

Chapter 10

Reading and Writing



Abstract This chapter focuses on two issues. Firstly, we consider the critical reading of research articles that use MLA, and secondly we explore the standards for writing up research that has used MLA. Critical reading is important both for people who do not regularly use MLA themselves and for those who are regular users. The irregular users need to be able to assess the methodology of studies using MLA, whilst regular users may find inspiration for new ways and strategies of data analysis and for ways to write up and present their own research, particularly the methods and results sections. So the reading and writing parts of this chapter are related. When a method of analysis is used that is relatively new to its field, there are no clear standards as to what should be included in the methods section or how the tables might be laid out.

Keywords Multilevel analysis · Critical reading · Reporting

Communication is an important part of the research process. Research results are important in themselves, but will only be used if they are communicated to the relevant audiences. In public health and health services research, we usually have two types of audiences: the research community and the users of research in policy and practice (Bensing et al. 2003).

The ‘end users’ of research probably will not read the research papers themselves, but intermediaries certainly will. Such intermediaries might be health scientists and epidemiologists who work in policy development positions within (public) health authorities. It is crucial that we as researchers should write up our research in a way that makes our methodological and statistical approach as clear as possible.

The research community enters the process when we submit a paper for publication. Some reviewers will be selected for their specialised statistical knowledge, whilst others will be selected for their substantive knowledge about the subject of the research. We cannot guarantee that the latter will be completely up-to-date with MLA. We therefore need to write about our approach and to present our results in a way that is understandable to many audiences.

Critical Reading

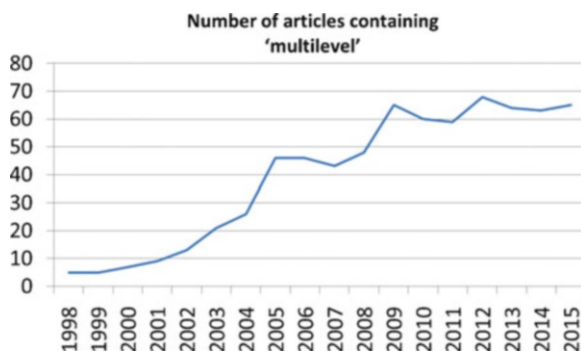
An increasing number of research articles in the area of public health and health services research are being published that use multilevel analysis. We have simply counted the number of articles that used the term ‘multilevel’ in a Pubmed search of the journals *Social Science and Medicine*, *Journal Epidemiology and Community Health* and *European Journal of Public Health* (see Fig. 10.1). This simple search may have missed some articles that used slightly different terminology such as ‘hierarchical’ instead of ‘multilevel’. However, the picture is clear and that is one of a huge increase in the use of multilevel analysis in our area of research: from 5 articles in 1998 to 65 articles in 2015 in just these three journals.

In the past the alternatives to multilevel analysis that we described in Chap. 3 were often used. However, it is now rare to see a published paper that analyses clustered data and does not use multilevel analysis. In fact, as early as 1998 we came across an article the authors of which—in a foot note—said that they initially submitted a ‘naïve’ (as they called it themselves) single-level analysis, but were asked by the reviewers to repeat the analysis using MLA (Matteson et al. 1998).

Given that currently so many research articles use MLA, it is important that researchers, even if they do not apply MLA themselves, are able to understand and critically appraise the work of others. When reading an article, we are inclined to focus more on the substantive results and less on the methodology, to the extent that we sometimes take the methodology for granted and skip the methods section. When relatively new and complicated methods are used, and we can still count MLA as such, the tendency to skip the methods section might be even stronger. However, it is also more dangerous to do so when the methods are new (Bingenheimer 2005). With new methods there will be no clear standards for reporting research results (see later in this chapter), researchers may make mistakes or debatable choices in their methodology, and reviewers are not always able to judge exactly what was done. It is therefore important for researchers and for users of research results to develop a way to read critically research articles that use MLA.

To help new users to read research articles critically and to understand the multilevel design employed, we have formulated a number of questions. You can

Fig. 10.1 Number of articles containing the term ‘multilevel’ in a Pubmed search of the journals *Social Science and Medicine*, *Journal Epidemiology and Community Health* and *European Journal of Public Health*



use these questions when reading and abstracting research articles. We will briefly elucidate them.

What Is the Research Question?

It might seem superfluous to draw attention to the research question. It is not, for two reasons. Firstly, we still occasionally stumble across published research articles that have no clear formulation of a research question or hypothesis at all. That means that as a reader you have to reconstruct the question yourself *after* reading the paper. Secondly, the research question determines the choice of method. It is therefore important to have a clear picture of what question the authors want to answer.

Increasingly researchers formulate an objective or aim instead of a research question. Usually an objective or aim will be less specific. Verstappen et al. (2005) formulated their objective in the abstract as *'To describe the variation in the numbers of imaging investigations requested by general practitioners (GPs) and to find likely explanations for this variation'*. In the introduction to the article they are a bit more specific without, however, making clear what the 'likely explanations' might be.

The present study measured the variation of imaging investigations among a large group of GPs and investigated the influence of professional and contextual determinants at three levels: the individual GP, local GP groups, and the region.

Compare this with an example of an explicit research question, as formulated by Turrell et al. (2007):

What is the relation between area-level socioeconomic disadvantage and mortality before and after adjusting for within area variation in individual level occupation? Does the relationship between mortality and individual level occupation differ by area level disadvantage? What is the variation in mortality at different geographical levels?

