is important because it preserves the benefits of randomization by preventing the *biased assessment of outcomes*. Blinding therefore prevents measurement bias (as opposed to randomization and concealment which prevent confounding bias and selection bias, respectively). Blinding also helps to reduce non-compliance and contamination or cross-overs - especially in the control group (since they are unaware that they are not getting the active treatment). Usually one thinks about either a single-blind study (where either the patient or the physician is blinded) or a double-blind study (where both the patient and the physician are blinded). However, ideally, blinding should occur at all of the following levels:

- Patients
- Caregivers
- Collectors of outcome data (e.g., research assistants or study physicians),
- Adjudicators of the outcome data (e.g., the adjudication committee or study physicians), and
- The data analyst

A placebo is any agent or process that attempts to mask (or blind) the identity of the true active treatment. It is a common feature of drug trials. The value of the placebo, and blinding in general, is especially important when the primary outcome measure being assessed is non-specific or "soft" – for example, patient self-reported outcomes like degree of pain, nausea, or depression. The placebo effect refers to the tendency for such "soft" outcomes to improve when a patient is enrolled in a treatment study, regardless of whether they are actual receiving an active treatment. The placebo effect can be regarded as the baseline against which to measure the effect of the active treatment.

A true placebo may not be justifiable if a known proven treatment is already the standard of care. For example, in a stroke prevention study looking at a new anti-platelet agent, a true placebo control would be very hard to justify given the proven benefit of aspirin (which would be regarded as the minimum standard of care).

Sometimes blinding (and/or the use of a placebo) is just not feasible – for example in a RCT of a surgical intervention it is very difficult to mask who got the surgery and who did not (at least to the patients and caregiver!). In such situations the study is referred to as an "open" trial. Blinding can also be hard to maintain – for example, when the treatment has very clear and obvious benefits, or side effects or other harms. In such cases it is important to try to choose a "hard" outcome – like death from any cause, and to standardize treatments and data collection as much as possible

7. Loss-to-follow-up, non-compliance, and contamination

It is important that everyone assigned to either the intervention or control group receives equal follow-up and are ultimately all accounted for. Loss-to-follow-up (LTFU), non-compliance, and contamination are important potential problems in all RCT's. If they occur with any frequency the study will have reduced power (because there are fewer subjects to provide information), and if they occur in a non random fashion (meaning they occur in one arm more than the other -

which is often the case) then bias maybe introduced.

Loss to-follow-up can occur for many reasons most of which can be related to the outcomes of interest. For example, subjects are more likely to be lost-to-follow-up if they have side effects, or if they have moved away, got worse, got better or have simply lost interest. Death can be another cause of loss-to-follow-up. If the final outcome of the subjects LTFU remains unknown, then they cannot be included in the final analysis and so (if the rate of loss to-follow-up is high and/or differentially affects one group more than the other) they can have a significant negative effect on the study's conclusions (i.e., low power due to the smaller sample size and biased results due to the differential LTFU). Note that the negative effects of LTFU cannot be easily corrected by the intention-to-treat analysis, since without knowledge of the final outcome status these subjects have to be dropped from the analysis (See below for further details).

Similarly, poor compliance can be expected to be related to the presence of side effects, iatrogenic drug reactions, whether the patient got better or got worse, or whether the patient simply lost interest. People who do not comply with an intervention can be expected to have a worse outcome than those that do. An example of this phenomenon is shown in the table below from the Clofibrate trial – one of the earliest lipid lowering therapy RCTs. The table shows the 5-year cumulative mortality rate by treatment group (as assigned by randomization), and according to whether they complied with the trial protocol. Persons who did not comply had a much higher mortality rate at 5-years, regardless of the treatment group they were assigned to:

Clofibrate Trial	Cumulative 5-year mortality	
	Compliant	Non-compliant
Treatment	15.0%	24.6%
Control	15.1%	28.3%

So the degree to which non-compliance occurs in a study, and the degree to which it differentially affects one arm of the study more than the other, is important to assess.

