Chapter 9
Modelling Strategies

Abstract When devising a modelling strategy, researchers determine the steps they will take to answer their research question or test their hypothesis. Two general principles are important. Firstly, most of the steps that you would take in a single-level regression analysis are also relevant for MLA. Secondly, start with simpler models, for example in terms of the number of levels, and add further complexity as required. The statistical model used depends on the measurement level of the dependent variable. In a baseline model, the variances are estimated at each level. After that we can start to analyse the fixed effects in a more exploratory manner or a specific hypothesis can be tested. Disentangling context and composition and providing an indication of their relative importance are often the aims of the modelling strategy. As the number of higher level units is often small, it may not be possible simultaneously to analyse several contextual variables. We end this chapter by discussing the interpretation of results in the light of a number of common assumptions.

Keywords Multilevel analysis · Modelling strategy · Measurement level · Exploratory research · Hypothesis testing · Sample size · Assumptions

Before you actually start analysing your data, it is important to define a strategy for your analysis or modelling strategy. The modelling strategy describes what you intend to do when analysing the data and takes the form of a sequence of steps that lead to an answer to your research question. The modelling strategy naturally comes somewhere in the middle of the research cycle (Fig. 9.1). It is determined by the research questions of your study, the hypotheses (where these exist) and the nature of the data; as such, it reflects the logic of your research. After you have determined your modelling strategy, you will undertake the analysis and write up the results in tables and figures as necessary and in the main body of your report. The way that you write up your research should follow the steps of your modelling strategy (see also Chap. 10).

Many of the decisions you make when defining your modelling strategy are not specific to multilevel analysis but are appropriate for data analysis in general.

A. H. Leyland, P. P. Groenewegen, *Multilevel Modelling for Public Health and Health Services Research*, https://doi.org/10.1007/978-3-030-34801-4_9

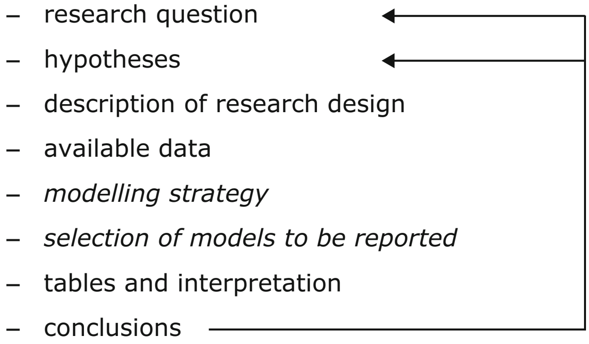
– research question
– hypotheses
– description of research design
– available data
– *modelling strategy*
– *selection of models to be reported*
– tables and interpretation
– conclusions

Fig. 9.1 The place of the modelling strategy in the research cycle

Everything that you have learned about single-level regression analysis is likely to be important when you undertake a multilevel regression analysis.

Some important general advices are to start simple and only make your analysis more complicated when you are happy that you have a clear understanding of the results of your simpler analysis. This is not to say that we would argue in favour of using inadequate statistical models purely on the grounds of simplicity. But, as an approach to improve your understanding of the data and the research problem, it is a useful step. How can you expect to understand and explain a complex model if you do not have an understanding of a simpler underlying model?

Define the Data Structure

We discussed multilevel data structures in Chap. 4. The simplest multilevel data structures are strict hierarchies with only two levels. Often our data structures in the real world are more complicated, but again it is useful to start simple.

Simplification could be based on the frequencies of the occurrence of certain combinations in the data. For example, although in reality your data might contain a level below individual patients, such as that of the separate contacts patients make with the health service, it may be that in your data 99% of patients only had one contact. Or, if we were analysing pregnancy outcomes in different hospitals, we would want to take into account that pregnancies are nested in women, with one woman possibly having more than one pregnancy. However, if we have hospital data from only 2 years, it could be that there is a very small number of women with more than one pregnancy in the data set. A way of keeping things simple would be to select initially only the first pregnancy that occurred of any women with two pregnancies in the data set or to select one at random. That would result in a two-level analysis instead of a three-level analysis with limited power to differentiate between the levels of women and pregnancies. After conducting the analysis for a two-level model, and once you are satisfied that the conclusions for this model are clear, you can run a three-level model to check whether that alters the results. Given that there would be little additional data—just the additional pregnancies of the few women who had more than one pregnancy during the 2-year study period—we would not expect substantial differences between the models. The most important additional information is likely to be the ability to partition the variation between that attributable to unexplained differences between women and that due to differences between pregnancies within women. This means that it would probably make more sense to report the results of the three-level analysis rather than the two-level analysis. However, the sparsity of the data structure (the vast majority of women only having one pregnancy during the study period and virtually no women with more than two) may cause computational problems and a need to resort to reporting the results from a two-level model.

