Quality & Quantity **38:** 381–389, 2004.

Longitudinal Meta-analysis

CORA J. M. MAAS*, JOOP J. HOX and GERTY J. L. M. LENSVELT-MULDERS
Department of Methodology and Statistics, Faculty of Social Sciences, Utrecht University, POB 80140, NL-3805 TC Utrecht, the Netherlands

Abstract. The goal of meta-analysis is to integrate the research results of a number of studies on a specific topic. Characteristic for meta-analysis is that in general only the summary statistics of the studies are used and not the original data. When the published research results to be integrated are longitudinal, multilevel analysis can be used for the meta-analysis. We will demonstrate this with an example of longitudinal data on the mental development of infants. We distinguish four levels in the data. The highest level (4) is the publication, in which the results of one or more studies are published. The third level consists of the separate studies. At this level we have knowledge about the degree of prematurity of the group of infants in the specific study. The second level are the repeated measures. We have data about the test age, the mental development, the corresponding standard deviations, and the sample sizes. The lowest level is needed for the specification of the meta-analysis model. Both the way in which the multilevel model has to be specified (the Mln-program is used) as the results will be presented and interpreted.

Key words: longitudinal analysis, meta-analysis, multilevel analysis.

1. Introduction

In the social and behavioural sciences, research results are often inconsistent. Human behaviour is complex and difficult to explain. Depending on characteristics of the studies, such as samples, contextual variables and operationalizations, different conclusions may be drawn on the same research topic. The aim of meta-analysis (Glass, 1976; Light and Pillemer, 1984; Lipsey and Wilson, 2001) is the integration of such different research results and the explanation of the prevailing inconsistencies. Characteristic of the meta-analysis procedure is that in general only summary statistics, as p-values, correlations or standard deviations, of the studies are used and not the original data. So, if the research question is about the correlation between two variables, meta-analysis is used to combine the correlations as found in the different studies to one 'overall correlation'. Also the deviations of the individual studies to this overall correlation are estimated. When the deviations are

* Correspondence concerning this article should be sent to Cora Maas, Department of Methodology and Statistics, University of Utrecht, Heidelberglaan 1, 3584 CS Utrecht, The Netherlands. Phone +31 30 2534594. Fax +31 30 253 5797. Electronic mail may be sent to c.maas@fss.uu.nl

substantial, characteristics of the studies can be used to explain these deviations (Cornell and Mulrow, 1999).

A simple approach to meta-analysis consists of combining the p-values of a number of studies into one global p-value. Methods for this are well-known (cf. Hedges and Olkin, 1985; Schwarzer, 1989). A more sophisticated approach is to estimate in each study an *effect size* for the outcome, and to combine the effect sizes into one global effect size, plus a significance test for the combined effect. Implicit in combining effects from different studies is the assumption that all studies estimate the same effect in an identical way; it is assumed that the effects are homogeneous across the studies. Part of a meta-analysis is a test of this assumption. If the homogeneity test is significant, the assumption of homogeneity is rejected, and the effects must be assumed heterogeneous. In the presence of heterogeneous effects, there is no common study effect. There is, of course, an average study effect, and variance of the study outcomes around that average. The next step is explaining this variance using study characteristics. In classical meta-analysis (cf. Hedges and Olkin, 1985), the approach towards explaining between-studies variance is to cluster the studies into homogeneous clusters, followed by a post hoc explanation of the cluster structure.

A very general meta-analysis model is the random effects model (cf. Hedges and Olkin, 1985, p. 189). The random effects model assumes that the study results differ because of sampling variation, plus additional variation that reflects real differences between the studies. Hedges and Olkin propose to use weighted least squares regression to analyze the differences between the studies. Raudenbush and Bryk (Raudenbush and Bryk, 1985; Raudenbush, 1994) point out that random effects meta-analysis is closely related to multilevel analysis. In meta-analysis, we have two nested levels of observations: studies, and individuals within studies. If we had access to the original data, we could perform a standard multilevel analysis on these data, including the study characteristics as explanatory variables at the study level. However, generally we do not have access to the original data. In this case, we use an adapted multilevel procedure, which uses only the sufficient statistics available from the publications (cf. Goldstein, 1995). The flexibility of multilevel analysis allows us to include not only study characteristics, but also certain design characteristics, such as the occurrence of repeated measures. In the remainder of this article, we show how multilevel analysis can be used to carry out a meta-analysis of longitudinal data.

2. The Model

A classical meta-analysis model is the random effects model, as formulated by Hedges and Olkin (1985). In this model, the different research results are the result of not only random sampling fluctuations within the studies, but also of systematic differences between the studies. The specification of this model is:

$$O_j = \mu_0 + r_j + e_j \tag{1}$$

where O_j is the research result of study j, μ_0 is the mean effect of all studies, r_j is the residual prediction error, and e_j is the sampling error.