Research questions also differ regarding how specific they are. Some research questions ask whether there is a relationship between two variables, without specifying the direction. Others ask whether a particular relationship will be found. These are basically hypotheses formulated as research questions. An example of this is a study by Van Stam et al. (2014) on Sexual and Reproductive Health (SRH). They tested the hypothesis that the relationship between educational attainment and SRH differed according to the level of globalisation of the region where the subjects live (effect moderation). Hence, their research question can also be formulated as: Is this hypothesis confirmed or refuted in our data?

The combination of a research objective and a concrete hypothesis is also specific enough to guide the remainder of an article. For example, Agyemang et al. (2009) formulated as their objective *'to assess the effect of neighbourhood income and unemployment/social security benefit (deprivation) on pregnancy outcomes'*. Their hypothesis was *'that low neighbourhood income and deprivation [are] associated with poor pregnancy outcomes after adjustment for individual-level characteristics'*.

In a general analysis of research questions, Mayo et al. (2013) discussed the use of language, suggesting that words such as ‘explore’ and ‘describe’ should be avoided when formulating a research question because of the difficulty such words pose in determining whether or not the question has been answered. They stress how the correct formulation of the research question will assist the researcher in the choice of the optimal design for the study.

In many cases the multilevel nature of the problem is already indicated by the research question, such as when the question is about the relationship between variables at different levels. An example is the research question posed by Jat et al. (2011): *what are the effects of individual, community and district level characteristics on the utilisation of maternal health services?*

Which Levels Can Be Distinguished Theoretically?

It is important to be aware of the difference between the levels that one would like to be able to distinguish in an ideal situation and the reality with which one actually has to work. If the research question is to explain differences between hospitals in patients’ judgements about quality of care, the most obvious levels are probably patients at the lower level and hospitals at the higher level. However, if we analyse the research problem in terms of the actors involved and the opportunities and constraints they experience (see Chap. 2), we might come to the conclusion that the physician responsible for the treatment and the ward in which the patients are treated are likely to be the drivers of patients’ experiences. That might imply a three-level model of patients, physicians and hospitals, or possibly four levels with physicians nested within wards (or a cross-classification of physicians and wards, depending on the hospital structure).

Often the introduction of an article uses a theoretical notion of a relevant higher-level unit, connected to a mechanism that relates this context to individual behaviour or outcomes. The ‘data and methods’ section then moves to an operational definition of higher-level units, often chosen for practical reasons of data availability. This pragmatically chosen definition of the higher-level units might be different from the units implied by the theoretical reasoning in the introduction of the article. The results are therefore based on units that do not correspond to what was intended and this may lead to less clear effects. Returning to the previous example regarding patients’ judgements of quality of care, if the physician is the true driver of the patients’ experiences but this level is unobservable, then the extent to which there will be differences between hospitals will depend on the degree to which physicians assessed as providing high or low quality cluster within the same hospitals. Often in the discussion the emphasis moves back from the pragmatic context of the available data that were used in the analysis, to the theoretical notions from the introduction.

We illustrate this with research examples that have studied the effect of neighbourhood characteristics on health or health behaviour. Ball et al. (2007) moved from ‘local neighbourhoods’ in the abstract to suburbs of between 4000

and 30,000 inhabitants in the methods section, and back to neighbourhoods in the last paragraph of the discussion. In a study on obesity in New York City, Black et al. (2010) used United Hospital Funds areas as neighbourhoods. NYC has 34 of these areal units. Given the population of the city (over eight million people), these must be huge areas and it is doubtful that we could really call them neighbourhoods. The article gives the average sample size per area, but not the number of inhabitants. Sellström et al. (2008) studied environmental influences on smoking during pregnancy. Citing the importance of peer groups in adolescent smoking, they state that social influences are apparently important in explaining why pregnant women keep on smoking. The actual units they use in their analysis to capture these social influences are neighbourhoods with between 4000 and 10,000 inhabitants. This is quite far from the idea of peer group influences that they brought up in their theoretical reasoning.

Another example of the connection between the theoretical reasoning in the introduction of an article and the definition of spatial units in the methods section is provided in a paper by Karvonen et al. (2008) on smoking patterns. They state: '*An ideal spatial context for an exploration of smoking patterns by small area would comprise a reasonably stable and homogeneous population with relatively low variation of disadvantage*'. Subsequently in the methods section, they rationalise their use of 107 neighbourhoods in Helsinki: '*These areas are of the size that most residents could walk across them in 15–20 min and have an average population of 4000*'.

These examples—and there are many more—illustrate the importance of theorising the contexts that are being used as higher-level units and of being aware of the fact that there is often a gap between the theoretically interesting units and what is actually available or used. This gap may be part of the explanation for the finding that the influences of contextual variables on individual outcomes are sometimes weak, and it is important that any such gap should be acknowledged in the paper.

What Is the Structure of the Actual Data Used?

Apart from the issue discussed in the previous section, there are often reasons why there is a discrepancy between the levels that would be relevant on theoretical grounds and those actually used. One reason is that information may be lacking on some relevant levels.

