

Translating Evidence into Efficacy: Evaluating Strengths and Weaknesses of Different Study Designs

Nathan D. Wong, PhD, FACC
Associate Professor and Director
Heart Disease Prevention Program
University of California, Irvine, CA USA

Why are papers rejected for publication? (The Top 11 Reasons)

1. The study did not address an important scientific issue
2. The study was not original
3. The study did not actually test the authors' hypothesis
4. A different type of study should have been done
5. Practical difficulties led the authors to compromise on the original study protocol (e.g., recruitment, procedures)
6. The sample size was too small
7. The study was uncontrolled or inadequately controlled
8. The statistical analysis was incorrect or inappropriate
9. The authors drew unjustified conclusions from the data
10. There is a significant conflict of interest among authors
11. The paper is so badly written that it is incomprehensible

Critical Appraisal

1. **Why was the study done, and what clinical question is being asked?** (a brief background, review of the literature, and aim / hypothesis should be stated)
2. **What type of study was done?** (experiment, clinical trial, observational cohort or cross-sectional study, or survey)

Critical Appraisal (cont.)

3. Was the design appropriate for the research?
- Clinical trial preferred to test efficacy of treatments (e.g., HPS simvastatin trial)
 - Cross-sectional study preferred for testing validity of diagnostic/screening tests or risk factor associations (e.g., NHANES III)
 - Longitudinal cohort study preferred for prognostic studies (e.g., Framingham)
 - Case-control study best to examine effects of a given agent in relation to occurrence of an illness, esp. rare illnesses (e.g., cancer)

Elements of a Formulated Question

- **Patient or Population:** Who is the question about? (e.g., pts with diabetes mellitus)
- **Intervention or Exposure:** What is being done or what is happening to the patient/population? (e.g., tight control)
- **Outcome(s):** How does the intervention affect the patient/population (mortality, CHD incidence)
- **Comparison(s):** What could be done instead of the intervention? (e.g., standard management)

Was the study original?

- Few studies break entirely new ground
- Many studies add to the evidence base of earlier studies which may have had other or more limitations
- Meta-analyses depend on literature containing multiple studies addressing a question in a similar manner

Features Distinguishing New vs. Previous Studies

- Is the study in question bigger in sample size, or with longer-follow-up (e.g., adding to meta-analyses of previous studies)?
- Is methodology more rigorous (e.g., having addressed criticisms of previous ones)?
- Is the population studied different from that of previous studies (ages, gender, ethnic groups)?
- Does the new study address a clinical issue of sufficient importance so it is politically desirable even if not scientifically necessary?

Greenhalgh T, BMJ 1997; 315: 305-8

Study Population Issues

- **How were the subjects recruited?** Is there potential recruitment bias (e.g., from taking respondents of advertisements), or is survey done in a random (e.g., random digit-dialing) or consecutive sample?
- **Who was included?** Many trials exclude those who have co-morbidities, do not speak English, or take other medications—may provide scientifically clean results, but may not be representative of disease in question.

Study Population (cont.)

- **Who was excluded?** Study may exclude those with more severe forms of disease, therefore limiting generalizability
- **Were subjects studied in “real-life” circumstances?** Is the consenting process describing the benefits/risks, access to study staff, equipment available, etc. be similar to that in an ordinary practice situation?

Was the study design reasonable?

Two basic questions can be asked:

- What intervention or maneuver is done, and what is the comparison group? (e.g., active therapy vs. placebo or usual care)
- What are the outcome(s) and how are they measured? (e.g., standard definition for outcome, validation of questionnaires)

Examples of problematic description of methods

- We determined smoking status from medical records: **assumes medical records are 100% accurate**
- We surveyed doctors' treatment of hypercholesterolemia: **assumes that what doctors say they do is what they actually do**
- We approached 200 patients (50% male), and 120 of them (40% male) agreed to participate: **failure to adequately describe study population and how respondents may have differed from non-respondents**
- We randomized subjects to aggressive management vs. usual care: **failure to give details about each study group—should be reproducible so others can repeat**

Was systematic bias avoided or minimized?

- Systematic bias is defined as anything that erroneously influences the conclusions about groups and distorts comparisons.
- Aim is to have the groups compared as similar as possible except for the difference being examined (same study procedures, contacts with study staff, same methods to assess outcomes in blinded manner)

Randomized Controlled Trials (RCTs)

- Randomized controlled trial eliminates systematic bias (in theory) by allocating treatments among participants in a random fashion
- The allocation process eliminates selection bias in group characteristics (check comparability of baseline characteristics such as age, gender, severity of disease and covariate risk factors)
(selection bias)

RCT's (cont.)

- Need to check for any biases in treatments or care provided between the groups (**performance bias**)
- Need to check for differences in follow-up and withdrawals between the groups— large differences in loss to follow-up can compromise validity of trial (**exclusion bias**)
- Need to check for any differences in how the outcomes were ascertained between the groups (**detection bias**)

Advantages of RCT's

- Allows rigorous evaluation of a single intervention in a well-defined population
- Prospective design (events occur after the intervention)
- Presumably eradicates bias by comparing two identical groups (but see below)
- Allows for meta-analysis

Disadvantages of RCT's

- Expensive and time-consuming
- Often performed on too few patients, or undertaken for too short a period
- Often funded by large research bodies or pharmaceutical companies which dictate the research agenda
- Often involves many inclusion and exclusion criteria to recruit those who will respond to intervention, thus limiting generalizability to a more general patient population.

