SUNTEXT REVIEWS

SunText Review of Pharmaceutical Sciences
ISSN: 2766-5232

Open Access
Research Article
Volume 1:1

# Multilevel Analysis: Building and Testing Model a Systematic Review of Theoretical Evidences

**Mohammed SA[1,*], Workneh BD[1], Belayneh YM[1] and Hailu AD[2]**

[1]Department of Pharmacy, College of medicine and Health Sciences, Wollo University, Dessie, Ethiopia
[2]Department of Pharmacy, Dessie Health Science Collage, Dessie, Ethiopia

[*]**Corresponding author:** Mohammed SA, Department of Pharmacy, College of medicine and Health Sciences, Wollo University, Ethiopia; E-mail: ahmedsolomon21 [at] gmail [dot] com

**Received date:** 24 September 2020; **Accepted date:** 27 September 2020; **Published date:** 30 September 2020

**Citation:** Mohammed SA, Workneh BD, Belayneh YM, Hailu AD (2020). Multilevel Analysis: Building and Testing Model a Systematic Review of Theoretical Evidences. SunText Rev Pharm Sci 1(1): 105.

**DOI:** https://doi.org/10.51737/2766-5232.2020.005



## Abstract

**Background:** Multilevel modelling is an approach that can be used to handle clustered or grouped data using a statistical model that specifies an estimated relationship between variables that have been observed at different levels of a hierarchical structure. The presence of two random variables; the measurement level random variable and the subject level random variable are the feature that distinguishes the multilevel model from an ordinary regression model. The aim of this review was to review types, assumptions and model building of multilevel analysis.

**Method:** Relevant kinds of literature were searched from Google Scholar, PubMed, Hinari, Web of Science, Scopus, and Science Direct. A total of 426 kinds of literature were searched. After the exclusion of redundant and irrelevant literature, 15 kinds of literature were reviewed.

**Result:** The full multilevel regression model assumes that their hierarchal data set, with one single dependent variable at all existing levels. The multilevel model can be built by estimating regression lines separately for each unit or within unit relationships for each unit and summarizes unit relationships between group variables variance in intercepts and slopes predicted by between unit variables. Multilevel models have similar assumptions with major general linear models except some of the assumptions are modified for the hierarchical nature of the design.

**Conclusion:** Multilevel modelling is generalized linear modelling where regression coefficients themselves given a model. The models of parameters vary at more than one level that have a hierarchical structure in which the dependent variable is measured at the lowest level and the independent variables are measured at all available levels.

**Keywords:** Multilevel; Model; Assumption; In combination

## Introduction

Social research regularly involves problems that investigate the relationship between individuals and society where individuals interact with the social groups, to which they belong, that individual persons and the social groups are influenced by each other. The individuals and the social groups are conceptualized as a hierarchical system of individuals nested within groups, with individuals and groups defined at separate levels of this hierarchical system. Naturally, such systems can be observed at different hierarchical levels, and variables may be defined at each level [1]. This leads to a statistical model that specifies and estimates relationships between variables that have been observed at different levels of a hierarchical structure referred to as multilevel research. Multilevel modelling is an approach that can be used to handle clustered or grouped data [2].

Multilevel models (also known as hierarchical linear models, nested models, mixed models, random coefficient, random-effects models, random parameter models, or split-plot designs) are statistical models that vary more than one level. The feature that distinguishes the multilevel model from an ordinary regression model is the presence of two random variables; the measurement level random variable and the subject level random variable. Because multilevel models contain a mix of fixed and random effects, they are sometimes known as mixed-effects models [2].

Research questions demand multilevel analysis when health simultaneously affected at the level of individuals and at the level of contexts. Multilevel relates to the levels of analysis in public health research consists of individuals (at lower level) who are nested within spatial units (at higher levels). The observations that are being analysed are correlated or clustered along spatial, non-spatial, or/and temporal dimensions; or the causal processes are thought to operate simultaneously at more than one level; and/or there is an intrinsic interest in describing the variability and heterogeneity in the population, over and above the focus on average relationships [3]. The objective of this review was to systematically review types, assumptions and model building of multilevel analysis.