The random effect meta-analysis model specified in Equation (1) (cf. Hedges and Olkin, 1985, p. 189) is formally equivalent to the multilevel intercept only model, presented below in the usual notation (cf. Bryk and Raudenbush, 1992):

$$d_j = \gamma_0 + u_j + e_j \tag{2}$$

The variance of the residual prediction error (σ_j^2) reflects the heterogeneity of the research results. The null hypothesis that this variance is zero, is identical to the null hypothesis tested in the usual homogeneity tests in meta-analysis. When the null hypothesis is rejected, attempts can be made to explain the observed variance with characteristics of the studies. Adding explanatory variables at the study level the model becomes:

$$d_j = \gamma_0 + \gamma_1 Z_{1j} + \gamma_2 Z_{2j} + \cdots + \gamma_p Z_{pj} + u_j + e_j \tag{3}$$

where d_j is the research result of study j, γ_0 is the regression intercept of all studies, γ_1 to γ_p are regression coefficients, Z_1 to Z_p are characteristics of the studies, u_j is the residual prediction error, and e_j is the sampling error.

When the published research results to be integrated are longitudinal, and the studies only publish the effects at the different time points (meaning no growth curves), multilevel analysis can be used for the meta-analysis. An advantage of the multilevel meta-analysis is, that characteristics of the study level can be entered as explanatory variables in the regression equation, while in the meta-analysis procedure only clusters of studies, with comparable characteristics, can be distinguished. An additional advantage is that multilevel meta-analysis does not assume that all studies report on the same time points, in fact, all time points may be different.

3. The Specification of a Longitudinal Meta-analysis Model

Starting-point for the specification of the longitudinal meta-analysis model is the 'normal' longitudinal model. If we, just for a moment, consider the outcome of a study as observed (and not as a series of means of a group with a specific standard deviation), we can specify a simple longitudinal model as follows. The 'observed score' (Y_{ts}) of Study s at timepoint t is a linear function of the time plus random error. Therefore the lowest level is the level of the repeated measures, and this level 1 model is specified as follows:

$$Y_{ts} = \pi_{0s} + \pi_{1s} time_{ts} + e_{ts} \tag{4}$$

where Y_{ts} is the observed score at timepoint t in study s, π_{0s} is the intercept of study s, $time_{ts}$ is the timepoint t in study s, π_{1s} is the effect of the time in study s on the observed score, and e_{ts} is the random time effect, that is the deviation of observed scores from the study mean.

If we specify no explanatory variables at the second level (level of the study) the specification of the second level becomes:

$$\pi_{0s} = \beta_{00} + r_{0s} \tag{5}$$

$$\pi_{1s} = \beta_{10} + r_{1s} \tag{6}$$

where π_{0s} is the intercept of study s, π_{1s} is the effect of time in study s on the observed score, β_{00} is the intercept, β_{10} is the effect of time on the observed score, r_{0s} is the random study effect of the intercept, and r_{1s} is the random study effect of time on the observed score.

These equations can be written as one single regression equation by substituting the Equations (5) and (6) into (4):

$$Y_{ts} = \beta_{00} + \beta_{10} time_{ts} + r_{0s} + r_{1s} time_{ts} + e_{ts} \tag{7}$$

The difference between the longitudinal multilevel model specified in Equation (7), and the longitudinal meta-analysis model, is that in the latter we have not access to the original observed scores but only to the means, standard deviations and sample size of the observed scores from each study. However, assuming normal distribution, these are sufficient statistics that contain all information in the sample. With respect to the specification of the model, the consequence is that the response variable Y_{ts} represents the mean of the observed scores in study s, and that the error variance of the within study sampling errors e_{ts} in Equation (7) is assumed known. Assuming normality, the standard error of a mean is estimated by the well-known formula:

$$SE = \frac{SD_{ts}}{\sqrt{n_{ts}}} \tag{8}$$

In this equation, SD and n are known quantities, so the standard error SE and the corresponding sampling variance $V = SE^2$ are known. Thus, the sampling variance of the means in each study is produced by:

$$V_{ts} = \frac{SD_{ts}^2}{n_{ts}} \tag{9}$$

The unknown parameters in Equation (7) are the fixed regression coefficients β_{00} and β_{10}, and the study-level variances V_{0s} of the r_{0s} and V_{1s} of the r_{1s} terms. The null hypothesis for the fixed regression coefficients, which correspond to tests of specific study characteristics, is that they are equal to zero. This is tested by a t-ratio formed by the ratio of the estimated coefficient to its standard error. When the number of level two units is large, this ratio follows a standard normal distribution, otherwise a Student distribution can be used (cf. Bryk and Raudenbush, 1992). The study-level variances V_{0s} and V_{1s} can also be tested for significance. The usual

Maximum Likelihood procedure includes an asymptotic test based on the standard error of V computed from the inverse of the information matrix (Goldstein, 1995). Bryk and Raudenbush (1992, page 55) prefer a χ^2-test using the standardised OLS-residuals, where:

$$\chi^2 = \sum_{s=1}^{S} \left(\frac{(Y_{ts} - \hat{Y}_{ts})}{SE} \right)^2 \tag{10}$$

with: $df = s - q - 1$ (q = number independent variables).

