Chapter 6

Meta-Analysis and Meta-Regression: An Alternative to Multilevel Analysis When the Number of Countries Is Small

Aart C. Liefbroer and Mioara Zoutewelle-Terovan

6.1 Introduction

Cross-national comparative research often is based on the analysis of hierarchically nested data structures containing information on multiple levels. A common situation integrates data at two levels, with micro-level (level-1) information about individuals and macro-level (level-2) information about countries. In the social sciences, the most popular way of analyzing such hierarchical cross-national data is by means of multilevel techniques (Hox et al. 2010; Rabe-Hesketh and Skrondal 2008; Snijders and Bosker 1999).

Multilevel analysis is very effective when dealing with data at multiple levels because it allows the estimation of effects occurring at all these levels (e.g. individual effects, country effects) simultaneously, as well as the estimation of interactions between variables at different levels (cross-level effects). A multilevel model may have the following structure:

$$y_{ic} = X_{ic}\beta + Z_c\gamma + X_{ic} * Z_c\theta + e_{ic} + u_c$$

A. C. Liefbroer (✉)

Netherlands Interdisciplinary Demographic Institute (NIDI), The Hague, The Netherlands

Department of Epidemiology, University Medical Centre Groningen, University of Groningen, Groningen, The Netherlands

Department of Sociology, Vrije Universiteit, Amsterdam, The Netherlands

e-mail: liefbroer@nidi.nl

M. Zoutewelle-Terovan

Netherlands Interdisciplinary Demographic Institute (NIDI), The Hague, The Netherlands

University of Groningen, Groningen, The Netherlands

e-mail: zoutewelle@nidi.nl

© The Author(s) 2021

A. C. Liefbroer, M. Zoutewelle-Terovan (eds.), *Social Background and the Demographic Life Course: Cross-National Comparisons*,
https://doi.org/10.1007/978-3-030-67345-1_6

101

where the outcome y_{ic} (for person i in country c) depends on observed individual characteristics (X_{ic}), observed country-level characteristics (Z_c), cross-level interactions between observed characteristics at the individual and country level ($X_{ic} * Z_c$), unobserved individual effects (e_{ic}) and unobserved country effects (u_c), under the assumption that unobserved effects are normally distributed and uncorrelated with observed effects.

Nowadays, most software packages offer a broad suite of multilevel models that are easy to use by social scientists. However, one considerable problem in estimating multilevel models concerns a low number of level-2 observations in a sample. For instance, many multi-country datasets contain large numbers of individuals per country (often hundreds or thousands), but include only a small number of countries (often less than 30 or even less than 20). With few level-2 units, the use of multilevel models may result in unreliable inferences because of biased estimates (coefficients and variance components) and inaccurate (often underestimated) standard errors (Arend and Schafer 2019; Austin 2010; Bell et al. 2010; Bryan and Jenkins 2016; Hox 1998; Maas and Hox 2004; McNeish and Stapleton 2016; Van der Leeden et al. 2008). When an increase in level-2 units is not feasible, one could consider alternative analytical tools such as the implementation of completely different analyses techniques or the use of multilevel simulation-based models able to surmount estimation bias and provide accurate statistical tests (Bryan and Jenkins 2016; Goldstein 2011; Hamaker and Klugkist 2010; Maas and Hox 2004; McNeish and Stapleton 2016). Yet, many of these techniques may not be feasible for various substantive or practical reasons.

This paper discusses aspects related to the analysis of nested data structures with a small level-2 sample size in the context of cross-national research. Specifically, we will discuss several alternative analytical tools one can apply to overcome problems of standard multilevel modeling, as well as their limitations. However, our main goal is to propose and illustrate a viable alternative technique – what we term *the 2-step meta-analytic approach* – suited for the analysis of multi-country datasets with a small number of countries (but the method can be easily applied to any type of analysis of nested data with few level-2 sample size). This method provides accurate estimators and standard errors (*SEs*) and allows for reliable inference when one is interested in modeling both individual and country effects. Next to providing accurate estimations, the method we propose is highly info-graphic ensuring a fast and clear information communication, and is accessible to the average social scientist (not skilled in using more advanced simulation techniques).

6.2 Unreliability of Estimates in Multilevel Models with a Small Level-2 Sample Size

The reliability of multilevel estimates may be questioned when the number of level-2 units (e.g. countries) is low (Arend and Schafer 2019; Bell et al. 2010; Bryan and Jenkins 2016; Hox et al. 2010; Hox 1998; McNeish and Stapleton 2016). This warning has been evoked regularly in multilevel textbooks (Rabe-Hesketh and Skrondal 2008; Snijders and Bosker 1999), but in practice it has been often disregarded for several reasons. For one thing, many of these warnings were quite abstract and not accompanied by clear explanations and guidelines about which number of level-2 units is considered too low and which problems a researcher may encounter if model assumptions are violated. General rules of thumb regarding the minimum number of level-2 units required for accurate estimation in multilevel analyses varies considerably between authors, and range from 10 to 100 level-2 units (Hox 1998; Kreft and de Leeuw 1998; Rabe-Hesketh and Skrondal 2008; Snijders and Bosker 1999), with 30 units as the most common recommendation (Hox 1998; Maas and Hox 2004).

In essence, the standard multilevel models rely on maximum likelihood estimation methods which are based on the assumption that errors are normally distributed and variances across groups are heterogeneous (Seco et al. 2013). When the level-2 distributional assumption is violated (which may be the case when dealing with few units), multilevel estimates and their standard errors (especially for the variance components) may not be accurate. Several Monte-Carlo simulation studies have shown that the minimum sample size for obtaining unbiased estimates in multilevel analysis depends on the type of dependent variable (e.g. continuous, categorical), the type and number of predictor variables, the use of (un)balanced group sizes, the specific model parameters of interest (fixed, random or variance components), the potential interest in cross-level interactions, the specification of the random and fixed parts, or the choice of estimation method (Austin 2010; Bell et al. 2010; Bryan and Jenkins 2016; Maas and Hox 2004, 2005; McNeish and Stapleton 2016; Schmidt-Catran and Fairbrother 2015; Stegmueller 2013; Van der Leeden et al. 2008). For example, to obtain unbiased point estimates of coefficients of model predictors, Maas and Hox (2004) recommend a minimum of 10 level-2 units, for good variance estimates at least 30 units, and for accurate *SEs* a minimum of 50 units. In practice, it remains hard to draw general conclusions from the existing studies that are directly applicable to many complex research designs in multi-country studies.