In the example of patients' judgements about quality of hospital care that we gave at the end of Chap. 2, the researchers might for pragmatic reasons have chosen hospitals to be their higher level. For some indicators of quality of care this may be appropriate (such as those that reflect hospital policies) but for others—think of whether the treatment by hospital personnel is polite—the more appropriate level might be wards, teams or even individual nurses and doctors. One reason to use only

the hospital level is that there is no information available about the levels in between (Hekkert et al. 2009; Sixma et al. 2009).

Another reason might be that the numbers at a certain level are too small. The extreme case is when there is only one unit at one level within each higher-level unit. The household might be a relevant level from a theoretical point of view, but if only one member of each household has been interviewed then the household and individual levels are indistinguishable. In the example dataset used in the tutorial in Chap. 12, the authors collapsed four levels into two for pragmatic reasons, concentrating on patients and GPs but leaving out the practice level (most GPs were single-handed) and the episode of care level (most patients had only one episode of care during the study period). Researchers might also simplify their data structure by choosing only one observation from a (theoretically larger) dataset. For example, Jat et al. (2011), in their study of environmental influences on pregnancy outcomes, only chose the last pregnancy of each woman in their sample. In so doing the level of the women who gave birth and the level of the newborn infants collapsed into one level. Another example is Van Berkestijn et al. (1999) who only used the first consultation in each episode of care. This meant that they could restrict their model to just two levels: the GPs in their study and the episode of care which coincides with the consultation.

A good reason to opt for fewer levels than are actually available is that this may make the analysis less complicated. It is, however, important to be aware that leaving out a higher level is less problematic than leaving out an intermediate level. In the former case, the variation at the omitted level is simply added to that at the new highest level. When an intermediate level is omitted, the variation will in general be split between the higher and lower levels (see Chap. 6 and also the section on variation at different levels later in this chapter).

Whatever the reason for omitting levels, it is important to be aware of the difference between the levels that were theoretically postulated and the levels that were actually used. It is elucidative to draw a simple diagram of the levels and the numbers used at all levels. Chapter 4 on multilevel data structures gives examples of such diagrams.

It is also important to consider the numbers at the different levels and the average number of lower-level units per higher-level unit. The number of higher-level units is sometimes quite small. As we pointed out in Chap. 3, the higher-level units are treated as a sample and there should be sufficient numbers of units at this level for it to make sense to estimate an average and variance. The number of units is also important if authors want to include characteristics of these units in their analysis. If so, the numbers should be sufficient to estimate the coefficients associated with these characteristics in addition to the mean and variance. We have come across several examples where the authors (and reviewers) were apparently not aware of this. Some of these studies are international comparisons with the countries as higher-level units and a characterisation of welfare state regimes in the form of a set of dummy variables as independent variables. Even though the welfare state regime might be seen as a single concept, it is usually operationalised as a series of dummy variables. Eikemo et al. (2008) included 23 countries, their higher-level units, but added

4 dummy variables at this level. Witvliet et al. (2012) had 46 countries and 6 dummy variables for welfare state regimes. And Rathmann et al. (2015) analysed data for 27 countries and included 4 dummy variables indicating welfare state typology.

The problem of trying to include more contextual variables than the data can support is, however, not restricted to the analysis of welfare states. Friele et al. (2006) had one analysis with 80 hospitals and another with 40 hospitals which included 7 independent variables at the hospital level. With a simple rule of thumb of 10 cases for each independent variable, the first analysis was reasonable but not the second. For the estimation of contextual effects, the number of lower-level units becomes irrelevant; the authors were attempting to estimate 9 quantities (a mean, 7 regression coefficients and a variance) from 40 contextual observations. Further examples include Huizing et al. (2007) who had 15 wards in nursing homes and included 6 independent variables at this level, and Nicholson et al. (2009) who included four independent contextual variables with just 22 higher-level units.

What Statistical Model Was Used?

Most statistical models that can be run as single-level analysis can also be used in MLA (see Chap. 4). Questioning what statistical model was used and whether this was appropriate is therefore as relevant when reading a multilevel article as when reading about a single-level analysis. If the authors specify the algebraic form of their model in the article or in a technical appendix, a useful check is to see whether the subscripts correspond to the levels that have been included.

To as great an extent as possible (within the space constraints imposed by journals), the methods section of a paper should provide sufficient information to enable other researchers to reproduce the analysis reported in an article. This includes the type of model (linear, logistic, Poisson, etc.), details of the levels used (including the specification of any which are cross-classified or multiple membership), the variables included in each model in the fixed and random parts (including interactions), and details of the software and estimation procedures used. Published descriptions of the model used and estimation techniques are sometimes so brief that these cannot even be deduced from the software that was used.

Some authors have compared their results of MLA with a single-level model. As we argued in Chap. 3, in cases where the units for whom the outcomes are measured are nested within higher-level units, MLA is the preferred approach. The examples provided here illustrate again that using a single-level model in circumstances that indicate that a multilevel model is appropriate may lead to false conclusions about the effect of higher-level variables. In Chap. 3, we discussed the example of an intervention study in GP practices (Renders et al. 2001) where the intervention effect was significant in a single-level (patients) model, but not in a multilevel model. We also referred to Mauny et al. (2004) who analysed the occurrence of the malaria parasite in blood samples taken from people living in villages in Madagascar. In the single-level

model, they found a significant coefficient for the size of villages which they did not find in a MLA. This was due to the misestimated precision when the village size was assigned to all individuals and treated as a series of independent individual-level observations. A similar example that we have previously mentioned in this chapter is the article by Matteson et al. (1998). In a footnote they state that, in the single-level analysis which they initially submitted, more county variables were significant.

