Lecture 22

Meta-Analysis: Resolving Inconsistencies across Studies

Thought Question 1:

Suppose new study involving 14,000 participants found a relationship between particular food and certain type of cancer.

Report on study notes “past studies of this relationship have been ***inconsistent***. Some have found relationship, others have not.”

What might be ***explanation for inconsistent results*** from different studies?

Thought Question 2:

Suppose **ten similar studies**, all on the same kind of population, were conducted to determine the relative risk of heart attack for those who take aspirin and those who don't.

To get an overall picture of the relative risk we could **compute a separate confidence interval** for each study ***or*** **combine all the data** to create one confidence interval.

Which method do you think would be **preferable**? Explain.

Thought Question 3:

Suppose two studies compare surgery versus relaxation for sufferers of chronic back pain. One was done at back specialty clinic and the other at a suburban medical center. Result was relative risk of further back problems following surgery versus following relaxation training.

To get overall picture, we could compute **separate confidence interval** for each study ***or*** **combine data** to create one confidence interval.

Which method preferable? Explain.

Thought Question 4:

Refer to Thought Questions 2 and 3.

If two or more studies have been done to measure the same relative risk, give one reason why it would be better to combine the results and one reason why it would be better to look at the results separately for each study.

25.1 Need For Meta-Analysis

Most relationships we now study are moderate in size, researchers often fail to find a statistically significant result.

Number of participants in a study is a **crucial factor** in determining whether the study finds a "significant" relationship or difference, and many studies are simply too small to do so.

➡ reports often published that **appear to conflict** with earlier results, confusing public and researchers.

The Vote-Counting Method

Find all studies conducted on topic and ***simply count how many had found a statistically significant result***. Often discount entirely all studies that had not, and attempt to explain any remaining differences in study results by subjective assessments.

This **vote-counting method** is seriously flawed unless number of participants in each study taken into account.

Example: If ten studies of effect of aspirin on heart attacks had each been conducted on 2200 men, rather than one study on 22,000 men, none of ten studies would be likely to show a relationship.

What Is Meta-Analysis?

Meta-analysis is a collection of statistical techniques for combining studies.

These techniques focus on the *magnitude of the effect* in each study, rather than on either a vote-count or a subjective evaluation of the available evidence.

Term *meta-analysis* coined by Gene Glass in 1976 paper entitled “Primary, secondary and meta-analysis of research.”

What Meta-Analysis Can Do:

Find definitive answers to questions about small and moderate relationships by taking into account data from a number of different studies.

25.2 Two Important Decisions For The Analyst

When reading the results of meta-analysis, you should know …

1. Which studies were included.
2. Whether the results were compared or combined.

Q1: Which Studies Should Be Included?

Types of Studies

Include **all** studies or only published in reviewed journals?

Timing of the Studies

Has technology changed so results of studies conducted many years ago may not be comparable to those today?

Quality Control

Require studies meet stringent quality criteria – e.g. use of control group, randomization, etc.

Accounting for Differences in Quality

Include all studies but account for differences in quality in the process of analysis.

Assessing Quality

Have two independent (blinded) assessments of study quality.

Q2: Results Be Compared or Combined?

Meta-Analysis and Simpson's Paradox

Two studies comparing surgery to relaxation as treatments for chronic back pain. One conducted at back-care specialty clinic, other at suburban medical center. Patients with most severe problems tend to seek out specialty clinic. Relaxation training may be sufficient and preferable for suburban medical center, but surgery preferable for specialty clinic.

Populations Must Be the Same and Methods Similar

Were the same populations sampled and similar methods used?

Case Study 25.1: Smoking and Fertility

20 studies that examined whether men who smoked cigarettes had lower sperm density than men who did not smoke. ***Studies differed*** …

- Ten reported person measuring sperm density was **blind** to the smoking status of the participants.
- Six **defined "smoker"** as someone who smoked at least ten cigarettes per day.
- Thirteen **used infertility clinics** as a source of participants, seven did not.

None of these factors resulted in a difference in outcome.

Source: Vine et al., January 1994.

Case Study 25.1: Smoking and Fertility

Authors conducted an analysis using all data, as well as separate analyses on two types of populations: those attending infertility clinics and those not attending.

Results: Using all data combined, giving more weight to studies with larger sample sizes … ***reduction in sperm density for smokers compared with nonsmokers was 12.6%*** (95% CI 8.0% to 17.1%).

Test of null hypothesis that the reduction in sperm density for smokers in the population is actually zero resulted in a ***p-value less than 0.0001***.