Contamination refers to the situation when subjects cross-over from one arm of the study into the other – thereby contaminating the initial randomization process. Perhaps the most famous example of this was in the early AIDS treatment RCTs, where participants were able to get their assigned treatments "assayed" by private laboratories to find out whether they were receiving the placebo or active drug (AZT) (the folks in the placebo group would then seek the active AZT drug from other sources!).

As one can imagine excessive loss to follow-up, poor compliance, and/or contamination (see Figure 2) can translate into the slow prolonged death of your trial! Essentially as these three effects take their toll, the original RCT degenerates into a mere observational (cohort) study because the active and compliant participants in each arm of the study are no longer under the control of the study investigator!

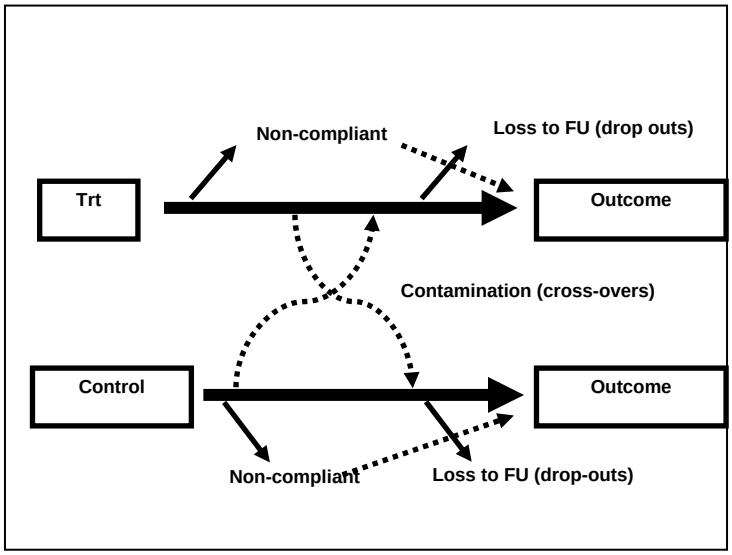
The problem of LTFU is particularly problematic for trials which are attempting to do the gold-standard **intention-to-treat analysis (ITT)** (which is the principle that **all** participants are analyzed according to their original randomization group or arm - regardless of protocol violations). So, ideally all subjects should be accounted for in both the denominator and the numerator of the groups event rate. But with lost to follow-up, if the subject is included in the denominator but not the numerator then the event rate in that group will be underestimated. However, if the subjects are simply dropped from the analysis (a common approach) then the analysis can be biased (in fact if subjects are dropped from the analysis because their outcome status is unknown then one is de facto doing a per protocol (PP) analysis). To mitigate the problems of LTFU some trials will impute an outcome based on a missing data protocol. Techniques include using the last or the worst observation, or attempting to predict the unobserved outcome based on the characteristics of the subjects. However, regardless of the technique used, it is ultimately unverifiable, and so the results should be viewed with caution.

One technique to assess the likely impact of poor compliance or LTFU is the "*5 and 20 rule*" which states that if LTFU or compliance affects < 5% of study participants then bias will be minimal, whereas if it affects >20% then bias is likely to be considerable. One can also assess the potential impact of LTFU by doing a "*best case worst case*" *sensitivity analysis*. In the best case scenario, the subjects LTFU are assumed to have had the best outcome (e.g., none of them had the adverse outcome) and the event rates in each group are calculated after counting all of the LTFU in the denominator but not in the numerator. In the worst case scenario, all of the subjects LTFU are assumed to have had the adverse outcome so the LTFU are counted in both the numerator and the denominator of the event rates. The overall *potential impact* of the LTFU is then gauged by comparing the actual results with the range of findings generated by the sensitivity analysis. Here's an example:

> In a RCT of asthma patients visiting the emergency department (ED), only 100 of 150 subjects assigned to the treatment arm (which involved an educational intervention requiring the subjects to attend 2 sessions at a local asthma treatment center) complied with the treatment and were available for follow-up. The rate of LTFU was therefore 33% clearly exceeding the 5 and 20 rule. The primary outcome, which was defined as a repeat visit to the ED over the next 6 months, occurred in 15% of the 100 subjects who remained in follow-up (i.e., 15 of 100). The results of the best/worst case sensitivity analysis were: best case 15/150 = 10% and worst case 65/150 = 43%. So, clearly the study's findings are questionable given the high loss to follow-up (33%) and the wide range of estimates for the repeat ED visit rate in the treatment arm.