The article of Bryan and Jenkins (2016) came as a real wake-up call for the multilevel community conducting cross-national research. Their Monte-Carlo simulations showed the conditions under which multilevel estimates and their standard errors (*SEs*) may be unreliable or biased, and provided guidelines for what should be considered a minimum number of level-2 units when conducting multilevel analysis in multi-country studies. Table 6.1 presents a summary of their findings for both linear and logit models. In short, when analyzing continuous outcomes,

Table 6.1 Overview of multilevel estimator performance for continuous and binary outcomes (based on Bryan and Jenkins 2016)

Models	Multilevel models	Estimate	Variability in estimates	SE
Linear (continuous outcomes)	<i>Individual-level fixed effects</i>	Reliable	No	Reliable
	<i>Individual-level variance</i>	Reliable	No	Reliable
	<i>Country-level fixed effects</i>	Over- or under-estimated; high uncertainty if $N_c < 25$	Yes	Underestimated if $N_c < 20$
	<i>Country-level variance</i>	Underestimated if $N_c < 20$	Yes	Underestimated if $N_c < 10$
Logistic (binary outcomes)	<i>Individual-level fixed effects</i>	Sometimes underestimated if $N_c < 20$	Sometimes	Reliable
	<i>Country-level fixed effects</i>	Relatively reliable	Sometimes	Substantially underestimated if $N_c < 25$
	<i>Country-level variance</i>	Substantially underestimated if $N_c < 30$	Yes	Substantially underestimated if $N_c < 20$

individual-level estimates (fixed effects or variance components) are reliable regardless of the number of level-2 units. However, a minimum of 25 level-2 units should be available when analyzing country-level effects. Fitting multilevel logit models with a low number of countries brings up even more problems than the linear models, and biased estimates can be found also for fixed effects. The general recommendation is to have at least 30 level-2 units when fitting logit models.

6.3 Common Solutions for Modeling Nested Data with Few Level-2 Units

When concluding that a level-2 sample is too small to apply standard multilevel models, the next step is to identify viable alternative methods to answer the same multilevel-like research questions. Several authors have discussed alternative modeling approaches which include common frequentist techniques (e.g. regression models), correction estimators (e.g. Huber/White sandwich estimators or non-linear transformations of the dependent variable) and more versatile resampling procedures for statistical inference such as Bootstrapping and Bayesian approaches (Bryan and Jenkins 2016; Cheah 2009; Goldstein 2011; Hamaker and Klugkist 2010; McNeish and Stapleton 2016; Seco et al. 2013). However, drawbacks of many of these approaches are that they may not be suited for testing more complex

cross-level hypotheses, that they are not easily available in commonly used software packages, or that they require advanced statistical skills and/or computational power which most applied researchers do not possess. As a result, empirical research has continued to use multilevel models, even when level-2 sample sizes were questionable. Below, we briefly discuss various suggested methods to analyze multi-country datasets with few level-2 units. We will focus on frequentist methods and sampling techniques and do not discuss correction estimators as they have been proven to perform unsatisfactory with small sample sizes (Diggle et al. 2002). For the sake of parsimony, in this discussion we restrict ourselves to solutions for models with continuous outcome variables. However, most solutions would also apply to models with other types of dependent outcomes, such as binary ones.

Regression with (Country-Specific) Clustered Standard Errors on Pooled Data

If we would analyze nested data with the most commonly used regression method – OLS – we may end up making inaccurate statistical inferences. Individual-level model errors within the same country may be correlated and if we fail to control for the within-country error correlation we may obtain downwardly biased *SEs*, shrunken confidence intervals, large *t*-statistics and small *p*-values (Cameron and Miller 2015; Hox 1998). In addition, if OLS regression models would use a country-level predictor (continuous or dichotomous), the country *SEs* may be biased as well (Cameron and Miller 2015). Given this situation, regression with clustered *SEs* may be used instead as it accounts for the dependence between individual observations. This method is now widely used and incorporated in most common statistical software packages (e.g. STATA). However, this method only controls for within-country correlation, but it does not specifically model it (Bryan and Jenkins 2016). Moreover, estimation of *SEs* may be inaccurate with less than 20 level-2 units for balanced designs and less than 50 level-2 units for unbalanced designs (Cameron and Miller 2015). Additionally, because we are often specifically interested in cross-national variation of effects, a multitude of interaction terms between variables of substantive interest and country dummies are required to test specific hypotheses. The overload of interactions and high incidence of multicollinearity in the resulting variables makes many analyses of interest unfeasible when using this method.

Regression with Country-Specific Fixed-Effects on Pooled Data

Whereas the previous method controls for intra-country correlations, the effects estimated are not country-specific, but assumed to be equal across countries (i.e. level-2 units). This is problematic as many analyses aim to specifically model country effects. An alternative is to use the pooled data and fit distinct country intercepts (as fixed parameters). With this technique, the unobserved factors of each country are not separately modeled but are integrated in the intercepts of each country (Bryan and Jenkins 2016). However, given that we model fixed parameters for each country, country-level factors cannot be included as additional predictors. Similar to the previous method, cross-national variation in certain effects can be analyzed only through interactions between country indicators and individual factors and again we are confronted with the issue of estimating a large number of parameters and an

overload of interactions difficult to interpret. Moreover, (Cameron and Miller 2015) warn that by introducing country-specific fixed effects, our estimations lose precision and estimation bias may still occur when the number of countries is small.

Two-Step Approach

Bryan and Jenkins (2016) proposed a more exploratory approach in which regressions are fitted in two steps. The first step is performed at the individual level using country-specific fixed effects. Thus, regular regression models are fitted separately for each country. The second step is conducted at the country level, and country effects are analyzed by regressing the country intercepts on the country-level predictors. Although this technique is advantageous as it reveals the sources of variation, the small number of countries continues to be a problem in implementing a regular regression model in the second step (Combs 2010; Green 1991; Nunnally 1978).

Multilevel Bootstrapping

The three methods described above represent variations on the classical regression models as alternatives for multilevel modeling. However, to obtain unbiased estimates and correct *SEs* in complex research designs in which distributional assumptions are not met, many authors recommend the use of resampling techniques and one such technique is multilevel bootstrapping (Goldstein 2011; Goldstein et al. 2002; Seco et al. 2013). Three different bootstrap strategies have been used to correct for estimates bias and inaccurate *SEs* (Carpenter et al. 2003; Seco et al. 2013; Thai et al. 2013; Van der Leeden et al. 2008):

- (a) *parametric residual bootstrapping* – new data is generated by keeping the predictors fixed and resampling with replacement of the residuals at the two levels from a normal distribution;
- (b) *non-parametric residual bootstrapping* – new data is generated by keeping the predictors fixed and resampling with replacement residuals at both levels from the observed basic residuals;
- (c) *case bootstrap* – new data is generated from the original sample before any modeling is performed (for an overview of different options for cases bootstrap see Roberts and Fan 2004; Van der Leeden et al. 2008).