The decision to simplify might also be based on a preliminary analysis of variation, if this were to show that the variation at one of the levels in your dataset was trivial. With simple hierarchical data, the inclusion of additional levels is not a big problem, but with the more complicated data structures (such as cross-classified and multiple membership models), it might be a wise first step at least to consider leaving out levels that do not really contribute to the variation in the outcomes.

Often there are also deviations from strict hierarchies. A multiple membership model could be simplified if only a few cases belong to more than one higher level unit. If most patients usually see their own GP and only occasionally another GP, you could assign them to their usual GP. (If there is a list system, then this would be the GP to whose list that patient belongs.) Doing this simplifies the data structure to a strict hierarchy and keeps the analysis simple.

The first steps in the analysis of a cross classified data structure could be to analyse the two hierarchies separately first, as was done for example by Chum and O'Campo (2013). They studied the determinants of cardiovascular disease in residential neighbourhoods and the neighbourhoods where people worked. This gave a first impression of the variation at different levels. The prevalence of CVD clustered more strongly in residential than in work neighbourhoods. Their strategy was to estimate the variance attributable to each level in three models (individuals nested in residential neighbourhoods, work neighbourhoods and the cross classification of the two). Their next step was then to analyse the fixed effects associated with the characteristics of the two contexts in this cross-classified structure.

The information that can be gained through the use of a cross-classified data structure depends to some extent on the degree of overlap between the two hierarchies. If there is considerable overlap, then the results from the two-level models are unlikely to differ since there would be little difference between the hierarchical data structures used in each. However, when there is less overlap, the results may differ if one context is more important than the other. In either case, using a cross-classified model will help to gain an understanding of the relative importance of the contexts, which may in itself relate to one of your research questions.

Measurement Level and Distribution of the Dependent Variable

The measurement level and the distribution define the statistical model that should be used. If the dependent variable is continuous and approximately normally distributed, then linear regression is appropriate. It may be that a transformation is necessary to make the outcome follow an approximate normal distribution; you should remember that such transformations make your job of explaining the model and the parameter estimates more difficult. With a dichotomous variable, you will normally choose logistic regression. Often an ordinal dependent variable, such as self-rated health, can be dichotomised to make the analysis simpler. It should be noted, however, that this results in a loss of information. It is up to you as the researcher to decide whether this loss of information is acceptable; this will in part depend on the field of research and what is currently seen as 'good practice'. Often we only find out whether this loss of information is important after comparing the analysis of a dichotomised dependent variable with, for example, an ordered logit analysis. Such analyses are often best undertaken as a form of sensitivity analysis (in this case it is the sensitivity to the choice of analytical model that you are testing). When the results of two competing analyses are not materially different, it can be enough to say so in a sentence or two. The choice of which set of results to present as your main results then amounts to a trade-off between the need to explain a more complex model and the added information that such a model may bring.

The results of a linear regression model are often not seriously affected by violations of the distributional assumptions. As a consequence, a first step in your analysis could again be to use a simpler model, such as linear regression, and only when you have a fuller understanding of your data and the relationships between variables progress to more complicated models, such as ordered logits in the case of ordinal variables or Poisson models in the case of count variables.

The Baseline Model

Defining the baseline model comes early in your modelling strategy. It is often called the null model or empty model. This suggests that the baseline, against which we will evaluate further models, is always a model that contains no individual variables. This is, however, not necessarily the case. For example, if the main focus of your analysis is the relationship between income and access to specialised care, and if you know that access to specialised care is also dependent on age, you might decide to use a model including only age as the baseline.

