APPLICATIONS OF ESTIMATING TREATMENT EFFECTS IN META-ANALYSES WITH MISSING DATA

by

I. Elaine Allen
Ingram Olkin

July 2000

Department of Statistics
STANFORD UNIVERSITY
Stanford, California
APPLICATIONS OF ESTIMATING TREATMENT EFFECTS
IN META-ANALYSES WITH MISSING DATA

by

I. ELAINE ALLEN
Babson College

INGRAM OLKIN
Stanford University

July 2000

Department of Statistics
Sequoia Hall
STANFORD UNIVERSITY
Stanford, California 94305-4065

http://www-stat.stanford.edu
Applications of Estimating Treatment Effects in Meta-Analyses with Missing Data

I. Elaine Allen, PhD

Ingram Olkin, PhD

Corresponding Author: Dr. I. Elaine Allen, Associate Professor
Math/Science Division
Babson Hall
Babson College
Babson Park, MA 02457
Phone: 781-239-6413
Fax: 781-239-6416
e-mail: allenie@babson.edu

1Babson College, Babson Park, MA 02457
2Stanford University, Stanford, CA 94305

July 6, 2000
Abstract

When screening publications for inclusion in meta-analyses it is not always possible to restrict the analysis to randomized controlled trials (RCTs). Using this restriction often reduces the number of eligible trials to a small percentage of all publications on a particular treatment or in a specific disease and, in some cases, may preclude meta-analysis entirely. This is an ongoing and growing problem for anyone synthesizing research and performing meta-analyses. It is due in part to the exploding number of clinical journals reporting results of clinical trials and also to the variable quality of the studies being reported. If many studies are excluded, the resulting meta-analysis including only RCTs may produce biased effect estimates not representative of the total published research. This paper examines three instances of missing data and illustrates methodologies to address these problems with examples. The three types of missing information examined in this paper are: (1) including uncontrolled comparative studies or single-arm trials with RCTs in comparative meta-analyses; (2) testing treatment differences when randomized controlled trials exist only between treatments and placebo; and (3) imputing treatment differences in multicenter trials with inconsistent comparative groups over sites when individual patient data is available.
Introduction

When evaluating studies and extracting data for meta-analyses of clinical trials, studies are often excluded when primary treatment comparisons do not exist or data are from non-randomized controlled trials (non-RCTs). These missing comparisons can occur in a variety of situations. It is not unusual to start with an initial pool of more than 5,000 abstracts, find fewer than 100 potential studies and finally to analyze less than 30 randomized controlled trials (RCTs) in the final meta-analysis. Data documenting this reduction from abstracts to final eligible studies from a large number of meta-analyses, and an algorithm for estimating the final number of eligible studies, are discussed by the authors elsewhere (Olkin and Allen, 1999). Three instances of missing comparative data in meta-analyses are discussed below.

In some therapeutic areas, few RCTs are performed. In cancer studies, for example, one may have fewer than 10 RCTs for analyses because few Phase III trials have been performed for a given treatment comparison. However, for the same treatments, many small or medium sized non-RCTs and uncontrolled studies may exist. Data can be imputed for the missing treatment arms by using the information from studies (both RCTs and non-RCTs) where these treatments can be compared. Weights can also be incorporated into the imputation based on the degree of heterogeneity in the studies. A random effects model using comparisons based on historic controls (Begg and Pilote, 1991) is used to impute missing comparator arms under certain assumptions for meta-analyses using RCTs for imputation. This method was extended to include RCTs and non-RCTs for imputation and applied in a meta-analysis of interleukin-2 for the treatment of metastatic melanoma (Allen, et al, 1998).
Next, randomized controlled trials may be available between treatment and placebo control but no head-to-head comparative drug trials exist between treatments. From these well-controlled RCTs of each treatment vs. identical or similar controls, comparisons between the treatments are estimated using ANOVA under certain assumptions. A meta-analysis of the safety and efficacy of GIIb/IIIa inhibitors for patients undergoing percutaneous coronary interventions (Ross, 2000) is used to illustrate this technique.

