Detecting and Accommodating Outliers in Meta-Analysis for Evaluating Effect of Albendazole on *Ascaris lumbricoides* Infection

Hamid Alavi Majd 1; Khadijeh Najafi Ghobadi 1; Alireza Akbarzadeh Baghban 2; Nayebali Ahmadi 3; Elham Sajjadi 4

1Department of Biostatistics, School of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, IR Iran
2Department of Basic Sciences, School of Rehabilitation Sciences, Shahid Beheshti University of Medical Sciences, Tehran, IR Iran
3Department of Lab Sciences and Proteomics Research Center, School of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, IR Iran
4Department of Hematology, Tarbiat Modares University, Tehran, IR Iran

*Corresponding Author: Khadijeh Najafi Ghobadi, Department of Biostatistics, School of Paramedical Sciences, Shahid Beheshti University of Medical Sciences, Tehran, IR Iran. Tel/ Fax: +98-2122707347, E-mail: kh.najafi420@yahoo.com

*Received: January 22, 2014; Revised: March 20, 2014; Accepted: April 9, 2014*

**Background:** Meta-analysis is a statistical technique in which the results of two or more independent studies, with similar objectives, are mathematically combined in order to improve the reliability of the results. The outliers, which may exist even in random models, can affect the validity and strength of meta-analysis results. One of the factors that can negatively affect the validity of results in a meta-analysis computation is the presence of outlier, an element of a data set that distinctly stands out from the rest of the data. The most thorough method of identifying distant data in terms of outliers has been formulated by Hedges and Olkin (2). Numerous graphic methods have been introduced for the evaluation of unusual samples, but these methods can just be used for fixed models (3, 4). However, bearing in mind that there are mixed models and random models as well, it is better to investigate outliers through these models too (5, 6). Gumedze and Jackson (7) introduced methods of detecting and accommodating outliers in a meta-analysis work by a random effects variance shift model. Therefore, the current study uses Gumedze and Jackson’s model to meta-analytically combine and analyses the results of different studies to achieve a more reliable outcome.

**Objectives:** The current study uses “random effects variance shift model” to evaluate and correct the outliers in performing a meta-analysis study of the effect of albendazole in treating patients with *Ascaris lumbricoides* infection. One of the factors that can negatively affect the validity of results in a meta-analysis computation is the presence of outlier, an element of a data set that distinctly stands out from the rest of the data. The most thorough method of identifying distant data in terms of outliers has been formulated by Hedges and Olkin (2). Numerous graphic methods have been introduced for the evaluation of unusual samples, but these methods can just be used for fixed models (3, 4). However, bearing in mind that there are mixed models and random models as well, it is better to investigate outliers through these models too (5, 6). Gumedze and Jackson (7) introduced methods of detecting and accommodating outliers in a meta-analysis work by a random effects variance shift model. Therefore, the current study uses Gumedze and Jackson’s model to meta-

**Patients and Methods:** The study used data from 14 clinical trials; each article was composed of two groups, a treatment group and a placebo group. These articles compared the effect of single dose intakes of 400 mg albendazole in treating two groups of patients with *Ascaris lumbricoides* infection. The articles were published in a number of internationally indexed journals between 1983 to 2013. For both groups in each article, the total number of participants, the number of those with *Ascaris lumbricoides* infection, and the number of those recovered following the intake of albendazole were identified and recorded. The relative risk (RR) and variance were computed for each article individually. Then, using meta-analysis, the RR was computed for all the articles together. In order to detect outliers the “random effects variance shift model” and “likelihood ratio test” (LRT) were used. Adopting the bootstrap method, the accuracy rates for sampling distribution of the tests, which were used for multiple testing, were obtained and the relevant graphs were depicted. For data analysis, STATA and R software were used.

**Results:** According to meta-analysis results, the estimate for RR was 2.91, with a 95% confidence interval of 2.6 to 3.25. According to the method used in this study, three articles (articles number 4, 7, and 12) were outliers and, as such, they were detected in the graphs.

**Conclusions:** We can detect and accommodate outliers in meta-analysis by using random effects variance shift model and likelihood ratio test.

