

Review Article

Application of Hierarchical/Multilevel Models and Quality of Reporting (2010–2020): A Systematic Review

Killian Asampana Asosega (),^{1,2} Atinuke Olusola Adebanji (),¹ Eric Nimako Aidoo (),¹ and Ellis Owusu-Dabo ()³

¹Department of Statistics and Actuarial Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana ²Department of Mathematics and Statistics, University of Energy and Natural Resources, Sunyani, Ghana ³Department of Global and International Health, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana

Correspondence should be addressed to Killian Asampana Asosega; killian.asosega@uenr.edu.gh

Received 5 January 2024; Revised 14 February 2024; Accepted 24 February 2024; Published 8 March 2024

Academic Editor: Luigi Cerulo

Copyright © 2024 Killian Asampana Asosega et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Introduction. Multilevel models have gained immense popularity across almost every discipline due to the presence of hierarchy in most data and phenomena. In this paper, we present a systematic review on the adoption and application of multilevel models and the important information reported on the results generated from the use of these models. Methods. The review was performed by searching Google Scholar for original research articles on the application of multilevel models published between 2010 and 2020. The search strategy involved topics such as "multilevel models," "hierarchical linear models," and "mixed models with hierarchy." The search placed more emphasis on the application of hierarchical models in any discipline but excluded software methodological development and related articles. Results. A total of 121 articles were initially obtained from the search results. However, 65 articles met the inclusion criteria for the review. Out of the 65 articles reviewed, 46.2% were related to health/epidemiology, 15.4% to education and psychology, and 16.9% to social life. The majority of the articles (78.5%) were two-level models, and most of these studies modelled univariate responses. However, the few that modelled more than one response modelled them separately. Moreover, 83.1% were cross-sectional design, and 9.2% and 6.2% were longitudinal and repeated measures, respectively. Moreover, a little over half (55.4%) of articles reported on the intraclass correlation measure, and all articles indicated the response variable distribution where most (47.7%) were normally distributed. Only 58.5% of articles reported on the estimation methods used as Bayesian (20%) and MLE (18.5%). Again, model validation measures and statistical software were reported in 70.8% and 90.8% articles, respectively. Conclusion. There is an increase in the utilization of multilevel modelling in the last decade, which could be attributed to the presence of clustered and hierarchically correlated data structures. There is a need for improvement in the area of measurement and reporting on the intraclass correlation, parameter estimation, and variable selection measures to further improve the quality of the application of multilevel models. The integration of spatial effects into multilevel models is very limited and needs to be explored in the future.

1. Introduction

Multilevel analysis is a collection of statistical techniques used to examine the relationship between variables characterizing individuals and those characterizing groups [1]. Multilevel modelling usually has to do with data that are hierarchically structured in nature. In multilevel modelling, related research variables can be defined and declared at any level of the hierarchy based on the focus and data structure. While some variables can be measured at their own natural level, others have to be moved from one level to another by either aggregation or disaggregation [1]. In multilevel modelling, the finer scale at which the response or dependent variable is measured is called the lower level, whereas the aggregate scale is referred to as the higher level. Multilevel models usually anticipate both differences between the higher-level units and dependence within those units. Within observations, dependences are mostly anticipated because their members are assumed to be influenced by the same aggregate effects [2]. Moreover, the existence of group dependence among lower-level units violates the fundamental assumption of independence in a standard regression analysis, resulting in an increase in the risk of inefficient model estimation and inappropriate inferences with respect to parameter estimates [2–4].

Multilevel models, also known as hierarchical linear models (HLM), have in recent times received attention and utilization from several disciplines, especially in the social, educational, biological, and medical fields where datasets are usually nested [5]. Hierarchical linear models are also very useful in longitudinal data structures, where measurements measured at different points in time are nested within the observations or units on which those measurements were made. Modelling of the outcome variable in these situations presents a flexible way to appropriately capture and account for the nested data structure to ensure that standard errors and model parameters are accurately estimated [4, 6].

A study by El-Horbaty and Hanafy [7] argued that most problems in social sciences often involve investigations of the relationship between individual and society-level characteristics. The general notion is that individuals and their social groups are conceptualized as a hierarchical system, where the individuals and groups are defined at separate levels of this hierarchical system. There exist several forms of hierarchical models, which vary in terms of the type of design (random intercept, random slopes, and random coefficients regression model), number of levels, measurement scale of the dependent variable (continuous and categorical), and number of measured responses (univariate and multivariate).

Hierarchical analyses are performed on data that have some form of nested structure. Data with nested structures are often associated with some form of dependency. For instance, in a corporate working environment, individuals in the same department or unit often showcase similar performance and provide similar responses to questions about aspects of the work environment. According to [8], the presence of nonindependence in any given dataset may be considered either a nuisance variable or something to be fundamentally appreciated, but the prevalence and growing presence of nested data require a variety of statistical tools to easily handle nested data.