## Methodology

### Search strategy, inclusion and exclusion criteria

The review was conducted by adapting a search strategy in identified databases. Books that were peer-reviewed and non-reviewed with the time period of 1991 to present and those written in English were included in the review.

The search for literatures was conducted in three separate ways: searches in electronic databases on the internet, hand searches and iterative reviews of reference lists of papers. The databases searched were PubMed, Hinari and Goggle Scholar. The search was conducted using the following search terms: ‘Multilevel’, ‘Model’, ‘Assumption’, and in combination. All searches were performed from September 3-8, 2020.

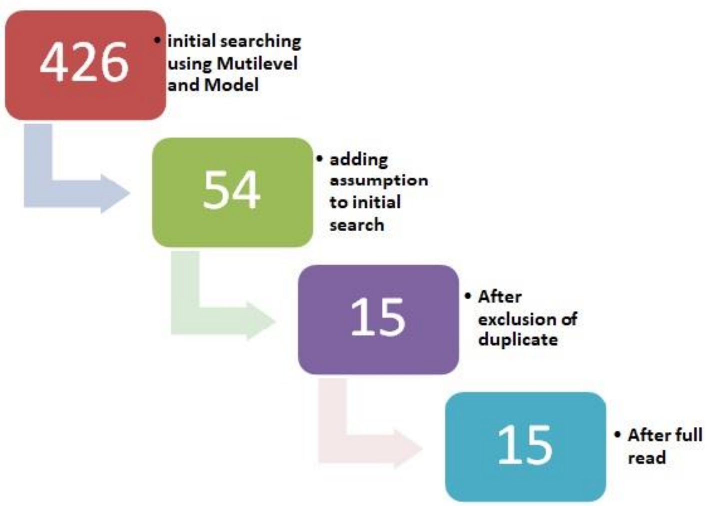


***Figure 1:*** *Data searching process.*

When the electronic search using ‘Multilevel’, and ‘Model’ was done, as shown in (Figure 1) 426 studies were available from all sources. Then, adding the search engines ‘Assumption’ 54 studies were availed. From this study, 38 were excluded because of duplication and non-fulfilment of inclusion criteria. Finally, a total of 15 separate studies were selected for inclusion in the systematic review, out of these 4 is published articles in peer-reviewed journals and 11 are books.

### Assessment of methodological quality

Methodological validity was checked prior to inclusion of selected articles and during the review by undertaking critical appraisal using preferred reporting items for systematic reviews and meta-analysis (PRISMA) flow diagram and guidance set out by the centre for reviews and dissemination.

There were four reviewers in this review and three reviewers appraised the full text of each study independently. Any discrepancies between the three reviewers were resolved through discussion and/or involving a fourth reviewer as arbiter. Finally, fourth reviewers validate the final selection of publications.

## Results and Discussion

### Type and nature of variables

Variables can be defined at any level of the hierarchy in the multi-level analysis. At each level in the hierarchy, we may have several types of variables. In multilevel data, there is not one ‘proper’ level at which the data should be analysed. Rather, all levels present in the data are important in their own way. This becomes clear when we investigate cross-level hypotheses, or multilevel problems [4].

The multilevel models assume hierarchical data, in which the dependent variable is measured at the lowest level and the independent variables are measured at all available levels [5]. Structural data are hierarchical or nested data are likely to be correlated [6]. Clustered data (nested data) are measurements are taken on subjects that share a common category that leads to correlation. Clustered longitudinal data outcome is measured repeatedly for the same subject over time, and subjects are clustered within some unit. In a clustered data, each “level” represents a factor that can be thought of as a random sample from a larger population, otherwise classification [7].