**Keywords:** Meta-Analysis; Albendazole; Outliers; Random Effects Variance Shift Model; *Ascaris lumbricoides*
sure the degree or size of outliers in a meta-analysis of the
effect of albendazole on patients with *Ascaris lumbricoides*
infection. *Ascaris lumbricoides* is one of the most soil-trans-
mitted helminthes (STH) in the world. It is estimated that
4.5 billion individuals are at risk of STH infection (Ascaris
lumbricoides, hookworms, and Trichuris trichiura) and as
many as 1.2 billion individuals might be infected with As-
caris lumbricoides, with *Ascaris lumbricoides*, close to 800
million with *Trichuris trichiura*, and more than 700 million
with hookworms (8, 9). The majority of STH infected in-
dividuals are children and the infections are an important
factor contributing to malnutrition in this age group (10).

2. Objectives

The current study used "random effects variance shift
model" to evaluate and correct the outliers in performing
a meta-analysis study of the effect of albendazole in
 treating patients with *Ascaris lumbricoides* infection.

3. Patients and Methods

The current study used data from 14 clinical trial ar-
ticles, investigating the effect of albendazole in treating
patients with *Ascaris lumbricoides* infection (11-24). The
articles had already been published in internationally
referred journals from 1983 to 2013. The articles were
first obtained through different sources like the internet,
data banks, and internationally recognized journals with
some special criteria indicated below and then were sub-
jected to the relevant meta-analyses. We used the terms
"albendazole" in combination with "trial" or "study",
"ascariasis", and "Ascaris lumbricoides". Bibliographies of
identified articles were screened for additional relevant
studies. The other criteria such as sample size, age, di-
agnostic method, and dosage were checked in selected
articles. The patients under study had been matched in
terms of age in the studies under meta-analysis. Besides,
all the studies had used a similar definition for recovery,
the same amount and frequency of albendazole (a daily
single dose of 400 mg of oral medication), and a similar
binary response variable for recovery versus nonreco-
very. For each of the 14 articles, the total number of partici-
pants, the number of those infected with *Ascaris lumbrici-
oides* as well as those recovered following the intake of
albendazole (for each of the two groups), the effect size,
and variance of the intervention were computed. Bearing
in mind that each study was composed of both the alben-
dazole and the placebo groups, the responses produced
would follow a dichotomous variable. To compare the ef-
effect of albendazole on *Ascaris lumbricoides*, the cure rates
were used to compare two groups under study. The priori-
ity index of the effect of albendazole as cure ratio in inter-
vention group to the placebo group was considered the
relative risk (RR). The effect size or RR is shown by θ; then
test statistic has to be defined for the significance of the
effect size. Test statistics is defined by \( Q = \sum W_i (Y_i - \bar{Y})^2 \)
in which \( W_i = 1 / V_i \), and \( \bar{Y} = \sum (W_i Y_i) / \sum (W_i) \). Under the
null hypothesis, for all the effect sizes, which were similar
or symmetrical, the distribution of \( Q \) statistics was chi-
square with k-1 degree of freedom (25). To detect outliers
in the data of current study, the random effect Variance
Shift Outlier Model was used. As such, this method was
used to detect and test the outliers. For meta-analysis,
the STATA software was used and the R software was em-
ployed to administer this method (26). A brief account of
the method is provided below.

3.1. Random Effect Variance Shift Outlier Model (RVSOM)

Following Gumedze and Jackson (7), basic model on
standard random effects for meta-analysis is as follows:
\[
y = \mu_i + \delta_i d_i + u + e (1)
\]
where \( y \) is a n-vector of estimated treatment effects
for the n independent studies, \( \mu \) is the unknown overall
treatment effect, \( \delta_i \) is a n-vector of ones, \( u \) is a n-vector
of unknown random effects, \( u \sim N(0, \tau^2 I_n) \), where \( \tau^2 \) is
the between-study variance, which is unknown, \( e \) represents
residual errors with \( e \sim N(0, R) \) where \( R = \text{diag} (\delta_1^2, \delta_2^2, \ldots, \delta_n^2) \). The elements of \( R \), the study variances, are regarded
as known. The variance-covariance matrix of \( (1) \) can then
be written as \( \text{var}(y) = V = \tau^2 I_n + R \) with the variance of the i
th study treatment effect given as \( \text{var}(y_i) = \tau^2 + \delta_i^2 \).

