Current Issues with Meta-analysis in Medical Research

Demissie Alemayehu
Pfizer & Columbia

Semanade Bioinformatica, Bioestadistica Analists de Supervivencia
Instituto Panamericano de Estudios Avanzados en Probabilidad y Estadistica
May 2-8, 2010
Abstract

In the absence of definitive trials on the comparative safety and effectiveness of drugs, a systematic synthesis of available data may provide critical information to help decision making by medical professionals, patients and other stakeholders. However, uncritical and lopsided use of pooled data to inform decision about important health-care issues may have consequences that adversely impact public health, stifle innovation, and confound medical science. In this talk, current methodological issues in meta-analysis are highlighted, and advantages and disadvantages of alternative techniques are evaluated. The need for a more integrated strategy toward the synthesis, interpretation and dissemination of pooled data is suggested.

The talk, intended for applied statisticians and others involved in medical research, will involve review of case studies to illustrate relevant issues.
Outline

• Part I: Standard Meta-analysis
  – A Case Study: The Avandia Analysis
  – Procedures for Meta-analysis
  – Criteria for Causality
  – The Avandia Analysis Revisited

• Part II: Non-standard Meta-analysis
  – Cumulative Meta-analysis
  – Indirect Comparison
  – Observational Data Analysis
  – Comparative Effectiveness Research

• Concluding Remarks
Avandia Meta-analysis


“Diabetes drug called heart death risk”

Steve Sternberg Source: *USA Today*: May 22, 2007
Definition: Meta-analysis

• Meta-Analysis is a statistical approach for combining information from independent studies to address a pre-specified hypothesis of interest.
  – Based on systematic literature review
  – Provides precise estimate of treatment effect, giving due weight to studies included
Benefits of Meta-analysis

• Precision / Power
  – Individual clinical trials may lack power to provide conclusive results
  – Particularly useful for rare events

• To assess consistency (generalizability) of results
  – To settle controversies arising from conflicting studies

• Answer questions not posed by the individual studies
  – generate new hypotheses
When to do a Meta-analysis

• Data are available from more than one study
• There are no major differences in the study characteristics
• Outcome has been measured in similar ways
When **not** to do a Meta-analysis

- ‘Garbage in - garbage out’
  - A meta-analysis is only as good as the studies in it

- Beware of ‘mixing apples with oranges’
  - Compare like with like

- Beware of reporting biases
Steps in Meta-analysis

1. Detailed written protocol development
2. Comprehensive / systematic search for eligible studies
3. Presentation of results
4. Sensitivity analysis / test for heterogeneity performed
5. Discussion of limitations of analysis
6. Reporting of results in light of available body of knowledge
Protocol Development

• Formulation of problem to be addressed
• Eligibility criteria for studies to be included defined *a priori*
• Statistical methods for combining the data
Selection of Studies

- Minimize Sources of Bias
  - Publication Bias
    - Positive studies more likely to be published than negative studies
  - Reviewer Bias
    - Tendency to include only studies that favor the hypothesis of interest

- Define universe of studies to be considered
- Define explicit & objective criteria for study inclusion/ rejection based on quality grounds

Hierarchy of Evidence

Selection of Studies, cont’d

- **Popular Databases**
  - PubMed / MedLine
  - Cochrane Review
  - Trial result registries

- **Other Sources**
  - Hand Search: FDA, library
  - Personal references, emails
  - Web: Google, etc.
Publication Bias

• Decision to publish is influenced by study result
  – Small studies of no effect unlikely to be published
  – Overestimation of treatment effect.

• Detection:
  – No universally accepted methods
    • Funnel plots treatment effect against its precision (sample size or variance)
    • Egger’s Weighted Regression Statistic
    • Begg’s Test
  – Most methods heuristic.
    • Operating characteristics not well-studied
Publication Bias

Funnel Plot: Plots of Variability or Sample Size vs. Effect Size

- Precision: increasing function of study size
- In absence of bias, results from small studies scatter widely at bottom of graph
- Publication bias: asymmetrical funnel plots
Publication Bias (cont.)