The fixed effects are regression coefficients; in which values of interest are all represented in the dataset while random effects are variations of regression coefficients between levels (variance components) ever-existing natural heterogeneity among subjects. Cross-level interaction effects are fixed effects of the joint effect of variables at level one in conjunction with variables at level two. Mixed-effects models combine both factor(s) [8].

### Types of multilevel models

**Random intercepts model:** In this model, the intercepts are allowed to vary, so the dependent variable for each individual observation will be predicted by the intercept that varies across groups [9]. The model also assumes that slopes are fixed (the same across different contexts). In addition, this model provides information about intra class correlations, which are helpful in determining whether multilevel models are required in the first place [10].

**Random slopes model:** A random slopes model is a model in which slopes are allowed to vary, and therefore, the slopes are different across groups. This model assumes that intercepts are fixed (the same across different contexts) [10].

**Random intercepts and slopes model:** In this model, both intercepts and slopes are allowed to vary across groups, meaning that they are different in different contexts [10].

## Statistical model analysis

The multilevel model can be conceptualized as a two-stage equation.

Step 1: Estimates separate regression equations within units. Relationships within units (intercept and slopes).

Step 2: Uses "summaries" of between unit relationships as group variables.

## Mathematically

Level 1: Regression lines estimated separately for each unit or within unit relationships for each unit. Analyse the model with no explanatory variables. The intercept only model is given by the model of equations (1).

$$Y = ß_{0j} + ß_{1j} X_{ij} + \varepsilon_{ij} \qquad (1)$$

Where: $ß_{0j}$ and $ß_{1j}$ = regression coefficients

Level 2: Variance in intercepts and slopes are predicted by between unit variables. Level 2 models variance in level-1 parameters (intercepts and slopes) with between unit variables. Analyse the model with all lower level explanatory variables.

$$ß_{0j} = \gamma_{00} + \gamma_{01} (\text{Group j}) + U_{0j} \qquad (2)$$

$$ß_{1j} = \gamma_{10} + \gamma_{11} (\text{Group j}) + U_{1j} \qquad (3)$$

Adding cross level interaction between explanatory group level variables and those individual level explanatory variables leads a full model formulated in equation (4). So, substitute equations (2) and (3) in equation [1]:

$$Y_{ij} = \gamma_{00} + \gamma_{01g\,j} + \gamma_{10} I_{i\,j} + \gamma_{11g\,j}\, I_{ij} + U_{0j} + U_{1j}\, I_{ij} + \varepsilon_{ij} \qquad (4)$$

Where:

$\gamma_{00}$ is variances of group intercepts (over all intercept)

$\gamma_{10}$ is regression coefficient (slop of individual variable)

$\gamma_{11}$ is slop of interaction (how large Z affect X)

$\gamma_{01}$ is variances of group slopes (group regression coefficient)

$U_{0j}$ and U 1j are errors in the group-level equations;

$U_{0j}$ is group level deviation of each intercept from all intercept ($\gamma_{00}\gamma_{00}$)

The full multilevel regression model assumes that their hierarchal data set, with one single dependent variable at all existing levels. For example, Joop Hox's multilevel analysis [4] assume data from J classes, with a different number of pupils nj in each class. On the pupil level, the outcome variable 'popularity' (Y), measured by a self-rating scale that ranges from 0 (very unpopular) to 10 (very popular). They have two explanatory variables on the pupil level: pupil gender (X1: 0 = boy, 1 = girl) and pupil extraversion (X2, measured on a self-rating scale ranging from 1 to 10), and one class level explanatory variable teacher experience (Z: in years, ranging from 2 to 25). Using variable labels the equation is:

$$\text{Popularity}_{ij} = \beta_{0j} + \beta_{1j}\, \text{gender}_{ij} + \beta_{2j}\, \text{extraversion}_{ij} + e_{ij} \qquad (5)$$

In this regression equation, β0j is the intercept; β1j is the regression coefficient (regression slope) for the dichotomous explanatory variable gender, β2j is the regression coefficient (slope) for the continuous explanatory variable (extraversion), and the usual residual error term is $e_{ij}$. The subscript j is for the classes (j = 1 . . . J) and the subscript i is for individual pupils (i = 1 . . . nj).