Among these three bootstrapping procedures, residual bootstrapping has been established as providing the most accurate estimations (Carpenter et al. 2003). Still, Seco and colleagues (2013) showed that residual bootstrapping does not perform very well for small group sizes. In other words, bootstrapping remains incapable to solve the problems of regular multilevel modeling with few level-2 units. In addition, bootstrapping is also procedurally quite difficult for most social researchers as it is not typically integrated as an automated option in the commonly used software packages and often requires advanced programming skills.

Bayesian Multilevel Models

Bayesian estimation for multilevel data is considered to be one of the best analytical approaches when dealing with small samples (Hamaker and Klugkist 2010). In

essence, the Bayesian approach builds on the regular multilevel approach in specifying the models at each level, but it deviates by introducing an additional step in which prior distributions are defined for the model parameters (Hamaker and Klugkist 2010; Stegmueller 2013). In other words, the Bayesian estimation approach focuses on obtaining a posterior distribution for model parameters starting from a prior distribution and the observed data. Compared to classical frequentist methods, the Bayesian approach has the advantage that it is not based on the normality assumption or asymptotic results, which is important when dealing with small sample sizes (Hamaker and Klugkist 2010). However, with this approach, the specification of priors is crucial for obtaining unbiased estimates, especially with a small number of level-2 units (Austin 2010), and arriving at proper specifications of these priors remains challenging for any user.

In conclusion, the first two methods are good alternatives to multilevel modeling if modeling level-2 information is not explicitly the focus of the research. If the latter is the case, resampling multilevel techniques (bootstrapping and Bayesian) are recommended. Still, these methods are not widely implemented in research software packages, their use requires advanced statistical and programming skills, specialist software and computational performance (to reduce long computational time in exploratory analysis) – elements which are often not available to most social researchers.

6.4 An Alternative Stepwise Approach for Testing Individual, Country and Cross-Level Effects

A general issue in cross-national research is that it has been centered primarily on individual or country-level effects, whereas cross-level effects have received rather little attention. This is unfortunate, as these types of effects are often very interesting from a substantive point of view. In many comparative projects, the main interest is in examining whether individual-level effects vary across countries and whether we can explain this type of variation with cross-level effects in which individual-level variables are interacted with country-level variables of interest. Multilevel models may answer such questions very well. However, as our overview in the previous section has made clear, they cannot be implemented when the number of countries is low (often below 30). In addition, we listed several reasons which make the alternative methods recommended by literature unfeasible for research. In this section, we present an alternative (*the 2-step meta-analytical approach*), a stepwise approach which includes the use of meta-analysis and meta-regression to analyze variations across different effects as well as moderating country-level factors. Such a stepwise approach can replicate effects estimated in multilevel analysis, is reliable with few level-2 units and is easy and straightforward to apply without requiring very advanced analytical and programming skills. This method is described and illustrated below.

6.4.1 *The 2-Step Meta-Analytical Approach*

Meta-analysis and meta-regression are often applied in medical research to summarize or combine results on specific relationships that have been tested in multiple separate studies (Borenstein et al. 2009). In these instances, studies constitute the second level of analysis. In such approaches the aims are (1) to generate an overall estimate for the strength of the relationship under consideration, (2) to assess whether significant cross-study variation in the overall effect estimate exists, and (3) to determine which study-level factors could explain the variation (if cross-study heterogeneity is encountered). Cross-studies meta-analytical research in the medical field in the majority of cases includes few studies (often 10 or fewer) and much attention has been paid to develop methods providing reliable and unbiased estimates and correct confidence intervals for estimations (Friede et al. 2017; Rover et al. 2015; Wiksten et al. 2016). However, results of the meta-analytical approach should be interpreted with caution with very few studies, which in the medical research is considered to be less than 5, or even 3 studies (for specific information see Seide et al. 2019; Rover et al. 2015).

If one would replace studies as the level-2 units by countries, it is relatively straightforward to see how this procedure could be used in analyzing cross-national differences in the strength of particular individual-level relationships. It basically entails two steps.

Step 1. Separate Regression Models for Each Country

In the first step, separate regression models are fitted for each country. Suppose one has information on 15 countries, this would lead to 15 country-specific estimates of the relationship of interest. Compared to the common use in the medical field, the advantage in this particular case is that the study design and methodology are very similar across countries, thus reducing the extent to which variation (or heterogeneity as it is usually called in the meta-analytical literature) in the estimates could be due to differences in initial approach (Friede et al. 2017).

Step 2. Meta-Analysis and Meta-Regression

In the second step, a *meta-analysis* is performed on the set of country-specific estimates of the relationship of interest. Two different types of meta-analyses have been developed: fixed-effects and random-effects. Fixed-effects meta-analysis assumes a common effect of a risk factor for a certain outcome and provides an average estimate (Borenstein et al. 2009; Friede et al. 2017; Palmer and Sterne 2016). Random-effects meta-analysis assumes that the ‘true’ effect of interest may vary across level-2 units (Harbord and Higgins 2008; Palmer and Sterne 2016), and this seems a much more reasonable assumption in most studies on country-effects. Random-effects meta-analysis separates real differences in the effect of the predictor on the outcome from sampling variability/chance. In the meta-analysis community, much attention has been paid to developing and testing methods that estimate confidence intervals that are reliable and unbiased, even with very small numbers of level-2 units (Rover et al. 2015; Seide et al. 2019; Wiksten et al. 2016). Simulation studies

showed that certain estimation methods such as Knapp-Hartung – although more conservative – may be implemented with few level-2 units (Friede et al. 2017). The random-effects meta-analysis approach also offers a test of whether the estimate of interest shows significant variation across countries. If the level of variation is low (and not statistically significant), the conclusion is that the relationship of interest is country-invariant. If the level of variation is substantial, one could proceed and use meta-regression to try to explain this variation. In addition to providing reliable estimates of the overall strength of an effect of interest and its cross-country variability, this method provides powerful opportunities for visualization of the variation in the strength of effects across countries (information that is much more difficult to attain if using multilevel analysis or other methods).

As mentioned above, if the results of the meta-analysis suggest variability in country-effects, *meta-regression* can be used to identify factors that may explain this heterogeneity. Meta-regression (Harbord and Higgins 2008; Thompson and Higgins 2002; Thompson and Sharp 1999) can be used to analyze the moderating role of a factor by regressing the country-effects on country-level predictors. The advantage of using meta-regression instead of OLS regression is twofold (Palmer and Sterne 2016). First, when using multi-country data, we need to ensure that the data are properly weighted. By assigning weights to studies, we ensure that large studies are less likely to dominate the analysis and small studies are not seen as unimportant. Second, in situations including few units of analysis/countries, meta-regression applications offer solutions to accurately establish the statistical significance of an effect such as the Knapp-Hartung modification (Knapp and Hartung 2003) or the permutation-based resampling (Harbord and Higgins 2008; Gagnier et al. 2012). These characteristics make the method eminently suited for studying which country-level variables could explain cross-national differences in relationships of interest.