The term "multilevel analysis" has been used most often to describe the set of statistical analyses, also referred to as random coefficient models, random-effects, and mixedeffects models [9–11]. Hierarchical models are generally suitable for dealing with nonindependence.

Several studies [12–14] have shown that most geographical processes or events often display strong differences between locations or regions. Empirical results from physical geography have established that this form of *spatial heterogeneity* has long been known as very vital to the accuracy of models [15]. According to Browne and Goldstein [12], the most effective way to address spatial heterogeneities that have been observed extensively in geography as well as other areas where clustered or hierarchical structured data are observed is multilevel modelling. In a related study, Gelman [16] also observed that random-effects models are very desirable across several disciplines of science precisely due to their ability to explain between and within-group heterogeneity.

Again, Wolf et al. [17] emphasized that multilevel models have been used for quite a long time across a variety of geographical problems. For instance, related studies in epidemiology employ multilevel models to estimate the effect between conditions such as smoking, asthma, mental health, and cancer [18–21]. Moreover, multilevel models are used in areas such as economic and urban structure analysis [22–24]. In economics, multilevel models have been used to highlight the impact of economic and individual indicators on the well-being of citizens [25], country and industry factors on job demand [26], and on privatization of state enterprises [27]. Multilevel models have also been utilized to examine country- and individual-level influences on sanitation and environmental-related issues [28, 29].

In real life problems, the usual assumption of statistical independence of observations does not hold for nested structured data or data with an underlying hierarchy. More often than not, however, the populations from which these data are collected or generated usually have multifaceted structures where measurements on data units are not mutually independent but depend on each other through some form of underlying complex structural relationships. These multifaceted structured relationships cannot be effectively captured and accounted for by models that assume independence of sampled observations from a given population [1, 30, 31].

Methodological details in scientific research are very important as they have a significant impact on the validity of statistical results and inferences arising from the generated results. Hence, insufficient methodological details in scientific articles have significant setbacks in terms of study replication, inaccurate parameter estimates, and subsequent invalid inferences. For instance, according to Hox et al. [32], analysis of variables at the wrong level leads to statistical and conceptual problems, which affect the statistical power of the analysis performed. In addition, insufficient methodological details particularly in multilevel modelling could result in ignoring the nested structure of datasets and could yield inaccurate and imprecise parameter estimates and their corresponding standard errors [33, 34]. Lack of sufficient details of the methodology used could affect the trustworthiness of study findings [4].

Despite the continuous increase in utilization of multilevel models across diverse fields of endeavour, there are no clear standardized directions and guidelines on the usage and reporting of the generated results. Moreover, there is limited review of literature on the development and applications of multilevel models in the last decade, a gap this study seeks to bridge.

This study seeks to provide the current update by means of a systematic review of some relevant information on the use of multilevel models in the existing literature in the last decade (2010–2020) reported across several areas of application. The time scale was purposely chosen to understand how reporting standards in the utilization of multilevel models have been and evolved over the period which will inform and direct further review in subsequent decades. Moreover, the review will help identify the limitations and strengths of these models in the literature which could be leveraged upon to improve the quality of statistical information reported in the use of multilevel models.

Multilevel models are accredited and desirable across a wide range of disciplines due to their ability to account for group heterogeneity [16]. Multilevel models further acknowledge correlations in observations within the same group which is popularly known in geography as platial or horizontal dependence [35]. However, platial dependence for correlation among observations from the same place is different from the Anselin [36] spatial dependence concept which accounts for the fact that nearby observations are more related irrespective of regional boundaries [37]. Accounting for both spatial and platial dependence in one model results in improvements in the precision and accuracy of parameter estimates and enhances the validity and reliability of inferences [35, 38]. Also, Wolf et al. [17] highlight the powerfulness of multilevel models to account for spatial dependence. They further recommend the blend of spatial and platial effects to maximize the advantages presented in both. Generally, MLMs are utilized to yield narrow interval estimates with some biasedness, whereas spatial models produced unbiased estimates but wide intervals, and hence, the integration of both multilevel and spatial effects provides improvements in results generated. Moreover, ignoring spatial effects in multilevel models could result in more extreme estimates but narrower confidence intervals. This stems from the underestimation of standard errors which could lead to serious consequences as nearby spillover effects could remain unaccounted for in the model [32, 39].

2. Methods

This systematic review was carried out in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) statements [40, 41]. This study further reported the review in line with the PRISMA guidelines.

This review considered original articles published between 2010 and 2020 inclusive, and the search was performed using Google Scholar. The search strategy adopted the following topics: "hierarchical linear model," "multilevel model," and "mixed model." However, the area of application of these models was not restricted.