Finally, another type of missing treatment comparator data can occur when synthesizing results from a large multicenter trial when sites include different treatment and control groups. Synthesizing the information from these studies, and imputing comparator arms, can be done using hierarchical mixed models when individual patient data is available. An example of synthesizing comparisons from a large multicenter trial on assessing the efficacy of vocational training techniques is given.

**Estimated Rates Using Historic Controls**

A random effects model with historic controls was used to estimate differences between treatment groups using a random effects model for a meta-analysis of cancer studies in melanoma. In this analysis, at least one randomized controlled trial included both treatments but many single-arm uncontrolled trials included only one of the treatments of interest. These comparative studies (which included both RCTs and non-RCTs) were analyzed in combination with many other non-comparative single-arm studies. The difference between treatments in the comparative studies was used to estimate treatment differences in non-comparative studies. These calculations of effect estimates follow the methods outlined by Begg & Pilote (1991).
From 4,580 citations and an initial pool of 210 studies, a total of 154 publications including 7,341 patients were deemed eligible for this meta-analysis. These publications were found by applying prospectively defined eligibility criteria when screening for clinical trials of Interleukin-2, cisplatin, DTIC, or α-Interferon in the treatment of metastatic melanoma. These publications included only 2 RCTs, 65 non-RCTs, and 87 uncontrolled studies. Following extraction of the data and consensus by clinical extractors, the clinical studies were summarized by treatment arm in a design matrix showing the missing comparator arms. A generalization of this design matrix is given below where it is evident that some comparisons between treatments can be calculated directly (TxA vs. TxC in Study 1) but some will be imputed where data in the trial exist on only one of the two comparative treatments (i.e.; TxA only in Study 5):

Unlike the random effects model of Dersimonian and Laird (1986), in this model Begg and Pilote assume that the treatment effect is constant from study to study and the between-study heterogeneity is reflected in the random effects. This accounts for the differences in the case selection and, thus, the differences in baseline means from study to study. The data input to the model are the following:

(1) n comparative trials: \((x, y)\), where \(x\) is the observed effect of treatment 1, \(y\) is the observed effect of treatment 2;

(2) \(k\) uncontrolled studies of treatment 1 with observed effects \(u\);

(3) \(m\) uncontrolled studies of treatment 2 with observed effects \(v\).

The standard errors of all the effects \((x, y, u, v)\) must be known (i.e.; extractable from the publication) or derived from the studies (i.e.; given the appropriate variances, standard
deviations or results from hypothesis tests and given the treatment arm size). The input data to the random effects model are the treatment outcomes \((x, y, u, \text{ and } v)\) and their standard errors \((\text{se}_x, \text{se}_y, \text{se}_u, \text{se}_v)\) where \(u\) are in the column \(x\) without the comparative treatment arm \(y\), and \(v\) are in the column \(y\) without the treatment arm \(x\). In this example, the outcomes are response rates to treatment and the risk difference is the effect size. Weights are given to the single arm studies are dependent on the degree of heterogeneity between the studies. If the case mix in all the studies is similar (in this example, the treatment responses \(u\) and \(v\) are similar to \(x\) and \(y\), respectively), the studies with imputed effect sizes (risk differences) are given weights similar to the controlled studies. If the between study variance is large, indicating a high degree of heterogeneity between the uncontrolled and controlled studies, the uncontrolled studies are given very small weight in the overall risk difference.

The input data used for estimating the overall risk difference are given in Table 2. The resulting random effects estimates of the risk difference for these studies using all the studies and for only the comparative studies are given in Table 3. For these estimates, which are examining the response rate in the treatment of metastatic melanoma comparing IL-2 given before or after chemotherapy, 86 studies were used. Fifty studies contained data for IL-2 given prior to chemotherapy and 47 studies contained data for IL-2 given following chemotherapy. If all non-comparative studies were excluded, only 11 studies would be eligible to estimate the risk difference between the response rates for IL-2 given pre as compared to post-therapy. Comparing the results for the analysis using all the data to the analysis including only comparative studies, a similar overall risk difference is shown but with a much larger confidence interval because of the loss of information contained in the non-comparative studies.
Testing Treatment Differences Between Placebo Controlled Studies

For new chemical entities, comparative testing against therapeutic competitors may not be possible because none may be approved for sale or the number of patients required for an RCT may be prohibitively large. However, large RCTs comparing each new compound to placebo may exist. For this missing data problem, comparisons of the random effects estimates from each trial may be compared in an ANOVA model after comparability of control groups has been tested. The resulting model can also include covariates to control for differences in patient mix and baseline values for individual trials.

