Critical Appraisal of a Randomized Controlled Trial
Are the Results of the Study Valid?

Objectives:

1. Compare and contrast observational and experimental studies.
2. Define study population and inclusion criteria.
3. Define randomization and allocation concealment and differentiate between the two.
4. Define block and stratified randomization
5. Define blinding, and recognize studies where blinding may not be possible.
6. Define intention-to-treat analysis, and describe advantages.
7. Identify baseline data in a clinical trial.
8. Identify attrition, and discuss effects on clinical trial.
9. Compare and contrast efficacy and effectiveness, and internal and external validity.
10. Use above concepts to critically appraise a clinical study.
11. Define bias.

Three questions to ask when reviewing a clinical study:

1. Are the results of the study valid?

    - validity is the degree to which the study answers the question being asked, or measures what it intends to measure or 'do the results represent an unbiased estimate of the effect of treatment?'
    - if the results of the study are not valid, then it doesn't matter what the results are

2. What are the results?

    - what is the size and precision of the treatment effect?

3. How do I apply these results to my patients?

    - is my patient similar to the patients in the trial?
    - what are the risks and benefits of the therapy?
    - what are my patient's preferences?

Lois Champion
lois.champion@lhsc.on.ca

Observational and experimental studies of treatment effects

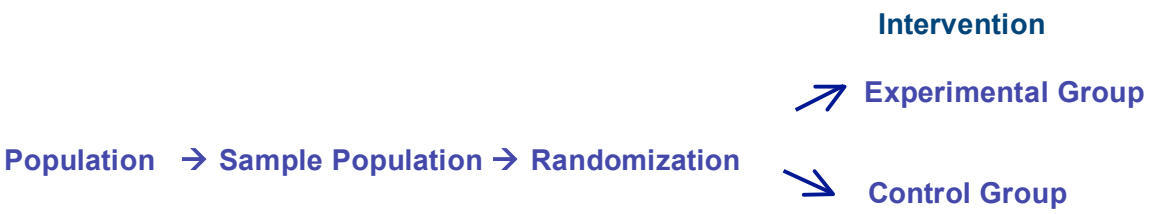
1. observational studies of interventions:

- investigators observe what happens to patients who do or do not have an intervention
- the advantage is simplicity and feasibility
- the disadvantage however is that there may be differences between the two groups of patients that affect outcome apart from the intervention or therapy and conclusions may therefore be misleading
- well-designed observational studies may provide valid information that is later corroborated by experimental studies *(for example mammography and breast cancer)* however results from observational studies have also differed substantially from experimental studies as well *(examples include vitamin E and cardiac disease, hormone replacement and cardiac disease)*
- for some clinical questions observational studies may be the only evidence that is ethically possible, or practical (because of cost and sample size issues)

2. experimental studies

- clinical trials that specify conditions of study, intervention and outcomes
- randomized controlled clinical trials have treatment randomly allocated

Randomized controlled clinical trials



- experimental (intervention) group allocated to treatment or intervention in addition to standard (usual) therapy
- control (comparison) group allocated to standard therapy, and may receive a placebo

Study population (sampling)

- clinical trials require that patients meet inclusion criteria to be eligible for entry into the study
- exclusion criteria remove individuals from eligibility of the study
- some common types of exclusion criteria:
    1. *patients with diseases other than the one being studied (ie. comorbidities that may impact on outcome)*
    2. *patients who are not expected to survive regardless of therapy for duration of anticipated follow-up*
    3. *patients with contraindications to the experimental or standard therapy*
    4. *patients who do not consent to participate*
    5. *quite often pregnancy is an exclusion criteria*

Allocation of treatment – randomization

- randomization of patients to experimental and control groups is done to try to make the groups comparable
- randomization ensures that each patient has a known chance of receiving the experimental therapy, and with allocation concealment that group assignment cannot be predicted (randomization is usually done 1:1 so that there are approximately equal numbers in the experimental and control groups, in some studies however other randomization ratios are used, for example 2:1)
- factors that may affect outcome (baseline characteristics) will tend to be equally distributed between the groups
- randomization does not guarantee that the two groups will be similar with respect to baseline characteristics, any differences however will have occurred by chance
- patients in the treatment and control groups can be compared after the study is completed to see if there are important differences that may have affected results of the study (*these differences may be corrected for statistically*)
- clinical trials in which patients are <u>not</u> randomized tend to show larger treatment effects than trials with randomization