2.1. Selection of Studies Included in the Review. Articles obtained from the search were eligible for inclusion if they were original research articles and must have been written in English peer-reviewed journals reporting on the application of multilevel models. Articles on statistical development of the methodology were also excluded from the review.

2.2. Identification of Studies. The selected articles for the review were used to extract some information presented in Tables 1 and 2. The PRISMA flowchart (Figure 1) was used to

summarize all article selection process [40, 41]. During the first review phase, a total of 121 articles were identified, out of which 11 were duplicates. Moreover, after a careful examination of the abstracts of the selected studies, some were excluded because they were nonoriginal articles (conference proceedings) and those that did not have either multilevel or hierarchical modelling as a keyword or abstract or in the title of the article.

In the second phase of the review, out of the 96 articles considered, only 76 pertained to multilevel/hierarchical modelling applications to hierarchically structured data. Hence, 11 articles were further excluded due to nonexistence and inconsistency.

The third aspect of the review looked at the full-text versions of the eligible articles included in this review. A comprehensive review of the 65 articles was conducted with emphasis on the number of levels, number of observations at the lowest and highest levels, method of estimation, method of model validation, presence of outliers in datasets, measurement of intraclass correlation (ICC), number of response variables, presence of spatial effects, and software (packages/macros) used, and others are discussed in the results section of this review.

2.3. Study Characteristics. For the study design, the review considered several aspects of each article. Some of the key aspects include the number of levels in the hierarchy, estimation method, sample sizes at both the lowest and highest levels of the hierarchy, type of response variable number or response variables, type of study, data source and sampling techniques used, model validation measures, presence of outliers and how they are handled, and measurement of ICC which is very fundamental in hierarchical modelling framework.

In addition, the review also looked at whether or not the distribution of the response variable was indicated as well as assumptions concerning the MLM and whether those assumptions were met or not based on empirical statistics presented in the articles selected for this review. The study also investigated whether spatial components were included in the multilevel and how they were captured in the modelling framework to help direct future focus on the topic.

2.4. Inferential Information on Studies. This section presents some important inferential indicators concerning multilevel models in general. The hierarchical model is characterized by random and fixed effects or components, where the random components are usually associated with grouping (higher)level variables within the hierarchy which account for the grouping-level variables in the model.

Moreover, the review reported on the estimation techniques as well as the software employed in each selected study since these usually have effects on the validity of parameter estimates and inferences [42, 43]. Besides, the computational times, flexibility, and user-friendliness of these software programs vary considerably. Several software programs are available for implementing hierarchical

TABLE 1: Characteristics of reviewed articles.

Area of application	<i>n</i> = 65
Climate science	1 (1.5%)
Economics	6 (9.2%)
Education/psychology	10 (15.4%)
Health/epidemiology	30 (46.2%)
Human development	2 (3.1%)
Planning	2 (3.1%)
Social life	11 (16.9%)
Sports	3 (4.6%)
Number of levels	
Two	51 (78.5%)
Three	13 (20.0%)
Four	1 (1.5%)
Number of response variables	
One	58 (89.2%)
Two	4 (6.2%)
Three	1 (1.5%)
Four	2 (3.1%)
Lower-level units	
Individuals/human	45 (69.2%)
Location	5 (7.7%)
Time period	3 (4.6%)
Firms	2 (3.1%)
Publications/manuscripts	3 (4.6%)
Households	3 (4.6%)
Properties	2 (3.1%)
Measurements	1 (1.5%)
Angles	1 (1.5%)
Study design	
Cross-sectional	54 (83.1%)
Experimental	1 (1.5%)
Longitudinal	6 (9.2%)
Repeated measures	4 (6.2%)
Measurement of ICC	
Yes	36 (55.4%)
No	29 (44.6%)

modelling. These programs include SPSS (version 19 and above) which includes a GLMM framework in the GEN-LINMIXED procedure [44], SAS macro GLIMMIX which was the very first software introduced for HLM based on the penalized quasi-likelihood (PQL) estimation procedure, lme4 [45], glmml, MASS(glmm PQL functionality), most recent sjstats [46], R2jags in *R*, and xtmixed and gllamm functions in STATA [47, 48].

The significance or hypothesis testing regarding the fixed and random effects in multilevel models is usually tested and assessed separately. The fixed-effects components of MLM are usually assessed by the Wald test, while those of the random-effects variances are tested through LRT or by comparing the goodness-of-fit measures of the models using the Bayesian and Akaike information criteria [49].

The variable selection strategy used in the respective studies selected under review is also essential and worth noting and considering. The most commonly used variable selection is the stepwise (forward or backward) selection strategy. This is usually based on the model fit indices such as AIC, BIC, or DIC and the Wald test on model parameters to examine their respective contributions in the fitted models. Notwithstanding, some models do not employ any variable

TABLE 2: Characteristics of statistical inference and estimation methods.

