Comparison of Methods of Handling Missing Data in Individual Patient Data Meta-analyses: An Empirical Example on Antibiotics in Children with Acute Otitis Media

Laura Koopman1, Geert J. M. G. van der Heijden1, Diederick E. Grobbee1, and Maroeska M. Rovers1,2

1 Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, the Netherlands.
2 Department of Otolaryngology, Wilhelmina Children’s Hospital, University Medical Center Utrecht, Utrecht, the Netherlands.

Received for publication March 28, 2007; accepted for publication October 25, 2007.

What is the influence of various methods of handling missing data (complete case analyses, single imputation within and over trials, and multiple imputations within and over trials) on the subgroup effects of individual patient data meta-analyses? An empirical data set was used to compare these five methods regarding the subgroup results. Logistic regression analyses were used to determine interaction effects (regression coefficients, standard errors, and p values) between subgrouping variables and treatment. Stratified analyses were performed to determine the effects in subgroups (rate ratios, rate differences, and their 95% confidence intervals). Imputation over trials resulted in different regression coefficients and standard errors of the interaction term as compared with imputation within trials and complete case analyses. Significant interaction effects were found for complete case analyses and imputation within trials, whereas imputation over trials often showed no significant interaction effect. Imputation of missing data over trials might lead to bias, because association of covariates might differ across the included studies. Therefore, despite the gain in statistical power, imputation over trials is not recommended. In the authors’ empirical example, imputation within trials appears to be the most appropriate approach of handling missing data in individual patient data meta-analyses.

imputation; meta-analysis; missing data; review [publication type]

Abbreviations: AOM, acute otitis media; IPD, individual patient data.

Individual patient data (IPD) meta-analyses, that is, meta-analyses that use IPD rather than simply the overall results of each trial, have been proposed as a major improvement in meta-analyses (1–3). As IPD meta-analyses generally include more detailed data, they usually have greater statistical power to carry out informative subgroup analyses. Moreover, IPD meta-analyses allow accurate classification of patients based on individual characteristics and may, therefore, allow a more thorough assessment as to whether subgroup differences are spurious or not (1–3). The assessment of subgroup effects is relevant for clinical practice, as most physicians would like to use the specific characteristics of a patient to decide on a patient’s individual treatment (4–6).

Missing data complicate the analyses of IPD meta-analyses, as in any study. For IPD meta-analyses, the same approaches for handling missing data might be used as in a single study. However, not only the frequency of missing data but also the missingness process may vary across studies from which individual data are pooled. Different methods of handling missing data may, therefore, have a different impact on the results of IPD meta-analyses.
Moreover, because of pooling in IPD meta-analyses, another type of missing data may occur in the pooled data set; namely, some variables might not be measured at all in a specific included trial. This type of missingness may further complicate the handling of missing data in IPD meta-analyses.

Conventional complete case analyses, that is, removing subjects with a missing value from the analyses, may not only reduce precision because only part of the data is used (7–9) but also produce biased results because the results may improve when missing data are imputed (7, 10). Common methods of imputation are single and multiple imputations. With single imputation, the available data of subjects without missing values in the study are used in a regression model to estimate the distribution of the variables for which values are missing (11). A random value of the estimated distribution will replace the missing values for the particular variable. With multiple imputations, regression techniques are used to estimate multiple distributions of the variable for which values are missing. Bootstrap techniques are used to draw a value from the estimated distributions to replace the missing value. Each missing value is, thus, imputed several times; consequently multiple data sets are created (11).

An essential difference between imputing data in a single study and imputation in IPD meta-analyses is that imputation in IPD meta-analyses can be performed within the data set of each trial before these data are pooled into one data set or for the final data set after pooling (over trials). In particular, when IPD are handled as if they belong to one trial, it might seem logical to impute missing data over trials. However, most published IPD meta-analyses used the so-called two-stage approach, where each trial is analyzed separately, by use of its raw data before the summary results from each trial are pooled and analyzed with conventional meta-analyses techniques (12). In this two-stage approach, imputation within trials might be most suitable. With imputation within trials, the variables not measured in specific trials are not imputed. With imputation over trials, all data are imputed; that is, variables that were not measured in a specific trial are imputed on the basis of estimates from other trials. Furthermore, imputation of missing data over trials will result in a gain in statistical power. However, the imputation of missing variables over trials might be biased because some variables might be associated with each other in one of the included studies, whereas this association may differ for the other studies. This might result in biased effect estimates (13).