Specific types of randomization

- particularly in small trials it may be useful to add additional randomization techniques to try to ensure that the study groups are similar for important prognostic factors
- if these techniques are not used then the randomization process is known as 'simple randomization'

- types of specific randomization techniques include

1. **block randomization:**
    - block randomization is used to balance the numbers in each group during the study, patients are randomized in blocks
    - for example the first 4 patients may be allocated to control, the next 4 to experimental group etc.
    - ideally the size of the block will change at random so that allocation is not predictable *(this is known as permuted block design)*

2. **stratified randomization:**
    - stratified randomization is used to ensure that the groups are balanced with respect to an important prognostic marker by randomizing patients separately for that marker
    - for example in a trial of a cancer therapy patients might be randomized separately depending on the stage of their cancer
    - this is particularly useful in small studies

3. *treatment allocation by minimization:*
    - *technically not a randomization technique, but is a process used to ensure balance between groups for several prognostic factors, even with a small sample size*
    - *each patient is allocated to treatment or control depending on which would lead to a better balance between the groups for the prognostic factors*

Concealment of treatment allocation

- allocation concealment means that investigators have no way of knowing whether a patient will be assigned to the experimental or control group
- the person generating the allocation sequence should not be the person identifying patients for the study; this will help to prevent selection bias
- allocation concealment is always possible in a randomized trial, even if it may not be possible to blind patients or investigators to the actual assigned group
- analyses of studies with inadequate concealment show a larger treatment effect due to the bias that is introduced
- *deterministic methods of allocation: A method of allocating participants to interventions that uses a pre-determined rule without a random element (eg. alternate assignment based on day of week, hospital number, birth date etc.). Because group assignments can be predicted in advance in deterministic methods allocation may be manipulated causing selection bias.*

Blinding

- blinding prevents study patients and investigators from determining the groups to which the individual has been assigned (experimental or control) after allocation
- blinding, unlike allocation concealment, is not always possible
- blinding may involve any, or all, of the following: patients, treating physicians and caregivers, and investigators who ultimately analyze the data
- blinding helps to prevent bias and also may improve adherence to therapy and prevent attrition (for example if a patient knows they are in the control group, they may drop out of the study)

- some definitions:
  - open label: all participants in study are aware of treatment being received after randomization
  - single blind: either patient or clinician/investigator unaware of treatment assignment
  - double blind: patient and investigator unaware of treatment assignment
  - triple blind: patient, investigator and study analysts unaware of allocation (double blind and triple blind are often used synonymously); in fact it may be unclear when reading a published study who in fact is 'blinded'
  - unblinding refers to disclosure of allocation (criteria for unblinding are usually built into study protocols for safety reasons – for example if there is an adverse event, then caregivers may be given information about whether the patient is in the control or experimental group and the study medication discontinued)

Intention-to-Treat (ITT) analysis

- intention-to-treat is the analysis of study groups according to which treatment they were assigned, and not according to which treatment (if any) they actually received
- advantages of ITT:
  - retains balance in prognostic factors arising from the original random treatment allocation
  - unbiased estimate of treatment effect
  - admits non-compliance and protocol deviations, which is what happens in the 'real world'
- limitations of ITT:
  - can make analysis confusing if there are a large number of crossovers between the experimental and control arms of the study
  - estimate of treatment effect is more conservative because of dilution of effect by noncompliance

Baseline data in clinical trials

- knowing the baseline characteristic of the study participants is important because this:
  1. allows you to determine how similar the study participants are to your patients (the generalisability of the results, or external validity of the study)
  2. shows how similar the groups are with respect to factors that may impact prognosis or study outcomes (this is how balanced the groups are as a result of the randomization process, this affects the internal validity of the study)
- information that should be provided with respect to baseline data includes:
  1. typical demographic factors (age, sex)
  2. factors that are part of standard therapy and which may affect outcomes (for example medications taken)
  3. important co-morbid diseases
  4. factors that could predict adverse effects of therapy
  5. any factors which were used in stratified randomization
  6. prespecified subgroups