To illustrate this problem, a meta-analysis was proposed to compare the efficacy and safety of parenteral GPIIb/IIIa receptor inhibitors in patients undergoing PCI (percutaneous coronary interventions). A search of the published literature (all languages, from 1/90 to 4/99) based on a prospectively defined protocol attempted to find eligible comparative RCTs of abciximab, eptifibatide, or tirofiban. Random effects meta-analyses were planned comparing abciximab to the two other treatments, eptifibatide and tirofiban.

Fifteen studies were found including 19,074 patients. All studies were treatment vs. placebo with no studies containing any pair of treatments. In order to estimate and test the differences between treatment groups, several analyses were carried out:

(1) Data were grouped by treatment, odds ratios (ORs) comparing individual treatments to placebo were estimated and random effects meta-analyses were performed for each treatment group separately.

(2) Placebos were examined for all studies for all three treatments to ensure their comparability between treatments.
An ANOVA model with multiple comparisons of the odds ratios from (1) comparing treatment to placebo were analyzed over treatment groups including baseline covariates to adjust for case mix differences and including variability measures to adjust for heterogeneity of studies within treatment groups.

This method allows for the comparison of treatments when no head-to-head studies exist by first examining individual treatment meta-analyses, testing the comparability of placebo groups between treatments to minimize bias in case mix, and, finally, comparing the treatments using ANOVA with covariates and multiple comparisons. The initial meta-analyses by treatment group can be summarized and examined visually, as given in Figure 1 for the twelve studies with the 30-day outcome of Death or Myocardial Infarction (MI). Table 4 gives the individual numerical estimates from the random effects model and the results of the ANOVA comparing treatments. An ANOVA was also used to compare the placebo groups within and between treatments and no statistically significant differences were seen within or between studies. Additionally, Cochran’s Q statistic examined between study (within treatment) heterogeneity and no statistically significant differences were found. For comparisons between treatments when no direct comparative studies exist, this is a useful method for comparing random effect estimates from the meta-analysis for each treatment in an ANOVA model.

Estimating Missing Effect Sizes with Complete Patient Data

Within a large multicenter trial, each site usually includes identical treatment and control groups with all other sites. However, if comparisons between new and standard practices are different between sites, it may be that neither treatments nor controls were
applied consistently across sites. A meta-analytic synthesis of the data comparing treatments can be completed when individual patient data are available. An ongoing multicenter trial is used to illustrate this method.

The study, which is ongoing, examines the effects of different vocational training models on ability to obtain and retain employment. Subjects were randomized to treatment or control groups but sites were able to define their own groups based on standard practice in their region of the country. With these missing study-level comparisons, this gives an incomplete block design. This design matrix is given below showing eight sites, three treatments, and one control. This differs from the previous example in that not all sites used a control and because individual patient data were available.

A mixed hierarchical model was used for this estimation. This model has the advantage that all patients, regions and interventions are used in fitting the model and all comparisons are estimable using the information from sites with a given intervention to estimate the effect in the missing cells of the table. Predictions of the intervention effect for a region are estimated, adjusted for important subject, vocational training, and site covariates if necessary. A variance for the estimated effect is calculated in order to give an interval (95% confidence interval) estimate.

The model is called a mixed model because it includes both fixed and random effects, and both micro (subject) and macro (region, intervention) level data. Here an effect is called fixed if the levels in the study represent all possible levels of the factor, or at least all levels about which inference is to be made. Factor effects are called random if the levels of the factor that are used in the study represent only a random sample of a larger set of potential
levels. Depending on a trial’s design and objectives, study sites or regions could be considered fixed or random effects.