What Was the Modelling Strategy?

This relates to the steps that the authors say they are going to take when analysing their data in order to answer their research question and/or to test their hypotheses. Ideally the modelling strategy should follow on from the research question and hypotheses. One typical sequence might be to start by examining the variation at different levels in a null model and reporting the intraclass correlation. The next step would be to introduce individual-level variables, evaluating the changes in variation at all levels. A reduction in the higher-level variation at this stage indicates compositional effects. The next step may then be to introduce higher-level variables and evaluate the decrease in variation at that level. Of course, the modelling strategy should reflect the hypotheses that one wants to test.

It is important that the modelling strategy is a systematic and logical sequence of steps and that the modelling strategy as described in the methods section is indeed executed and reported in the results section. Many research papers do not include a modelling strategy at all or else report their results in a different order to that suggested by the strategy. Tables should reflect the modelling strategy as far as possible; however, it is often not necessary to document every step in the tables. This might easily lead to large and unclear tables (for example, see the four page landscape table in Béland et al. 2002).

Examples of clear modelling strategies accompanied by results sections that follow the steps outlined in the methods section include those presented by Van Yperen and Snijders (2000), Ball et al. (2007) and Merlo et al. (2005).

Van Yperen and Snijders studied Karasek's job demand-control model. The main hypothesis of this model is that the job stress that workers experience depends on the interaction between the demands that are made of them and the amount of control they experience over their own job. Strong demands lead to particularly high levels of job stress when workers have less control over their work. They test this hypothesis and look at demand and control both at the individual level and the group level. Removing the group effects (by including them) means that individual-level demands and control are then relative to those experienced by co-workers. Their modelling strategy neatly follows the hypotheses.

Ball et al. studied educational variation in walking for women and whether this can be explained by intrapersonal and social characteristics and by perceived and

objectively assessed facets of the physical environment. Their modelling strategy consisted of four steps. In the first step, only education was included in the model. In subsequent steps, environmental variables, social variables and finally personal variables were added.

Merlo and colleagues studied differences between hospitals in neonatal mortality for low risk and high risk pregnancies against the background of regionalisation and concentration of services. They used four steps, starting with an empty model; they then added characteristics of the hospitals where the deliveries took place. In step 3, maternal and delivery characteristics were added. In the final model, these characteristics were replaced by a propensity score to take confounding by indication into account.

A more specific issue when evaluating the modelling strategy is the completeness of the individual-level model. This is particularly important in studies of composition and context and when forming league tables. In studies of context and composition, the researcher may wish to explore whether variation at the higher or contextual level remains when relevant individual characteristics have been taken into account. The range of individual variables available is often quite small, especially when using routinely collected or register data. In a study on the use of tranquilizers by Groenewegen et al. (1999), only the age and sex of the users were known. In a study of the socio-economic determinants of compliance to colorectal cancer screening (Pornet et al. 2011), the individual model consisted of only age, sex and insurance type. The risk is then that the clustering of people with, for example, a low socio-economic status in certain neighbourhoods leads to apparent neighbourhood-level variation that would have disappeared if socio-economic status had been measured at the individual level.

The completeness of the individual-level model is especially important when creating 'league tables' as a measure of institutional performance. The individual characteristics then act as a means of correcting for differences in case-mix. With good case-mix correction, the higher-level residuals reflect, to as great an extent as possible, the 'true' differences between higher-level units such as nursing homes. Patients or their representatives can use that information to inform their choice of care site (Arling et al. 2007).

Does the Paper Report the Intercept Variation at Different Levels?

Sometimes researchers only report fixed effects. In this case, they are apparently only using MLA in order to have appropriate estimates of the confidence intervals or other measures of uncertainty around the regression coefficients. This may for example be the case when the data are collected using a two-stage sample and the authors want to adjust for that. Nevertheless, it would be interesting to see the extent

to which the dependent variable clusters within higher-level units. As we discussed in Chap. 6, an estimate of the higher-level variance is necessary for power calculations. We usually obtain these estimates from published research about similar problems or data sets. However, some of the estimation procedures used (such as generalised estimating equations—GEE) will only correct the standard errors of the estimates without explicitly estimating the variance at the different levels.

Sometimes the variation is of central importance to the research question at hand; even if this is not the case, the reporting of variation can be seen as a service to the academic community because of its potential interest to readers of the article. As such, the intercept variance should be reported as well as the individual variance, enabling the reader to calculate the intraclass correlation coefficient if this was not reported in the article. In some cases the intercept variance is reported for the empty model, whilst in other cases it is more relevant to report the intercept variation only after taking into account some individual-level variables. If treatment outcomes in different hospitals are analysed, and the hospitals differ in composition according to the age, sex and severity of illness of the patients treated, it might be more relevant to report the between-hospital variation after these case-mix variables have been taken into account.

If slope variance is also important, this should be reported alongside the covariance between the intercept and the slope. Remember that the variance of the intercept and the covariance are dependent upon where the slope variable has been centred, so any non-standard centring (that is if the location has been changed so that a value of 0 on the transformed slope variable does not correspond to a value of 0 on the original variable) should also be reported as an aid to interpretation. We provided an introduction to random slopes in Chap. 5 along with a guide to the interpretation of different patterns of covariance.