- comparability of the experimental and control group using baseline data:
  - if randomization and allocation concealment have been performed correctly any difference in the baseline characteristics of the groups will have occurred by chance
  - P values are <u>not</u> useful for comparing baseline characteristics *(the differences will have occurred by chance and we are not considering a hypothesis that there is a difference between the groups, as well because there are usually many baseline characteristics presented then chance alone dictates that at least one characteristic will show a statistically significant difference)*
  - if there are imbalances between the groups that are potentially important this must be discussed in the study, and may be accounted for using an adjusted analysis of the data

Attrition

- patients and/or data may be lost to follow-up in clinical studies
- attrition can introduce bias if the characteristics of the study participants lost to follow-up are different between the experimental and control groups
- in some studies a large proportion of patients may be lost to follow-up and this must be taken into account reviewing the study, and should be discussed by the authors
- statistical adjustments for various scenarios can be used to help with interpretation of the study *(for example a sensitivity analysis may be used with an assumption that all the experimental patients lost to follow-up had died, and all the control group patients lost had good outcomes, a kind of 'worst-case scenario' analysis)*

Compared to what?

1. Placebo controlled trials:

- placebo controlled trials use the null hypothesis ('the difference between the experimental and control therapies for the outcome of interest is zero')
- statistical analysis then looks at whether the data is consistent with the null hypothesis
- placebo control is used to control for natural variations in disease, as well as bias

2. Active control trials:

- NB: the concept of equivalence and noninferiority trials is introduced here, however this is to provide some background for you. There are some quite specific methodological issues involved with these trials, which will not be covered.
- placebo controls are not ethical if a known effective therapy exists
- active control trials are often used to study a new therapy that is cheaper, easier to administer, or has potentially less adverse effects than the standard therapy
- active control trials may be:
    1) noninferiority trials: designed to show that the new therapy is not worse than the active control by more than a prespecified amount (Δ)
    2) equivalence trials: designed to show that the new treatment is not different from the control therapy by more than a specific amount (equivalence margin) (this is a 2-sided test; the true treatment effect is between - Δ and +Δ). Most trials are noninferiority trials rather than equivalence.
- results of a trial that shows no difference between two therapies (A and B) may be due to:
    1) A and B are equally effective or
    2) A and B are equally ineffective or
    3) the trial was too small to detect a difference between A and B
- this means that with an active-control trial we need to know that the active control is effective, and by how much
- the study population of the active-control trial should be similar to the study population that was used to show that the active-control therapy was effective
- the decision of what difference between the two therapies would be clinically acceptable, and yet have them considered equivalent must also be decided (a priori)

Validity – external vs internal validity

- validity refers to the soundness of a study, a study is internally valid if it is well-designed and unbiased
- a study is externally valid if the results are applicable to patients seen in ordinary practice

Efficacy and effectiveness

1. **efficacy**
    - study of efficacy is a study of whether a therapy has the ability to bring out the intended effect, in an ideal world
    - efficacy trials are designed to maximize the potential for detecting an effect of the experimental therapy (ie. answer the question – 'does this therapy work under optimal conditions?')
    - efficacy trials are randomized, using a homogeneous study population
2. **effectiveness**
    - effectiveness studies ask the question 'does this therapy work under usual circumstances?'

Standard for reporting of clinical trials (CONSORT):

- CONSORT (CONsolidated Standards Of Reporting Trials) is a statement, checklist and standardized flowchart designed for improving reporting of clinical studies
- the CONSORT guidelines have been adapted by a number of journals and require published trials to conform to the guidelines
- see the flowchart for study participants that is used to describe the flow of patients in the study

Trial Registration

- a number of high-profile journals now require clinical trials to be registered with a central database before the trial begins, if not the study will not be able to be published in the journal
- this initiative is designed to decrease problems such as incomplete reporting of data, changing study outcomes, not publishing negative trials etc.

Publication bias:

- publication bias occurs when the publication of research depends on the direction of the study results, and whether they are statistically significant
- publication bias includes 'positive' results being:
    - more likely to published
    - published rapidly
    - in more than one source; duplication (this can cause overlap of information, making it seem as though there is more trial data than there actually is)
    - cited by others

    - the medical literature therefore may be a selective and biased subset of studies and outcomes
    - as a result of publication bias, and instances where negative trials, or trials with adverse events have not been published there is a move to a mandated trial registry (several of the core clinical journals have required trial registration as a prerequisite for publication of trials in the journal in the future)

References:

1. Treatment allocation in controlled trials: why randomize? BMJ 1999;318:1209.

2. Statistics notes: How to randomize. BMJ 1999;319:703.

3. Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA 2001;286:821.

4. Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA 2001;286:821-830.

5. Subverting randomization in controlled trials. JAMA 1995;274:1456.

6. Statistics notes: Treatment allocation by minimization. BMJ 2005;330:843.

7. Statistics notes. Concealing treatment allocation in randomized trials BMJ 2001;323:446.

8. Adequacy and reporting of allocation concealment: review of recent trials published in four general medical journals. BMJ 2005;330:1057.

9. The landscape and lexicon of blinding in randomized trials. Ann Int Med 2002;136:254.

10. Physician interpretations and textbook definitions of blinding terminology in randomized controlled trials. JAMA 2001;285:2000.

11. Therapy and validity. The principle of intention-to-treat. Chapter 2B1 Users' Guide to the Medical Literature. AMA Press, 2002.

12. Inclusion of patients in clinical trial analysis: the intention-to-treat principle. MJA 2003;179:438.

13. Post-randomization exclusions: the intention to treat principle and excluding patients from analysis. BMJ 2002;325:652.

14. Intention-to-treat principle. CMAJ 2001;165:1339.

15. What is meant by intention-to-treat analysis? Survey of published randomized controlled trials. BMJ 1999;319:670.

16. Baseline data in clinical trials. MJA 2003;179:105.

17. Reporting attrition in randomized controlled trials. BMJ 2006;332:969-971.

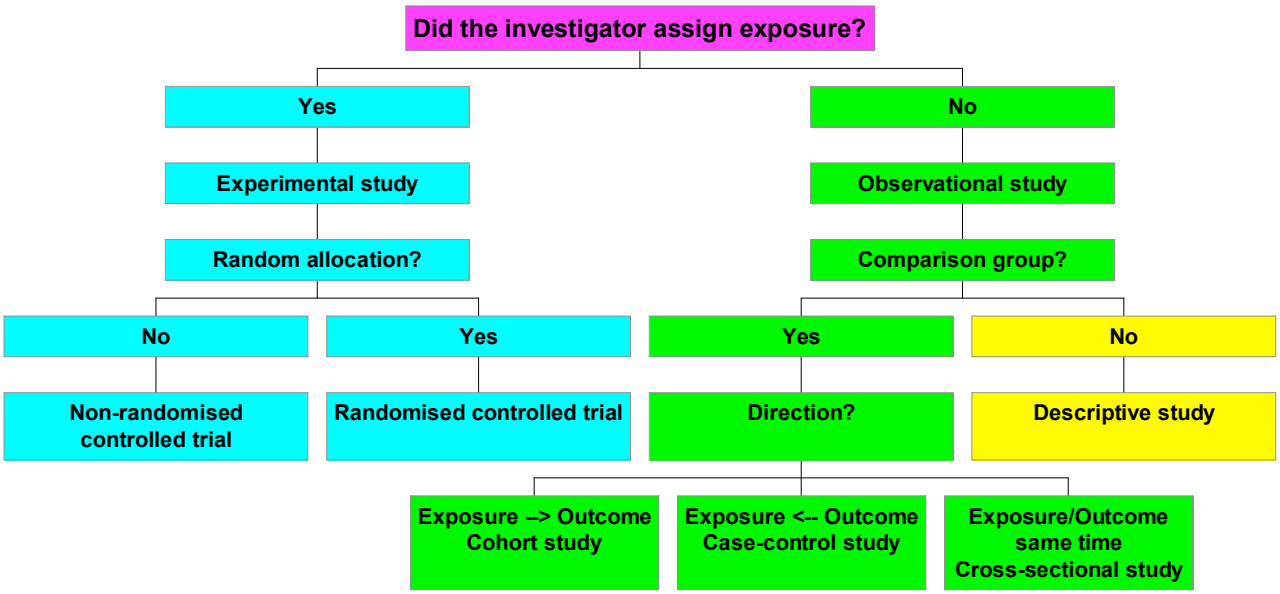
18. Reporting of noninferiority and equivalence randomized trials. JAMA 2006;295:1152-1160.