To determine the best strategy to handle missing data in IPD meta-analyses, we explored the impact of various methods of handling missing data on the subgroup effects of IPD meta-analyses. Using empirical data, we compared complete case analyses, single imputation within trials, single imputation over trials, multiple imputations within trials, and multiple imputations over trials. Conventionally, significance (p < 0.05) of the interaction term between treatment and subgrouping variables is considered conditional for studying treatment effects stratified for these subgroups (14). Therefore, we assessed the impact of the five methods of handling missing data on the results of the interaction tests and the treatment effects in the stratified subgroup analyses.

**MATERIALS AND METHODS**

For this study, the data of an IPD meta-analysis were used, which evaluated the effect of antibiotics versus placebo or no treatment in children with acute otitis media (AOM) as described elsewhere (15). In our empirical example, the primary outcome measure was pain, fever, or both at 3–7 days, and age, bilateral AOM, and otorrhea were the subgrouping variables.

**Imputation techniques**

Single (conditional mean) imputation was performed by use of the Missing Value Analysis function available in SPSS (SPSS for Windows, version 14.0; SPSS, Inc., Chicago, Illinois). This method fits a prediction model for each variable with a missing value, the variable with a missing value is the outcome, and all other variables (i.e., all measured covariates, a variable for study, and the outcome variable (15)) are included as predictors. Missing values are replaced by estimates resulting from the prediction model (10, 11).

Multiple imputation was done by use of the aregImpute algorithm (16) in S-PLUS (S-PLUS for Windows, version 6.2; Lucent Technologies, Inc., Murray Hill, New Jersey). aregImpute is a technique that uses additive regression, bootstrapping, and predictive mean matching for multiple imputation. Bootstrap techniques are used to impute missing data by drawing predicted values from a full Bayesian predictive distribution. Different bootstrap resamples are used for each of the multiple imputations, in which a flexible additive regression model is fitted on a sample with replacement from the original data. This model takes the uncertainty in the imputations caused by having to fit imputation models into account and is used to predict all of the original missing and nonmissing values for the target variable. Thereby, aregImpute uses predictive mean matching with optional weighted probability sampling (17, 18). The same variables, used as predictors in the single imputation process, were used for the multiple imputation process. The imputation process was repeated five times. Consequently five data sets were created.

Since the dichotomous variables were coded as 0 or 1, the imputed values of these variables were rounded to 0 or 1, and the imputed values of continuous variables were rounded to the nearest observed integer.

Although it is likely that a different process gives rise to missing data for each study, we assume similarity of the missingness process across studies.

**Subgroup analyses**

Fixed logistic regression analyses, including a dummy variable for study, were used to determine the interaction effect, that is, the regression coefficient (β), standard error, and p value of the interaction term: subgrouping variable × treatment. The interaction effects investigated were
age $\times$ treatment, bilateral AOM $\times$ treatment, otorrhea $\times$ treatment, and age $\times$ bilateral AOM $\times$ treatment. Stratified subgroup analyses were performed to determine the treatment effects in the subgroups, that is, risk difference, relative risk, and their 95 percent confidence intervals.

The multiple imputation process both within and over trials resulted in five data sets. We analyzed these data sets separately. To combine the results of the interaction tests (i.e., $\beta$ and standard error of the interaction terms) and the stratified subgroup analyses (i.e., relative risk and risk difference) for the five data sets, we used the formulas of Rubin as described by Schafer (9) to combine point estimates, taking into account the variance of the estimates within and between studies.

To examine the influence of the various methods to handle missing data, we compared the interaction effects and the subgroup effects among complete case analyses, single imputation within and over trials, and multiple imputations within and over trials.

RESULTS

In table 1, the distribution of missing data of subgrouping variables and outcomes is presented. Of those variables that were available in all trials, only 3 percent were missing. These missing variables could be imputed both within and over trials. The variables bilateral AOM and otorrhea were not measured in one ($n = 316$) and four ($n = 1,118$) trials, respectively. That is, for 19 percent and 66 percent of all the cases, bilaterality and otorrhea were missing. These missing variables could be imputed only over trials.