Method of estimation	
Bayesian	13 (20.0%)
GEE	1 (1.5%)
GLS	2 (3.1%)
MLE	12 (18.5%)
PQL	1 (1.5%)
ReML	3 (4.6%)
ReML and Bayesian	1 (1.5%)
ReML and MLE	3 (4.6%)
Weighted least squares	2 (3.1%)
NI	27 (41.5%)
Distribution of response	
Binary	24 (36.9%)
Multinomial	1 (1.5%)
Normal	31 (47.7%)
Normal and binary	2 (3.1%)
Ordinal and binary	1 (1.5%)
Ordinal	5 (7.7%)
Poisson	1 (1.5%)
Model validation measures	
AIC	3 (4.6%)
BIC	2 (3.1%)
AIC and BIC	7 (10.8%)
AIC, BIC, and LRT	11 (16.9%)
DIC	7 (10.8%)
<i>R</i> -square	3 (4.6%)
	5 (7.7%)
RMSEA/MSE	6 (9.2%)
HDI/QIC	2(3.1%)
	19 (29.2%)
Statistical software	(0.20)
	6(9.2%)
Malua	$\delta(12.5\%)$
Mipius D	4(0.2%)
	13(20.0%)
SAS SAS and Mlwin	9(13.8%)
	1(1.3%)
SF 55	5(7.7%)
Stata State and Mlwin	9(13.0%)
WinBLICS	2(3.1%)
WIIIDOG5	2(3.170) 6(0.2%)
Statistical packages	0 (9.2%)
GLIMMIX	5 (77%)
GLIAMM	1(1.5%)
HSAP	1(1.5%)
IACS (riage runiage and coda)	1(1.570) 2(31%)
lme4/lmer	$\frac{2}{4}(6.2\%)$
lme4 spden and sistats	2(31%)
Nmle	2(3.1%)
PROC MIXED/NLMIXED	4(62%)
VCMMR estimation	1(1.5%)
xtlogit	1(1.5%)
xtmixed	1(1.5%)
NI	41 (63.1%)

selection procedure as in the case of confirmatory analysis where a specific hypothesized model of interest is fit, usually based on previous research or recommendations from theory and experience [33].



FIGURE 1: Flowchart for the selection of articles. Source: BMJ 2021; 372: n71. doi: 10.1136/bmj.n71. https://www.prisma-statement.org/.

3. Results

The development and application of multilevel models across several disciplines have received and continuous growing interest in recent times in the scientific literature. From the initial 121 articles obtained from the search, the detailed results presented are regarding the 65 articles included in the final detailed review.

Out of the final sixty-five (65) articles, the highest and second highest of 10 (15.4%) and 8 (12.3%) of the articles were published or recorded in 2020 and 2019, respectively (Figure 2). Thirty articles (46.2%) were applied in health/epidemiology, 11 (16.9%) in social life, 10 (15.4%) in education/psychology, and the least of 1 article in the area of climate science. On the number of levels in the multilevel structure of the models used, 51 articles (78.5%) were at 2-level models and 13 (20.0%) were 3-level hierarchical models. Moreover, over eighty percent were cross-sectional studies, whereas 15.4% (10 articles) were longitudinal/repeated measure studies, and the only article was an experimental study.

A total of 55 articles (84.6%) reported on the sample sizes at the lowest level (level-1) of the multilevel structure, which ranged from 6 to 5,700,000 observations with a corresponding median of 7,079 (Q1 = 686; Q3 = 44,906). With respect to 2-level multilevel models, 46 (90.2%) articles reported the sample size at the lowest level, which ranged from 244 to 5,700,000 (Q1 = 1077; Q3 = 49,062), while the 3-level hierarchical model recorded sample sizes ranging from 6 to 231, 326 with a median of 2,498 (Q1 = 59; Q3 = 17,296) observations at the lower level.



FIGURE 2: Number of reviewed articles by years.

The lower-level units in the reviewed articles selected mostly (69.2%) referred to individuals (such as patients, customers, students, nurses, employees, and women) irrespective of the number of levels in HLM. The number of response variables in the selected studies varied from one (1) response (n = 58; 89.2%) to 4 (3.1%) response variables. However, all two or more response variables in the articles were modelled separately as univariate variables one after the other rather than as a multivariate variable. The distribution of the response variables was reported in all articles as normal (n = 31; 47.7%), binomial/binary (n = 24; 36.9%), ordinal (n = 5; 7.7%), normal and binary (n = 2; 3.1%), and

one (1) each for ordinal and binomial, Poisson, and multinomial distributions as available software can fit or model a variety of distributions as response variables.