Trialist's go to great lengths to attempt to reduce these problems by enrolling subjects who are more likely to be compliant and not lost-to-follow-up. To do this, studies will sometimes use two screening visits prior to actually performing the enrollment (to weed out the "time-wasters"), and in drug trials will often use pre-randomization run-in periods (using placebo and active drug) to weed out the early non-adheres. Once subjects are enrolled into a study it is very important to minimize LTFU as much as possible. This means that you should attempt to track and follow-up on all the subjects who were non-compliant, ineligible, or chose to drop out of your study for any reason. This is not easy to do!

Figure 2. Effects of non-compliance, loss-to-follow-up, and contamination on a RCT.



8. Measuring Outcomes, Sub-group analyses and Surrogate end points

The primary and secondary study outcomes – with associated definitions should be defined before the study is started (these are termed *a priori* or pre-specified comparisons). The best outcome measures are hard, clinically relevant end points such as disease rates, death, recovery, complications, or hospital/ER use. It is important that all outcomes are measured with accuracy and precision, and, most importantly, that they are measured in the same manner in both groups or arms.

Sub-group analyses i.e., the examination of the primary outcome among study sub-groups defined by key prognostic variables such as age, gender, race, disease severity etc, can provide important information by identifying whether the treatment has a different effect within specific sub-populations. However, it is critical that all sub-group analyses be pre-specified ahead of time. This is because there is a natural tendency among authors of trials that did not show a positive treatment effect for the primary outcome of interest to go "fishing" for positive results by conducting all manner of sub-group comparisons. This approach naturally leads to the problem of multiple comparisons and the potential for false positive findings. Thus all sub-group comparisons which were not pre-specified should be regarded as "exploratory findings" that should be re-examined in future RCT's as pre-panned comparisons.

Hard clinical relevant outcomes cannot always be used, for example it usually takes too long to measure disease mortality. To reduce the length and/or the size of the intended study surrogate end points may be used under the proviso that they are validated biologically relevant endpoints. Obviously one should carefully consider whether a surrogate end point used is an adequate measure of the real outcome of interest. Ideally, a prior RCT should be proven that the end point is a valid surrogate measure for the real outcome of interest. For example, in a study designed to reduce stroke incidence, the degree of blood pressure reduction would be considered a valid surrogate end point given the known causal relationship between blood pressure and stroke risk, whereas the reduction in intimal thickness of peripheral arteries would not be.

9. Statistical Analyses

Statistical analyses of trials are on the face of it typically very straight forward, since the design has created balance in all factors, except for the intervention per se. So, oftentimes it's a simple matter of comparing the primary outcome measure between the two arms. For continuous measures the t-test is commonly used, while the Chi-square test is used for categorical outcomes. Non-parametric methods maybe used when the study is relatively small or when the outcomes are not normally distributed. In survival type studies, Kaplan Meire survival curves or advanced Cox regression modeling may be used to study the effect of outcomes which occur over time (and to help assess the problems of loss-to follow-up and censoring).