3.2. Extending the Random Effects Model to the RVSOM

According to Gumedze and Jackson (7), the random ef-
effects variance shift outlier model (RVSOM) for the ith
study (which allows an inflated variance for the ith study)
takes the form:
\[
Y = \mu_i + \delta_i d_i + u + e (2)
\]
This adds an extra term \( \delta_i d_i \) to model (1), where \( d_i \) is the
ith unit vector of length n, i.e. with value 1 in the ith posi-
tion and zero elsewhere, and is an unknown random coef-
ficient with \( \delta_i \sim N(0, \omega^2 i^2) \) for \( \omega^2 \geq 0 \).

The subscript \( j \) indicates which study has an inflated
variance. Model (2) has the form of a simple linear mixed
model with \( \delta_i \) as a random effect with variance \( \omega^2 i^2 \).
The variance-covariance matrix for the data under the RVSOM
for the jth observation is:
\[
\text{var}(y) = \omega^2 i^2 d_i d_i' + V
\]
An extension of model (2), which allows different inflat-
ed variances for more than one study, can be written as:
\[
Y = \mu_i + D_i \delta_i + u + e
\]
Where I is a subset \{1, 2, ..., r\} of studies considered to be
outliers, D = \{d_i\} is an n \times r matrix containing entires of 0
and 1, where an entry of 1 in the ith row and jth column
indicates that study i has the jth of r inflated variances, and
\( \delta_i \) is a r \times 1 vector of unknown random effects. We referred
to this model as an 'extended RVSOM' (7).

3.3 Administering the Random Effect Variance Shift
Outlier Model

At First, we used forest plot diagram to detect outliers in
our data, then we entered the outliers detected in forest
plot in the RVSOM Model as the jth observation. Then the model was fitted to the data and the degree of $\omega^2_j$, for the jth was computed; the larger size of $\omega^2_j$, the more likely for it to detect as an outlier. The likelihood ratio test (LRT) was used to measure the size or magnitude of $\omega^2_j$.

The null hypothesis was $H_{0,j}: \omega^2_j = 0$ against the alternative hypothesis was $H_{A,j}: \omega^2_j > 0$ for a RVSOM for observation j. Stram and Lee (27, 28) showed that the asymptotic null distribution of the test statistic for testing this type of hypothesis was a mixture of two chi-squared distributions on zero and one degree of freedom. However, Gumedze and Jackson (7) showed that for the RVSOM conditions it cannot be met; hence, following Gumedze et al. (29), we had to use a parametric bootstrap procedure to obtain the distribution of our test statistic.

3.4. Empirical Distribution of the LRT Statistic and Multiple Testing

Under the null hypothesis, when there are no outliers in the data, empirical distribution of the likelihood ratio test statistics by a parametric bootstrap procedure is as follows:

Step 1. Fit the null model (1) to the data to obtain estimates $\hat{\mu}$ and $\hat{\tau}^2$.
Step 2. Generate a new data vector from model (1) and estimates $\hat{\mu}$ and $\hat{\tau}^2$.
Step 3. Compute the likelihood ratio test statistics $LRT_j$, $j = 1, ..., n$, by fitting the model (2) to the simulated data and compute and save the order statistics of the set $LRT_j$ for $j = 1, ..., n$.
Step 4. Repeat steps 2 and 3 R times. This step generates an empirical distribution of size R for each order statistic.
Step 5. Calculate the 100 (1-α)th percentile for each order statistic for the required significance level α. The percentiles, using $\alpha = 0.05$ and $k = 1$ for largest order statistic, $k = 2$ for second largest order statistic, are shown in the plots given in the results.

4. Results

The data used in this study were taken from 14 clinical trial articles that investigated the effect of albendazole on patients with *Ascaris lumbricoides* infection; $y_i$ indicated the relative risk in the ith article. As the forest plot diagram (Figure 1) indicated, articles 4, 7 and 12 were different from the rest of studies. The results of RVSOM model are shown in Figure 2 and Table 1.