Cross-Level Interactions

If there is an explicit hypothesis about the interaction between variables at different levels, this can be tested by introducing a cross-level interaction. In a more exploratory analysis or when the hypothesis is about variation in the slopes, one would estimate the slope variance and the covariance between the slope and the intercept. You will, however, have more power to test for a specific cross-level interaction than for a random slope.

In general, interaction terms are not always easy to interpret. It may be helpful to illustrate them using a figure. Several nice examples can be found in the published literature; for example, see any of Turrell et al. (2007), Joshi et al. (2008), Stafford et al. (2008) and Mohnen et al. (2012). From this last publication, we show the interaction between neighbourhood social capital (higher level) and household composition (individual level) on self-rated health (Fig. 10.2).

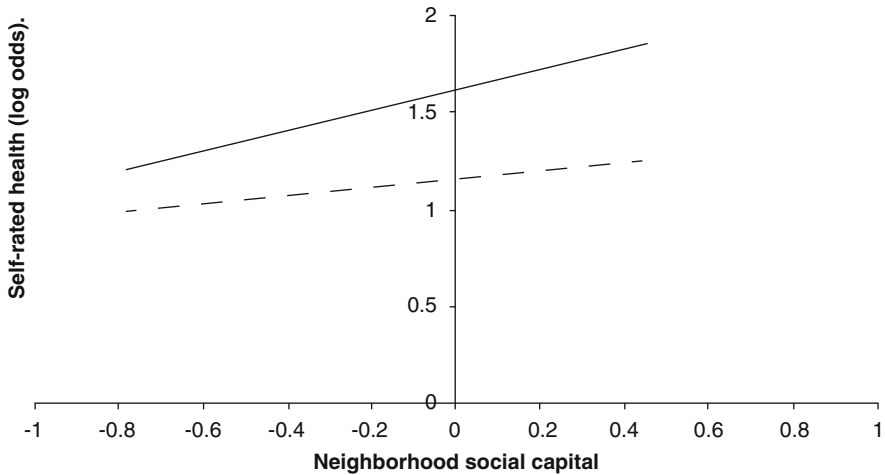


Fig. 10.2 Interaction of neighbourhood social capital and whether (black line) or not (dashed line) there are young children in the household on self-rated health (reproduced with permission from Oxford University Press, the European Journal of Public Health)

What Are the Shortcomings and Strong Points of the Article?

Try to summarise the points of criticism and try to weigh their consequences for the value of the results of the analysis that was presented. Try also to identify a number of positive points from the article you have been reading. The shortcomings are important in critical reading and they are very important in forming your overall judgement as to how confident you can be that the results of the study are indeed a valuable addition to our knowledge. However, the strong points of an article may help you in improving the formulation of your own research.

Writing Up Your Own Research

It is impossible to come up with a single form of presentation that will suit all types of analysis. The information that you need to show depends on your research question (and this is another reason for considering study design carefully before starting). Moreover, all general advice about how to write a research paper applies to papers that report on MLA and this will not be repeated here.

The Introduction or Background Section

The introduction or background section of your research paper should contain a clearly formulated research question—a grammatically well-formed sentence that

ends with a question mark. In the ‘reading’ part of this chapter, we noted the tendency of some research papers only to state an objective, which is often less clearly specified than a research question or a hypothesis.

Previous literature, where available, should be used to develop your research question and the hypotheses you intend to test. As an aid to focusing your arguments when writing the introduction, it is advisable to consider using ‘what is known about this subject?’ bullet points as required by some journals. It is important to identify the gaps in current knowledge and not just to tread a well-worn path.

Specifically when writing an article using multilevel analysis, the introduction should contain a theoretical argument as to why different levels or contexts are relevant to the particular research question. We started Chap. 1 by stressing the importance of context as an influence on people’s health, well-being, health behaviour and healthcare utilisation. This should be reflected in the attention that is given to discussing the relevant aspects of the context. In some cases the context might seem self-evident, such as in a study of health outcomes among hospitalised patients. The relevant context would then be the hospital. Even so, health outcomes are probably more strongly influenced by the particular department in which a patient was treated than the hospital as a whole. In the case when the context is a geographical unit, the link between geographical scale and area type on the one hand and the mechanism that is supposed to cause the outcome at the individual level is particularly important. If, for example, we want to analyse the relationship between social capital and health, the way in which we conceptualise social capital and the type of mechanism that we assume will influence the areal unit that we would want to use. When we conceptualise social capital as the social networks of people living in the same area, supplying each other with emotional and instrumental support, we would require smaller areal units than for a conceptualisation of social capital in terms of community resources, norms and trust (Moore et al. 2005). When the discrepancy between the size of the units used and the supposed mechanism that links the units to the outcomes is too large, it becomes increasingly difficult to draw conclusions based on your analysis of the data.

The Methods Section

The methods section firstly makes the step from the theoretical and conceptual discussion of context as it appears in the introduction or background to the concrete levels actually to be used in the data analysis. Especially when you use existing data at any of the levels, it is likely that there will be discrepancies between the theoretical context and the levels that you use in practice. It is important to describe this discrepancy and to discuss the consequences in the final section of the paper.

In the methods section, you should detail the units or levels used and the data structure. These provide the rationale for the use of MLAs. The relevant numbers (for example, the population of the areas and sample drawn from these) should be detailed.