In the next section, we will illustrate this method with an empirical example and compare the results with those from a ‘classic’ multilevel analysis.

6.4.2 Example: The Relationship Between Parental Education and Teenage Parenthood Across 15 European Countries

To illustrate the 2-step meta-analytical approach, we examine the relationship between parental education and teenage parenthood across 15 European countries. It is well-known that children from a lower social-class background run a higher risk of teenage pregnancy and thus of teenage parenthood than children from a higher social-class background (Pirog et al. 2018). What is less known, is whether this risk varies across countries. We expect that it does, and more specifically, that the risk is weaker in countries that offer better opportunities for individual agency and development. In such countries, institutional, cultural and economic factors are thought to buffer the potentially negative consequences of family disadvantage.

As continuous dependent variables are most common in social science applications, we will first use OLS regression to derive parameter estimates of the country-specific effects of social-class background on the risk of teenage pregnancy. In this way, we will illustrate both the ‘traditional’ multi-level approach and the 2-step meta-analytic approach. However, this method can also be applied if logistic regression is used for the within-country regressions (although the specifics of the method are a bit more complicated). In a second example we will briefly illustrate how our method can be used in the latter case.

Data

We use data on 15 countries from the Gender and Generations Project (see Fokkema et al., 2016 for details). These data were collected between 2004 and 2009. To make results as comparable as possible across countries, we select men and women born between 1966 and 1975, leaving us with between 1000 and 2000 respondents per country. The following countries are included: Austria, Australia, Bulgaria, Belgium, Czech Republic, France, Georgia, Germany, Lithuania, the Netherlands, Norway, Poland, Romania, Russia, and Sweden. Our final sample consists of 29,022 individuals.

Variables

The key dependent variable of interest (*Teenage parenthood*) is whether the respondent had a first birth before the age of 20 (0 = no, 1 = yes). The key individual-level independent variable is the level of education of the parents. Information on the educational attainment of both parents was available, scored according to ISCED. To facilitate comparison across countries, these were converted into the newly developed continuous ISLED-scaling (Schröder & Ganzeboom, 2014; Brons and Mooyaart 2018). The mean of the ISLED scores of both parents was used as the indicator of *Parental education*. If information on only one parent was available, the ISLED score of that parent was used. ISLED scores vary between 0 and 100. To facilitate interpretation, we divided scores by ten. A number of additional individual-level variables were included in the analyses (*Gender*, *Age*, *Number of siblings*, *Without BIOparents < 15* – whether respondents grew up most of their youth before age 15 with both parents or not, and *Unknown parental status* – unknown whether they grew up with both biological parents).

The country-level variable of interest is the Human Development Index (HDI), developed by the UN. This is a composite measure based on life expectancy (indicating people’s ability to live a long and healthy life), educational attainment (indicating people’s ability to acquire knowledge) and living standards (indicating people’s ability to acquire a decent standard of living). We use the HDI score of the 15 countries in the year 2000 as this is the earliest date for which HDI scores are available for all countries included (ideally, we would have wanted scores for the period 1990–1995 as this comes closer to the period in which our respondents made fertility decisions). Fig. 6.1 shows the HDI scores of the countries in our sample.

Fig. 6.1 HDI scores for GGP countries in the year 2000

6.4.2.1 Example for Continuous Outcomes

The ‘Classic’ Multilevel Approach

In the first example we analyze data by estimating linear probability models, effectively treating our binary outcome variable as a continuous one. We do so to facilitate the comparison of model estimates across models and across countries. The more complicated logit model estimations and comparisons (see also Mood 2010) will be presented in Sect. 3.2.2. A further advantage of the linear probability model is that we can interpret the effect estimate of our parental education variable as the shift in the percentage of respondents experiencing a teenage birth resulting from a ten-point difference in the ISLED score of a respondent’s parents. We ran two multilevel models. The first is a random-slope model, in which both the intercept and the slope of ISLED are allowed to vary across countries. The second is another random-slope model in which HDI is added as a country-level indicator and the interaction between parental education and HDI as a cross-level indicator. The results from both models are presented below as Stata output.

Output 1 shows that, across all 15 countries, there is a negative effect of parental education on the risk of experiencing teenage parenthood. A ten-point increase in ISLED is associated with a 1.6% decrease in the risk of teenage parenthood. In addition, Output 1 shows that there exists considerable cross-country variation in the effect of parental education. The random slope for parental education is .0001896, with an estimated standard error of .0000747, so the estimate is more than 2.5 times its standard error. In Output 2, HDI and the interaction between parental education and HDI are added. HDI has a statistically significant negative effect, suggesting that teenage parenthood is less common the higher the HDI score of countries is. In addition, the interaction between parental education and HDI also is statistically significant. The negative parental education gradient becomes weaker the higher the HDI score of countries is. This is in line with our expectations.

Output 1 Random-slope multilevel model of the relationship between parental educational attainment and teenage parenthood

<i>Teenage parenthood</i>	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Gender	.1131694	.0032397	34.93	0.000	.1068197	.1195191
Age	-.0002281	.0000472	-4.83	0.000	-.0003205	-.0001356
Parental education	-.0165465	.0037315	-4.43	0.000	-.0238602	-.0092329
Without BIOparents < 15	.0298401	.0052786	5.65	0.000	.0194943	.040186
Unknown parental status	-.016978	.0264223	-0.64	0.521	-.0687648	.0348087
Number of siblings	.0113589	.0010416	10.91	0.000	.0093174	.0134005
_cons	-.0114718	.0262477	-0.44	0.662	-.0629164	.0399728

Random-effects Parameters		Estimate	Std. Err.	[95% Conf. Interval]		
Country: Independent						
var(Parental education)		.0001896	.0000747	.0000876	.0004103	
var(_cons)		.0036049	.0013345	.001745	.0074473	

var(Residual)		.0747446	.0006208	.0735377	.0759713	

Output 2 Random-slope multilevel model of the relationship between parental educational attainment and teenage parenthood including the macro-level indicator and a cross-level interaction