![Figure 1. Forest Plot Diagram Used to Investigate the Effect of Albendazole on Patients With Ascaris lumbricoides Infection](image-url)
Figure 2a shows the estimates $\omega_j^2$ from the jth RVSOM and the next two plots, Figures 2b and 2c, show the corresponding estimates of the between study variance and the treatment effect. The plot 2d shows the likelihood ratio statistics from which we see that observations 4, 7 and 12 are clearly detected as expected outliers; in particular, its LRT statistic is around three times the threshold for the first order statistic. All these figures refer to the fact that articles 4, 7 and 12 had served as outliers.

Figure 2. Random Effect Variance Shift Outlier Model Statistics Plotted Against Study Number for the Effect of Albendazole on Patients With Ascaris lumbricoides Infection

Table 1. Results From Clinical Trial Articles Investigating the Effect of Albendazole on Patients With Ascaris lumbricoides Infection

<table>
<thead>
<tr>
<th>Trial</th>
<th>Authors</th>
<th>Year</th>
<th>Albendazole</th>
<th>Placebo</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>A1+</td>
<td>A1-</td>
</tr>
<tr>
<td>1</td>
<td>Oyediran and Oyejide</td>
<td>1983</td>
<td>22</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>El-Masry et al.</td>
<td>1983</td>
<td>11</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>Bwibo and Pamba</td>
<td>1984</td>
<td>36</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>Ovedoff</td>
<td>1984</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>Chien et al.</td>
<td>1989</td>
<td>37</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>Upatham et al.</td>
<td>1989</td>
<td>74</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>Stephenson et al.</td>
<td>1990</td>
<td>7</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>Sinniah et al.</td>
<td>1990</td>
<td>51</td>
<td>5</td>
</tr>
<tr>
<td>9</td>
<td>Beach et al.</td>
<td>1999</td>
<td>61</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>Olds et al.</td>
<td>1999</td>
<td>179</td>
<td>40</td>
</tr>
<tr>
<td>11</td>
<td>Patrick P et al.</td>
<td>2009</td>
<td>63</td>
<td>25</td>
</tr>
<tr>
<td>12</td>
<td>J. Ndibazza et al.</td>
<td>2010</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>13</td>
<td>Speich B et al.</td>
<td>2012</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>Wiria AE et al.</td>
<td>2013</td>
<td>144</td>
<td>65</td>
</tr>
</tbody>
</table>

$^a$ A1+ refers to those recovered following the treatment and A1- refers to those who did not recover following the treatment.
Table 2. Estimated Parameters for Models Fitted to Investigate the Effect of Albendazole on Patients With Ascaris lumbricoides Infection

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Model M_0</th>
<th>95% CI</th>
<th>Model M_1</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu )</td>
<td>11.247376</td>
<td>(8.8,13.7)</td>
<td>10.29457445</td>
<td>(5.16,15.43)</td>
</tr>
<tr>
<td>( \tau )</td>
<td>20.083981</td>
<td>-</td>
<td>85.23284239</td>
<td>-</td>
</tr>
<tr>
<td>( \omega_j^2 )</td>
<td>-</td>
<td>-</td>
<td>140.53558626</td>
<td>-</td>
</tr>
<tr>
<td>( \omega_j^2 )</td>
<td>-</td>
<td>-</td>
<td>302.27428826</td>
<td>-</td>
</tr>
<tr>
<td>( \omega_{12}^2 )</td>
<td>-</td>
<td>-</td>
<td>0.01746554</td>
<td>-</td>
</tr>
</tbody>
</table>

*As lumbricoides data: Overall treatment effect (\( \mu \)), Variance shift estimates for \( j \)th the study (\( \omega_j^2 \)), and between-study variance (\( \tau^2 \)); M_0: Random effects; model M_1: Extended RVSOM for study 4, 7, and 12.*

Acknowledgements

This manuscript was the result of a master degree thesis in the Shahid Beheshti University of Medical Sciences. We appreciate the School of Paramedical Sciences head and the deputy manager of the research department.

Authors’ Contributions

All authors participated equally.

Financial Disclosure

There is no conflict of interest.

References