For the estimation method used in each of the reviewed articles, the maximum likelihood (MLE) was implemented in 12 articles: Bayesian through MCMC (n = 13), restricted maximum likelihood (ReML; n = 3), generalized least squares (GLS, n = 2), weighted least squares (WLS, n = 2), ReML and MLE (n = 3), ReML and Bayesian (n = 1), and generalized estimation equations (GEE; n = 1). These estimation techniques were mostly implemented in R (n = 13), SAS (n = 9), Stata (n = 9), Mlwin (n = 2), HLM (n = 6), SPSS (n = 5), MPlus (n = 5), and Stata and Mlwin (n = 2). Articles in which R-software was used employed packages such as lme4, Sjstats, runjags, lmer, spdep, Jags, and HSAR. For SAS, the packages used included Glimmix and PRO MIXED, and in Stata, glamm, xtlogit, and xtmixed were used.

Moreover, 55.4% of the articles selected presented intraclass correlation (ICC) measures, which quantify the group (cluster)-level effects on the response measured at the lowest level. Model validation (goodness of fit) measures were indicated in 45 (73.8%) articles using indices such as Akaike information, Bayesian information, deviance information, quasi-likelihood information criteria, and likelihood ratio test (LRT) with some reporting at least one of these fit indices. Spatial effects were considered in only 4 (6.6%) out of the 61 articles reviewed. Moreover, the presence of outliers was reported in 3 articles whereby one of such removed the outliers before model fitting.

4. Discussion

Multilevel models have in recent times gained popularity across several disciplines such as geography [12, 13], education and psychology [50-53], sports [54, 55], public health and epidemiology [56-60], and economics [26, 38, 61]. However, the application of these models is somehow scarce or nonexistent in engineering and climate science as observed in this review. According to Casals et al. [33], the validity of conclusions and or inferences in any scientific paper is associated with the appropriateness of the statistical methods employed to obtain the results. Therefore, the relevance to adequately describe the statistical techniques used in the analysis which accords the reader the opportunity to determine whether the conclusions and inferences made are right or otherwise. As evident in the results, most of the important information on HLMs as recommended in [49, 62] was not reported or presented which consequently undermines the reliability of the generated results as well as the validity of conclusions in such studies. For instance, nearly half (42.6%) of the articles selected failed to report on the estimation procedure used in generating the results. Some also did not report on the software used for the analysis of data.

There are divergent views on the issue of sample size per cluster and their impact on multilevel models. For example, Clarke [63] opined that cluster sizes below five per cluster result in overestimation of cluster-level variance in 2-level MLMs but recommend a mean of five observations per cluster to yield more valid and reliable parameter estimates. However, Hox and McNeish [39] argue that the smaller cluster sizes are problematic but often dissipate with a large number of clusters. Regarding parameter estimation, the maximum likelihood (ML) estimation and weighted least squares (WLS) are efficient in large sample cluster sizes, whereas the Bayesian method is efficient and recommended for small sample sizes to ensure precision in estimates. Despite the nonexistent clear rule on the minimum sample size requirement, recommendations vary based on the complexity of the model and intra-class correlation. Hox and McNeish [39] further opined that accurate parameter estimates and standard errors linked with few clusters could be obtained by using restricted maximum likelihood estimation instead of full MLE.

Besides the impact of small sample sizes on parameter estimates and standard errors, small samples also have significant effects on the statistical power of tests as the result in low power to reject an incorrect null hypothesis [64]. The different recommendations on cluster sizes include Kreft's [65] minimum of 30 observations and 30 groups, Maas and Hox [66] more than 30 groups and less than 30 observations, Hox's [1] at least 20 observations for 50 groups, Clarke and Wheaton [67] minimum of at least 10 observations per cluster and at least 200 clusters to generate unbiased and efficient estimates, and Snijders and Bosker [68] suggestion of at least one observation per group with some groups with larger sizes. Vallejo et al. [69] highlight the importance of a larger number of groups than a larger number of observations per group and recommend at least 50 groups with at least 20 observations per group to produce robust and efficient estimates. These different suggestions on the number of groups and observations per group underscore the need for standardization of these areas to enhance the validity and reliability of generated results in the utilization of MLMs.

Moreover, the inferential information such as confidence interval estimation, hypothesis testing, and model validation measures are usually related to the estimation technique used. Therefore, the absence of information regarding the estimation method used often results in the difficulty for readers to evaluate the methods used for the inferences and conclusions. Again, some estimation methods have limitations depending on the situation as in the case of MLE which produces biased and inconsistent parameter estimates when small sizes are small and for the PQL, which also generates biased parameter estimates when the sample standard deviations of random components are large in binary data [49].