This meta-analysis model is formulated as a special case of a multilevel (hierarchical data) model in which the highest level is that of the region and the lowest level is that of an observation on an individual patient (Goldstein, 1995). The regions are combined within a single model where outcomes occur at different levels of the data hierarchy and efficient estimates can be obtained. The advantage of this model for meta-analysis is that it utilizes all the data available and region and intervention level data can be estimated in cases where patient data is available for some regions but not for all the study regions. Although in the example here, only sites with individual patient data are combined, this procedure is an efficient way of combining published studies, which provide only summary data, with studies for which the individual patient data are available.

Using the notation of Bryk & Raudenbush (1992), this analysis assumed a random effects model for region. The model can be written as:

\[ y_{ijk} = \mu_i + \tau_i + L_j + R(L)_{jk} + (\tau L)_{ij} + e_{ijk}, \]

where \( y_{ijk} \) is the outcome for the \( ijk^{th} \) patient  
\( \mu_i \) is the \( i^{th} \) intervention mean  
\( \tau_i \) is the intervention effect  
\( L_j \) is the region effect  
\( R(L)_{jk} \) is the effect of subject within a region  
\( (\tau L)_{ij} \) is the intervention-by-region effect  
\( e_{ijk} \) is the error and the \( k \)-subscript indicates patient within region.

Intervention effects are then tested using the F-Test,
\[ F = \frac{MS(\text{intervention})}{MS(\text{region} \times \text{intervention})} \]

Estimates of individual intervention effects can be estimated and contrasts can be identified which compare pairs of effect sizes can also be tested. Prediction intervals for missing intervention effects can also be calculated. For this example, the hierarchical models were estimated using HLM (Bryk & Raudenbush, 1996). It is also possible to estimate these models using SAS Proc Mixed (Normand, 1996, Westfall, et al, 1999) and MIXOR (Hedeker & Gibbons, 1996). Following estimation of effect sizes, meta-analyses can be performed on these imputed effect sizes and synthesized over regions, as is illustrated by the Figure 2 where control values were imputed for Regions D and F.

**Discussion**

Three different instances of commonly occurring missing data have been presented with examples illustrating statistical methods to impute the missing information. All the methods extend standard statistical tools and are easily applied. The first method, using non-RCTs and single-arm studies in meta-analyses, allows from much broader application of meta-analysis to published data. The second technique, comparing treatments when RCTs exist against common control arms but without direct comparative trials, allows for meta-analyses comparing compounds to occur earlier in the drug approval process. The final application illustrates imputing comparative treatments when individual patient data exist. All three of these methods extend the pool of studies eligible for a meta-analysis to a wider pool of published and unpublished studies that would have been eliminated previously.
References


Table 1: Example of Design Matrix for Controlled and Uncontrolled Studies

<table>
<thead>
<tr>
<th>Study</th>
<th>TxA</th>
<th>TxB</th>
<th>TxC</th>
<th>TxD</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>T</td>
<td>U</td>
<td>M</td>
<td>S</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>D</td>
<td>G</td>
<td>D</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