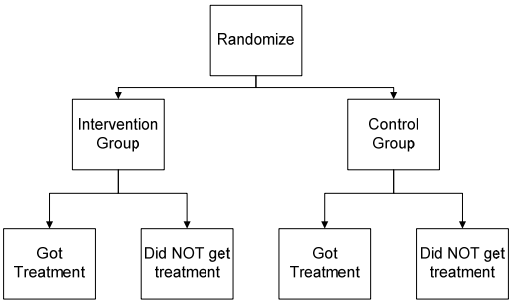
The most important concept to understand in terms of the statistical analysis of RCT's is the principle of **Intention-to-Treat Analysis (ITT)**. This refers to the analysis that compares outcomes based on the *original treatment arm that each individual participant was randomized to regardless of protocol violations.* Protocol violations include ineligibility (i.e., subjects who should not have been enrolled in the study in the first place), non-compliance, contamination or LTFU. The ITT analysis results in the **most valid but conservative estimate** of the true treatment effect. The ITT is the approach that is truest to the principles of randomization - which seeks to create perfectly comparable groups at the outset of the study. It is important to note that not even the ITT analyses can fix the problem of loss-to-follow-up unless the missing outcomes are imputed using a valid method (which can never be fully verified). Thus, in my opinion, no amount of fancy statistics can fix the problem that the final outcome is unknown for a sub-set of subjects.

Two other alternative analytical approaches may be used (as shown the diagram below) but they are both **fundamentally flawed**. The first is the **Per Protocol (PP)** analysis in which subjects in the treatment arm who did not comply with the treatment and control subjects who got treated (i.e., cross-overs) are simply dropped from the analysis. Only those subjects who complied with the original randomization scheme are therefore analyzed. The PP analysis answers the question as to whether the treatment works among those that comply, but it can never provide an unbiased assessment of the true treatment effect (because the decision to comply with a treatment is unlikely to occur at random). Also, as mentioned previously, if subjects are dropped from the ITT analysis because their outcome status is unknown, then one is de facto doing a PP analysis.

The second alternative analysis approach is the egregious **As Treated (AT)** analysis, in which subjects are analyzed according to whether they got the treatment or not (regardless of which group they were originally assigned to). So the non-compliant subjects in the intervention arm are simply moved over to the control arm, and vice versa. This analysis is akin to analyzing the trial results as if a cohort study had been done (i.e., everyone had decided for themselves whether to get treated or not), and completely destroys any of the advantages afforded by randomization. You will see examples of published studies that used the AT analysis (typically they are studies that did not show a positive effect under an ITT analysis), but you have to ask yourself – "what was the point of doing the trial in the first place if you ended up dong an AT analysis?" The approach is wholly without merit![2]

[2] The FF text refers to this analysis as an explanatory trial, which in my opinion this does not disparage this approach sufficiently! Other descriptions for analyses that are <u>not</u> intention-to-treat include exploratory, or effectiveness.

The analysis of RCT's
Intention-To-Treat versus Per Protocol or As Treated



Interntion-to-Treat	YES	YES	NO	NO
Per protocol	YES	DROP	DROP	NO
As Treated	YES	NO	YES	NO

Where:
YES means that the group is included in the analysis as the group that got treatment.
NO means that the group is included in the analysis as the group that DID NOT get treatment.
DROP means that the group is not included in the analysis (it is simply ignored).

In the table below is a real life example of the application of these three analytical approaches to a trial that compared surgical treatment (CABG) to medical treatment for stable angina pectoris in 768 men (European Coronary Surgery Study Group. Lancet 1979;i:889-93). 373 men were randomized to medical treatment but 50 ended up being treated by surgery. Of the 394 subjects randomized to surgery treatment, 26 did not receive it. A further subject was lost to follow-up and was dropped from these analyses which are based on 767 subjects. The death rates according to the actual treatments received are calculated and then the absolute risk reduction (ARR) (for surgery vs. medical) with 95% CI are calculated using the three different approaches.

Note the very high death rate in the 26 subjects who should have gotten surgery but received medical treatment – clearly these are a very sick group of patients who either died before surgery or were too sick to undergo it. Note also the 52 subjects who should have gotten medical treatment but somehow got surgery. Their mortality (4%) is much lower than the rest of the medically treated group (8.4%) – however, these men could have been healthier at baseline and so its impossible to judge the relative merits of surgery based on these two figures.