For the complete case analyses, information on otorrhea, bilateral AOM, age, and pain or fever or both at 3–7 days was available for 32, 81, 100, and 97 percent, respectively. After single and multiple imputations within trials, information on otorrhea, bilateral AOM, age, and pain or fever or both at 3–7 days was available for 34, 81, 100, and 100 percent, respectively. Expectedly, after single and multiple imputations over trials, the percentage of available information increased to 100 percent for all variables.

Results of interaction tests

The results of the interaction test (i.e., regression coefficient ($\beta$), standard error, and $p$ value of the interaction terms) differed among the five methods of handling missing data (table 2). After single imputation within trials, the $\beta$ and standard error of the interaction terms were comparable to those of the complete case method. After multiple imputation within trials, the standard errors of the interaction terms of both “bilateral AOM $\times$ treatment” and “otorrhea $\times$ treatment” were slightly larger than those of the complete case analyses, whereas the $\beta$ and standard error for the combined subgrouping variable age and bilateral AOM and the standard error for age were smaller.

After single imputation over trials, the $\beta$ of the interaction term “age $\times$ treatment” was larger than the $\beta$'s of the complete case analyses and imputation within trials. Both the $\beta$'s of the other interaction terms and the standard error of otorrhea were smaller after single imputation over trials.

TABLE 1. Distribution of variables and missing data in the data sets after five methods of handling missing data.

<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>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Aged &lt;2 years</td>
<td>286</td>
<td>35</td>
<td>280</td>
<td>34</td>
<td>5</td>
<td>0</td>
<td>287</td>
<td>35</td>
<td>280</td>
<td>34</td>
<td>0</td>
<td>287</td>
<td>35</td>
<td>280</td>
<td>34</td>
<td>0</td>
<td>287</td>
<td>35</td>
</tr>
<tr>
<td>Bilateral AOM</td>
<td>220</td>
<td>33</td>
<td>236</td>
<td>35</td>
<td>316</td>
<td>19</td>
<td>220</td>
<td>33</td>
<td>236</td>
<td>35</td>
<td>315</td>
<td>19</td>
<td>220</td>
<td>33</td>
<td>236</td>
<td>35</td>
<td>315</td>
<td>19</td>
</tr>
<tr>
<td>Otorrhea</td>
<td>61</td>
<td>22</td>
<td>46</td>
<td>18</td>
<td>1,118</td>
<td>68</td>
<td>65</td>
<td>23</td>
<td>51</td>
<td>19</td>
<td>1,088</td>
<td>66</td>
<td>63</td>
<td>22</td>
<td>50</td>
<td>19</td>
<td>1,088</td>
<td>66</td>
</tr>
<tr>
<td>Pain and/or fever at 3–7 days</td>
<td>293</td>
<td>37</td>
<td>189</td>
<td>24</td>
<td>50</td>
<td>0</td>
<td>303</td>
<td>37</td>
<td>194</td>
<td>24</td>
<td>0</td>
<td>300</td>
<td>36</td>
<td>194</td>
<td>24</td>
<td>0</td>
<td>306</td>
<td>37</td>
</tr>
</tbody>
</table>

* The covariates included in the imputation model were age, gender, treatment, smoking in the household, having been breastfed, family history of AOM (AOM in the household, history of acute ear infections, and history of recurrent AOM), ear, nose, or throat surgery, fever at baseline, bilateral AOM at baseline, red tympanic membrane at baseline, pain at baseline, otorrhea at baseline, crying at baseline, cough at baseline, vomiting at baseline, perforation at baseline, and fever or ear pain or both at 3–7 days.
as compared with complete case analyses and imputation within trials. After multiple imputation over trials, the standard error of the interaction term “bilateral AOM × treatment” was larger than the standard error of the complete case analyses and imputation within trials, whereas the β’s of all interaction terms and the standard error of age and bilateral AOM combined were considerably smaller.

