THE ARCSINE DIFFERENCE EFFECT MEASURE IN METÁNALYSIS WITH APPLICATION TO ADVERSE EVENTS FROM LONG-ACTING BETA-AGONISTS IN ASTHMA PATIENTS

A Thesis in
Statistics
by
Krista Foster

© 2011 Krista Foster

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Master of Science

August 2011
The thesis of Krista Foster was reviewed and approved* by the following:

Vernon Chinchilli
Chair of the Department of Public Health Sciences
Professor of Statistics
Thesis Adviser

Bruce Lindsay
Head of the Department of Statistics

Runze Li
Professor of Statistics

*Signatures are on file in the Graduate School.
Abstract

This thesis first addresses the statistical issues that are associated with the use of meta-
analysis in the presence of rare events. Limitations of commonly used effect measures, such
as the log odds ratio and log relative risk, are described. Then, the benefits of using the
arcsine difference as the effect measure in meta-analysis are investigated, and the various
effect measures are compared in a simulation study. Finally, real data on the adverse
events from long-acting beta-agonists in asthma patients is considered. Using this data,
the results of a meta-analysis with the arcsine difference effect measure are compared to
previously published results using the log odds ratio.
# Table of Contents

List of Figures vi

List of Tables viii

1 Introduction to Meta-Analysis 1
   1.1 Preliminary Steps ................................. 1
   1.2 Fixed and Random Effects .......................... 3

2 The Problem: Zero Events 6

3 Proposed Methods for Zero Events 9

4 Simulation Study 15
   4.1 Methods ............................................. 16
   4.2 Results ............................................. 17

5 Example: Adverse events from long-acting beta-agonists in asthma patients 25
   5.1 Background ........................................... 25
   5.2 Analysis ............................................. 26
      5.2.1 Peto Odds Ratio with No Zero Event Trials ............ 27
      5.2.2 Arcsine Difference with No Zero Event Trials ............ 29
      5.2.3 Arcsine Difference with Summary Data for Zero Event Trials . . . 31
List of Figures

3.1 Graphical representation of the arcsine difference (as presented by Rucker et al, 2009) .......................................................... 10
4.1 Monte Carlo estimates of bias when the true odds ratio is 2 ............... 18
4.2 Monte Carlo estimates of bias when the true odds ratio is 1.33 ............ 19
4.3 95% confidence intervals: Coverage of true treatment effect over 1000 meta-analyses when true odds ratio is 2 ........................................ 19
4.4 95% confidence intervals: Coverage of true treatment effect over 1000 meta-analyses when true odds ratio is 1.33 ........................... 20
4.5 95% confidence intervals: Coverage of zero over 1000 meta-analyses when true odds ratio is 1.33 ............................... 21
4.6 Monte Carlo estimates of bias when the true odds ratio is 2 under the random effects model ......................................................... 22
4.7 Monte Carlo estimates of bias when the true odds ratio is 1.33 under the random effects model ......................................................... 22
4.8 95% confidence intervals: Coverage of true treatment effect over 1000 meta-analyses when true odds ratio is 2 under the random effects model) .... 23
4.9 95% confidence intervals: Coverage of true treatment effect over 1000 meta-analyses when true odds ratio is 1.33 under the random effects model .... 24
5.1 Forest plot for the fixed effects meta-analysis model with the arcsine difference effect measure (zero event trials excluded) ......................... 31
5.2 Forest plot for the fixed effects meta-analysis model with the arcsine difference effect measure (zero event trials included) ......................... 33
5.3 Forest plot for the random effects meta-analysis model with the arcsine difference effect measure (zero event trials included) . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 34
List of Tables

2.1 Summary data for a single trial with a binary outcome ..................... 6
2.2 Standard effect measures for binary outcomes ............................ 7
2.3 Behavior of sample estimates and variance estimates with zero events ... 7
4.2 Ten different meta-analyses performed ...................................... 17
5.1 Summary of studies testing long-acting β-agonists with concomitant inhaled corticosteroids versus inhaled corticosteroids alone ..................... 27
5.2 Summary data for trial i .......................................................... 27
5.3 Summary of meta-analysis using Peto odds ratio effect measure .......... 29
5.4 Summary of meta-analysis using the arcsine difference effect measure (zero event trials excluded) .................................................. 30
5.5 Summary of meta-analysis using the arcsine difference measure (zero event trials included) ...................................................... 32
Chapter 1

Introduction to Meta-Analysis

Meta-analysis, in general, refers to the analysis of data from multiple already-existing studies. These studies address the same or similar research questions, and the goal is to review and summarize the results over all of the studies. Meta-analysis is simply a method of analyzing the data; there is no need to conduct additional experiments or generate new data. Oftentimes, the data used to perform a meta-analysis consists of only the summary statistics for each study, not the raw data. When we consider several studies, it is likely that the studies or experiments differ in design and quality. Sample sizes may vary widely from one study to another, and furthermore, the results may differ by study. By analyzing the results of all of the different studies at once, more reliable conclusions can be drawn.

1.1 Preliminary Steps

When performing a meta-analysis, there are some basic steps that must be followed. First and foremost, a focused clinical research question must be defined. This is crucial because all of the studies chosen to be included in the meta-analysis need to address this same research question. If the question is too vague, an overwhelming number of studies may exist that ask this same question. Additionally, each study may ask slightly different questions that all fall into some broad category. A question that is too narrowly focused may also cause issues when it comes to meta-analysis because there are too few studies available. The
definition of the clinical research question will determine the number of studies that are ultimately included in the meta-analysis.

The second step in performing a meta-analysis, which is likely the most difficult and time-consuming, is the literature search. It is vital to conduct a thorough search of both published and unpublished reports to find relevant studies. Sometimes reports remain unpublished if the treatment or intervention is not found to be effective or as effective as some other treatment or intervention. In this case, publication bias exists. One way to protect against performing a meta-analysis in the presence of publication bias is to compare sample sizes with p-values (or estimated effects) across all of the relevant studies. Since it is more difficult to find statistical significance with small sample sizes, some of the studies with small sample sizes will likely yield results that are negative, or not statistically significant. If all or nearly all of the p-values are small, it may raise some questions. In order to check for publication bias, a plot of sample size versus p-value (or the magnitude of the effect) for each study is used. If the resulting plot is a vertical band of points, all with small p-values, then publication bias is suspected. Conversely, if some of the p-values for the studies with small sample sizes are relatively large, then researchers can proceed to the third step.

After conducting a thorough search of both published and unpublished results, some inclusion-exclusion criteria must be applied to the various relevant studies. These criteria may include the design of the experiment and the population among other things. For instance, studies may be included only if they are described as being randomized or only if they are described as being double blind.

After applying the inclusion-exclusion criteria to the results of the literature search, data must be obtained from the included studies. In most cases, the reports contain the relevant descriptive statistics, including sample sizes, means or proportions, and standard errors at the very least. While it would be more ideal to have the raw data for each included study, this is extremely rare, and meta-analyses are generally performed using the descriptive statistics.
1.2 Fixed and Random Effects

Since the included studies may have various differences in design, quality, sample size, and population, it is common to account for this in the meta-analysis model. Two types of models are used: fixed effects models and random effects models. Fixed effects models assume that there is no inter-study variability; that is, all of the involved studies would yield the same results if sample sizes were sufficiently large. The fixed-effects model is computationally easier, but its underlying assumptions may be unrealistic in many cases. These models only account for variation within each individual study and assume homogeneity across the studies. On the other hand, random effects models are more conservative and account for both inter- and intra-study variability. They assume that while there is variation within each individual study, there may also exist some variation, or heterogeneity, from study to study. If there exists any study heterogeneity at all, the estimate of the random-effects coefficients will be more conservative, the standard errors will be larger, and thus the confidence intervals will be wider. In this way, it is possible to find statistical significance with a fixed-effects model but not with a random-effects model. If desired, we can also test for heterogeneity, but we don’t usually use the results to determine which model to use, and it is actually common to perform the analysis using both models regardless of the amount of heterogeneity. This is because the test for heterogeneity is extremely powerful and very sensitive when we use a large number of studies but very weak and insensitive when we have a small number of studies (Borenstein et al, 2009; DerSimonian & Laird, 1986).