The Bryk and Raudenbush test has the advantage that in the intercept only model it is equivalent to the usual homogeneity test employed in meta-analysis (Hedges and Olkin, 1985). The problem with this test is that it is only available in the program (VK)HLM developed by Bryk and Raudenbush (Bryk, Raudenbush and Congdon, 1994), and not in the program MLn developed by Goldstein (Rasbash and Woodhouse, 1995). However, it is possible to carry out multilevel meta-analysis using standard multilevel software. Since this approach is more general, we will restrict ourselves to the results of the general multilevel analysis using the program MLn (Rasbash and Woodhouse, 1995).

The model parameters β_{00}, β_{10}, V_{0s} and V_{1s} can be estimated using multilevel software, provided that it is possible to put constraints on the variances in the random part of the model. Specifically, the variance of the e_{ts}, which is assumed known, is specified by including the standard error as given in Equation (8) at the lowest level as a predictor variable. The standard error is used as a predictor in the random part only:

$$Y_{ts} = \beta_{00} + \beta_{10} time_{ts} + r_{0s} + r_{1s} time_{ts} + SD_{ts} x\, e_{ts} \tag{11}$$

Using the explanatory variable SD and constraining the level one variance to one, we obtain the required sampling variance (Goldstein, 1995, p. 98). The estimation of the model parameters β_{00}, β_{10}, V_{0s} and V_{1s}, which are specified at the study level, then proceeds in the usual way.

4. Example

We demonstrate the specification and interpretation of a longitudinal meta-analysis multilevel model, using an example from a longitudinal meta-analysis of the development of infants. The data used in the example are so-called Bayley-scores, which are standardized scores on the mental development of infants. We have access to 28 articles, in which 40 studies are reported, with a total of 67 points of time. The articles are published in the period between 1973 and 1990. The difference with the above specified multilevel model and the data used in this example is that we now have to specify a four level model. We need three levels for substantive reasons, because we have not only longitudinal data nested within studies, but also

an additional level of studies nested within in articles. Therefore the longitudinal model, without other explanatory variables, becomes:

$$Y_{tsa} = \gamma_{000} + \gamma_{100} Age_{tsa} + r_{0sa} + r_{1sa} Age_{tsa} + u_{00a} + SD_{tsa} x\, e_{tsa} \qquad (12)$$

For the specification of the meta-analysis model we need another level. This level exists for technical reasons; at this level we constrain the estimate of the variance of the random sampling errors to equal 1. Multiplying this estimate with the predictor that contains the standard errors, gives us the desired 'known' variances of the first level. The specification of the model in Mln is given in Appendix 1.

The results of the intercept-only model show that there is no variance on the level of the studies. This means that studies in one publication are exchangeable; they are more alike than studies reported in different articles. As a result, we may collapse our data over this level. Leaving out the study level, which means that the variables observed at the study level are now treated as if they were observed at the time-points level, gives as estimate for the mean Bayley-score 101.7. The total variance is decomposed into two parts, 74.65% is at the study level, 25.35% at the time-points level. The study level variance is clearly significant ($p < 0.00$). In addition to the significance test, Hunter and Schmidt (1990) suggest that the between study variance should be considered important when it is at least 25% of the total variance. Therefore, the conclusion is that there are substantial differences between the studies. Adding the age (in months) of the infants (centred around the grand mean) gives the following equation:

$$Bayley = 101.7 + 0.2551 Age$$

We should expect no effect of the variable '*Age*', because the Bayley-scores are standardized for '*Age*'. Nevertheless, the results show a positive effect of '*Age*'. This means that premature infants perform less well. Adding the predictor variable '*Prematurity*' (also centred about the grand mean) to the model gives the following:

$$Bayley = 97.37 + 0.18 Age - 9.509 Prematurity$$

The effect of '*Age*' is no longer significant. The more an infant is premature the lower his Bayley-score. Adding the interaction of '*Age*' and '*Prematurity*' to the model gives the following result:

$$Bayley = 97.11 + 0.07204 Age - 9.593 Prematurity + 0.2891 Age\ x\ Prematurity$$

The effect of '*Age*' is again not significant. The interpretation of the regression equation is that the less premature the higher the Bayley-score, but when the premature infants become older they make up their backwardness. The variance explained by the variables in the last model is 83.49% at the study level, and at the level of the time points 6.02%.