Table: Effect of different analysis approaches on an RCT of coronary artery bypass surgery versus medical treatment in 767 men with stable angina. (Lancet 1979;i:889-93).

	Allocated (vs. actual) treatment				
	Medical (medical)	Medical (surgical)	Surgical (surgical)	Surgical (medical)	ARR (95% CI)
Num. subjects	323	50	368	26	
Deaths	27	2	15	6	
Mortality (%)	8.4%	4.0%	4.1%	23.1%	
ITT analysis	7.8% (29/373)		5.3% (21/394)		2.4% (-1.0, 6.1)
PP analysis	8.4% (27/323)		4.1% (15/368)		4.3% (0.7, 8.2)
AT analysis	9.5% (33/349)		4.1% (17/418)		5.4% (1.9, 9.3)

When subjects were analyzed according to the groups they were randomized to (ITT), the results show that surgery had a small non-significant benefit vs. medical treatment. However, when the data was analyzed according to those that complied (PP), or according to the final treatment received (AT), the results show a larger and now statistically significant benefit for surgery. However, both estimates are **biased** in favour of surgery because the 26 high risk subjects were either dropped from the surgery group or were moved into the medical group – so clearly this analysis is stacked against medical treatment!!

10. RCTs and Meta-analyses - assessing trial quality, trial reporting, and trial registration

Meta-analyses of RCTs have fast become the undisputed king of the evidence-based tree. The preeminence of meta-analysis as a technique has had three important implications for the RCT.

i) Assessment of study quality – Given the variability in the quality of published RCTs, meta-analysts will often attempt to assess their quality in order to determine whether the quality of a trial has an impact on the overall results. While there are several approaches to conducting quality assessment of RCTs (example: Jadad, 1996), they all essentially focus on the same criteria: a description of the randomization process, the use of concealment, the use of blinding, and a description of the loss-to-follow-up and non-compliance rates.

ii) Trial reporting – Quality assessment of published trials using the Jadad scale or other similar tools often indicates that the trials are of marginal or poor quality – in part, because they did not report information on the key quality criteria (i.e., randomization, concealment, blinding, LTFU). One is therefore left unsure as to whether the trial did not follow these basic steps, or whether the authors simply failed to report these steps in the paper. This problem has led to the development of specific guidelines for the reporting of clinical trials, called the CONSORT Statement (see Moher, JAMA 2001). This statement aims to make sure that trials are reported in a consistent fashion and that specific descriptions are included so the validity criteria can be independently assessed.

iii) Trial registration – A big problem for meta-analyses is the potential for publication bias. The results of any meta-analysis can be seriously biased if there is a tendency to not publish trials with negative or null results. Unpublished negative trials have the tendency to be either small studies (that had insufficient power to detect a clinically important difference), or on occasion, large trials that

were sponsored by drug companies who do not want to release the information that shows their drug or gadget did not live up to their expectations. Several controversies in the past few years has led to the International Committee of Medical Journal Editors (which includes JAMA, NEJM, Lancet etc) to require that for a trail to be published in any of these journals it must have been registered prior to starting it!. The idea here is that the scientific community will then have a registry of all trials undertaken on a topic, rather than the current situation where the published literature represents all trials that were deemed worthy of publication.

III. Advantages and disadvantages of RCT's - Summary

Advantages:

High internal validity

Control of exposure – including the amount, timing, frequency, duration

Randomization - ensures balance of factors that could influence outcome i.e., it "controls" the effect of known and unknown confounders

A true measure of efficacy.

Disadvantages:

Limited external validity

Artificial environment, that includes the strict eligibility criteria and the fact that they are conducted in specialized tertiary care (referral) medical centers limits generalizability (external validity).

Difficult/complex to conduct, take time, and are expensive.

Because of ethical considerations they have limited scope – mostly therapeutic / preventive only

Finally, one should be sure to re-familiarize yourself with the measures of effect that are commonly used to describe the results of RCT's. These were covered in the Frequency lecture and include the CER or baseline risk, the TER, the RR, the RRR, the ARR and the NNT.