A fixed-effects model simply involves calculating a weighted average of the treatment effect across all of the included studies. The treatment effect can be measured in a variety of different ways depending on whether the response variable is continuous or binary. For the purposes of this project, my focus will be on meta-analyses in which we are dealing with a binary response, so we assume a binary response in the following example.

Let’s assume that we have $K$ studies and that our treatment effect is the natural logarithm odds ratio. For $k = 1, \ldots, K$, we have that the estimated treatment effect for the $k$th study is given by $\hat{\theta}_k$, and that the standard error of this estimate is $s_k$. The weight that we assign to the
estimate from the $k$th study is given by $w_k = \frac{1}{s_k^2}$. Thus, the overall treatment effect estimate is given by
\[
\hat{\theta} = \frac{\sum_{k=1}^{K} w_k \hat{\theta}_k}{\sum_{k=1}^{K} w_k},
\]
and its estimated standard error is given by
\[
s = \sqrt{\frac{1}{\sum_{k=1}^{K} w_k}}.
\]

We can then test the null hypothesis of no treatment effect and construct confidence intervals for the treatment effect. We can test the null hypothesis of study homogeneity as well. The test statistic is
\[
Q = \sum_{k=1}^{K} w_k (\hat{\theta}_k - \hat{\theta})^2,
\]
where $Q$ has an asymptotic $\chi^2(K-1)$ distribution.

The aforementioned model corresponds to the linear model,
\[
\hat{\theta}_k = \theta + \epsilon_k \text{ with } \epsilon_1, \ldots, \epsilon_K \sim N(0, \sigma_k^2) \text{ independent, } k = 1, \ldots, K
\]
where $\theta$ is the true population natural logarithm odds ratio, $\epsilon_k$ is the error associated with the $k$th study, and $\sigma_k^2$ is the intra-study variability associated with the $k$th study (estimated by $s_k^2$) (Borenstein et al, 2009). Note that there are other methods of performing meta-analyses specifically for odds ratio, namely the Mantel-Haenszel method (1959) and Peto’s method (Yusuf et al, 1985).

The random-effects model then corresponds to the linear model,
\[
\hat{\theta}_k = \theta + \tau_k + \epsilon_k \text{ with } \epsilon_1, \ldots, \epsilon_K \sim N(0, \sigma_k^2) \text{ independent, } \tau_1, \ldots, \tau_K \text{ iid } N(0, \omega^2), \text{ } k = 1, \ldots, K
\]
where $\tau_k$ is the random effect of the $k$th study and $\omega^2$ is the inter-study variability. It follows that the random-effects model has the same weighted average form as the fixed-
effects model, but the weights will be different. Now, we have \( w_k^* = \frac{1}{s^2_k + \omega^2} \) where

\[
\hat{\omega}^2 = \max(0, Q - K + 1) \frac{\sum_{k=1}^{K} w_k}{\left(\sum_{k=1}^{K} w_k\right)^2 - \sum_{k=1}^{K} w_k^2}.
\]

Thus, for the random-effects model, the estimate of the treatment effect is given by

\[
\hat{\theta} = \frac{\sum_{k=1}^{K} w_k^* \hat{\theta}_k}{\sum_{k=1}^{K} w_k^*},
\]

and its estimated standard error is given by

\[
s = \sqrt{\frac{1}{\sum_{k=1}^{K} w_k^*}}.
\]

Again, we can use these estimates to test the null hypothesis of no treatment effect and to construct confidence intervals (Borenstein et al, 2009).

Since meta-analyses are widely used, routines to perform meta-analyses exist in both R and SAS as well as in other software packages. Therefore, I will implement the already-existing routines to focus on the problem at hand.
Chapter 2

The Problem: Zero Events

When studies involve binary outcomes, a variety of effect measures may be used, but all of these measures are simply transformations of the risk difference. These include risk difference, log risk ratio, and log odds ratio, with the log odds ratio being the most commonly used. These measures are not only easy to calculate mathematically, but they also have logical interpretations that can be explained to the general population.

Consider Table 2.1 which summarizes the results of a single study with a categorical response.

<table>
<thead>
<tr>
<th>Event</th>
<th>Yes</th>
<th>No</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment</td>
<td>a</td>
<td>b</td>
<td>n_T</td>
</tr>
<tr>
<td>Control</td>
<td>c</td>
<td>d</td>
<td>n_C</td>
</tr>
<tr>
<td>Total</td>
<td>a+c</td>
<td>b+d</td>
<td>n</td>
</tr>
</tbody>
</table>

Table 2.1: Summary data for a single trial with a binary outcome

Table 2.2 summarizes the corresponding effect measures with sample estimates and the behavior of those estimates when zeros are present. Notice that when the event in question is rare, there is a high probability of observing zero events and statistical problems arise. There does not seem to be a problem with the sample risk difference, but the log relative risk and log odds ratio will both be infinite if just one zero occurs in the table. That means that if an event is extremely rare or conversely, extremely frequent, it is quite possible to
get infinite values as our sample estimates, making further analysis with hypothesis testing or confidence intervals impossible. Additionally, both of these measures can be undefined.

<table>
<thead>
<tr>
<th>Effect measure</th>
<th>Sample Estimate</th>
<th>Variance Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk Difference</td>
<td>$\frac{a}{a+b} - \frac{c}{c+d}$</td>
<td>$\frac{ab}{(a+b)^3} + \frac{cd}{(c+d)^3}$</td>
</tr>
<tr>
<td>Log Risk Ratio</td>
<td>$\log\left(\frac{a}{a+b}\right) - \log\left(\frac{c}{c+d}\right)$</td>
<td>$\frac{1}{a} - \frac{1}{a+b} + \frac{1}{c} - \frac{1}{c+d}$</td>
</tr>
<tr>
<td>Log Odds Ratio</td>
<td>$\log\left(\frac{a}{b}\right) - \log\left(\frac{c}{d}\right)$</td>
<td>$\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}$</td>
</tr>
</tbody>
</table>

Table 2.2: Standard effect measures for binary outcomes

If we consider the variance estimates for these measures, even more problems arise. Table 2.3 summarizes the behavior of the variance estimates. In the case where $a = c = 0$, we cannot perform hypothesis testing or construct confidence intervals using any of these measures without some type of correction. If this was a single trial, it would not be able to provide us with any meaningful or significant results, and if it was one study in a meta-analysis, it likely would be excluded from the analysis. One accepted practice is to use a continuity correction. For instance, if we added 0.5 to all of the cell counts, none of our cell counts would be zeros, and thus both the effect measure estimates and variance estimates would be defined, nonzero, and finite. The problem with a continuity correction is that it results in bias.