• **Egger’s Weighted Regression Statistic**
  - Test for asymmetry of the funnel plot.
  - Effect size regressed on precision (inverse of se)

• **Begg and Mazumdar**
  - Test for interdependence of variance and effect size using Kendall’s tau
  - If publication bias, expect that high se’s (i.e., small studies) would be associated with larger effect sizes.
• **Summary of Treatment Effects**
  - Mean Difference
  - Odds Ratio
  - Risk Difference
  - Number Needed to Treat (NNT)

• **Inference:**
  - Different approaches exist, but there is no single "correct" method
  - Two General Approaches
    - Fixed Effect Models
    - Random Effects Models
  - Bayesian Meta-analysis
• Fixed Effect Models
  – If it is reasonable to assume underlying effect is the SAME for all studies
    • Only one source of sampling error:
      - within study
  – Problems with ignoring heterogeneity:
    • confidence intervals too narrow

\[ Y_i \sim N\left(\theta_i, \sigma_i^2\right) \]

\[ \theta_i = \theta \]
Fixed Effect Models, \textit{cont’d}

- Inverse-variance

\[ \hat{\theta} = \frac{\sum_{i=1}^{r} \hat{\theta}_i w_i}{\sum_{i=1}^{r} w_i} \]

\( r = \) number of studies in the meta-analysis, and

\[ w_i = \frac{1}{s.e.(\hat{\theta}_i)^2} \]
Fixed Effect Models, *cont’d*

- **Mantel-Haenszel**
  - Confidence interval for the Mantel-Haenszel odds ratio calculated using the Robins, Breslow and Greenland variance formula ([Robins et al., 1986](#))

- **Peto-Method**
  - Often used for rare events
Statistical Methods, cont’d

• Random Effects Models
  - True effect could vary from study to study
    - E.g., effect size higher in older subjects
  - Sources of sampling error:
    - within study
    - between studies
  - Gives wider confidence interval

\[ Y_i \sim N\left(\theta_i, \sigma_i^2\right) \]

Random Effects \[ \theta_i \sim N(\theta, \tau^2) \]

\[ \Rightarrow Y_i \sim N(\theta_i, \sigma_i^2 + \tau^2) \]
Random Effects Models

- Common method: DerSimonian-Laird
Bayesian Meta-analysis
Bayesian Hierarchical Model

Priors on \((\theta, \tau^2)\):

- Usually noninformative normal prior on \(\theta\)
  - For both fixed and random effects
- For random effects:
  - Noninformative inverse gamma or uniform prior on \(\tau^2\)
  - Inferences sensitive to prior on \(\tau^2\)
Bayesian Meta-analysis

• Advantages

  – Allows probability statements:
    • E.g. Post probability median survival of drug A > that of drug B.
  • Multiplicities not an issue:

  – For decision making
    • Can incorporate probabilistic loss function
Bayesian Meta-analysis

• Disadvantages
  – Use of subjective prior beliefs
  – Eliciting prior beliefs is a non-trivial exercise
    • Few guidelines to help the Bayesian analyst.
  – Different prior distributions can generate varying results
  – Computational complexity
Meta-analysis of Rare Events

• Large-sample procedures may not be appropriate for pooling rare events

  – Event rates < 1%:
    • Peto one-step odds ratio
      – Reasonable bias, power
      – Bias substantial, unbalanced case
  – Event rates 2%-5%:
    • Logistic Regression, Exact Method, MH OR better performance (in terms of bias) than Peto
  – Risk difference
    • Conservative confidence intervals
    • No need to exclude studies with 0 events in both groups
Meta-analysis: Heterogeneity

- If study results differ greatly, it may not be appropriate to combine
  - Differences in **patient groups** studied
  - Differences in **interventions** studied
  - Differences in **primary outcome** studies
  - Studies carried out in **distinct settings**
    e.g., different countries

- **Cochran’s Q Test**
  - Low power, esp. when # of studies is small
  - Liberal if number of studies is large

- **Graphical Exploration**
Heterogeneity

• Cochran’s Q Statistic
  – Weighted sum of squared differences between individual study effects and the pooled effect across studies
  – Distributed as a chi-square statistic with k-1 df
  – Low power, especially when the number of studies is small
  – Too much power, if the number of studies is large
Heterogeneity

• $I^2$ Statistic
  – Percentage of variation across studies due to heterogeneity rather than chance
  – Higher percentage (greater than 50%) indicates heterogeneity
  – Operating characteristics influenced by number of studies and study sizes.
Handling Heterogeneity

• If heterogeneity is suspected:
  – Stratify the studies into homogeneous subgroups and then fit separate fixed effects
    • Post-hoc nature
    • Multiplicities
  – Construct a random effects estimate across all studies.
    • Concern: if heterogeneity exists among studies, a summary measure across those studies should not be provided?
  – Fit a meta-regression model that explains the heterogeneity in terms of study-level covariates.
Meta-Regression with Summary Data