Non-randomized Controlled Trials

- Treatment intervention may be applied in one group of patients (hospitalized), and “control” intervention in a separate group of patients from another source (outpatient clinic)
- May be done when randomization is unethical or inappropriate (e.g., trial examining exposure to cigarette smoking)
- Need to check for any self-selection biases—are there any baseline differences between the two groups that could invalidate the effects of the intervention? (e.g., treated group could have more severe confounding risk factors)

Observational Studies

- Cohort and case-control studies seldom can identify two groups of subjects (exposed vs. unexposed or cases vs. controls) that are similar (e.g., in demographic or other risk factors).
- Much of the controlling for baseline and/or follow-up differences in subject characteristics occurs in the analysis stage (e.g., multivariable analysis as in Framingham)

Observational Studies (cont.)

- While statistical procedures may be done correctly, have we considered all possible confounders?
- Some covariates may not have been measured as accurately as possible, and more often, may not be even known or measured.

Prospective (Cohort) Studies

- Cohort studies begin with identification of a population, assessment of exposure (e.g., lipid or BP levels)
- Follow-up to the occurrence of outcomes (CHD events)-- temporal sequence to events is known

Cohort Studies (cont.)

- Difficult to ascertain effect of exposure because of many differences between exposed and unexposed groups (confounding factors).
- Statistical adjustment for known risk factor differences can help, but unknown factors that may differ between exposed and unexposed groups will never be adjusted for.

Prospective Cohort Example: Framingham Heart Study

- Longest running epidemiologic study
- Began with 5209 persons aged 30-62 at baseline in 1948, studied biennially to date (most are deceased now)
- Risk factors measured at each examination, some began later (e.g., HDL-C around 1970) or done only at certain exams (echocardiography, CRP)
- Event ascertainment/adjudication involves panel of 3 physicians reviewing medical records

Retrospective (Case-Control) Studies

- Case-Control studies often match cases (esp. of a rare disease) with controls by age, sex, and other factors
- Retrospective ascertainment of exposure status – may be determined by unreliable means (past memory, medical records)

Case-Control Studies (cont.)

- Cannot determine for sure whether exposure preceded development of disease
- Also difficult to identify all differences between cases and controls that can be statistically adjusted for

Examples where observational studies have taken us down the wrong path.....

- Meta-analysis of observational studies have shown a 50% lower risk of CHD among estrogen users vs. non-users (which may have had many unknown differences that were not adjusted for), but recently randomized trials (HERS, WHI) show no benefit
- Numerous prospective studies show a 25-50% lower risk of CHD among those taking vitamin E and other antioxidants vs. placebo— recent randomized trials (e.g., HOPE, HPS) show no benefit.

Cross-Sectional Surveys

- Examples: NHANES III, CHIS (telephone)
- Surveys should include a representative, ideally randomly-chosen (rather than a small sample of approached subjects who actually agree to be surveyed) sample.
- Data collected at a single time, but may relate to questions about habits in the past
- Can statistically adjust for confounders, but difficult to establish the temporal nature of exposure and disease.

Case Reports and Series

- Provides “anecdotal” evidence about a treatment or adverse reaction
- Often with significant detail not available in other study designs
- May generate hypotheses, help in designing a clinical trial.
- Several reports forming a “case series” can help establish efficacy of a drug, or thru adverse reports, cause its demise (example: Cerivastatin fatal cases of rhabdomyolysis).

Duration of Follow-up

- Is the planned follow-up reasonable and practical for the study question and sample size utilized?
 - effect of a new painkiller on degree of pain relief may only require 48 hours
 - effect of a cholesterol medication on mortality may require 5 years)

Completeness of Follow-up

- Conclusions of study can be at jeopardy if there are more unknown subjects lost to follow-up than explain the differences in outcome.
- Ignoring those withdrawals will often bias results in favor of the intervention, so standard to analyze results on an “intention-to-treat” basis, including all who were originally randomized.

Follow-up (cont.)

- Patient withdrawal may be caused by:
 - Incorrect entry of patient into a trial
 - Suspected adverse reaction to a drug (although many drug AE's are similar to placebo AE's)
 - Loss of patient motivation
 - Withdrawal by clinician for clinical reasons
 - Loss to follow-up
 - Death

Hierarchy of Evidence

(for making decisions about clinical interventions)

1. Systematic reviews and meta-analyses
2. Randomized controlled trials with definitive and clinically significant effects
3. Randomized controlled trials with non-definitive results
4. Cohort studies
5. Case-control studies
6. Cross-sectional surveys
7. Case reports

Summary

- Is the study original?
- Whom is the study about?
- Was the design of the study sensible?
- Was systematic bias avoided or minimized?
- Was the study large enough and continued long enough to make the results and conclusions credible?