Estimate of reduction in sperm density for the normal (not infertile) men only was even higher at 23.3%.

Case Study 25.1: Smoking and Fertility

The results of this meta-analysis indicate that smokers' sperm density is on average 13% [when studies are weighted by sample size] to 17% [when studies are given equal weight] lower than that of nonsmokers. . . . The reason for the inconsistencies in published findings with regard to the association between smoking and sperm density appears to be the result of ***random error and small sample sizes in most studies****. Consequently, the* ***power is low*** *and the chance of a false negative finding is high.*

(*Source:* Vine et. al., January 1994, p. 40)

Be cautious: these were obviously observational studies, so there may be other confounding factors.

25.3 Some Benefits of Meta-Analysis

1. Detecting small or moderate relationships
2. Obtaining a more precise estimate of a relationship
3. Determining future research
4. Finding patterns across studies

Determining Future Research

Example 1: Designing Better Experiments

Early meta-analysis to study extrasensory perception (Honorton, 1985) had many flaws: improper randomization of target pictures and potentially sloppy data recording. Early studies used photographs that sender was supposed to transmit mentally to a receiver. If sender was holding photograph of true target and receiver was later asked to pick which of four pictures had been true target, receiver could have seen the sender's ***fingerprints***, and these—and not some psychic information—could have provided the answer.

New experiments, reported in Case Study 22.1, designed to be **free of flaws** identified in the first meta-analysis.

Finding Patterns across Studies

Example 3: Grouping Studies by Orientation

Bruvold (1993) compared adolescent smoking prevention programs characterized as having one of four orientations: "**rational**" orientation: lectures and displays of substances; "**developmental**" orientation: lectures with discussion, problem solving, and some role playing; "**social norms**" orientation: participation in community and recreational activities; "**social reinforcement**" orientation used discussion, role playing, public commitment not to smoke.

Individual studies are likely to focus on one orientation only. But a meta-analysis can group studies according to which orientation they used and do a comparison.

25.4 Criticisms of Meta-Analysis

1. Simpson’s Paradox
2. Confounding variables
3. Subtle differences in treatments of the same name
4. The file drawer problem
5. Biased or flawed original studies
6. Statistical significance versus practical importance
7. False findings of “no difference”

Possibility of Confounding Variables

Meta-analysis essentially observational, various treatments cannot be randomly assigned across studies. May be differences across studies confounded with treatments used.

Subtle Differences in Treatments of Same Name

Example: chemotherapy may be applied weekly in one study but biweekly in another.

The File Drawer Problem

Some studies may not be discovered by the meta-analyst. Likely studies that did not achieve statistical significance and thus were never published, called the **file drawer problem** (studies filed away somewhere but not publicly accessible). If statistically significant studies more likely to be available, meta-analysis may overestimate size of relationship.

Biased or Flawed Original Studies

If original studies flawed or biased, so is the meta-analysis. **Example:** Meta-analysis on whether oat bran lowers cholesterol. Many studies included were partly financed by Quaker Oats. (*Source:* Ripsin et al., 1992, p. 3317)

Statistical Significance vs Practical Importance

Meta-analysis likely to find statistically significant results because typically combined studies provide very large sample sizes. Important to learn magnitude of an effect in meta-analysis and not just if it is statistically significant.

False Findings of “No Difference”

Meta-analysis can erroneously reach conclusion of no difference/relationship—when simply not enough data to find one that was statistically significant. Be careful not to confuse overall sample size with one used for any particular subgroup.

Case Study 25.2: Mammograms

- February 1993, National Cancer Institute convened an international conference. ***Meta-analysis conducted on the effectiveness of mammography*** as a screening device.
- Results for women 40–49 years was: "For this age group it is clear that in the first 5–7 years after study entry, there is no reduction in mortality from breast cancer that can be attributed to screening" (Fletcher et al., 20 October 1993, p. 1653).
- Results lead to withdrawal of National Cancer Institute's support for mammograms for women under 50. American Cancer Society refuted study and announced that it would not change its recommendation.

Case Study 25.2: Mammograms

Meta-analysis considered eight randomized experiments conducted over 30 years, including 500,000 women. Sounds impressive. But not all studies included women under 50.

Even pooling the data from all eight randomized controlled trials produces insufficient statistical power to indicate presence or absence of benefit from screening. In the eight trials, there were ***only 167,000 women*** *(30% of participants) aged 40–49,* ***a number too small to provide a statistically significant result****.* *(Sickles and Kopans, 20 October 1993, p. 1622)*

Results are inconclusive because of small sample sizes and possible methodological flaws.