<table>
<thead>
<tr>
<th>Effect measure</th>
<th>Sample Estimate</th>
<th>Variance Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk Difference</td>
<td>0 if $a = c = 0$</td>
<td>0 if $a = c = 0$</td>
</tr>
<tr>
<td>Log Risk Ratio</td>
<td>$\pm\infty$ if one zero</td>
<td>$\infty$ if $a = 0$ or $c = 0$</td>
</tr>
<tr>
<td></td>
<td>undefined if $a = c = 0$</td>
<td></td>
</tr>
<tr>
<td>Log Odds Ratio</td>
<td>$\pm\infty$ if one zero</td>
<td>$\infty$ if at least one zero</td>
</tr>
<tr>
<td></td>
<td>undefined if two zeros in same column</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.3: Behavior of sample estimates and variance estimates with zero events
In the paper, “Why add anything to nothing? The arcsine difference as a measure of treatment effect in meta-analysis with zero cells”, Rucker, et al (2009) suggest one alternative effect measure that will eliminate these problems. With their proposed method, a continuity correction is not necessary in order to include studies where zero events are observed in the analysis.
Chapter 3

Proposed Methods for Zero Events

Rucker et. al. (2009) propose using the arcsine difference as the effect measure in meta-analyses in which the event of interest is rare. They claim that this method is a good alternative when there are a number of zero event studies because the sample arcsine difference and its variance estimate are both defined regardless of zero events.

The sample estimate for the arcsine difference is defined as

$$\text{arcsin} \left( \sqrt{\frac{a}{a+b}} \right) - \text{arcsin} \left( \sqrt{\frac{c}{c+d}} \right).$$

Like the risk difference, this value always is defined and only zero when $a = c = 0$. Figure 3.1 shows the graphical representation of the arcsine difference.

The one-armed asymptotic variance estimate is given by $\frac{1}{4a}$, and the two-armed asymptotic variance estimate is given by $\frac{1}{4(a+b)} + \frac{1}{4(c+d)}$, both of which are always finite. Thus, if we have zero events in a trial, instead of deleting the trial from our meta-analysis or using a continuity correction, it may be beneficial to use the arcsine difference. We will now verify both of these variance estimates.

First, consider only the treatment group. Assume that the number of events in the treatment group, $Y$, is a binomial($n_T, p_T$) random variable, where $n_T$ is the number of patients in the treatment group, and $p_T$ is the probability of an event in the treatment group. We will estimate $p_T$ with $\hat{p}_T = \frac{a}{a+b}$. Using the normal approximation to the
binomial distribution, we then know that the sampling distribution of \( \hat{p}_T \) is approximately normal with mean \( p_T \) and variance \( \frac{p_T(1-p_T)}{n_T} \). Thus, we can write this as

\[
\sqrt{n_T}(\hat{p}_T - p_T) \stackrel{D}{\rightarrow} N(0, p_T(1-p_T)).
\]

The delta method then states that for some transformation \( g \),

\[
\sqrt{n_T}(g(\hat{p}_T) - g(p_T)) \stackrel{D}{\rightarrow} N(0, p_T(1-p_T)[g'(p_T)]^2).
\]

Setting \( g(\hat{p}_T) = \arcsin(\sqrt{\hat{p}_T}) \), we know \( g'(p_T) = \frac{1}{2\sqrt{p_T(1-p_T)}} \). Utilizing the delta method, we can easily see that

\[
\sqrt{n_T}(\arcsin(\sqrt{\hat{p}_T}) - \arcsin(\sqrt{p_T})) \stackrel{D}{\rightarrow} N \left(0, \frac{1}{4} \right).
\]

Thus, the variance of \( \arcsin(\sqrt{\hat{p}_T}) \) is approximately \( \frac{1}{4n_T} \) or \( \frac{1}{4(a+b)} \). While this only corresponds to a one-armed trial, we can extend the same idea to a control group. Think of the number of events in the control group, \( X \), as a binomial(\( n_C, \hat{p}_C \)) random variable.
Then, following these steps, we get

\[ \sqrt{n_C} (\arcsin(\sqrt{\hat{p}_C}) - \arcsin(\sqrt{\hat{p}_C})) \overset{D}{\rightarrow} N \left(0, \frac{1}{4}\right). \]

Now, we have that \( \arcsin(\sqrt{\hat{p}_T}) \) has estimated variance \( \frac{1}{4(a+b)} \), and \( \arcsin(\sqrt{\hat{p}_C}) \) has estimated variance \( \frac{1}{4(c+d)} \).

With a two-armed trial we are interested in the arcsine difference, \( \arcsin \sqrt{\frac{a}{a+b}} - \arcsin \sqrt{\frac{c}{c+d}} \), or equivalently, \( \arcsin(\sqrt{\hat{p}_T}) - \arcsin(\sqrt{\hat{p}_C}) \). Using the basic properties of variance, we know that the variance of this estimate is given by

\[ \text{Var}(\arcsin(\sqrt{\hat{p}_T})) + (-1)^2 \text{Var}(\arcsin(\sqrt{\hat{p}_C})) - 2\text{Cov}(\arcsin(\sqrt{\hat{p}_T}), \arcsin(\sqrt{\hat{p}_C})). \]

Since we are assuming that the number of events in the control group, \( X \), and the number of events in the treatment group, \( Y \), come from independent binomial distributions, this reduces to

\[ \text{Var}(\arcsin(\sqrt{\hat{p}_T})) + \text{Var}(\arcsin(\sqrt{\hat{p}_C})) = \frac{1}{4(a+b)} + \frac{1}{4(c+d)}. \]

These calculations confirm that the variance estimates for the arcsine difference effect measure will be defined in the presence of zero events in a trial in both one-armed and two-armed studies.

While both the point estimate and the variance estimate will always be finite, Rucker et al (2009) discuss ways that we can improve upon the approximation to the true variance when an event is rare. We might replace the arcsine transformation with a version of the transformation that includes a continuity correction. This would yield an almost constant variance. Additionally, we might include additional terms in our variance approximation to improve upon the first order approximation, but this turns out to be worse than simply ignoring the additional terms. Thus, Rucker et al focus on evaluating the variance function analytically.
We know the exact variance of a transformation \( f \) is computed as follows:

\[
\text{Var}[f(X)] = \text{E}[f^2(X)] - [\text{E}[f(X)]]^2.
\]

Then, if \( X \) is a binomial\((n, p)\) random variable, and \( f(x) = \arcsin(\sqrt{x}) \), we find that

\[
\text{Var}[f(p)] = \sum_{k=0}^{n} \binom{n}{k} p^k(1 - p)^{n-k} \left[ f \left( \frac{k}{n} \right) \right]^2 - \left[ \sum_{k=0}^{n} \binom{n}{k} p^k(1 - p)^{n-k} f \left( \frac{k}{n} \right) \right]^2,
\]

where the sums are finite. The maximum value of this variance function for a given sample size generally occurs when \( p \) is very small, and this maximum is generally above the value of the asymptotic variance. Taking the first derivative and setting it equal to zero, we can iteratively solve to find the maximum value of the variance. Rucker et al showed that for a large \( n \), the maximum value of the variance can be approximated as about \( 0.42/n \). While the asymptotic variance is given by \( 1/4n \) or \( 0.25/n \), for a small \( p \), say under \( 0.2 \), a more conservative estimate is \( 0.42/n \). Thus if events are rare in both treatment arms and \( n \) is large, this conservative arcsine method may be used. Conversely, if no events are observed, the asymptotic variance approximation of \( 1/4n \) is adequate. Therefore, the analytical arcsine method is as follows. If a study arm has at least one event, we can use the arcsine transformation with its calculated analytical variance for a given observed event proportion. Then, for study arms with no events, we simply use the approximation \( 1/4n \) (Rucker et al, 2009).