In addition, most of the articles, besides failing to report on the estimation procedure, did not measure the intraclass correlation (ICC), which is very fundamental and important in the multilevel modelling framework. In [70–72], the value of the ICC indicates the existence of a multilevel structure and therefore recommends multilevel modelling to adequately account for clustering effects which could result in underestimation of standard errors and subsequent loss of statistical power when ignored. The absence of the ICC in a study to quantify the level of clustering effects could undermine results and findings since the ICC is very fundamental in informing the use of multilevel models. Moreover, failure to report the ICC value could affect the trustworthiness of findings. According to Merlo et al. [73], without the ICC value, the degree of clustering effect may not be ascertained. In Hayes [74], quantifying the grouplevel effect in terms of ICC is a basic requirement and recommends the implementation of MLMS with at least an ICC of 0.05. Based on Hayes' suggestion, ICC values in studies with clustering effects should be very fundamental to ensuring transparency and reproducibility.

Model validation and selection indices were not reported in some articles and therefore can affect the trustworthiness of findings in such articles. Casals et al. [33] further confirmed that fulfilment of model assumptions results in valid statistical inferences of generated results and conclusions. It is therefore imperative for authors to report on these model assumptions and validation measures as well as how they work and are assessed for replication and reproducibility of these studies.

5. Conclusion

The application of hierarchical models over the years has increased significantly across a wide spectrum of areas, most especially in public health [56–58, 60, 75], education and psychology [50–53], and geography, economics, and development [26, 38, 43, 61]. This is a result of the underlying hierarchical structure observed in most available data in these areas [1, 5, 35].

This review shows an increase in the adoption of multilevel models across a wide range of disciplines, which could be attributed to the presence of clustered surveyed data [1, 5, 35]. However, the review observed a low reportage of the clustering/grouping effect in terms of the ICC or variance partition coefficient which validates the application of multilevel models. Again, the method of parameter estimation and model adequacy/assessment are not clearly indicated to ensure that the underlying assumptions are met or otherwise for reproducibility. We recommend the standardization of reporting guidelines aimed at improving the quality of reporting and validity of findings among researchers and editors [76].

6. Limitation

The review conducted encountered a few limitations which include (1) the possibility of potential bias as a result of search terms used in searching for articles might have been very sensitive and therefore could have ignored some quality-related articles and (2) the relatively small number of articles selected during this review in the 11-year period.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] J. J. Hox, *Multilevel Analysis: Techniques and Applications*, Routledge, Great Britain, UK, 2nd edition, 2010.
- [2] K. Jones, R. Johnston, D. Manley, D. Owen, and C. Charlton, "Ethnic residential segregation: a multilevel, multigroup, multiscale approach exemplified by London in 2011," *Demography*, vol. 52, no. 6, pp. 1995–2019, 2015.
- [3] G. Leckie, R. Pillinger, K. Jones, and H. Goldstein, "Multilevel modeling of social segregation," *Journal of Educational and Behavioral Statistics*, vol. 37, no. 1, pp. 3–30, 2012.
- [4] A. B. Snijders Tom and R. Bosker, Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modelling, Sage Publishers, London, 2nd edition, 2012.
- [5] J. M. Ferron, B. A. Bell, M. R. Hess, G. Rendina-Gobioff, and S. T. Hibbard, "Making treatment effect inferences from multiple-baseline data: the utility of multilevel modeling approaches," *Behavior Research Methods*, vol. 41, no. 2, pp. 372–384, 2009.
- [6] H. Finch, "Multilevel modeling in the presence of outliers: a comparison of robust estimation methods," *Psicológica*, vol. 38, pp. 57–92, 2017.
- [7] Y. S. El-Horbaty and E. M. Hanafy, "Some estimation methods and their assessment in multilevel models: a review," *Biostatistics and Biometrics Open Access Journal*, vol. 5, no. 3, pp. 69–76, 2018.
- [8] P. Bliese, "Multilevel: multilevel functions," R Package Version, vol. 2, p. 163, 2016.
- [9] T. S. Clark and D. A. Linzer, "Should I use fixed or random effects?" *Political Science Research and Methods*, vol. 3, no. 2, pp. 399–408, 2015.
- [10] S. A. Raudenbush and W. Stephen Bryk, "Hierarchical linear models," *Encyclopedia of social measurement*, 2004.
- [11] J. Pinheiro and D. Bates, *Mixed-effects Models in S and S-PLUS*, Springer science \& business media, Berlin, Germany, 2006.
- [12] W. Browne and H. Goldstein, "MCMC sampling for a multilevel model with nonindependent residuals within and between cluster units," *Journal of Educational and Behavioral Statistics*, vol. 35, no. 4, pp. 453–473, 2010.
- S. F. Reardon, D. Kalogrides, and K. Shores, "The geography of racial/ethnic test score gaps (cepa working paper 16-10)," 2017, http://cepa.stanford.edu/sites/default/files/wp16-10v201803.pdf.
- [14] C. C. Rosen, K. M. Kacmar, K. J. Harris, M. B. Gavin, and W. A. Hochwarter, "Workplace politics and performance appraisal: a two-study, multilevel field investigation," *Journal* of *Leadership and Organizational Studies*, vol. 24, no. 1, pp. 20–38, 2017.
- [15] L. Anselin, "Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity," *Geographical Analy*sis, vol. 20, pp. 1–17, 1988.
- [16] A. Gelman, "The connection between varying treatment effects and the crisis of unreplicable research: a bayesian perspective," *Journal of Management*, vol. 41, no. 2, pp. 632–643, 2015.
- [17] L. J. Wolf, L. Anselin, D. Arribas-Bel, and L. R. Mobley, "On spatial and platial dependence: examining shrinkage in spatially dependent multilevel models," *Annals of the Association* of American Geographers, vol. 111, pp. 1–13, 2021.