The next step in the hierarchical regression model is to explain the variation of the regression coefficients βj introducing explanatory variables at the class level:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} Z_j + u_{0j} \qquad (6)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11} Z_j + u_{1j} \qquad (7)$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21} Z_j + u_{2j} \qquad (8)$$

The model with two pupil-level and one class-level explanatory variable can be written as a single complex regression equation by substituting equations (6), (7) and (8) into equation (5) gives:

$$\text{Popularity}_{ij} = \gamma_{00} + \gamma_{10}\, \text{gender}_{ij} + \gamma_{20}\, \text{extraversion}_{ij} + \gamma_{01}\, \text{experience}_j + \gamma_{11}\, \text{gender}_{ij} \times \text{experience}_j + \gamma_{21}\, \text{extraversion}_{ij} \times \text{experience}_j + u_{1j}\, \text{gender}_{ij} + u_{2j}\, \text{extraversion}_{ij} + u_{0j} + e_{ij} \qquad (9)$$

## Sample size determination

It is generally accepted that increasing sample sizes at all levels estimates and standard errors improve. To be statistically safe, as "rule of thumb", researchers should use '30/30' rule, a sample of at least 30 groups with at least 30 individuals per group. On the other hand, the numbers should be modified as if there is strong interest in cross-level interactions, the number of groups should be larger, (a 50/20 rule-50 groups with 20 individuals/ group); if there is stronger interest in the random part, or in the variance and/ or covariance components, the number of involving groups should be larger, leading to a 100/10 rule (100 groups with 10 individuals/group). One should take into account the costs attached to data collection, so if the number of groups is increased, than the number of individuals per group might decreases [5].

## Assumptions of multilevel modeling

Multilevel models have the same assumptions as other major general linear models (e.g., Anova, regression), but some of the assumptions are modified for the hierarchical nature of the design (i.e., nested data). Accordingly, checking and improving the specification of a multilevel model in many cases can be carried out while staying within the assumption of the multilevel model.

### Linearity

The assumption of linearity states that there is a rectilinear (straight-line, as opposed to non-linear or U-shaped) relationship between variables [11]. However, the model can be extended to nonlinear relationships [12]. A regression analysis expected to fit the best rectilinear line that explains the most data given your set of parameters. Therefore, the base models rely on the assumption that the data follow a straight line (though the models can be expanded to handle curvilinear data). Graphically, by plotting the model residuals (the difference between the observed value and the model-estimated value) versus the predictor, linearity can be tested. If a pattern emerges, a higher-order term may need to be included or you may need to mathematically transform a predictor/response [13].

### Normality

The assumption of normality states that the error terms at every level of the model are normally distributed [11]. QQ plots which are obtained in standard regression modeling in R can provide an estimation of where the standardized residuals lie with respect to normal quintiles. Strong deviation from the provided line indicates that the residuals themselves are not normally distributed [13].

### Homoscedasticity

The assumption of homoscedasticity or homogeneity of variance assumes equality of population variances. However, different variance-correlation matrix can be specified and the heterogeneity of variance can itself be modelled [11]. In R, we extract the residuals from the model, place them in our original table, take their absolute value, and then square them (for a more robust analysis with respect to issues of normality. Finally, take a look at the ANOVA of the between-subjects residuals [13].

### Independence of observations

Independence of observation is an assumption which states that cases are random samples from the population and that scores on the dependent variable are independent of each other [11].

### Model fitness

One way of assessing model fit is the chi-square likelihood-ratio test, which assesses the difference between models. The likelihood-ratio test can only be used when models are nested. It can be used for examining what happens when effects in a model are allowed to vary, and when testing a dummy-coded categorical variable as a single effect. However, when testing non-nested models, comparisons between models can be made using the Akaike information criterion or the Bayesian information criterion, among others [10,14].