We should consider the bias of the arcsine difference measure as well. Again, we use the fact that \( \hat{p}_T \) is approximately normal with mean \( p_T \) and variance \( \frac{p_T(1-p_T)}{n} \), and we let \( g \) be a smooth transformation of \( p_T \). Then, we have that the following relationship holds where the second term is the bias term:

\[
\text{E}(g(\hat{p}_T)) \approx g(\hat{p}_T) + \frac{1}{2} g''(\hat{p}_T) \frac{\hat{p}_T(1 - \hat{p}_T)}{n_T}.
\]
Suppose \( g(\hat{p}_T) = \arcsin(\sqrt{\hat{p}_T}) \). Then,

\[
E(\arcsin(\sqrt{\hat{p}_T})) \approx \arcsin(\sqrt{\hat{p}_T}) + \frac{1}{2} \cdot \frac{(2\hat{p}_T - 1)}{4(\hat{p}_T(1 - \hat{p}_T))^{3/2}} \cdot \hat{p}_T(1 - \hat{p}_T)
\]

\[
= \arcsin(\sqrt{\hat{p}_T}) + \frac{(2\hat{p}_T - 1)}{8n_T\sqrt{\hat{p}_T(1 - \hat{p}_T)}}.
\]

As \( n_T \) increases, the bias decreases, and the bias term eventually becomes negligible. Additionally, if \( \hat{p}_T = 0.5 \),

\[
g''(\hat{p}_T) = \frac{(2\hat{p}_T - 1)}{4(\hat{p}_T(1 - \hat{p}_T))^{3/2}} = \frac{0}{4(0.5)^3} = 0,
\]

and the bias term vanishes. In general, the bias is smallest when \( \hat{p}_T \) is closest to 0.5, and the bias term increases with larger deviations of that sample estimate from 0.5.

Here, we have only considered the one-armed case. Considering the two-armed case, we can again just extend the previous properties to both the control and treatment groups. Thus, we have that

\[
E(\arcsin(\sqrt{\hat{p}_T})) \approx \arcsin(\sqrt{\hat{p}_T}) + \frac{1}{2} \cdot \frac{(2\hat{p}_T - 1)}{4(\hat{p}_T(1 - \hat{p}_T))^{3/2}} \cdot \hat{p}_T(1 - \hat{p}_T)
\]

and

\[
E(\arcsin(\sqrt{\hat{p}_C})) \approx \arcsin(\sqrt{\hat{p}_C}) + \frac{1}{2} \cdot \frac{(2\hat{p}_C - 1)}{4(\hat{p}_C(1 - \hat{p}_C))^{3/2}} \cdot \hat{p}_C(1 - \hat{p}_C).
\]

Considering the expectation of the arcsine difference, \( \arcsin(\sqrt{\hat{p}_T}) - \arcsin(\sqrt{\hat{p}_C}) \), the second and third terms make up the bias:

\[
E(\arcsin(\sqrt{\hat{p}_T}) - \arcsin(\sqrt{\hat{p}_C})) = E(\arcsin(\sqrt{\hat{p}_T})) - E(\arcsin(\sqrt{\hat{p}_C}))
\]

\[
\approx \left( \arcsin(\sqrt{\hat{p}_T}) - \arcsin(\sqrt{\hat{p}_C}) \right) + \frac{1}{2} \cdot \frac{(2\hat{p}_T - 1)}{4(\hat{p}_T(1 - \hat{p}_T))^{3/2}} \cdot \hat{p}_T(1 - \hat{p}_T)
\]

\[
- \frac{1}{2} \cdot \frac{(2\hat{p}_C - 1)}{4(\hat{p}_C(1 - \hat{p}_C))^{3/2}} \cdot \hat{p}_C(1 - \hat{p}_C).
\]
Here, the bias can be expressed as,

\[
\frac{1}{2} \left( \frac{(2\hat{p}_T - 1)}{4(\hat{p}_T(1 - \hat{p}_T))^{3/2}} \cdot \frac{\hat{p}_T(1 - \hat{p}_T)}{n_T} - \frac{(2\hat{p}_C - 1)}{4(\hat{p}_C(1 - \hat{p}_C))^{3/2}} \cdot \frac{\hat{p}_C(1 - \hat{p}_C)}{n_C} \right),
\]

so larger differences between \( \hat{p}_T \) and \( \hat{p}_C \) will yield more biased estimates of the arcsine difference. Additionally, for a given difference between \( \hat{p}_T \) and \( \hat{p}_C \), the bias will also increase as the difference between \( n_T \) and \( n_C \) increases. Thus, we can conclude that the arcsine difference effect measure performs best and is least biased for small differences between \( \hat{p}_T \) and \( \hat{p}_C \) as well as between \( n_T \) and \( n_C \).

While the arcsine difference effect measure is biased, any transformations of the event probability will also be biased. Thus, the log odds ratio and log risk ratio will both yield biased estimates as well. It may be advantageous to use the arcsine difference measure in the presence of zero event trials due to the behavior of both the point estimate and the variance estimate when zeros are present. Both of these values will always be defined and finite regardless of zero events.
Chapter 4

Simulation Study

Rucker, et al (2009) performed a simulation study to show that using the arcsine difference as a measure of treatment effect results in relatively small bias and relatively good coverage of 95% confidence intervals. Here, I attempt to emulate some of their results as well as develop some new results.

Initially, I used six of the already simulated scenarios by Rucker et al (2009) which are summarized in Table 4.1.

<table>
<thead>
<tr>
<th>Number of Trials</th>
<th>Distribution of Trial Sizes</th>
<th>$p_T$</th>
<th>Odds Ratio</th>
<th>$\tau^2$</th>
<th>% Treated</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>log-normal($\mu = 6, \sigma^2 = 0.7$)</td>
<td>0.5</td>
<td>2.00</td>
<td>0.05</td>
<td>50</td>
</tr>
<tr>
<td>2</td>
<td>log-normal($\mu = 6, \sigma^2 = 0.7$)</td>
<td>0.5</td>
<td>1.33</td>
<td>0.05</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td>log-normal($\mu = 6, \sigma^2 = 0.7$)</td>
<td>1.0</td>
<td>2.00</td>
<td>0.05</td>
<td>50</td>
</tr>
<tr>
<td>4</td>
<td>log-normal($\mu = 6, \sigma^2 = 0.7$)</td>
<td>1.0</td>
<td>1.33</td>
<td>0.05</td>
<td>50</td>
</tr>
<tr>
<td>5</td>
<td>log-normal($\mu = 6, \sigma^2 = 0.7$)</td>
<td>5.0</td>
<td>2.00</td>
<td>0.05</td>
<td>50</td>
</tr>
<tr>
<td>6</td>
<td>log-normal($\mu = 6, \sigma^2 = 0.7$)</td>
<td>5.0</td>
<td>1.33</td>
<td>0.05</td>
<td>50</td>
</tr>
</tbody>
</table>

Table 4.1: Six of the scenarios from Rucker, et al (2009)

Since Rucker, et al (2009) chose the number of trials and distribution according to specific data, I decided to choose these parameters according to the data we will later examine and analyze. The data is described in the paper, “Long-acting Beta-Agonists with and without Inhaled Corticosteroids and Catastrophic Asthma Events” by Salpeter, et al (2010). Before eliminating trials with zero events, there were 95 studies, so I use 95 as the number of trials. Of the trial sizes I was able to obtain, the sizes were approximately
distributed as log-normal($\mu = 7, \sigma^2 = 0.5$), but one study has a total sample size of 26,353. Using this information, I drew 94 of my trial sizes from the log-normal($\mu = 7, \sigma^2 = 0.5$) distribution and one trial size from the Normal($\mu = 26000, \sigma = 500$). In the next section, I describe the methods I used to perform the simulation study.