In a study of body mass index (BMI) among women in nearly 33,000 communities in 57 countries, Corsi et al. (2012) adjusted their baseline model for the age of the women. Given that BMI is known to be related to age, and the countries studied have a range of rather different demographic profiles (and there are probably even greater differences between the communities within those countries), it is only possible to interpret the variation in BMI at the levels of communities and countries after accounting for differences in the age structure.

It often makes sense to adjust the baseline model for age and sex when studying health outcomes. For example, Voigtländer et al. (2010) made such an adjustment to their baseline model when analysing the influence of regional and neighbourhood deprivation on self-rated health. Another example is provided by Deraas et al. (2014) who fitted a baseline model including age and sex in their study of the influence of primary care on unplanned hospital admissions.

Cole et al. (2009) studied mental health outcomes and musculoskeletal disorders in a cohort of healthcare workers. They had five measurements per worker. They adjusted their baseline model for year of observation to take changes in the prevalence of health problems over time into account when estimating the variance at hospital and regional level.

The baseline model consists of limited information such as the overall average of the dependent variable (and relationships with key variables of interest such as age and sex) and the variances at the different levels. In previous chapters, we have discussed how to interpret the variation at the different levels in the study (see Chap. 6: Apportioning variation in multilevel models).

Exploratory Research and Hypothesis Testing

The modelling strategy differs according to the aims of the research and the research questions. We distinguish here between exploratory research and hypothesis testing research.

In *exploratory research*, the research question is only partly specified. The dependent or outcome variable is specified, but the independent variables are not. An example of an exploratory research question would be: does hospital length of stay vary between hospitals and which characteristics of hospitals explain this variation? The dependent variable is length of stay and the independent variables are not specified. A useful modelling strategy in a case like this would be as follows:

1. Estimate a random intercept model to verify if there is indeed variation between hospitals in length of stay of the patients. This null or baseline model might already include some basic patient characteristics that are known to be related to length of stay and without which any analysis would be deemed to be incomplete: perhaps the patient's age and sex. In an exploratory analysis, it may be more appropriate not to include any covariates in the null model.
2. Then add the individual-level variables, such as diagnosis, comorbidities or treatment. Adding the individual-level variables might reduce the variation between hospitals because of differences in case-mix (differences in the composition of the patient population) between hospitals.

3. The next step is to add hospital characteristics to see which variables at this level relate to length of stay. These could include the size of the hospital or the degree of specialisation.
4. At this stage, it might be interesting to explore random slopes for some of the individual-level variables. For example, the relationship between the age of the patients and length of stay might vary between hospitals. In an exploratory analysis, the slope variation can be a source of new hypotheses about how hospitals influence length of stay.
5. Finally, you could consider introducing selected cross-level interactions. Your choice of the interactions to include might be informed by your findings regarding the random slopes. If, for example, you have seen that the effect of age on length of stay varies between contexts, then you could explore whether this was due to an interaction between the patient's age and a hospital characteristic such as the size of the hospital. Alternatively you may have a particular interest in examining cross-level interactions involving pre-specified individual or contextual variables. If this were the case, then these key variables would usually be mentioned in your research question, and it might be more appropriate to undertake this analysis *before* looking for random slopes in step 4.

Changes in the amount of variation at the different levels should be evaluated at each step. In an exploratory analysis, you might want to use a stepwise procedure, selecting those variables that matter for the outcome of your study, such as forward or backward selection of significant variables. As with any exploratory analysis, you should be aware that performing multiple tests at a given level of significance means that you are likely to encounter statistically 'significant' results by chance.

In *hypothesis testing research*, we specify not only the dependent variable but also one or more independent variables. An example of a research question related to a hypothesis could be: is more social capital in neighbourhoods related to better self-rated health among the people who live there? The first step is the same as in exploratory research: estimate an appropriate baseline model to see how the variation in self-rated health is apportioned between individuals and neighbourhoods. Again, this baseline model might include some variables that are known to be correlated with self-rated health. At this point you can either introduce the contextual variable of interest (social capital in this example) or the individual variables. In the following sequence, we start with the contextual variable(s) of interest.