The nature of the statistical model that you use will largely be determined by the dependent variable that you are analysing. As in any other empirical research paper,

it should be clear at what scale the dependent variable has been measured and consequently what the statistical model will be. Software packages that handle MLA differ and you should identify which package you have used.

In the days when MLA was relatively new to public health and health services research, authors used to give a general algebraic formulation of their multilevel model. Although by now more researchers are familiar with these models, it may still be useful to detail the actual model used. Particularly if the model that you are using is more complicated or in some way non-standard, providing the full formulation of the model used either in the methods section or in an appendix will aid other researchers understanding of your work and enhance its reproducibility.

The interpretation of the average outcome, variances and regression coefficients sometimes depends on the point of reference taken. Meaningful interpretation can be facilitated by centring independent variables around the mean or another relevant value. Studies do not always state whether or not they centred the data, but this should of course be mentioned.

An important element of the methods section is the description of the modelling strategy. The modelling strategy gives the steps that you are going to take in order to answer your research question or test your hypotheses. A sensible null model should be defined, and you should detail which variables are included in subsequent models and how these variables were selected.

The modelling strategy is not just a summary of the steps taken; it should contain a logical line of reasoning. Chapters 7 and 9 have discussed modelling strategies and working through the example datasets you can see modelling strategy in practice. Snijders and Bosker (2012) give helpful guidance in developing the modelling strategy.

The first step is the definition of your reference model. This might be either an empty model that only estimates the variances or a model including a few basic variables that are deemed necessary to give a fair picture of higher-level variance. The following steps introduce individual-level and/or higher-level variables. These steps are typically evaluated with reference to the first modelling step.

The methods section should enable the reader to replicate the study (at least in principle if not in reality).

The Results Section

The results section reports the findings from your study. You should give the necessary interpretation of your results, but you should also facilitate the reader's own interpretations. Consider, for example, that if variables are on different scales then the interpretation may be difficult. Some variables may be dummies, for example urbanicity may be coded as 0 (non-urban) and 1 (urban), and in the same regression analysis the proportion of the population over 65 may be included, ranging perhaps from 0.12 to 0.25. The coefficients for the two variables are then not comparable; whilst one provides an estimate of the difference between outcomes

in urban and non-urban areas, the other gives an estimate of the difference between two non-existent contexts containing no people over the age of 65 and one containing only people over 65.

In quantitative studies, tables play an important part. There are many very different ways of putting the results of an analysis into a table, without a gold standard for reporting multilevel analysis. A table (in general) should be self-contained and give an easy overview. If you want to show several consecutive models in the table, you might wish to avoid an empty column for the reference model by including the variance components in a separate table or as a footnote. If the emphasis is mainly on the higher level and you have a large number of individual-level variables, it might not be necessary to repeat this long list for each modelling step that only involves new higher-level independent variables. The coefficients of the individual-level variables may be largely invariant and could be included in a separate table or in an appendix.

The layout of any table should mirror the modelling strategy. However, it is not always necessary to present each and every step of your modelling strategy in the table. This is particularly the case if steps in the modelling process turn out not to add much information; it may be better to mention that you conducted the steps as intended but, for example, that the results or their interpretation do not differ from other reported models. This is particularly likely to be the case for sensitivity analyses. Again, full results may be reported in appendices or reported as being available from the author.

You should report the variance at the different levels. Even if variation is not at the heart of your study's research questions, it is important for other studies' power calculations. It may also be helpful for readers if you report the intraclass correlation. If your modelling strategy describes a number of subsequent models, you should probably detail changes in variance between models. If you are using logistic regression, you could consider converting variances to a meaningful scale (such as the median odds ratio or MOR; see Chap. 6).

If you report cross-level interactions, it is usually very helpful to your readers if you are able to present these graphically. An example was given earlier in this chapter in Fig. 10.2.

As the presentation of the results in tables is such an important element in terms of enabling your readers to follow and understand your results, we will give a few examples of how your results could be presented in tables. The best advice we can give is to take note when you find articles with a particularly nice presentation.

The first example is the presentation of a table for a two-level linear regression with (for example) an index of health as the dependent variable and independent variables at the individual level (such as age and gender) and at the context level (perhaps neighbourhood social capital). The table columns show the coefficients of the series of models that have been tested, starting from an empty model. The following models are one including only the individual-level variables (model 1), a model with only the contextual variables (model 2) and finally a model with both individual and contextual variables (model 3). Whether or not you need this particular sequence of models depends on your research question and hypotheses and the modelling strategy developed from your research question.

Table 10.1 Example of table layout for a two-level linear regression model

	Model 0 (empty model)	Model 1 (individual-level variables)	Model 2 (context variables)	Model 3 (individual + context variables)
<i>Fixed part</i>				
Intercept	x	x	x	x
<i>Individual variables</i>				
(e.g.) age		x		x
(e.g.) gender		x		x
<i>Context variables</i>				
(e.g.) social capital			x	x
<i>Random part</i>				
Individual-level variance	x	x	x	x
Higher-level variance	x	x	x	x

The table rows show first of all the fixed effects, starting with the overall intercept, followed by the regression coefficients for the variables at individual level and the regression coefficients at higher level. The lower part of the table shows the random part of each model. In the empty model, only the overall intercept and the two variances are estimated. The variances are the unexplained variance in our dependent variable. You could consider adding another row that shows the (change in) model fit. For a linear regression model, this could be the percentage of variance explained in subsequent (nested) models (Table 10.1).