### 4.1 Methods

For each combination of odds ratio and event probability in the treatment group, I simulated 42 (or 95) trials. The first step was to generate the number of patients in each of the trials from a lognormal distribution. In order to generate lognormal random variables, I first generated a standard normal random variable, $x$, and then transformed it.

\[
> x = \text{rnorm}(1, 0, 1) \\
> \text{size} = \text{ceiling}(\exp(\mu + \sigma^2 x))
\]

Then, I assigned approximately 50% of the total sample size to the treatment group and the rest to the control group.

\[
> u = \text{runif}(\text{size}, 0, 1) \\
> \text{treated} = \text{sum}(u \geq 0.5) \\
> \text{control} = \text{sum}(u < 0.5)
\]

I then needed to simulate some study heterogeneity among the trials. Since $\tau^2$ can be thought of as the variance of the treatment effects, I generated the true odds ratios for each trial from a Normal distribution with mean equal to the specified odds ratio and variance $\tau^2 = 0.05$.

\[
> \text{theta} = \text{rnorm}(95, \text{odds[i]}, \tau)
\]

Using the true event probabilities for the treatment and control groups and two more uniform(0,1) random variables, I then simulated the cell counts, $a, b, c,$ and $d$, that would appear in the $2 \times 2$ summary table.

\[
> \text{u.treat} = \text{runif}(\text{treated}, 0, 1)
\]
>a = sum(u.treat < pt)
>b = treated-a

c = sum(u.treat < pc)
d = control - c

After I had all of the summarized data for each trial, I was ready to perform the meta-analyses using R’s rma.uni function in the metafor package. Table 4.2 summarizes all of the effect measures I considered while performing the meta-analyses.

<table>
<thead>
<tr>
<th>Measure of Effect</th>
<th>Fixed or Random Effects</th>
<th>Continuity Correction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Odds Ratio</td>
<td>Fixed</td>
<td>No</td>
</tr>
<tr>
<td>Odds Ratio</td>
<td>Random</td>
<td>No</td>
</tr>
<tr>
<td>Odds Ratio</td>
<td>Fixed</td>
<td>Yes</td>
</tr>
<tr>
<td>Odds Ratio</td>
<td>Random</td>
<td>Yes</td>
</tr>
<tr>
<td>Risk Difference</td>
<td>Fixed</td>
<td>No</td>
</tr>
<tr>
<td>Risk Difference</td>
<td>Random</td>
<td>No</td>
</tr>
<tr>
<td>Risk Difference</td>
<td>Fixed</td>
<td>Yes</td>
</tr>
<tr>
<td>Risk Difference</td>
<td>Random</td>
<td>Yes</td>
</tr>
<tr>
<td>Arcsine Difference (analytic variance)</td>
<td>Fixed</td>
<td>No</td>
</tr>
<tr>
<td>Arcsine Difference (analytic variance)</td>
<td>Random</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 4.2: Ten different meta-analyses performed

For each effect measure, I then checked to see if the 95% confidence interval contained the true value of the treatment effect. I also checked to see if the confidence interval indicated that there was a treatment effect at all (did it contain 0 or not). For each combination of parameters, I repeated this 1000 times in order to find Monte Carlo estimates for all of the above effect measures as well as the bias of the estimates.

4.2 Results

In this section, I will only discuss the results of the latter six scenarios. The original six scenarios are described in detail by Rucker et al (2009), so I focus my attention to the new
results. Of these six new scenarios, three had a true odds ratio of 2, and three had a true odds ratio of 1.33. First, I consider the results of the fixed effect analyses according to the true odds ratios. In Figure 4.1, we see the Monte Carlo estimates of bias for each of the effect measures when the true odds ratio is 2. Notice that the arcsine difference appears to be the least biased measure across the three different event probabilities considered here.

Figure 4.1: Monte Carlo estimates of bias when the true odds ratio is 2

In Figure 4.2, we see the same graph when the true odds ratio is 1.33. Here, the arcsine difference does not perform quite as well. It appears to be the least biased with the smaller event probabilities, but it is the most biased when the true event probability in the treatment group is 0.05.

Figure 4.3 shows the proportion of times the confidence interval covered the true value of the treatment effect. None of the intervals covered the true odds ratio 95% of the time, and the arcsine difference is the only measure that ever covered more than 90% of the time. With the exception of the arcsine difference, all of the confidence intervals do best when the true event probability in the treatment group is 0.05, and they do worst when the true event probability is closest to zero, but the coverage is still not what would be expected.
Figure 4.2: Monte Carlo estimates of bias when the true odds ratio is 1.33

Figure 4.3: 95% confidence intervals: Coverage of true treatment effect over 1000 meta-analyses when true odds ratio is 2
The same results are shown for a true odds ratio of 1.33 in Figure 4.4. Here, we see the best coverage when the event probability in the treatment group is smallest, but again the coverage is not what we would expect for 95% confidence intervals as most of the confidence intervals cover the true value of the odds ratio in less than 90% of cases.

![95% CI Coverage when True OR=1.33](image)

Figure 4.4: 95% confidence intervals: Coverage of true treatment effect over 1000 meta-analyses when true odds ratio is 1.33

I also found the proportion of times the 95% confidence intervals covered zero because this would lead us to make the incorrect conclusion that the event probabilities in the treatment and control groups were the same. When the true odds ratio is 2, we never make the wrong conclusion when we perform a fixed effect meta-analyses, regardless of the effect measure and the true event probability in the treatment group. We also never make the wrong conclusion when the true event probability is 0.05 in the treatment group. Figure 4.5 shows that we are most likely to conclude no difference between the treatment and control groups when the true event probability is closest to zero. It also indicates that we are least likely to make that incorrect conclusion when we use the arcsine difference measure.

Now that we have seen how these estimates perform under the fixed effect models, let’s
Figure 4.5: 95% confidence intervals: Coverage of zero over 1000 meta-analyses when true odds ratio is 1.33

consider the random effect models. Figures 4.6 and 4.7 depict the bias of the estimates when the true odds ratios are 2 and 1.33, respectively. In both cases we see that the bias of the arcsine difference estimate appears to be comparable if not less than the bias of the other estimates.

Additionally, we see the performance of the 95% confidence intervals based on the random effect models. As would be expected, the confidence intervals based on the random effect models are more likely to cover the true effect measure than those based on the fixed effects models. Here, the confidence intervals for the risk difference both with and without a continuity correction and the arcsine difference cover the true value 100% of the time. This is true when the true odds ratio is either 1.33 or 2. Both of the log odds ratio confidence intervals perform much better when the true odds ratio is 1.33.

Finally, we can consider the probability that the confidence intervals include zero when we base them on the random effect models. The confidence intervals for risk difference both with and without a continuity correction always cover zero. That is, when the true odds ratio is 1.33 or 2, we always make the wrong conclusion if we perform a meta-analysis
Figure 4.6: Monte Carlo estimates of bias when the true odds ratio is 2 under the random effects model

Figure 4.7: Monte Carlo estimates of bias when the true odds ratio is 1.33 under the random effects model
using the risk difference as the effect measure. Additionally, the confidence interval for the arcsine difference included zero 100% of the time except when the true odds ratio was 2, and the true event probability in the treatment group was 0.05. In that case, the confidence interval never covered zero. On the other hand, the log odds ratio confidence intervals never covered zero when the true odds ratio was 2, and when the true odds ratio was 1.33, they only covered zero in the case when the event probability in the treatment group was 0.005. In general, the confidence intervals based on the random effect analyses were more likely to cover zero. Since random effect model are more conservative, this is just as we would expect.