19. Good enough: a primer on the analysis and interpretation of noninferiority trials. Ann Int Med 2006;145:62-69.

20. CONSORT. http://www.consort-statement.org/Explanation/explanation.htm#checklist

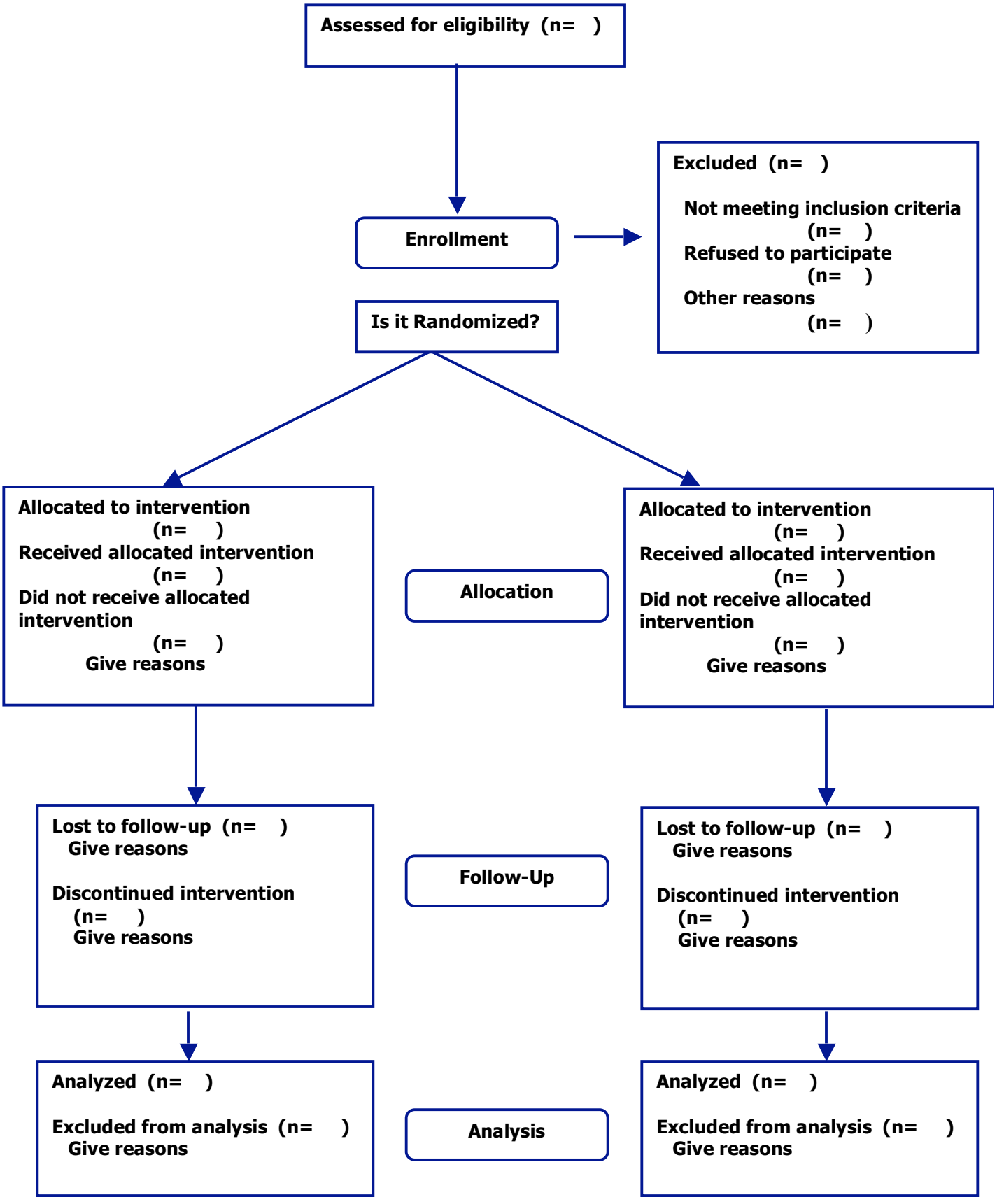
21. Trial Registration Report Card. NEJM 2005; 353:2809.