There is one variable at the article level: the year of publication. This variable is treated as a dummy variable. Articles published between 1973 and 1980 have score 0, and articles between 1981 and 1990 score 1. When the year of publication is missing, the mean of the dummy variable is imputed. We have checked the possibility, that the studies with the missing 'publication year' are different from the others, by adding a second dummy variable to the regression equation, which has the score 1 for the 'missing year code', and the score 0 for the studies with known publication year. Both the 'year of publication' variable, and the missingness dummy variable are not significant.

5. Discussion

The advantages of multilevel analysis for longitudinal meta-analysis are twofold. The first advantage is that the multilevel model and the accompanying software are very flexible, which means that many potential explanatory variables may be added to the model at various levels. These explanatory variables may have a substantive meaning, such as the *Age* and *Prematurity* variables used in our example, or serve as methodological control variables, like the publication year and missingness variables. The second advantage is that the multilevel model does not assume that the time points are the same in all studies, which means that there are very few restrictions to the studies to be added to the analysis.

For multilevel longitudinal meta-analysis it is required that each study publishes at least the actual time points, the sample size, and the mean and standard deviation of the main variable at each time point. Most studies in our example either published these, or gave information from which these statistics could be deduced. Meta-analysis methods are so far mostly used to combine the results of a set of studies that are already published. However, they can also be fruitfully applied in co-operative studies, where a number of research groups plan a set of longitudinal studies, with the aim of combining the results at a later stage. In such cases, the choice of variables and study design can be made before the meta-analysis, and the problems of integrating the results can be anticipated and solved in the planning stage.

Appendix 1

MLn commands to carry out multilevel longitudinal meta-analysis on Bayley scores.
(1) name c1 'article' c2 'study' c3 'testage' c4 'Bayley' c5 'SE' c6 'prem' c7 'year' c8 'dumyear' c9 'const'
(2) iden 4 'article' 3 'study' 2 'testage' 1 'const'
(3) resp 'Bayley'
(4) expl 'cons' 'SE'
(5) fpar 'SE'

(6) setv 1 'SE'
(7) setv 2 'const'
(8) setv 3 'const'
(9) setv 4 'const'
(10) input c11
(11) 0 0 0 1 1
(12) finish
(13) rcon c11

Explanation

(1): There are 9 variables in the data-file named article, study, testage, Bayley, SE, prematurity, year, dumyear and the constant
(2): There are 4 levels distinguished: the highest level is the article. In the articles are the results of one or more studies published, so the third level consists of the separate studies. The second level are the longitudinal data. The lowest level is needed for the specification of the meta-analysis model
(3): The dependent variable is the Bayley-score
(4): The independent variables are the constant and the standard error
(5): The standard error is only a predictor in the random part so we remove this variable from the fixed part
(6)–(9): The total variance of the 'Bayley-scores' is distributed over the three levels. The variance of the first level is constrained to one, so that the variance of the second level has the known variance of the 'standard error'
(10)–(13): To constrain the estimate of the random parameter at the first level equal 1, we specify a constraints vector: 0 0 0 1 1. The first three zeros are for the random parameters of the second, third en fourth level. These parameters have no constraints. The zeros in the fourth and fifth column specify the constrain of the estimate of the random parameter at the first level equal to 1

References

Bryk, A. S. & Raudenbush, S. W. (1992). *Hierarchical linear models*. Newbury Park, CA: Sage.
Bryk, A. S., Raudenbush, S. W. & Congdon, R. T. (1994). *HLM 2/3. Hierarchical linear modeling with the HLM/2L and HLM/3L programs*. Chicago: Scientific Software International.
Cornell, J. & Mulrow, C. (1999). Meta-analysis. In: H. J. Ader and G. J. Mellenbergh (eds.), *Research methodology in the social, behavioral, and life sciences*. London: Sage.
Glass, G. V. (1976). Primary, secondary and meta analysis of research. *Educational Researcher* 10: 3–8.
Goldstein, H. (1995). *Multilevel statistical models*. London: Edward Arnold.
Hedges, L. V. & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic Press.
Hunter, J. E. & Schmidt, F. L. (1990). *Methods of meta-analysis*. Newbury Park, CA: Sage.
Light, R. J. & Pillemer, D. B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.

Lipsey, M. W. & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
Rasbash, J. & Woodhouse, G. (1995). *MLn command guide*. London: Multilevel Models Project, University of London.
Raudenbush, S. W. (1994). Random effects models. In: H. Cooper & L. V. Hedges (eds.), *The handbook of research synthesis*. New York: Russell Sage Foundation.
Raudenbush, S. W. & Bryk, A. S. (1985). Empirical Bayes meta analysis. *Journal of Educational Statistics* 10: 75–98.
Schwarzer, R. (1989). Meta-analysis programs. Program manual. Berlijn: Institüt für Psychologie, Freie Universität Berlin.