Teenage parenthood	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
Gender	.1131956	.0032394	34.94	0.000	.1068464	.1195447
Age	-.0002243	.0000467	-4.80	0.000	-.0003159	-.0001327
Parental education	-.1130147	.0302013	-3.74	0.000	-.1722082	-.0538212
Without BIOparents < 15	.030191	.0052776	5.72	0.000	.019847	.040535
Unknown parental status	-.0166365	.0264173	-0.63	0.529	-.0684134	.0351405
Number of siblings	.0113262	.0010409	10.88	0.000	.0092861	.0133663
HDI	-.6833836	.1054858	-6.48	0.000	-.8901321	-.4766352
HDI x Parental education	.1192182	.037084	3.21	0.001	.0465348	.1919016
_cons	.540955	.0873755	6.19	0.000	.3697022	.7122078

Random-effects Parameters	Estimate	Std. Err.	[95% Conf. Interval]			
Country: Independent						
var(Parental education)	.0001037	.0000436	.0000455	.0002366		
var(_cons)	.0009233	.0003549	.0004346	.0019614		

var(Residual)	.0747455	.0006208	.0735386	.0759722		

Furthermore, Output 2 shows that the estimate for the random slope of parental education has dropped by almost half (from .0001896 to .0001037), suggesting that HDI can explain almost half of the country-level variation in the effect of parental education.

The 2-Step Meta-Analytical Approach

The alternative meta-analysis approach we propose starts with estimating a separate linear probability model per country, leading to 15 identically specified models overall. Output 3 shows the example for the Czech Republic. For the chosen country, the estimate of the association between parental education and the risk of teenage pregnancy is $-.030$, suggesting that a 10% increase in parental education leads to a 3.0% decrease in teenage parenthood. Estimates for all countries can be found in Table 6.2.

Output 3 Example of a country-specific model (the Czech Republic)

Source	SS	df	MS	Number of obs	=	1,735
Model	13.2519081	6	2.20865135	F(6, 1728)	=	24.60
Residual	155.159619	1,728	.089791446	Prob > F	=	0.0000
				R-squared	=	0.0787
				Adj R-squared	=	0.0755
Total	168.411527	1,734	.097123142	Root MSE	=	.29965

Teenage parenthood	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Gender	.1277083	.0144406	8.84	0.000	.0993854	.1560313
Age	.0002308	.000207	1.12	0.265	-.0001751	.0006367
Parental education	-.0295338	.0060878	-4.85	0.000	-.0414741	-.0175935
Without BIoparents < 15	.0879683	.0221639	3.97	0.000	.0444974	.1314392
Unknown parental status	.0094585	.0658958	0.14	0.886	-.1197854	.1387025
Number of siblings	.0182289	.0064057	2.85	0.004	.0056651	.0307926
_cons	-.1883895	.0893359	-2.11	0.035	-.3636074	-.0131716

Table 6.2 Country-level datafile to be used as input in meta-analysis and meta-regression procedures

	Association between Parental education and Teenage parenthood				HDI
	<i>OLS regression</i>		<i>Logistic regression</i>		
Country	B	<i>SE_B</i>	AME	<i>SE_AME</i>	
Russia	−0.01615	0.00563	−0.0175204	0.0058051	0.720
Bulgaria	−0.04893	0.00505	−0.0548612	0.0054996	0.713
Georgia	−0.02332	0.00596	−0.0265545	0.0062133	0.673
Romania	−0.03954	0.0051	−0.0560977	0.007007	0.708
Lithuania	−0.01393	0.00447	−0.0168735	0.0051631	0.757
Poland	−0.01039	0.00428	−0.0119415	0.004599	0.784
Czech Republic	−0.02953	0.00609	−0.0343537	0.0067771	0.821
Germany	−0.01313	0.00504	−0.0130292	0.0054126	0.86
France	−0.00538	0.00243	−0.0084212	0.0036588	0.849
Belgium	−0.00723	0.00221	−0.0105502	0.0037816	0.873
Norway	−0.01112	0.00208	−0.0155903	0.0031694	0.917
Austria	−0.01392	0.00383	−0.0140908	0.0039492	0.837
Sweden	0.0011	0.00267	0.0021482	0.0026105	0.877
Australia	−0.01147	0.0046	−0.0123621	0.0051695	0.899
Netherlands	−0.00223	0.00144	−0.0022191	0.0016706	0.878

In the second step, the country-specific estimates of interest (in this particular case, the estimates of the relationship between parental education and the risk of teenage parenthood) are collected into one dataset that is used as input for the meta-analysis. Table 6.2 shows an example of such a dataset, that includes additional parameters of potential interest as well as HDI as a country-level indicator. Using this dataset, we performed a meta-analysis (using the metan command in STATA 16). The results of this analysis are presented in Output 4 and graphically in Fig. 6.2.

Output 4 shows the estimates of the association for all countries, as well as their confidence intervals. The largest (negative) association is found in Bulgaria (−.049), whereas the smallest in found in Sweden (.001). At the bottom of Output 4,

Output 4 Meta-analysis of the country-variation in the association between parental education and the risk of teenage parenthood

Fig. 6.2 Graphical presentation of the results of a meta-analysis on cross-national variation in the association between parental education and the risk of teenage parenthood

Fig. 6.3 The association between parental education and the risk of teenage parenthood (Y-axis) and HDI scores (X-axis), based on Output 5

is very similar to the cross-level effect estimated in our ‘classic’ multilevel model (.1192, with a *SE* of .0370). The effect estimate suggests that the association between parental education and teenage parenthood is weaker in countries with a higher HDI score. To allow a better assessment of this finding, the regression line linking the association between parental education and teenage parenthood and HDI are plotted in Fig. 6.3. To facilitate interpretation, we limited the HDI scores (X-axis) to a range that is observed in our dataset. In addition to the regression line, also the 15 separate country data points are depicted in Fig. 6.3. This figure shows that in countries with low HDI scores (around .70), the association between parental education and teenage parenthood is quite strong (effect of around $-.03$), suggesting that a 10% point increase in ISLED scores leads to a decrease in the percentage of people experiencing teenage parenthood by about 3%. In countries with high HDI scores (around .90), the association between parental education and teenage parenthood is negligible. Thus, these findings are in line with our expectations.

6.4.2.2 Example for Binary Outcomes

Our example treated the dependent variable as continuous, thus allowing to use OLS regression. Clearly, this meta-analytic 2-step procedure can also be used with logistic regression as the first step in the analysis. In fact, the vast majority of the applications of meta-analysis in epidemiology use binary outcomes, and thus

perform meta-analysis and meta-regression with odds ratios from multiple clinical trials (e.g. Sattar et al. 2010) or observational studies (e.g. Jones et al. 2015) as the dependent variable of interest.