1. Add the contextual variable to the baseline model and see if there is a significant relationship with the outcome variable. If not the hypothesis is refuted. However, its effect could be masked by differences in the composition of the population of neighbourhoods. Hence, it might be worthwhile checking what happens to the effect if individual-level variables are added.
2. Add the relevant individual-level variables to the previous model and see whether the effect of the contextual variable stays the same or disappears. If there was an effect of the contextual variable and that disappears when individual variables are taken into account, then the apparent contextual effect was the result of differences in the composition of the neighbourhood populations.

3. In hypotheses testing research, you might also have specific ideas about cross-level interactions. Your hypothesis might be that the effect of social capital is stronger for people who have lived in their neighbourhood for a longer time. We would assume that the length of residence (an individual level variable) would already have been included in the model in step 2, in which case the next step would be to include the cross-level interaction between neighbourhood social capital (the contextual variable) and length of residence. It is not necessary first to fit a random slope model to test whether the effect of length of residence varies randomly between contexts.

Context and Composition

In Chap. 7 we discussed a very common modelling strategy, aimed at disentangling contextual effects and compositional effects. As is clear from the previous section, an attempt to make a distinction between contextual and compositional influences is a goal common to many modelling strategies in multilevel research.

Modelling the Effects of Higher Level Characteristics

In Chap. 3 we defined higher level units as units that can be sampled. Sample size is thus an issue not only at the lowest level but also at the higher levels. We have many lower level units nested within fewer higher level units. The number of higher level units is often restricted by the fact that in reality they form an entire population. Think of neighbourhoods within a city; the number of neighbourhoods is restricted by the size of the city and perhaps the administrative definitions with which we are working. The number of EU member states is equally restricted at any one time to the number of countries that are in the EU. Another restriction is more pragmatic; when the higher level units are organisations, such as schools, and you want to study students nested in schools, the effort needed to include more schools in a study is often considerable.

The number of higher level units has consequences if the focus of the research is on the effect of higher level characteristics. This number should then be sufficient to estimate a mean, a variance and the effect of the relevant variables of interest at that level. As a rule of thumb, the number of units that you need is approximately ten times the number of variables you want to include in the analysis. This means that if you want to include ten variables to test your hypothesis about the characteristics of hospitals and how they influence an outcome at patient level, you would need at least a hundred hospitals. Alternatively, if you want to analyse the effect of characteristics of the healthcare systems of EU member states on access to healthcare, the maximum number of higher level units (at the time of writing) is 28. As such, the number of country-level variables that could be included in an analysis is only two or three.

This limitation on the number of contextual variables that can be included in an analysis has consequences for the design of studies and for the modelling strategy. For the design of a study where the effects of higher level characteristics are important, it is more important to increase the number of higher level units (if this is possible) than the number of lower level units (Snijders and Bosker 2012). In terms of a modelling strategy, this means that we have to be careful not to include too many independent higher level characteristics at the same time. In the example of the analysis of 28 EU member states in which we wish to study the effect of healthcare systems on access to healthcare, we would probably want to include one confounder, such as the wealth of a country, along with one characteristic of the healthcare system at a time. We could repeat the analysis several times using each relevant healthcare system characteristic individually and compare the results. We would not be able to analyse the effects of several characteristics at the same time. This also excludes the possibility of adding a contextual variable with several categories since this would be operationalised by introducing a series of dummy variables. We would consequently have to be more careful in formulating our conclusions which would be based more on weighting the results against our hypotheses and background knowledge than on strict statistical criteria.

In Chap. 10 we will give some examples of studies where the authors were not sufficiently aware of this problem and, as a consequence, introduced more contextual variables than the available number of higher level units could support.

Random Effects at Higher Levels

In all of the models considered in this book, we have assumed that the higher level effects are all normally distributed. (This may be after an appropriate transformation; for example, in a multilevel logistic regression, we assume that the log odds ratios associated with membership of the higher level units are normally distributed.) This assumption is convenient but not always appropriate. Austin (2005, 2009) has considered the impact of this assumption and found that an inappropriate assumption of normality at the higher level does not appear to have implications for the estimation of fixed effects, but it may lead to biased or incorrect estimates of the variances. This then has consequences for assessment of the importance of different levels in a model or for studies in which the residuals themselves are of some importance (such as studies of institutional performance).