Overall, I think these results are consistent with the results from Rucker et al (2009). The arcsine difference does seem to be an effective way to measure treatment effect when we are dealing with zero events. The results indicate that the arcsine difference is a less biased estimate than using a continuity correction with either risk difference or log odds ratio. The confidence interval coverage appears to be comparable to other methods as well.
Figure 4.9: 95% confidence intervals: Coverage of true treatment effect over 1000 meta-analyses when true odds ratio is 1.33 under the random effects model

The real problem now is to deal with how to explain what the arcsine difference is to the general public. Are researchers willing to lose the intuitive nature of those more standard procedures for the ability to include more trials without adding bias? Where risk difference, odds ratio, and log odds ratio are easy to explain to the general population, the arcsine difference is definitely not intuitive, and researchers may be hesitant to use the procedure. Nonetheless, I will apply the arcsine difference measure to real data in the next section.
Chapter 5

Example: Adverse events from long-acting beta-agonists in asthma patients

As aforementioned, Salpeter, et al (2010) performed a meta-analysis on the adverse effects of long-acting β-agonists in asthma patients, both with and without inhaled corticosteroids.

5.1 Background

Long-acting β-agonists, specifically salmeterol and formoterol, are inhaled medications that are known to relax the muscle bands that surround the airways, allowing patients to breathe in and out more easily. While asthma patients use rescue inhalers to relieve the sudden onset of asthma systems, these β-agonists are prescribed by some physicians to be taken every day, regardless of the presence of asthma symptoms. They are considered to be controller medications which work slowly over time to prevent asthma symptoms from occurring. Oftentimes, these β-agonists are used in combination with inhaled corticosteroids. Inhaled corticosteroids are another type of controller medication known to reduce and prevent the swelling and excess mucus in the airway caused by inflammation, reducing the likelihood of future asthma attacks or flare-ups.
In their paper, Salpeter, et al (2010), consider the effects of these long-acting $\beta$-agonists, both with and without concomitant inhaled corticosteroids, on asthma-related hospitalizations, intubations, and deaths.

After their literature search, they identified 211 potentially relevant trials. Two trials were excluded due to the lack of randomization, and another 114 trials were excluded because they did not meet the appropriate study criteria. Of the 95 remaining trials, 83 did not have any adverse events. Since the odds ratio can only be estimated when at least one event is observed, Salpeter et al (2010) eliminated these 83 trials from the analysis. The authors proceeded to perform three different meta-analyses using only 12 of the trials. These 12 trials were split into two categories. The first five trials compared long-acting $\beta$-agonists with variable inhaled corticosteroids to a placebo, while the other seven trials compared long-acting $\beta$-agonists with concomitant inhaled corticosteroids to inhaled corticosteroids alone. The first two meta-analyses were performed considering the two groups of trials separately, and then, the groups were combined to perform a third meta-analysis (Salpeter et al, 2010). Note that the 95 aforementioned trials met the appropriate study criteria for this third meta-analysis. That is, these studies either compared long-acting $\beta$-agonists with variable inhaled corticosteroids to a placebo or compared long-acting $\beta$-agonists with concomitant inhaled corticosteroids to inhaled corticosteroids alone. Further information about which treatments were compared in the 83 studies which were eliminated was not available. The following analysis will focus on the second set of trials, long-acting $\beta$-agonists with concomitant inhaled corticosteroids compared to inhaled corticosteroids alone.

5.2 Analysis

First, I will repeat the same analysis that is reported. Then, using the same 7 studies, I will perform a meta-analysis using the arcsine difference effect measure. Finally, I will obtain the summary data for the eligible studies that were not included in the analysis and perform the meta-analysis using the arcsine difference effect measure. Table 5.1 summarizes the seven trials that were included in the Salpeter et al (2010) paper as well as the summarized data for all of the eligible trials that were excluded due to zero events.
<table>
<thead>
<tr>
<th>Study, Year or Subgroup</th>
<th>Treatment n/N</th>
<th>Control n/N</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSK pooled trials, 2008b</td>
<td>8/633</td>
<td>3/642</td>
</tr>
<tr>
<td>Ind et al, 2003</td>
<td>1/173</td>
<td>0/329</td>
</tr>
<tr>
<td>Kelsen et al, 1999</td>
<td>1/239</td>
<td>0/244</td>
</tr>
<tr>
<td>Kemp et al, 1998</td>
<td>1/126</td>
<td>0/128</td>
</tr>
<tr>
<td>O' Byrne et al, 2001</td>
<td>1/869</td>
<td>0/862</td>
</tr>
<tr>
<td>O' Byrne et al, 2005</td>
<td>1/1834</td>
<td>0/926</td>
</tr>
<tr>
<td>von Berg et al, 2003</td>
<td>1/165</td>
<td>0/83</td>
</tr>
<tr>
<td>Summarized data for all zero event trials</td>
<td>0/30961</td>
<td>0/25786</td>
</tr>
</tbody>
</table>

Table 5.1: Summary of studies testing long-acting β-agonists with concomitant inhaled corticosteroids versus inhaled corticosteroids alone

5.2.1 Peto Odds Ratio with No Zero Event Trials

Initially, I simply repeated the existing analysis in both SAS and R. The existing results used the Peto odds ratio as the effect measure. The Peto method, or one-step method, for calculating the odds ratio is as follows. Suppose for a given study, say study \( i \), we observe the following:

<table>
<thead>
<tr>
<th>Event</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group</td>
</tr>
<tr>
<td>Treatment</td>
</tr>
<tr>
<td>Control</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

Table 5.2: Summary data for trial \( i \)

Let \( O_i \) be the observed number of events in the treatment group. That is,

\[
O_i = a_i.
\]

Let \( E_i \) be the expected number of events in the treatment group, where

\[
E_i = \frac{(a_i + b_i) \times (a_i + c_i)}{n_i}.
\]
Finally, let
\[ I_i = \frac{(a_i + b_i) \times (c_i + d_i) \times (a_i + c_i) \times (b_i + d_i)}{n_i^2 \times (n_i - 1)}. \]

Then, we can estimate the log odds ratio in study \( i \) by
\[ Y_i = \frac{O_i - E_i}{I_i}. \]

It follows that the variance of the log odds ratio for study \( i \) is given by
\[ V_{Y_i} = \frac{1}{I_i}, \]
and each study is given weight,
\[ W_i = \frac{1}{V_{Y_i}} = I_i. \]

When performing a meta-analysis, we use the weighted average of the log odds ratio estimates:
\[ \frac{\sum_{i=1}^{k} W_i Y_i}{\sum_{i=1}^{k} W_i} \]
with variance,
\[ \frac{1}{\sum_{i=1}^{k} W_i}. \]

The 95% confidence interval for the summary log odds ratio is then given by
\[ \frac{\sum_{i=1}^{k} W_i Y_i}{\sum_{i=1}^{k} W_i} \pm 1.96 \times \frac{1}{\sum_{i=1}^{k} W_i}. \]

Exponentiating, we can obtain the summary odds ratio and the corresponding 95% confidence interval (Borenstein et al, 2009).

Following these steps, I was able to verify the results of the Salpeter et al (2010) paper. While I performed both a fixed effects analysis and a random effects analysis, there was minimal evidence of heterogeneity, and the results of the two analyses were identical. Table 5.3 summarizes these results.