- Can fit with standard weighted linear regression model

- With individual patient data, can fit by two-step process
Meta-regression: Issues

• Ecological fallacy
  – Group averages don't represent individuals well
  – Averages have little between-study variation

• Post-hoc specification of prognostic factors may lead to bias/spurious results

• Number of studies usually small

• Data may be unavailable (not conceived or not reported)

• Cannot handle factors that vary by patient within study
Meta-analysis: Data Presentation

Graphical display of results from individual studies on a common scale: Blobbograms

–Also referred to as forest plots
Meta-analysis: Sensitivity Analysis

• Examine Robustness of Findings
  – Examine effects in certain studies, certain groups of patients, or certain interventions
  – Robustness to departures from model assumptions
  – Sensitivity to departure from study selection criteria
Meta-analysis: Study Limitations

- Discuss Limitations of Analysis
  - Bias: Publication
  - Use of aggregate data: Confounding factors
  - Model assumptions
  - Heterogeneity
Meta-analysis: Communication of Findings

• Results summarized in light of available body of knowledge?
  – Association is not the same as causation
Assessment of Causality

• **Strength of Association**
  – Validity of measure: OR, RR, etc.
  – Clinical / statistical significance

• **Assessment of Consistency**
  – Review similar reports for drug
  – Review similar reports for class

• **Biological Plausibility**
  – Known potential biologic basis to suggest a causal link
  – Reference pharmacogenomic evidence
Assessment of Causality, *cont’d*

- **Epidemiology**
  - Consider the expected rate in the general population for comparable demographic groups

- **False Positive Findings**
  - Address issue of multiplicity
  - FDR

- **Imbalance (wrt Relevant Confounding Factors)**
  - Identify relevant risk factors

- **Pharmacovigilance Databases**
  - Recent advances in analysis of such data
  - Measures of disproportionality
QUOROM Statement

Major Limitations of Study
by:

Part II
Outline

• Cumulative Meta-analysis
  – Study Ordering
  – Limitations
• Indirect Comparison
  – Adjusted Indirect Comparison
  – Mixed Treatment Comparison
  – Network Meta-analysis
  – Issues with Indirect and MT Comparisons
• Observational Data Analysis
  – Sources of Bias
  – Methodological Issues
• Comparative Effectiveness Research
  – Background
  – Relation to Traditional Meta-analysis
  – Current Status
• Summary and Conclusion
Cumulative Meta-Analysis

• Commonly executed with chronologically ordered RCTs
  – Perform a new statistical pooling every time a new RCT becomes available
  – Impact of each study on the pooled estimate may be assessed
  – Reveals (temporal) trend towards superiority of the treatment or the control, or towards indifference
  – Performed *retrospectively*, the year when a treatment could have been found to be effective could be identified
  – Performed *prospectively*, an effective treatment may be identified at the earliest possible moment
Cumulative Meta-analysis: Study Ordering

• By control group event rate
  – Low control event rates require large n to establish efficacy
  – Arrangement by descending or ascending order helps assess study heterogeneity

• Study size
  – Small studies imply high variability, and publication bias
  – Pooling only small studies lead to overestimation of treat effect

• Study quality
  – Need a scoring algorithm for RCTs
  – To evaluate impact of poorly designed/conducted trials

• Effect size
  – Impacts of inconclusive trials on results of analysis
Cumulative Meta-Analysis: Issues

• Statistical:
  – Typically, frequentist approaches employed
  – Inflation of Type I error with repeated analyses.

• Evolution of treatment effect over time
  – Changes in patient demographics
  – Changes in healthcare delivery
  – Evolving medical practices
Indirect Comparison

• Well-conducted randomized controlled trials (RCTs) provide the most valid estimates of the relative efficacy of competing healthcare interventions
  – However, many interventions have not been directly compared in RCTs

• Comparison of different healthcare interventions using data from disparate studies
  – Often used because of a lack of, or sufficient, evidence from head-to-head comparative trials.
Indirect Comparison

\[ d_{AB} = d_{PB} - d_{PA} \]
Indirect Comparisons: Basic Assumptions

• Exchangeability/Similarity
  – All trials comparing pairs of tx arms estimate same effect
  – Different sets of trials being used are similar

• Independence between pairwise comparisons:
  – Not true for more than 2 arm trials in Bucher’s method.
Statistical Issues in Indirect Comparisons

• Reliable statistical procedures unavailable to validate assumptions of exchangeability/consistency