In some cases, it might be convenient to display the random effects in a separate table. This might be the case when your model includes random slopes. The random part will then contain the variance of the slope and the covariance between the slope and the intercept in addition to the variance of the intercept. In the event of a random slope being estimated for a categorical independent variable (such as gender), a useful option is to show the higher-level variance separately for the different categories. Table 10.2 provides an illustration of models showing different formulations of the random part. Note that if variances are shown for the different categories, in this example for men and women, the higher-level intercept variance is not estimated.

The Conclusion and Discussion Section

The conclusion and discussion section should start with a concise description of your main results and, if the study tests a hypothesis, whether or not the hypothesis was refuted. It is important to relate your results to the relevant literature, particularly

Table 10.2 Example of table layout for the random part in different models

Random part	Model 0 (empty model)	Model 3 (individual + context variables) + age random	Model 3 (individual + context variables) + gender random
Individual-level variance	x		
Higher-level intercept variance	x	x	
Slope variance for age		x	
Covariance between age and intercept		x	
Higher-level variance for males			x
Higher-level variance for females			x
Covariance between males and females			x

focusing on differences in results between your study and previous studies and the likely causes of such differences. Some journals ask for a few bullet points on ‘what this paper adds’. Even if the journal does not ask for these, it is often helpful to come up with these bullet points for yourself to help to focus the discussion.

This is normally followed by the strengths and weaknesses of the study; you may want to pay particular attention to your data, study design and analytical strategy. Of course, these should be seen against the background of the strengths and weaknesses of other studies in the fields. The strengths and weaknesses should be balanced; there is no reason why this should be an exercise in masochism. If there is a long list of weaknesses and only a few strong points, the authors should probably have undertaken a different (better) study.

It is important for you to provide an interpretation of the meaning of the study. You may come back to your theoretical framework as set out at the beginning of the article and you can discuss the mechanisms underlying the results that you have found and any implications for policy or practice. Finally, it may be worth pointing out any questions that remain unanswered and make suggestions for future research.

None of the above is specific to writing up a multilevel analysis. It is generic to well written research articles and based on an article in the *British Medical Journal* on structuring the discussion section of a research paper (Docherty and Smith 1999).

Specifically in relation to the discussion section of a multilevel study, it is important to return to the appropriateness of units (and the question as to whether the units that you have used are indeed relevant contexts) and the levels that you have included and excluded.

Conclusions

In this chapter, we have brought two subjects together: critical reading of papers written by others and writing up your own multilevel research. Even if you are only using the results of other people's research, it is important to understand the basics of the methods used. We have developed a number of questions that can help you to get to grips with the multilevel methods applied in published articles. As is true for our advice about writing up your research, our advice on reading other people's research is only in part specific to multilevel analysis. Whatever the methods used, the research questions should be clear and there should be a logical modelling strategy related to the research questions and hypotheses. However, there are also specific issues such as those related to the different levels that one may hypothesise in theory and those encountered in the actual data. When it comes to writing up your research, we have also given some examples of tables. However, there is also a link between reading and writing: look for the things you like about published research, such as understandable ways of putting complicated results into tables or concise ways of formulating conclusions, and avoid forms of presentation on which you are not so keen, such as a surfeit of regression models that add little to the conclusions.

References

- Agyemang C, Vrijkotte TGM, Droomers M, Van der VWal MF, Bonsel GJ (2009) The effect of neighbourhood income and deprivation on pregnancy outcomes in Amsterdam, the Netherlands. *J Epidemiol Community Health* 63:755–760
- Arling G, Lewis T, Kane RL, Mueller C, Flood S (2007) Improving quality assessment through multilevel modelling: the case of nursing home compare. *Health Serv Res* 42:1177–1199
- Ball K, Timperio A, Salmon J, Giles-Corti B, Roberts R, Crawford D (2007) Personal, social and environmental determinants of educational inequalities in walking: a multilevel study. *J Epidemiol Community Health* 61:108–114
- Béland F, Birch S, Stoddart G (2002) Unemployment and health: contextual-level influences on the production of health in populations. *Soc Sci Med* 55:2033–2052
- Bensing JM, Caris-Verhallen WCM, Dekker J, Delnoij DMJ, Groenewegen PP (2003) Doing the right thing and doing it right: toward a framework for assessing the policy relevance of health services research. *Int J Technol Assess Health Care* 19:604–612
- Bingenheimer JB (2005) Multilevel models and scientific progress in social epidemiology. *J Epidemiol Community Health* 59:438–439
- Black JL, Macinko J, Dixon LB, Fryer GE (2010) Neighborhoods and obesity in New York City. *Health Place* 16:489–499
- Docherty M, Smith R (1999) The case for structuring the discussion of scientific papers. *Br Med J* 318:1224–1225
- Eikemo TA, Bambra C, Joyce K, Dahl E (2008) Welfare state regimes and income-related health inequalities: a comparison of 23 European countries. *Eur J Public Health* 18:593–599
- Friele RD, Coppen R, Marquet RL, Gevers JKM (2006) Explaining differences between hospitals in number of organ donors. *Am J Transplant* 6:539–543
- Groenewegen PP, Leufkens HG, Spreeuwenberg P, Worm W (1999) Neighbourhood characteristics and use of benzodiazepines in The Netherlands. *Soc Sci Med* 48:1701–1711