Although it is common in epidemiology to use odds ratios as the dependent variables in meta-analysis, in modern social sciences such applications are regarded as problematic. Mood (2010) has shown that odds ratios resulting from logistic regressions of different samples (e.g. different population subgroups or different countries) cannot be compared to each other, as the unobserved heterogeneity in the model can vary across samples. However, the author suggests that average marginal effects (AME), that can be derived from the logistic regression results, can be meaningfully compared across samples. AME gives the average effect of an independent variable on the probability that the dichotomous dependent variable equals 1. As this quantity does not depend on the unobserved heterogeneity in the model, it can be used to compare effects across countries and the AMEs (and their standard errors) for different countries can be input in a meta-analysis and meta-regression, just as the B coefficients in an OLS regression can.

To illustrate this approach, we repeated the analysis presented above, but now used logistic regression rather than OLS regression, and calculated average marginal effects based on the logistic regression models. Next, we performed meta-analysis and meta-regression on these estimates. The results are presented in Output 6 and 7. The average effect of a ten-point increase in parental ISLED score is $-.010$,

Output 6 Results of a meta-analysis of the country-variation in the association between parental education and the risk of teenage parenthood, using AMEs from a logistic regression model as the dependent variable

metan AME SE_AME, lcols(Country) random second(fixed)				
Study	ES	[95% Conf. Interval]		% Weight
Russia	-0.018	-0.029	-0.006	6.35
Bulgaria	-0.055	-0.066	-0.044	6.45
Georgia	-0.027	-0.039	-0.014	6.20
Romania	-0.056	-0.070	-0.042	5.92
Lithuania	-0.017	-0.027	-0.007	6.56
Poland	-0.012	-0.021	-0.003	6.74
Czech Republic	-0.034	-0.048	-0.021	6.00
Germany	-0.013	-0.024	-0.002	6.48
France	-0.008	-0.016	-0.001	7.01
Belgium	-0.011	-0.018	-0.003	6.98
Norway	-0.016	-0.022	-0.009	7.14
Austria	-0.014	-0.022	-0.006	6.93
Sweden	0.002	-0.003	0.007	7.26
Australia	-0.012	-0.022	-0.002	6.56
Netherlands	-0.002	-0.005	0.001	7.42
D+L pooled ES	-0.019	-0.026	-0.011	100.00
I-V pooled ES	-0.010	-0.012	-0.008	100.00
Heterogeneity chi-squared = 181.63 (d.f. = 14) p = 0.000				
I-squared (variation in ES attributable to heterogeneity) = 92.3%				
Estimate of between-study variance Tau-squared = 0.0002				
Test of ES=0 : z= 5.07 p = 0.000				

Output 7 Results of a meta-regression of the association between parental education and the risk of teenage parenthood on HDI-scores, using AMEs from a logistic regression model as the dependent variable

metareg AME HDI, wsse(SE_AME)

Meta-regression			Number of obs =		15	
REML estimate of between-study variance			tau2		= .000131	
% residual variation due to heterogeneity			I-squared_res		= 86.57%	
Proportion of between-study variance explained			Adj R-squared		= 47.54%	
With Knapp-Hartung modification						
AME		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
HDI		.1431291	.0436092	3.28	0.006	.0489173 .237341
_cons		-.1351645	.0357209	-3.78	0.002	-.2123349 -.0579941

suggesting that on average, a ten-point increase in ISLED decreases the probability of a teenage birth by 1%. This effect hardly differs from the average effect (−.009) in the linear probability model. In addition, the pattern of country-variation in scores also very strongly resembles the one in Output 4. The results of the meta-regression also correspond quite closely with the ones resulting from the linear probability model (.143 versus .121).

6.5 Conclusion

The use of multilevel analysis in comparative research has recently been criticized as the number of countries involved in cross-national analysis is often viewed as too limited to allow reliable inferences and unbiased estimation of parameters of interest. This chapter proposes the *2-step meta-analytic approach* as an alternative to ‘classic’ multilevel analysis if one is interested in understanding cross-national variation in the link between individual-level variables as well as cross-level interactions. After a brief discussion of the main criticisms on the multilevel approach when using few level-2 units and an overview of existing modeling alternatives, the 2-step meta-analytical approach is outlined and illustrated using examples for both continuous and dichotomous outcomes. Still, although this method is discussed in the context of analyzing data for a small number of countries, it may be applicable to any type of research including a small number of level-2 units (e.g. schools, municipalities, hospitals).

The method we propose in this chapter as an alternative for multilevel analysis has several strengths. *First*, when using it, one can obtain reliable estimates and accurate *SEs* even when the number of countries is small (smaller than the 25–30 suggested as lower limit for multilevel analyses). Moreover, as a very small number of countries may lead to spurious findings even with meta-analytical techniques, with such techniques one may still be able to provide accurate inferences by an appropriate choice of estimation methods (e.g. Knapp-Hartung modification) and

permutation-based resampling. For example, Zoutewelle-Terovan and Liefbroer (2018) included 12 countries in their analyses and used a permutation test with adjustment for multiplicity – suited for a small number of countries and multiple covariates (Harbord and Higgins 2008). *Second*, when one is interested in specifically modeling individual effects, country-level effects and cross-level interactions, the 2-step meta-analytic approach provides great opportunities for such modeling. This method is superior to multilevel modeling as its graphic display allows a much more intuitive feel of what the findings mean in terms of the positioning of individual countries than is usually true for multilevel analysis. Also, whereas many of the alternative techniques to multilevel modeling presented in Sect. 6.2 encounter difficulties in explicitly modeling country-level effects and cross-level interactions, our method is capable to comprehensively do so. Additional examples of this approach can be found in other publications within the Context of Opportunities (CONOPP) project (see Brons and Harkonen 2018; Brons et al. 2017; Koops 2020; Zoutewelle-Terovan and Liefbroer 2018). *Third*, whereas our discussion and examples focus on modeling one random slope, multiple random slopes (e.g. how teenage parenthood is linked to parental education, parental separation and number of siblings) can be modeled as well with this method by repeating the 2-step meta-analytical approach for multiple associations. *Fourth*, in this chapter we center on a two-level nested design. However, the 2-step meta-analytical approach could also be extended to situations where individuals are nested within more than one level (e.g. cohorts or regions within countries). Methods to analyze 3-level data have been developed within epidemiology and psychometry, e.g. when multiple instruments are used within studies to measure the same underlying concept. In this approach, the instruments are viewed as a second level within ‘trials’. A country-design with an additional level consisting of regions (or cohorts) within countries can be viewed as a variation on this theme (Cheung 2014; Jackson et al. 2011; Van den Noortgate et al. 2015). *Finally*, our method is not only accessible to the average-skilled researcher (as it is easy to conduct and interpret and does not require any advanced simulation skills), but also only requires brief computational time and little computational power to run models, and can be performed with the most common software programs used in social sciences (e.g. STATA, R, SAS).