The interaction effects of “bilateral AOM × treatment,” “age × bilateral AOM × treatment,” and “otorrhea × treatment” were significant (p < 0.05) for both the complete case analyses and single imputation within trials. After multiple imputation within trials, the interaction effects reached borderline significance (p = 0.05–0.1). After imputation over trials, only a significant effect (p = 0.05) was found for bilateral AOM and age combined when single imputation over trials was used. For multiple imputations over trials, none of the interaction terms was significant.

Results of stratified subgroup analyses

Only small differences were found among the five methods of handling missing data regarding the relative risk, risk difference, and their 95% confidence intervals of the stratified subgroup analyses (table 3); that is, the conclusions with respect to the clinically relevant subgroups remained similar for all methods. The stratified subgroup analyses showed relevant treatment effects for the subgroups bilateral AOM, age and bilateral AOM combined, and otorrhea for all methods. For example, the risk differences for children aged less than 2 years with bilateral AOM were −27, −25, −26, −25, and −25 percent for complete case analyses, single imputation within, single imputation over, multiple imputations within, and multiple imputations over trials, respectively.

DISCUSSION

We examined the effect of different methods of handling missing data on the subgroup results in an IPD meta-analysis. The results of the five methods of handling missing data showed some remarkable differences with respect to the regression coefficients, standard errors, and p values of the interaction terms. However, no clinically relevant differences were found among the five methods regarding the subgroup effects.

For appreciation of our findings, some aspects deserve further discussion. First, it is remarkable that the interaction results differ among the various approaches. Despite the increased power, the interaction terms “bilateral AOM × treatment” and “otorrhea × treatment” were not significant after single and multiple imputation over trials, whereas the interaction term “age × bilateral AOM × treatment” was not significant after multiple imputation over trials. According to current recommendations (14), stratified subgroup analyses would not have been permitted for these subgroups. Subsequently, the clinically relevant subgroup effects found in the stratified analyses would have been missed.

Second, because some covariates might be associated with each other in one of the included studies yet this association may differ for the other studies, imputation over trials might lead to bias. Our results showed that the distribution of bilateral AOM was associated with age; that is, younger children were more frequently diagnosed with bilateral AOM. We could not, however, find any changes in this association after imputation. We did not, therefore, correct for this bias. However, other factors that were not measured in the trials might be associated with variables in the trials and might have introduced bias.