In both cases, the computed odds ratio is approximately 3.65, and the 95% confidence
<table>
<thead>
<tr>
<th>Study, Year or Subgroup</th>
<th>Treatment n/N</th>
<th>Control n/N</th>
<th>Peto OR</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSK pooled trials, 2008b</td>
<td>8/633</td>
<td>3/642</td>
<td>2.54</td>
<td>(0.77, 8.31)</td>
</tr>
<tr>
<td>Ind et al, 2003</td>
<td>1/173</td>
<td>0/329</td>
<td>18.21</td>
<td>(0.29, 1125.35)</td>
</tr>
<tr>
<td>Kelsen et al, 1999</td>
<td>1/239</td>
<td>0/244</td>
<td>7.55</td>
<td>(0.15, 380.34)</td>
</tr>
<tr>
<td>Kemp et al, 1998</td>
<td>1/126</td>
<td>0/128</td>
<td>7.51</td>
<td>(0.15, 378.39)</td>
</tr>
<tr>
<td>O’byrne et al, 2001</td>
<td>1/869</td>
<td>0/862</td>
<td>7.33</td>
<td>(0.15, 369.41)</td>
</tr>
<tr>
<td>O’byrne et all, 2005</td>
<td>1/1834</td>
<td>0/926</td>
<td>4.50</td>
<td>(0.07, 285.97)</td>
</tr>
<tr>
<td>von Berg et al, 2003</td>
<td>1/165</td>
<td>0/83</td>
<td>4.50</td>
<td>(0.07, 286.17)</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>14/4039</strong></td>
<td><strong>3/3214</strong></td>
<td><strong>3.65</strong></td>
<td>(1.39, 9.55)</td>
</tr>
</tbody>
</table>

Table 5.3: Summary of meta-analysis using Peto odds ratio effect measure.

Interval for the odds ratio is (1.39, 9.55). From these results, we have simply verified the conclusion that there is a statistically significant difference in the proportion of patients with at least one asthma intubation or death when we compare those taking long-acting β-agonists with concomitant inhaled corticosteroids and those taking inhaled corticosteroids alone. That is, since the entire confidence interval lies above one, we can conclude that there are fewer adverse events in the group taking inhaled corticosteroids alone. Now that I was able to replicate these results, I will analyze the same data using the arcsine difference effect measure.

### 5.2.2 Arcsine Difference with No Zero Event Trials

When using the arcsine difference as the effect measure, we now estimate the effect for study $i$ as

$$Y_i = \arcsin(\sqrt{a_i/n_{Ti}}) - \arcsin(\sqrt{c_i/n_{Ci}}),$$

and the variance is given by

$$V_{Y_i} = \frac{1}{4n_{Ti}} + \frac{1}{4n_{Ci}}.$$  

Each study is given weight

$$W_i = \frac{1}{\frac{1}{4n_{Ti}} + \frac{1}{4n_{Ci}}}.$$
For the meta-analysis, we simply use a weighted average of the arcsine difference estimates for each study:

\[ \frac{\sum_{i=1}^{k} W_i Y_i}{\sum_{i=1}^{k} W_i} \]

with variance,

\[ \frac{1}{\sum_{i=1}^{k} W_i} \]

The 95% confidence interval for the summary arcsine difference is then given by

\[ \frac{\sum_{i=1}^{k} W_i Y_i}{\sum_{i=1}^{k} W_i} \pm 1.96 \times \frac{1}{\sum_{i=1}^{k} W_i} \]

Following these steps, I performed both the fixed effects and random effects analyses. Since I used the same data as with the Peto odds ratio, there again was almost no evidence of heterogeneity, so the fixed and random effects analyses yielded exactly the same results. Table 5.4 summarizes those results, and Figure 5.1 shows the forest plot generated in R. Notice that the plot is specifically for the fixed effects model, but the plot for the random effects model is identical since there is no evidence of heterogeneity.

<table>
<thead>
<tr>
<th>Study, Year or Subgroup</th>
<th>Treatment n/N</th>
<th>Control n/N</th>
<th>Arcsine Diff</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSK pooled trials, 2008b</td>
<td>8/633</td>
<td>3/642</td>
<td>0.044</td>
<td>(-0.011, 0.099)</td>
</tr>
<tr>
<td>Ind et al, 2003</td>
<td>1/173</td>
<td>0/329</td>
<td>0.076</td>
<td>(-0.016, 0.168)</td>
</tr>
<tr>
<td>Kelsen et al, 1999</td>
<td>1/239</td>
<td>0/244</td>
<td>0.065</td>
<td>(-0.024, 0.154)</td>
</tr>
<tr>
<td>Kemp et al, 1998</td>
<td>1/126</td>
<td>0/128</td>
<td>0.089</td>
<td>(-0.034, 0.212)</td>
</tr>
<tr>
<td>O’byrne et al, 2001</td>
<td>1/869</td>
<td>0/862</td>
<td>0.034</td>
<td>(-0.013, 0.081)</td>
</tr>
<tr>
<td>O’byrne et al, 2005</td>
<td>1/1834</td>
<td>0/926</td>
<td>0.023</td>
<td>(-0.016, 0.063)</td>
</tr>
<tr>
<td>von Berg et al, 2003</td>
<td>1/165</td>
<td>0/83</td>
<td>0.078</td>
<td>(-0.054, 0.210)</td>
</tr>
<tr>
<td>Total</td>
<td>14/4039</td>
<td>3/3214</td>
<td>0.040</td>
<td>(0.017, 0.064)</td>
</tr>
</tbody>
</table>

Table 5.4: Summary of meta-analysis using the arcsine difference effect measure (zero event trials excluded)

The estimated arcsine difference is given by 0.040, and the estimated 95% confidence interval for the arcsine difference is (0.017, 0.064). Here, we notice that the confidence interval is entirely greater than zero. Thus, we again conclude that there is a statistically significant difference between the two groups, and we have evidence that there are fewer
adverse events in the group taking inhaled corticosteroids alone. This is the expected result given that the same data was used, and it shows that the results of the analysis using arcsine difference measure are matching up with the results using the Peto odds ratio. Now, we are ready to perform the analysis using the arcsine difference measure when we include the trials with zero observed events since these needed to be deleted in the initial analysis using the odds ratio.

5.2.3 Arcsine Difference with Summary Data for Zero Event Trials

An eighth trial is added in which we observe zero events across all patients, including 30961 in the treatment group and 25786 in the control group. Following the same steps as the previous analysis using the arcsine difference effect measure, I performed both fixed effects and random effects analyses. In this case, the two analyses yield different results. When we
test for homogeneity, we get a p-value of 0.08, therefore failing to reject the null hypothesis of homogeneity. Table 5.5 summarizes the results of the fixed and random effects analyses using the arcsine difference effect measure. Additionally, Figures 5.2 and 5.3, show the forest plots for the fixed effects and random effects models, respectively, as generated in R. Notice that in both cases, the eighth trial is weighted much more heavily than the others due to the large sample size.