At the same time, the meta-analytic method proposed is not a panacea. *First*, it remains difficult to establish a quantitative minimum for the level-2 sample size when conducting meta-analyses – such limits are rarely recommended. To date no consistent guidelines for minimum sample sizes exist. Some authors argue for a minimum of 3 level-2 units (Rover et al. 2015), others discuss a minimum of 8 units (Jenkins and Quintana-Ascencio 2020). However, such minima depend on aspects such as the amount of variance observed (a very small sample may be problematic with substantial statistical heterogeneity), the size of studies or the number of predictors used (Gagnier et al. 2012; Jenkins and Quintana-Ascencio 2020). *Second*, to deal with the small number of level-2 units, it has been recommended to use certain estimation techniques (e.g. Knapp-Hartung) or resampling options to establish significance (e.g. permutation test). However, such methods are recognized for being quite conservative and one may run the risk of obtaining false negatives. It is

difficult for us to establish the circumstances in which such situations occur (it was also not the goal of this chapter), but this is clearly one aspect that future research needs to clarify (Gagnier et al. 2012). Still, whereas in our research we may have marginally missed the reporting of some significant effects, the conservativeness of methods used increases our confidence in effects that reach the threshold for statistical significance. *Third*, the implementation of the method may become more difficult (although it remains feasible) if one of the individual-level variables has multiple categories. An important aspect of this approach is that the effects of additional variables are allowed to vary across countries (as separate analyses are performed for each country). On the one hand, this can be viewed as an advantage, as other variables might also have quite different effects across countries. On the other hand, this leads to the estimation of many parameters and one could view the multilevel model, with its assumption of fixed effects across countries, as a more parsimonious approach.

The 2-step meta-analytic approach is proposed as an alternative to multilevel modeling when the number of level-2 units is small and one is interested in modeling individual, country and cross-level effects. However, it is not our intention to claim that the method is superior to multilevel modeling. In fact, in our example we observed that multilevel analysis would still have led to accurate inferences. This suggests that in some situations of few level-2 units multilevel modeling still performs well and it was beyond the purpose of this chapter to demonstrate under which conditions it no longer does. Our main goal is to present a viable alternative when multilevel inferences are questionable. Our method may also be used as a sensitivity analysis to support results obtained from multilevel modeling. Furthermore, the 2-step meta-analytic approach may be preferable when one is interested in graphically displaying heterogeneity and making inferences based on the positions and characteristics of individual countries.

Acknowledgements The research leading to these results has received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant Agreement n. 324178.

References

- Arend, M.G., and T. Schafer. 2019. Statistical power in two-level models: A tutorial based on Monte Carlo simulation. *Psychological Methods* 24 (1): 1–19.
- Austin, P.C. 2010. Estimating multilevel logistic regression models when the number of clusters is low: A comparison of different statistical software procedures. *The International Journal of Biostatistics* 6 (1): 1–18.
- Bell, B.A., G.B. Morgan, J.D. Kromrey, and J.M. Ferron. 2010. The impact of small cluster size on multilevel models: A Monte Carlo examination of two-level models with binary and continuous predictors. *JSM Proceedings, Survey Research Methods Section 1* (1): 4057–4067.
- Borenstein, M., L.V. Hedges, J.P. Higgins, and H.R. Rothstein. 2009. *Introduction to meta-analysis*. UK: Wiley.

- Brons, M.D., and J. Harkonen. 2018. Parental education and family dissolution: A cross-national and cohort comparison. *Journal of Marriage and Family* 80 (2): 426–443.
- Brons, M.A., and J.E. Mooyaart. 2018. The Generations & Gender Programme: Constructing harmonized, continuous socio-economic variables for the GGS Wave. *Technical Paper GGS Methodology*: 1–22.
- Brons, M.D., A.C. Liefbroer, and H.B.G. Ganzeboom. 2017. Parental socio-economic status and first union formation: Can European variation be explained by the Second Demographic Transition theory? *European Sociological Review* 33 (6): 809–822.
- Bryan, M.L., and S.P. Jenkins. 2016. Multilevel modelling of country effects: A cautionary tale. *European Sociological Review* 32 (1): 3–22.
- Cameron, A.C., and D.L. Miller. 2015. A practitioner's guide to cluster-robust inference. *Journal of Human Resources* 50 (2): 317–372.
- Carpenter, J.R., H. Goldstein, and J. Rasbash. 2003. A novel bootstrap procedure for assessing the relationship between class size and achievement. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 52 (4): 431–443.
- Cheah, B.C. 2009. *Clustering standard errors or modeling multilevel data*, 2–4. New York: Research Note University of Columbia.
- Cheung, M.W.L. 2014. Modeling dependent effect sizes with three-level meta-analyses: A structural equation modeling approach. *Psychological Methods* 19 (2): 211–229.
- Combs, J.G. 2010. Big samples and small effects: Let's not trade relevance and rigor for power. *Academy of Management Journal* 53 (1): 9–13.
- DerSimonian, R., and R. Kacker. 2007. Random-effects model for meta-analysis of clinical trials: An update. *Contemporary Clinical Trials* 28 (2): 105–114.
- Diggle, P.J., P. Heagerty, K. Liang, and S.L. Zeger. 2002. *Analysis of longitudinal data*. Oxford University Press.
- Fokkema, T., A. Kveder, N. Hiekel, T. Emery, and A.C. Liefbroer. 2016. Generations and Gender Programme Wave 1 data collection: An overview and assessment of sampling and field-work methods, weighting procedures, and cross-sectional representativeness. *Demographic Research* 34: 499–524.
- Friede, T., C. Rover, S. Wandel, and B. Neuenschwander. 2017. Meta-analysis of few small studies in orphan diseases. *Research Synthesis Methods* 8 (1): 79–91.
- Gagnier, J.J., D. Moher, H. Boon, C. Bombardier, and J. Beyene. 2012. An empirical study using permutation-based resampling in meta-regression. *Systematic Reviews* 1 (1): 1–9.
- Goldstein, H. 2011. Bootstrapping in multilevel models. In *Handbook of advanced multilevel analysis*, ed. J.J. Hox and J.K. Roberts, 163–172. New York: Routledge.
- Goldstein, H., W. Browne, and J. Rasbash. 2002. Multilevel modelling of medical data. *Statistics in Medicine* 21 (21): 3291–3315.
- Green, S.B. 1991. How many subjects does it take to do a regression analysis. *Multivariate Behavioral Research* 26 (3): 499–510.
- Hamaker, E.L., and I. Klugkist. 2010. Bayesian estimation of multilevel models. In *Handbook of advanced multilevel analysis*, ed. J.J. Hox and J.K. Roberts, 137–161. New York: Routledge.
- Harbord, R.M., and J.P. Higgins. 2008. Meta-regression in stata. In *Meta-analysis in Stata: An updated collection from the Stata Journal*, ed. T.M. Palmer, J.A.C. Sterne, H.J. Newton, and N.J. Cox, vol. 8(4), 2nd ed., 493–519. College Station: Stata Press.
- Higgins, J.P.T., S.G. Thompson, J.J. Deeks, and D.G. Altman. 2003. Measuring inconsistency in meta-analyses. *BMJ* 327 (7414): 557–560.
- Hox, J. 1998. Multilevel modeling: When and why. In *Classification, data analysis, and data highways*, ed. I. Balderjahn, R. Mathar, and M. Schader, 147–154. New York: Springer.
- Hox, J.J., M. Moerbeek, and R. van de Schoot. 2010. *Multilevel analysis: Techniques and applications*. Routledge.
- Jackson, D., R. Riley, and I.R. White. 2011. Multivariate meta-analysis: Potential and promise. *Statistics in Medicine* 30 (20): 2481–2498.