Third, it may seem illogical that single imputation performed better than multiple imputations in our analyses. Single imputation commonly results in an overestimation of the precision of the study associations because too low estimates of the standard error are obtained, while correct estimates of the standard error are obtained with multiple imputations (7, 11). However, certain methodological problems became apparent during the multiple imputation process. The variable “study” could not be included in the multiple imputation process; because of the complete
<table>
<thead>
<tr>
<th>Subgrouping variable</th>
<th>Complete cases</th>
<th>Imputation within trials</th>
<th>Multiple imputations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Single imputation</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Imputation over trials</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Imputation within trials</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Multiple imputations</td>
<td></td>
</tr>
<tr>
<td></td>
<td>No. Risk difference 95% confidence interval Relative risk 95% confidence interval</td>
<td>No. Risk difference 95% confidence interval Relative risk 95% confidence interval</td>
<td>No. Risk difference 95% confidence interval Relative risk 95% confidence interval</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;2 years</td>
<td>541 –16 –24, –7 0.67 0.46, 0.89</td>
<td>567 –15 –23, –7 0.68 0.47, 0.89</td>
<td>567 –15 –23, –7 0.67 0.40, 0.90</td>
</tr>
<tr>
<td>≥2 years</td>
<td>1,047 –11 –17, –6 0.63 0.42, 0.85</td>
<td>1,076 –11 –16, –6 0.64 0.43, 0.85</td>
<td>1,076 –11 –16, –6 0.64 0.43, 0.85</td>
</tr>
<tr>
<td>Bilateral AOM*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>840 –6 –12, 0 0.79 0.57, 1.02</td>
<td>872 –6 –12, 0 0.80 0.58, 1.02</td>
<td>872 –5 –11, 0 0.82 0.53, 1.10</td>
</tr>
<tr>
<td>Yes</td>
<td>437 –22 –30, –13 0.55 0.29, 0.81</td>
<td>456 –20 –29, –11 0.57 0.32, 0.83</td>
<td>456 –21 –30, –13 0.55 0.29, 0.80</td>
</tr>
<tr>
<td>Bilateral AOM and age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unilateral, &lt;2 years</td>
<td>249 –4 –16, 8 0.90 0.57, 1.22</td>
<td>261 –5 –17, 6 0.87 0.55, 1.18</td>
<td>261 –3 –15, 8 0.91 0.48, 1.33</td>
</tr>
<tr>
<td>Unilateral, ≥2 years</td>
<td>591 –7 –14, 0 0.73 0.43, 1.04</td>
<td>611 –6 –13, 0 0.76 0.46, 1.06</td>
<td>611 –6 –13, 0 0.76 0.46, 1.06</td>
</tr>
<tr>
<td>Bilateral, &lt;2 years</td>
<td>259 –27 –39, –15 0.52 0.21, 0.83</td>
<td>273 –25 –36, –14 0.55 0.25, 0.84</td>
<td>273 –26 –38, –15 0.52 0.21, 0.82</td>
</tr>
<tr>
<td>Bilateral, ≥2 years</td>
<td>177 –13 –26, 0 0.64 0.16, 1.12</td>
<td>183 –12 –25, 1 0.65 0.18, 1.12</td>
<td>183 –14 –27, 0 0.63 0.14, 1.11</td>
</tr>
<tr>
<td>Otorrhea</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>412 –15 –24, –6 0.66 0.40, 0.92</td>
<td>439 –14 –23, –5 0.66 0.40, 0.93</td>
<td>441 –14 –23, –5 0.66 0.37, 0.95</td>
</tr>
<tr>
<td>Yes</td>
<td>106 –37 –55, –20 0.41 0.00, 0.94</td>
<td>116 –36 –53, –20 0.39 0.00, 0.93</td>
<td>113 –37 –54, –21 0.39 0.00, 0.94</td>
</tr>
</tbody>
</table>

* AOM, acute otitis media.
missingness of variables in some studies, the imputation algorithm could not determine what the distribution of the missing variable should have been in these studies when the variable study was included. Subsequently, it was not possible to take into account differences in distributions of variables according to study. Variables strongly skewed toward 0 or 1 could not be included in the imputation process for the same reason. To explore the influence of study in the single imputation process, we performed sensitivity analyses both with and without study. Because the results of imputation with or without study gave similar results, we presented only the results without study. Furthermore, combining the results of the analyses of five data sets after multiple imputations appeared to be laborious and time consuming. These methods should therefore be simplified and improved before they can be applied easily in IPD meta-analyses.

Fourth, to study the influence of two different adjustment methods, we have performed some additional analyses. First, we have added the covariate “study” to the regression analyses, as this is the best method to adjust for residual confounding. The results of these analyses were, however, in agreement with the earlier results. Second, we have analyzed a fully saturated model. The results of these analyses were also in agreement with the earlier findings; notably, there was virtually no difference in the regression coefficients of treatment and the interaction term, whereas the standard errors increased. It should be noted, however, that the odds ratios of the regression model are not the most relevant effect estimates in clinical practice. We, therefore, presented the stratified analyses without adjustment.

Finally, in the context of this empirical study, there is no “gold” standard available. Simulation studies are needed to further refine the relation between number and nature of missing values and to compare the results with a gold standard.

In conclusion, in our empirical example, imputation within trials appears to be the most appropriate approach of handling missing data in IPD meta-analyses. Despite a gain in statistical power, imputation over trials is not recommended, because it might lead to bias when associations between covariates differ across the included studies.

ACKNOWLEDGMENTS

This work was supported by the Netherlands Organization for Scientific Research (grant 916.46.090).

The authors thank Cees L. Appelman, Peter Burke, David P. McCormick, Roger A. Damoiseaux, Isabella Gaboury, and Paul Little for providing their data. They also thank Kristel J. M. Jansen for providing the S-PLUS script for multiple imputation.

Conflict of interest: none declared.

REFERENCES

2. Oxman AD, Clarke MJ, Stewart LA. From science to practice; meta-analyses using individual patient data are needed. JAMA 1995;274:845–6.