• Available procedures do not readily adjust for study heterogeneity
  – Adjusted indirect comparison, misleading term

• Rigorous study of the operating characteristics of commonly used techniques not available

• Low power, as a result of use of two separate variances,
  – Impacts reliability of non-inferiority conclusions
  – Often leads to indeterminate results, due to inflated Type II error
Challenges with Observational Data

• Sources of Bias with Observational Data
  – Overt and hidden biases
  – Data quality

• Controlling for overt biases
  – Regression approaches
  – Propensity score analysis
    • Post-hoc nature of analysis

• Controlling for hidden biases
  – Prior Events Rate Ratio (PERR) adjustment
    • Stringent assumptions

Challenges with Observational Data (cont.)

- **Basic notion of PERR**
  - Hazard ratio of the exposed to unexposed for a specific outcome before the start of the study reflects the combined effect of all confounders
  - **Corollary:** Division of the Incident Rate Ratio (IRR) during a study, by the IRR preceding the start of the study should correct for all confounders (both measured and unmeasured).

\[
PERR_{adj} \text{ IRR} = \frac{(R_E s / R_U s)}{(R_E p / R_U p)}
\]

Where \( R = \) rate, \( E = \) exposed group, \( U = \) Unexposed group, \( p = \) prior events, and \( s = \) study events
(Cf. Tannen et al. BMJ 2009;338:b81 doi:10.1136/bmj.b81)

- **Assumptions**
  - Constant covariate effect over time
  - No treatment–by-confounder interaction
  - Independence of pre- and post-treatment event rates
  - Events likely to occur both pre- and post therapy periods
Challenges with Observational Data (cont.)

• Research needed for
  – Controlling for hidden bias
  – Assessment of consistency between RCTs and Observational studies
  – Methods for combining RCTs and non-RCTs
  – Conditions for and appropriateness of combining RCTs and non-RCTs
Comparative Effectiveness Research (CER)
Background

• The American Recovery and Reinvestment Act (ARRA) appropriated $1.1B for comparative effectiveness research (CER):
  – $400M allocated to the NIH
  – $300M to the Agency for Healthcare Research and Quality (AHRQ), and
  – $400M to Office of the Secretary of HHS

• Objective: To conduct research comparing "clinical outcomes, effectiveness, and appropriateness of items, services, and procedures that are used to prevent, diagnose, or treat diseases, disorders, and other health conditions."
Comparative Effectiveness Research (cont.)

• Congressional Budget Office:
  – A rigorous evaluation of the impact of different options that are available for treating a given medical condition for a particular set of patients.

• Institute of Medicine (IOM):
  – Generation and synthesis that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care
    • The purpose is to assist consumers, clinicians, purchasers, and policy makers to make the informed decisions that will improve health care at both the individual and population levels.
Comparative Effectiveness Research

• A type of systematic review
  – Synthesizes available scientific evidence on a special topic

• Expands the scope of a typical systematic review
  – Goes beyond the effectiveness of a single intervention
  – Compares the relative benefits and harms among a range of available treatments or interventions for a given condition

• Parallels decisions facing clinicians, patients, and policy makers who must choose among a variety of alternatives in making diagnostic, treatment, and health-care delivery decisions
Limitations of Traditional Meta-Analysis also Apply to CER

• Quality
  – Publication bias
  – “Garbage in garbage out” phenomenon

• Heterogeneity
  – “Apples and oranges” phenomenon

• Methodological issues
  – Lack of uniform approaches
  – Handling heterogeneity
  – Assessing bias
  – Covariate issues
Current Status of CER

• NIH and AHRQ funding grants with ARRA CER funds
  
  http://www.effectivehealthcare.ahrq.gov/index.cfm/comparative-effectiveness-research-grant-awards/

• Initial focus on development of methods
  – AHRQ/NIH sponsored panels/workshops
  – Academic institutions
    • 2010 University of Pennsylvania Annual Conference on Statistical Issues in Clinical Trials:
A New Frontier for Statistical Research

• Opportunity to take another look at old problems
  – Issues with traditional meta-analysis

• Opportunity to apply existing results to new area
  – Role for decision theory and Bayesian approaches

• New problems that require fresh approaches.
  – Defining therapeutic indices for CER
Concluding Remarks

• Meta-analysis and CER pose considerable practical and methodological challenges

• Critical role for statisticians to engage in methodological work

• Without sound methodological foundation, potential for adverse impacts on public health
Selected References on Indirect Comparisons