One way in which the distribution of higher level residuals may appear non-normal is due to the presence of outliers. Multilevel data may contain outliers in the same way that the data for traditional regression models may be outlying; the difference is that in a multilevel model, the outliers may be at any level in the model. Methods have been developed for the detection and treatment of outliers at higher levels (Langford and Lewis 1998; Lewis and Langford 2001). These essentially rely on including a fixed effect for a context regarded as outlying; this removes the impact of this unit on the estimation of the higher level variance whilst including the lower level units (such as individuals) in the analysis.

Interpreting the Results in the Light of Common Assumptions

As we said at the beginning of this chapter, a number of assumptions are the same as in single-level regression analysis. We will briefly illustrate this with an example of a hypothetical intervention study. We have chosen the example of an intervention study to be able to address some assumptions that are typically made in such studies. The example is the evaluation of an intervention to reduce BMI. Individuals have been randomised to the intervention and control groups, and we have pre- and post-intervention measures for everyone in the study. Individuals are nested within communities (e.g. neighbourhoods or schools). A slightly different study design of a community intervention would be possible, in which it would be the communities (and all individuals within them) rather than the individuals that would be randomised to the intervention and control groups. The structure of the data is that of a three-level model with measurement occasions nested in individuals, clustered within areas (a repeated measures design). To make the intervention and control groups comparable, we adjust for age, sex and educational status (basic/higher). Algebraically the model can be written as shown in Eq. (9.1).

$$\begin{aligned} y_{ijk} &= \beta_0 + \beta_1 x_{1jk} + \beta_2 x_{2jk} + \beta_3 x_{3jk} + \beta_4 x_{4jk} + \beta_5 x_{5ijk} + \beta_6 x_{4k} x_{5ijk} + v_{0k} + u_{0jk} + e_{0ijk} \\ v_{0k} &\sim N\left(0, \sigma_{v0}^2\right) \\ u_{0jk} &\sim N\left(0, \sigma_{u0}^2\right) \\ e_{0ijk} &\sim N\left(0, \sigma_{e0}^2\right) \end{aligned} \tag{9.1}$$

Here y_{ijk} is the primary outcome, BMI, at measurement occasion (pre- or post-intervention) i for individual j in community k. x_{1jk}, x_{2jk} and x_{3jk} are individual-level covariates relating to the person's baseline age, sex and educational status; these do not change between measurement occasions. x_{4jk} denotes whether the individual is in the intervention (coded 1) or control (coded 0) groups, and x_{5ijk} indicates whether the measurement occasion was pre- (coded 0) or post- (coded 1) intervention. The term $x_{4jk}x_{5ijk}$ is then the cross-level interaction picking out the post-intervention measurement occasion in the intervention group. The coefficient associated with this term, β_6, is the parameter of interest, indicating the success or otherwise of the intervention. In addition to the individual characteristics, the model takes into account that there may have been a baseline difference in BMI between the intervention and control groups and that there may be a population change in BMI between the two

Table 9.1 Parameter estimates for the evaluation of a hypothetical intervention on BMI

Parameter		Coefficient	(SD)
	Fixed part		
Constant		25.155	(0.052)
Age		−0.510	(0.015)
Male		0.315	(0.042)
Higher education		−1.015	(0.042)
Intervention		−0.048	(0.044)
Time = post		0.018	(0.018)
Intervention ∗ (time = post)		−0.195	(0.025)
	Random part		
Community-level variance		0.090	(0.019)
Individual-level variance		1.961	(0.044)
Measurement occasion variance		0.396	(0.008)

measurement occasions; neither of these events should mistakenly be ascribed to an intervention effect. We also model the variances at the three levels.

First of all we will consider some assumptions underlying the fixed part of the model that was used to make the groups comparable. For these assumptions, it is irrelevant whether we are discussing an intervention study or an observational study. The parameter estimates are given in Table 9.1.

One assumption made in the model described in Eq. (9.1) is that the effect of age on BMI is linear for all ages. This is an assumption that can be tested easily by comparing this model with one where we also add age squared or a model where we recode age into a number of categories. Another assumption is that the effect of age on BMI is the same regardless of sex or education level and that the effect of education is the same for men and women. These assumptions can be tested by using interaction terms between these variables. Alternatively, if the study is powered for this, we could consider stratified analyses by key variables such as gender. Often a stratified analysis will give you a better impression of the size and direction of the interaction effect and whether this differs between groups. (This is at the cost of power; there will obviously be fewer observations in each of the strata than in the overall analysis.) However, as the stratified analysis takes more space in the tables, you may decide to report the version with the interaction effect and use the stratified analysis as a valuable step in your own interpretation of the interaction.