- Jenkins, D.G., and P.F. Quintana-Ascencio. 2020. A solution to minimum sample size for regressions. *PLoS One* 15 (2): 1–15.
- Jones, L., G. Bates, E. McCoy, and M.A. Bellis. 2015. Relationship between alcohol-attributable disease and socioeconomic status, and the role of alcohol consumption in this relationship: A systematic review and meta-analysis. *BMC Public Health* 15 (400): 1–14.
- Knapp, G., and J. Hartung. 2003. Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine* 22 (17): 2693–2710.
- Koops, J.C. 2020. *Understanding nonmarital childbearing in Europe and North-America. The role of socio-economic background and ethnicity in different societal contexts.* (PhD thesis). University of Groningen, Groningen.
- Kreft, I., and J. de Leeuw. 1998. *Introducing multilevel modeling.* Sage.
- Maas, C.J., and J.J. Hox. 2004. Robustness issues in multilevel regression analysis. *Statistica Neerlandica* 58 (2): 127–137.
- . 2005. Sufficient sample sizes for multilevel modeling. *Methodology* 1 (3): 86–92.
- McNeish, D., and L.M. Stapleton. 2016. Modeling clustered data with very few clusters. *Multivariate Behavioral Research* 51 (4): 495–518.
- Mood, C. 2010. Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European Sociological Review* 26 (1): 67–82.
- Nunnally, J.C. 1978. *Psychometric theory.* 2nd ed. New York: McGraw-Hill.
- Palmer, T.M., and J.A.C. Sterne. 2016. *Meta-analysis in Stata: An updated collection from the Stata Journal.* 2nd ed. College Station: Stata Press.
- Pirog, M.A., H. Jung, and D. Lee. 2018. The changing face of teenage parenthood in the United States: Evidence from NLSY79 and NLSY97. *Child & Youth Care Forum* 47 (3): 317–342.
- Rabe-Hesketh, S., and A. Skrondal. 2008. *Multilevel and longitudinal modeling using Stata.* STATA Press.
- Roberts, J.K., and X. Fan. 2004. Bootstrapping within the multilevel/hierarchical linear modeling framework: A primer for use with SAS and SPLUS. *Multiple Linear Regression Viewpoints* 30 (1): 23–34.
- Rover, C., G. Knapp, and T. Friede. 2015. Hartung-Knapp-Sidik-Jonkman approach and its modification for random-effects meta-analysis with few studies. *BMC Medical Research Methodology* 15 (1): 1–7.
- Sattar, N., D. Preiss, H.M. Murray, P. Welsh, B.M. Buckley, A.J. de Craen, S.R.K. Seshasai, J.J. McMurray, D.J. Freeman, J.W. Jukema, and P.W. Macfarlane. 2010. Statins and risk of incident diabetes: A collaborative meta-analysis of randomised statin trials. *The Lancet* 375 (9716): 735–742.
- Schmidt-Catran, A.W., and M. Fairbrother. 2015. The random effects in multilevel models: Getting them wrong and getting them right. *European Sociological Review* 32 (1): 23–38.
- Schröder, H., and H.B. Ganzeboom. 2014. Measuring and modelling level of education in European societies. *European Sociological Review* 30(1): 119–136.
- Seco, G.V., M.A. Garcia, M.P.F. Garcia, and P.E.L. Rojas. 2013. Multilevel bootstrap analysis with assumptions violated. *Psicothema* 25 (4): 520–528.
- Seide, S.E., C. Rover, and T. Friede. 2019. Likelihood-based random-effects meta-analysis with few studies: empirical and simulation studies. *BMC Medical Research Methodology* 19 (1): 16.
- Snijders, T.A.B., and R.J. Bosker. 1999. *Multilevel analysis: An introduction to basic and advanced multilevel modeling.* Sage.
- Stegmuller, D. 2013. How many countries for multilevel modeling? A comparison of frequentist and Bayesian approaches. *American Journal of Political Science* 57 (3): 748–761.
- Thai, H.T., F. Mentre, N.H. Holford, C. Veyrat-Follet, and E. Comets. 2013. A comparison of bootstrap approaches for estimating uncertainty of parameters in linear mixed-effects models. *Pharmaceutical Statistics* 12 (3): 129–140.
- Thompson, S.G., and J.P. Higgins. 2002. How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine* 21 (11): 1559–1573.

- Thompson, S.G., and S.J. Sharp. 1999. Explaining heterogeneity in meta-analysis: A comparison of methods. *Statistics in Medicine* 18 (20): 2693–2708.
- Van den Noortgate, W., J.A. Lopez-Lopez, F. Marin-Martinez, and J. Sanchez-Meca. 2015. Meta-analysis of multiple outcomes: A multilevel approach. *Behavior Research Methods* 47 (4): 1274–1294.
- Van der Leeden, R., E. Meijer, and F.M. Busing. 2008. Resampling multilevel models. In *Handbook of multilevel analysis*, ed. R. Van der Leeden and E. Meijer, 401–433. Springer.
- Wiksten, A., G. Rucker, and G. Schwarzer. 2016. Hartung–Knapp method is not always conservative compared with fixed-effect meta-analysis. *Statistics in Medicine* 35 (15): 2503–2515.
- Zoutewelle-Terovan, M., and A.C. Liefbroer. 2018. Swimming against the stream: Non-normative family transitions and loneliness in later life across 12 nations. *The Gerontologist* 58 (6): 1096–1108.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.