<p>| | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>576</td>
<td>1</td>
<td>45</td>
<td>0.0667</td>
<td>0.0378</td>
<td>0.0849</td>
<td>0.0327</td>
<td>723.21</td>
<td>0.0378</td>
<td></td>
<td></td>
</tr>
<tr>
<td>576</td>
<td>0</td>
<td>29</td>
<td></td>
<td></td>
<td>0.0697</td>
<td>0.0475</td>
<td>451.65</td>
<td>0.0475</td>
<td></td>
<td></td>
</tr>
<tr>
<td>578</td>
<td>1</td>
<td>29</td>
<td></td>
<td></td>
<td>0.0697</td>
<td>0.0475</td>
<td>451.65</td>
<td>0.0475</td>
<td></td>
<td></td>
</tr>
<tr>
<td>579</td>
<td>0</td>
<td>57</td>
<td>0.0175</td>
<td>0.0179</td>
<td></td>
<td></td>
<td>3307.02</td>
<td>0.0179</td>
<td></td>
<td></td>
</tr>
<tr>
<td>582</td>
<td>0</td>
<td>16</td>
<td></td>
<td></td>
<td>0.0330</td>
<td>0.0425</td>
<td>544.50</td>
<td>0.0425</td>
<td></td>
<td></td>
</tr>
<tr>
<td>587</td>
<td>0</td>
<td>14</td>
<td>0.0348</td>
<td>0.0348</td>
<td></td>
<td></td>
<td>420.50</td>
<td>0.0487</td>
<td></td>
<td></td>
</tr>
<tr>
<td>621</td>
<td>0</td>
<td>24</td>
<td>0.0201</td>
<td>0.0201</td>
<td></td>
<td></td>
<td>1200.50</td>
<td>0.0286</td>
<td></td>
<td></td>
</tr>
<tr>
<td>622</td>
<td>0</td>
<td>37</td>
<td>0.0273</td>
<td>0.0273</td>
<td></td>
<td></td>
<td>1407.03</td>
<td>0.0266</td>
<td></td>
<td></td>
</tr>
<tr>
<td>623</td>
<td>1</td>
<td>18</td>
<td></td>
<td></td>
<td>0.1111</td>
<td>0.0744</td>
<td>182.25</td>
<td>0.0747</td>
<td></td>
<td></td>
</tr>
<tr>
<td>626</td>
<td>0</td>
<td>12</td>
<td>0.1667</td>
<td>0.1667</td>
<td>0.0843</td>
<td>0.0371</td>
<td>86.40</td>
<td>0.1078</td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>
Table 3: Results of Random Effects Models

<table>
<thead>
<tr>
<th>Comparative</th>
<th>PRE-only Studies</th>
<th>POST-only Studies</th>
<th>Total Studies</th>
<th>Risk Difference</th>
<th>SE</th>
<th>CI L</th>
<th>CI U</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>11</td>
<td>50</td>
<td>47</td>
<td>86</td>
<td>0.0791</td>
<td>0.0106</td>
<td>0.0365</td>
<td>0.1214</td>
</tr>
</tbody>
</table>

Begg & Filote Random Effects Model
Overall Risk Difference and 95% Confidence Interval

<table>
<thead>
<tr>
<th>Total Studies</th>
<th>Risk Difference</th>
<th>SE</th>
<th>CI L</th>
<th>CI U</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>0.0766</td>
<td>0.0324</td>
<td>0.0005</td>
<td>0.1527</td>
<td>0.048</td>
</tr>
</tbody>
</table>

Dersimonian & Laird Random Effects Model
Overall Risk Difference and 95% Confidence Interval
Table 4: Individual Treatment ORs and Overall ANOVA Comparisons

<table>
<thead>
<tr>
<th>Study Drug</th>
<th># of Patients</th>
<th># Studies</th>
<th>Odds Ratio</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abciximab</td>
<td>9,606</td>
<td>5</td>
<td>0.45</td>
<td>(0.37, 0.55)</td>
</tr>
<tr>
<td>Eptifibatide</td>
<td>5,834</td>
<td>4</td>
<td>0.82</td>
<td>(0.69, 0.98)</td>
</tr>
<tr>
<td>Tirofiban</td>
<td>3,634</td>
<td>3</td>
<td>0.73</td>
<td>(0.57, 0.92)</td>
</tr>
</tbody>
</table>

One-way ANOVA comparing treatments:
- Abciximab vs. Eptifibatide p < 0.001
- Abciximab vs. Tirofiban p < 0.01
Table 5: Incomplete Block Design with Complete Individual Patient data

<table>
<thead>
<tr>
<th>Region</th>
<th>Standard</th>
<th>Type1</th>
<th>Type2</th>
<th>Type3</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>X</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>E</td>
<td>X</td>
<td></td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td></td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>
Figure 1
OR (95% CI) 30 Day Death/MI

Tirofiban
# Tx Arms: 3
# Patients: 3534

Eptifibatide
# Tx Arms: 4
# Patients: 5634

Aldefibatide
# Tx Arms: 5
# Patients: 9606

0.1  1.0  10.0
Favors Treatment  Favors Placebo
Figure 2

Mean Difference in Work Hours: Intervention vs Control