Next consider the impact of the intervention itself. An assumption here is that the intervention is equally effective regardless of age, sex or education level. It is conceivable that, and may be worth testing whether, the intervention is differentially effective for older and younger people, men and women or more and less educated people. Knowing not just whether an intervention has worked but for which groups it appears to be more or less successful is important if we subsequently want to improve or tailor the intervention and if we are interested in the impact of the intervention on inequalities. We can examine differential impacts on subgroups by introducing the appropriate interaction terms (between the intervention and the subgroup of interest) into the model.

There are also some assumptions implicit in the way the random part of the model has been formulated. In this model we have assumed that the variance in BMI is the same regardless of age, sex and educational level. This can be tested by estimating the variances separately for age groups, men and women and educational categories. Another assumption is that the variance is unchanged by the intervention. The model that was estimated, shows a decrease in mean BMI in the intervention group, but it is possible that the intervention has changed the variance. An example would be if the intervention had a greater impact on those with higher BMI; this would result not just in the decrease in BMI seen in the intervention group following the intervention but also a reduction in variance in the same group.

All of the above assumptions may be reasonable and may be supported by the data. But if the data does not support these assumptions, then fitting the alternative models may impact on estimates in unpredictable ways. In an example such as this, we have an extremely important single parameter—the intervention effect—and cannot say with certainty that changes to the model would not alter the magnitude or statistical significance of this estimate. In short, it is unlikely that your modelling strategy will test every aspect of your model, but it is important that you are aware of your underlying assumptions.

Conclusions

The modelling strategy for a multilevel analysis begins with the research question and hypothesis that the study is addressing. Simplifications to the model that you are fitting will help you to gain a better understanding of the data and an idea of your answer, with further detail being provided by the complexity that you subsequently add. There will inevitably be assumptions underlying any choices that we make during the construction of a modelling strategy, including which models we consider and which we do not. Whilst it may not be necessary formally to test every assumption, it is important that we are aware of the assumptions that we have made and what their consequences might be—even if the answer is that their consequences may be unpredictable.

References

Austin PC (2005) Bias in penalized quasi-likelihood estimation in random effects logistic regression models when the random effects are not normally distributed. Commun Stat Simul Comput 34:549–565

Austin PC (2009) Are (the log-odds of) hospital mortality rates normally distributed? Implications for studying variations in outcomes of medical care. J Eval Clin Pract 15:514–523

Chum A, O'Campo P (2013) Contextual determinants of cardiovascular diseases: overcoming the residential trap by accounting for non-residential context and duration of exposure. Health Place 24:73–79

Cole DC, Koehoorn M, Ibrahim S, Hertzman C, Ostry A, Xu F, Brown P (2009) Regions, hospitals and health outcomes over time: a multi-level analysis of repeat prevalence among a cohort of health-care workers. Health Place 15:1046–1057

Corsi DJ, Finlay JE, Subramanian SV (2012) Weight of communities: a multilevel analysis of body mass index in 32,814 neighborhoods in 57 low- to middle-income countries (LMICs). Soc Sci Med 75:311–322

Deraas TS, Berntsen GR, Jones AP, Førde OH, Sund ER (2014) Associations between primary healthcare and unplanned medical admissions in Norway: a multilevel analysis of the entire elderly population. BMJ Open 4:e004293

Langford IH, Lewis T (1998) Outliers in multilevel data. J R Stat Soc Ser A 161:121–160

Lewis T, Langford IH (2001) Outliers, robustness and the detection of discrepant data. In: Leyland AH, Goldstein H (eds) Multilevel modelling of health statistics. Wiley, Chichester, pp 75–91

Snijders TAB, Bosker RJ (2012) Multilevel analysis: an introduction to basic and advanced multilevel modeling. Sage, Los Angeles

Voigtländer S, Berger U, Razum O (2010) The impact of regional and neighbourhood deprivation on physical health in Germany: a multilevel study. BMC Public Health 10:403