Appendix I
Observational Studies



Study Design	Advantages	Disadvantages
Case-control study	• compare prevalence of suspected causal factors in cases and controls, and identify associations, can look at wide range of risk factors • quick • relatively inexpensive • no loss to follow-up • only feasible method for very rare disorders or if there is a long lag time between exposure and outcome	• relies on recall, or records for exposure status • prone to selection and recall bias • not useful for rare exposures • confounders • no randomization • cannot measure risk (odds ratio used in analysis)
Cohort study	• measures risk of disease association with exposure to a factor prospectively • can look at exposure to factors that are rare • can establish timing and directionality of events • outcome assessment can be standardized • easier and less expensive than RCT	• blinding difficult • no randomization • for rare diseases large sample size and long follow-up required (expensive and time consuming) which may not be feasible • exposure may be linked to a hidden or unknown confounder • does not prove causality
Cross-sectional survey	• inexpensive, simple • can document co-occurrence of disease and possible risk factors in individuals • may be useful to studying chronic disease with a high prevalence, but low incidence where cohort study may not be feasible	• establishes association but not causality • subject to problems of information and measurement bias • uncontrolled confounders possible

Appendix II
CONSORT Guidelines Flowchart



Appendix III
Sources of Bias

Definition of bias:

- systematic introduction of error into a study that can distort the results; bias affects the internal validity of the study
- when reviewing a study you need to think about the effect of possible bias on the results of the study

Observational studies

- all observational studies have built-in sources of bias
- these include:
  1. selection bias
     - selection (sampling) bias occurs when patients are selected in a way that will influence the outcome of the study
     - in observational studies ask 'are groups similar in all important respects except for the exposure (cohort study) or outcome/disease (case-control study)?'
     - *for example: if you are studying the outcome of laparoscopic surgery, are the patients who had laparoscopic surgery less likely to be obese?, are people with a diet high in fibre possibly less likely to smoke and more likely to exercise?*
  2. information bias (also known as measurement or observation bias)
     - occurs when methods of measurement are different in the different patient groups
  3. confounding – could the results be accounted for by some other factor associated with both the exposure and the outcome, but not directly involved causally?
  4. chance

Selection bias:

some specific types and examples of selection bias: (from Lancet Handbook of Essential Concepts in Research, 2006) – <u>these are provided for background information only</u>

membership bias: *members of a group (like joggers) might differ in other important respects from others*

incidence-prevalence (Neyman) bias: *may occur in diseases that are quickly fatal or transient. For example a hospital-based case-control study of snow shoveling and heart attacks would miss everyone who died in the driveway and never got to hospital.*

unmasking or detection bias*: an exposure may lead to a search for the disease/outcome in addition to the outcome so that the apparent risk is increased. For example hormonal therapy may cause endometrial bleeding, and this might result in further diagnostic tests, and the detection of endometrial cancer.*

non-respondent bias: *some population groups are less likely to return questionnaires and surveys. For example smokers are less likely to return questionnaires if there are questions about smoking, healthy people are more likely to return questionnaires.*

referral bias*: studies performed in tertiary centres will include patients referred as subjects, but patients with less severe disease may not be referred*

Information bias: synonyms measurement, classification, and observation bias

- information about outcomes (cohort study) or exposure (case-control study) should be gathered the same way for the groups
- *the different types of information bias below are provided for information only*

subject bias*: this is bias introduced by the subject (for example trying to please by not reporting side effects, reporting being compliant with medications even if they aren't, etc.)*

recall or reporting bias: *when patients who experience an adverse outcome have a different chance of recalling an exposure, independent of the extent of the exposure. Recall bias occurs most often in retrospective cohort and case-control studies.*

Hawthorne effect: *this is a change in behaviour that may occur when people are part of a study and are aware they are being observed.*

detection bias: *the tendency to look more carefully for an outcome in one of the groups.*

interviewer bias: *asking questions in such a way that the answer regarding exposure or outcome is more likely in one group.*

Confounding:

- factors that distort the true relationship of the study variable of interest because they are also related to the outcome of interest. Confounders are often unequally distributed between cohort or case-control groups. For example vitamin E was associated with a better outcome for cardiovascular disease, but people who took vitamin E also were less likely to smoke, and had higher socioeconomic status. These are confounders.
- confounders can be dealt with statistically using multivariate analyses to adjust for these variables, but that implies that we know what they are.

Outcome reporting bias:

- this is the selective reporting of some results but not others in trial publications (for example if there are three primary outcomes being studied, only one (more likely the positive outcome) may be published