- [18] R. J. Wright and S. V. Subramanian, "Advancing a multilevel framework for epidemiologic research on asthma disparities," *Chest*, vol. 132, no. 5, pp. 7578–769S, 2007.
- [19] C. Mobley, W. M. Vagias, and S. L. DeWard, "Exploring additional determinants of environmentally responsible behavior: the influence of environmental literature and environmental attitudes," *Environment and Behavior*, vol. 42, no. 4, pp. 420–447, 2010.
- [20] L. R. Mobley, T. Kuo, M. Urato et al., "Spatial heterogeneity in cancer control planning and cancer screening behavior spatial heterogeneity in cancer control planning and cancer screening behavior," *Annals of the Association of American Geographers*, vol. 102, 2012.
- [21] G. Moon, J. Pearce, and R. Barnett, "Smoking, ethnic residential segregation, and ethnic diversity: a spatio-temporal analysis," *Annals of the Association of American Geographers*, vol. 102, no. 5, pp. 912–921, 2012.
- [22] G. Dong, L. Wolf, A. Alexiou, and D. Arribas-Bel, "Inferring neighbourhood quality with property transaction records by using a locally adaptive spatial multi-level model," *Computers, Environment and Urban Systems*, vol. 73, pp. 118–125, 2019.
- [23] A. Petrović, M. Van Ham, and D. Manley, "Multiscale measures of population: within- and between-city variation in exposure to the sociospatial context," *Annals of the Association of American Geographers*, vol. 108, no. 4, pp. 1057–1074, 2018.
- [24] S. Comber, D. Arribas-Bel, A. Singleton, G. Dong, and L. Dolega, "Building hierarchies of retail centers using bayesian multilevel models," *Annals of the Association of American Geographers*, vol. 110, no. 4, pp. 1150–1173, 2020.
- [25] M. Novak and M. Pahor, "Using a multilevel modelling approach to explain the influence of economic development on the subjective well-being of individuals," *Economic Research-Ekonomska Istraživanja*, vol. 30, no. 1, pp. 705–720, 2017.
- [26] D. Anxo, S. Hussain, and G. Shukur, "The demand of parttime in European companies: a multilevel modelling approach," *Applied Economics*, vol. 44, no. 8, pp. 1057–1066, 2012.
- [27] R. Ramamurti, "A multilevel model of privatization in emerging economies," *Academy of Management Review*, vol. 25, no. 3, pp. 525–550, 2000.
- [28] S. T. Marquart-Pyatt, "Contextual influences on environmental concerns cross-nationally: a multilevel investigation," *Social Science Research*, vol. 41, no. 5, pp. 1085–1099, 2012.
- [29] B. Hannibal, X. Liu, and A. Vedlitz, "Personal characteristics, local environmental conditions, and individual environmental concern: a multilevel analysis," *Environ Sociol*, vol. 2, pp. 286–297, 2016.
- [30] J. Rasbash and W. J. Browne, "Non-hierarchical multilevel models," in *Handb Multilevel Anal*, pp. 301–334, Springer US, New York, NY, USA, 2008.
- [31] B. R. Saville, *Bayesian Multilevel Models and Medical Applications*, The University of North Carolina at Chapel Hill, North Carolina, CL, USA, 2008.
- [32] J. J. Hox, M. Moerbeek, and R. de Schoot, *Multilevel Analysis: Techniques and Applications*, Routledge, England, UK, 2017.
- [33] M. Casals, M. Girabent-Farrés, and J. L. Carrasco, "Methodological quality and reporting of generalized linear mixed models in clinical medicine (2000-2012): a systematic review," *Public Library of Science One*, vol. 9, no. 11, Article ID e112653, 2014.
- [34] H. Finch, "Multilevel modeling in the presence of outliers: a comparison of robust estimation methods," *Psicologica*:

International Journal of Methodology and Experimental Psychology, vol. 38, pp. 57–92, 2017.