- Hekkert KD, Cihangir S, Kleefstra SM, Van den Berg B, Kool RB (2009) Patient satisfaction revisited: a multilevel analysis. *Soc Sci Med* 69:68–75
- Huizing AR, Hamers JPH, De Jonge J, Candel M, Berger MPF (2007) Organisational determinants in the use of physical restraints: a multilevel approach. *Soc Sci Med* 65:924–933
- Jat TR, Ng N, San Sebastian M (2011) Factors affecting the use of maternal health services in Madhya Pradesh state of India: a multilevel analysis. *Int J Equity Health* 10:59
- Joshu CE, Boehmer TK, Brownson RC, Ewing R (2008) Personal, neighbourhood and urban factors associated with obesity in the United States. *J Epidemiol Community Health* 62:202–208
- Karvonen S, Sipilä P, Martikainen P, Rahkonen O, Laaksonen M (2008) Smoking in context—a multilevel approach to smoking among females in Helsinki. *BMC Public Health* 8:134
- Matteson DW, Burr JA, Marshall JR (1998) Infant mortality: a multi-level analysis of individual and community risk factors. *Soc Sci Med* 47:1841–1854
- Mauny F, Viel JF, Handschumacher P, Sellin B (2004) Multilevel modelling and malaria: a new method for an old disease. *Int J Epidemiol* 33:1337–1344
- Mayo NE, Asano M, Barbic SP (2013) When is a research question not a research question? *J Rehabil Med* 45:513–518
- Merlo J, Gerdtam U-G, Eckerlund I, Håkansson S, Otterblad Olausson P, Pakkanen M, Lindqvist P (2005) Hospital level of care and neonatal mortality in low and high risk deliveries: reassessing the question in Sweden by multilevel analysis. *Med Care* 43:1092–1100
- Mohnen SM, Völker B, Flap H, Subramanian SV, Groenewegen PP (2012) You have to be there to enjoy it? Neighbourhood social capital and health. *Eur J Public Health* 23:33–39
- Moore S, Shiell A, Hawe P, Haines VA (2005) The privileging of communitarian ideas: citation practices and the translation of social capital into public health research. *Am J Public Health* 95:1330–1337
- Nicholson A, Rose R, Bobak M (2009) Association between attendance at religious services and self-reported health in 22 European countries. *Soc Sci Med* 69:519–528
- Pornet C, De Jardin O, Morlais F, Bouvier V, Launoy G (2011) Socioeconomic determinants for compliance to colorectal cancer screening: a multilevel analysis. *J Epidemiol Community Health* 64:318–324
- Rathmann K, Ottova V, Hurrelmann K, de Looze M, Levin K, Molcho M, Elgar F, Gabhainn SN, van Dijk JP, Richter M (2015) Macro-level determinants of young people's subjective health and health inequalities: a multilevel analysis in 27 welfare states. *Maturitas* 80:414–420
- Renders CM, Valk GD, Franse LV, Schellevis FG, van Eijk JT, Van der Wal G (2001) Long-term effectiveness of a quality improvement program for patients with type 2 diabetes in general practice. *Diabetes Care* 24:1365–1370
- Sellström E, Arnoldsson G, Bremberg S, Hjern A (2008) The neighbourhood they live in—does it matter to women's smoking habits during pregnancy? *Health Place* 14:155–166
- Sixma H, Spreuwenberg P, Zuidgeest M, Rademakers J (2009) [Consumer Quality Index hospital stay]. NIVEL, Utrecht
- Snijders TAB, Bosker RJ (2012) Multilevel analysis: an introduction to basic and advanced multilevel modeling. Sage, Los Angeles
- Stafford M, Gimeno D, Marmot MG (2008) Neighbourhood characteristics and trajectories of health functioning: a multilevel prospective analysis. *Eur J Public Health* 18:604–610
- Turrell G, Kavanagh A, Draper G, Subramanian SV (2007) Do places affect the probability of death in Australia? A multilevel study of area-level disadvantage, individual-level socioeconomic position and all-cause mortality, 1998–2000. *J Epidemiol Community Health* 61:13–19
- van Berkestijn LG, Kastein MR, Lodder A, de Melker RA, Bartelink ML (1999) How do we compare with our colleagues? Quality of general practitioner performance in consultations for non-acute abdominal complaints. *International Journal for Quality in Health Care* 11:475–486
- Van Stam M-A, Michielsens K, Stroeken K, Zijlstra BJH (2014) The impact of education and globalization on sexual and reproductive health: retrospective evidence from eastern and southern Africa. *AIDS Care* 26:379–386

- Van Yperen NW, Sijnders TAB (2000) A multilevel analysis of the demands-control model: is stress at work determined by factors at the group level or the individual level? *J Occup Health Psychol* 5:182–190
- Verstappen W, ten Riet G, van der Weijden T, Hermesen J, Grol R (2005) Variation in requests for imaging investigations by general practitioners: a multilevel analysis. *J Health Serv Res Policy* 10:25–30
- Witvliet MI, Kunst AE, Stronks K, Arah OA (2012) Assessing where vulnerable groups fare worst: a global multilevel analysis on the impact of welfare regimes on disability across different socioeconomic groups. *J Epidemiol Community Health* 66:775–781

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.