### Statistical tests and power

The types of statistical tests in multilevel models depend on whether one is examining fixed effects or variance components. When examining fixed effects, the tests should be compared with the standard error of the fixed effect, which results in a Z-test. A t-test can also be computed. When computing a t-test, it is important to consider degrees of freedom. For a level one predictor, the degrees of freedom are based on the number of level one predictor, the number of groups and the number of individual observations. For a level two predictor, the degrees of freedom are based on the number of level two predictors and the number of groups [10].

Statistical power for multilevel models differs depending on whether it is level one or level two effects that are being examined. Power for level one effect is dependent upon the number of individual observations, whereas the power for level two effects is dependent upon the number of groups [14].

### Benefits of multilevel analysis

The multilevel approach offers several advantages. First, the result can generalize to a wider population. Second, fewer parameters are needed. The dummy variables approach would require 25 additional parameters. In the handling of more complex models and a limited amount of data, a reduction in the number of parameters is important. Third, information can be shared between groups. This is due to assuming that the random effects resulted from a common distribution. As a result, the precision of predictions for groups that have relatively little data improves. Finally, it can deal with data in which the times of the measurements vary from subject to subject [2].

### Limitation of multilevel analysis

Analysing variables from different levels at one single common level leads to two distinct types of problems. The first problem is statistical. If data are aggregated, the result is that different data values from many sub-units are combined into fewer values for fewer higher-level units. As a result, statistical analysis loses power due to information loss. On the other hand, for a larger number of sub-units, few data values from a small number of super-units will ‘blown up’ into many more values, if data are disaggregated. The second problem is conceptual. That is analysing the data at one level and formulating conclusions at another level [4].

## Acknowledgements

We are thankful to Wollo University for providing the necessary facilities to conduct this review.

## Declaration of Interest

The author has no relevant affiliations or financial involvement with a financial interest in or financial with the subject matter or materials discussed in the manuscript.

## Availability of Data and Materials

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

## Authors' Contributions

All authors were involved in searching for literatures and write up of the manuscript. All authors read and approved the final draft of the manuscript.

## Conflicts of Interest

There is no conflict of interest.

## References

1. Hox J. Applied multilevel analysis. TT-publikaties. 1995.
2. Richard B. Statistics: Multilevel modeling. Mathematics learning support system. 2008.
3. Levels and Variables. Multilevel Modeling.
4. Hox J, Moerbeek M, Van de Schoot R. Multilevel analysis: Techniques and applications. Routledge. 2010.
5. Jula NM. Multilevel model analysis using R. Romanian Statistical Review. 2014; 62: 55-66.
6. Wang W, Gelman A. Difficulty of selsecting among multilevel models using predictive accuracy. Statistics and Its Interface. 2015; 8: 153-160.
7. Raudenbush S, Bryk A. Hierarchical linear models: Applications and data analysis methods. Sage. 2002.
8. Muthén B, Asparouhov T. Beyond multilevel regression modeling: Multilevel analysis in a general latent variable framework. Handbook of advanced multilevel analysis. 2011; 15-40.
9. De Leeuw J, Meijer E, Goldstein H. Handbook of multilevel analysis. Springer. 2008.
10. Cohen J, Cohen P, West S, Aiken L. Applied multiple regression/correlation analysis for the behavioral sciences. Routledge. 2013
11. Harvey G. Nonlinear Multilevel Models, with an Application to Discrete Response Data. Biometrika. 1991; 78: 45-51.
12. Palmeri M. Testing the assumptions of multilevel models.
13. Kreft I, De Leeuw J. Introducing multilevel modeling. Sage. 1998.
14. Andrew G. Multilevel (Hierarchical) Modeling: What It Can and Cannot Do. American Statistical Association and the American Society for Quality Technometrics. 2006; 48: 432-435.