<table>
<thead>
<tr>
<th>Study, Year or Subgroup</th>
<th>Treatment n/N</th>
<th>Control n/N</th>
<th>Arcsine Diff</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSK pooled trials, 2008b</td>
<td>8/633</td>
<td>3/642</td>
<td>0.044</td>
<td>(−0.011, 0.099)</td>
</tr>
<tr>
<td>Ind et al, 2003</td>
<td>1/173</td>
<td>0/329</td>
<td>0.076</td>
<td>(−0.016, 0.168)</td>
</tr>
<tr>
<td>Kelsen et al, 1999</td>
<td>1/239</td>
<td>0/244</td>
<td>0.065</td>
<td>(−0.024, 0.154)</td>
</tr>
<tr>
<td>Kemp et al, 1998</td>
<td>1/126</td>
<td>0/128</td>
<td>0.089</td>
<td>(−0.034, 0.212)</td>
</tr>
<tr>
<td>O’byrne et al, 2001</td>
<td>1/869</td>
<td>0/862</td>
<td>0.034</td>
<td>(−0.013, 0.081)</td>
</tr>
<tr>
<td>O’byrne et al, 2005</td>
<td>1/1834</td>
<td>0/926</td>
<td>0.023</td>
<td>(−0.016, 0.063)</td>
</tr>
<tr>
<td>von Berg et al, 2003</td>
<td>1/165</td>
<td>0/83</td>
<td>0.078</td>
<td>(−0.054, 0.210)</td>
</tr>
<tr>
<td>All zero event trials</td>
<td>0/30961</td>
<td>0/25786</td>
<td>0.000</td>
<td>(−0.008, 0.008)</td>
</tr>
<tr>
<td>Total (fixed)</td>
<td>14/35000</td>
<td>3/29000</td>
<td>0.004</td>
<td>(−0.003, 0.012)</td>
</tr>
<tr>
<td>Total (random)</td>
<td>14/35000</td>
<td>3/29000</td>
<td>0.029</td>
<td>(0.005, 0.054)</td>
</tr>
</tbody>
</table>

Table 5.5: Summary of meta-analysis using the arcsine difference measure (zero event trials included)

In the initial fixed effects analysis, I found that the estimate for the arcsine difference is given by 0.0044, which is very close to zero. Additionally, the 95% confidence interval is (−0.0034, 0.0122), which contains zero. This is quite alarming since the results of this analysis suggest that the two groups, patients taking long-acting β-agonists with concomitant inhaled corticosteroids and patients taking inhaled corticosteroids alone, are not statistically different. That is, there is not a significant difference in the proportion of patients experiencing at least one asthma intubation or death, contradicting the previously published study that did not include any zero event trials.

The random effects analysis yields a much different estimate for the arcsine difference, 0.0294, and the confidence interval is given by (0.0049, 0.0539), which lies completely above zero. The results of this analysis are consistent with the previously published results in terms of statistical significance, but it appears that the treatment effect may have been somewhat exaggerated when the zero event trials were excluded. Notice that the previous
Figure 5.2: Forest plot for the fixed effects meta-analysis model with the arcsine difference effect measure (zero event trials included)

The arcsine difference estimate of 0.040 is greater than 0.0294, and the previous confidence interval is further from zero.

Usually, it is the case that the random effects analysis is more conservative since it yields wider confidence intervals. Here, we do see a wider confidence interval, but we find statistical significance in the random effects model and not with the fixed effects model. It is more common to find the opposite situation, statistical significance in the fixed effects setting and not in the random effects setting.

Now, we must decide which model to choose and what exactly these results mean. We know that we should not base our model choice on the test for homogeneity (Borenstein et al, 2009). In general, we should choose the fixed effects model when we believe all of the studies included in the analysis are functionally identical, and when our goal is to compute the common effect size for the identified population without the goal of generalizing to other
populations. If either of these is not true, we should choose the random effects model. The issue here is that we have a small sample size, which means that even if the random effects model is appropriate, we do not have enough information to apply it correctly. In this case, we have a few options. One option is simply to consider each trial individually, but if we do choose to perform a meta-analysis, it is appropriate to use the fixed effects model as long as we consider it as a descriptive analysis. That is, we are only describing the included studies, and we are not able to make inferences about a larger population (Borenstein et al, 2009). Nonetheless, in this case the fixed effects analysis using the arcsine difference effect measure provides some doubt to the validity of the previously published results. Now, we will consider what this might mean for patients and pharmaceutical companies.

Upon the publication of the Salpeter et al (2010) article, pharmaceutical companies that produce these long-acting β-agonists with variable and concomitant inhaled corticosteroids
are now required to include warning labels on these drugs. These labels warn patients that they may have an increased risk of asthma-related death when taking these medications, and consequently, some patients are hesitant to take these drugs. We have already addressed some of the implications of the methods of analysis. Specifically, the elimination of all trials with fewer than one event from the analysis may have skewed their results. It has also been argued that the risk of an adverse event is very tiny, and the benefits of the drugs outweigh the risks. Additionally, others have expressed concerns that the patients in these trials were not monitored appropriately and may not have been taking the medications correctly. While these may be valid concerns, the US Food and Drug Administration (FDA) continues to require the warnings based on the results of the 2010 article.

Now, using the arcsine difference effect measure, we were able to include all of the trials, even those with zero events. Considering the previously discussed fixed effects meta-analysis, we obtain results that contradict these warning labels. Thus, we can conclude that the risk of death for patients taking long-acting $\beta$-agonists with concomitant inhaled corticosteroids is no different than the risk of death for patients taking inhaled corticosteroids alone among the patients in these eight trials. We must keep in mind that these conclusions are based only on the fixed-effects analysis, so the conclusions cannot be extended to a wider population. Regardless, the results of this meta-analysis will most likely be welcomed by pharmaceutical companies and may put patients’ minds at ease.
After introducing meta-analysis, we focused on binary responses with zero events. Effect measures that are generally used with binary outcomes do not perform well in the presence of zero events. Specifically, the estimates of these measures may be infinite or undefined. Furthermore, the variance estimates may be infinite or zero, making the construction of confidence intervals impossible. The arcsine difference measure offers a solution to these problems.

While the arcsine difference measure may not be as easy for the general population to interpret as other effect measures such as the odds ratio, when zero event trials are present, there are major computational advantages to using this measure. We verified that both the point estimates and the variance estimates will always be defined and finite, which is not the case with the other effect measures for binary outcomes. Therefore, using the arcsine difference effect measure completely eliminates the need to either remove zero event trials from the analysis or use a continuity correction. Consequently, the arcsine difference measure is beneficial to use when considering rare events as we will not lose the information from zero event trials.

The performance of the arcsine difference measure was examined through a simulation study, and finally, we performed a meta-analysis of asthma clinical trials using the arcsine difference as the effect measure. The results of this meta-analysis were then compared to the results of a meta-analysis using the Peto odds ratio published by Salpeter et al in 2010. The
published results indicated that patients taking long-acting β-agonists with concomitant inhaled corticosteroids had a higher risk of asthma intubations and death than patients taking inhaled corticosteroids alone, but this meta-analysis did not include any zero-event trials. After adding an additional trial that combined the data from all of the zero-event trials that were ignored, we performed a meta-analysis using the arcsine difference rather than the Peto odds ratio.

The random-effects analysis yielded similar results to those that were published, indicating that the risk of adverse events was higher for the group taking long-acting β-agonists with concomitant inhaled corticosteroids. The estimated arcsine difference is 0.0294, and the confidence interval is (0.0049, 0.0539). On the other hand, the fixed effects analysis yielded the much smaller arcsine difference estimate of 0.0044 and a 95% confidence interval that contains zero: (−0.0034, 0.0122). These results provide some evidence that there is not as great of a difference in the risk of asthma intubations or death between the two groups of patients as is indicated when zero event trials are ignored. In particular, the fixed effects model suggests that among these eight trials, there is not a difference in the risk of asthma intubations or death among patients taking long-acting β-agonists with concomitant inhaled corticosteroids and patients taking inhaled corticosteroids alone.

There a few limitations that we need to keep in mind when considering this meta-analysis. Here, we did not perform our own literature search. Instead, information from Salpeter et al (2010) was used. Because we did not perform the search, we were only able to obtain information for individual trials if they were included in the previously published meta-analysis. Thus, the many trials that were deleted were not included as individual trials. Instead, we used the summarized data from the article to create one large zero event trial. Had we included each trial separately, the trials would have been weighted differently, and results may have differed both in terms of the heterogeneity and the actual estimates and confidence intervals. Additionally, if we had performed a new literature search, the included studies may have been slightly different, yielding slightly different results.
Bibliography