- [35] M. Arcaya, M. Brewster, C. M. Zigler, and S. V. Subramanian, "Area variations in health: a spatial multilevel modeling approach," *Health and Place*, vol. 18, no. 4, pp. 824–831, 2012.
- [36] L. Anselin, "Lagrange multiplier test diagnostics for spatial dependence and spatial heterogeneity," *Geographical Analy*sis, vol. 20, pp. 1–17, 1988.
- [37] G. Owen, R. Harris, and K. Jones, "Under examination: multilevel models, geography and health research," *Progress in Human Geography*, vol. 40, no. 3, pp. 394–412, 2016.
- [38] G. Dong, R. Harris, K. Jones, and J. Yu, "Multilevel modelling with spatial interaction effects with application to an emerging land market in Beijing, China," *Public Library of Science One*, vol. 10, no. 6, pp. 01307611–e130818, 2015.
- [39] J. Hox and D. McNeish, "Small samples in multilevel modeling," Small Sample Size Solutions, pp. 215–225, 2020.
- [40] M. J. Page, D. Moher, P. M. Bossuyt et al., "PRISMA 2020 explanation and elaboration: updated guidance and exemplars for reporting systematic reviews," *British Medical Journal*, vol. 372, p. n160, 2021.
- [41] B. Hutton, G. Salanti, A. Chaimani et al., "The quality of reporting methods and results inNetwork meta-analyses: an overview of reviews andSuggestions for improvement," *Public Library of Science One*, vol. 9, no. 3, Article ID e92508, 2014.
- [42] P. C. Austin, "Estimating multilevel logistic regression models when the number of clusters is low: a comparison of different statistical software procedures," *International Journal of Biostatistics*, vol. 6, no. 1, 2010.
- [43] Y. Li and Y. D. Wei, "The spatial-temporal hierarchy of regional inequality of China," *Applied Geography*, vol. 30, no. 3, pp. 303–316, 2010.
- [44] G. D. Garson, "Fundamentals of hierarchical linear and multilevel modeling," in *Hierarchical Linear Model*, pp. 3–26, Guid. Appl, Cupertino, CA, USA, 2014.
- [45] D. Bates, M. Mächler, B. M. Bolker, and S. C. Walker, "Fitting linear mixed-effects models using lme4," *Journal of Statistical Software*, vol. 67, no. 1, 2015.
- [46] D. Lüdecke, "Package 'sjstats'," 2019, https://github.com/ strengejacke/sjstats.
- [47] S. Rabe-Hesketh, A. Skrondal, and A. Pickles, "Reliable estimation of generalized linear mixed models using adaptive quadrature," *Statistics and Data Journal: Promoting communications on statistics and Stata*, vol. 2, pp. 1–21, 2002.
- [48] Sophia and A. S. Rabe-Hesketh, Multilevel and Longitudinal Modeling Using Stata, STATA press, College Station, TX, USA, 2006.
- [49] B. M. Bolker, M. E. Brooks, C. J. Clark et al., "Generalized linear mixed models: a practical guide for ecology and evolution," *Trends in Ecology and Evolution*, vol. 24, no. 3, pp. 127–135, 2009.
- [50] J. H. Bolin, W. H. Finch, and R. Stenger, "Estimation of random coefficient multilevel models in the context of small numbers of level 2 clusters," *Educational and Psychological Measurement*, vol. 79, pp. 217–248, 2019.
- [51] L. Sun, K. D. Bradley, and K. Akers, "A multilevel modelling approach to investigating factors impacting science achievement for secondary school students: PISA Hong Kong sample," *International Journal of Science Education*, vol. 34, no. 14, pp. 2107–2125, 2012.
- [52] D. B. McCoach and B. Kaniskan, "Using time-varying covariates in multilevel growth models," *Frontiers in Psychology*, vol. 1, pp. 17–12, 2010.

- [53] B. R. Subedi, B. Swan, and M. C. Hynes, "Are school factors important for measuring teacher effectiveness? A multilevel technique to predict student gains through a value-added approach," *Educational Research International*, vol. 2011, Article ID 532737, 10 pages, 2011.
- [54] H. M. Carvalho, "Multilevel models for the analysis of anglespecific torque curves with application to master athletes," *Journal of Human Kinetics*, vol. 49, no. 1, pp. 25–35, 2015.
- [55] E. Guerra, H. Zhang, L. Hassall, J. Wang, and A. Cheyette, "Who cycles to work and where? A comparative multilevel analysis of urban commuters in the US and Mexico," *Transportation Research Part D: Transport and Environment*, vol. 87, Article ID 102554, 2020.
- [56] Y. M. Park and Y. Kim, "A spatially filtered multilevel model to account for spatial dependency: application to self-rated health status in South Korea," *International Journal of Health Geographics*, vol. 13, pp. 6–10, 2014.
- [57] E. Garcia-Garzon, P. Zhukovsky, E. Haller et al., "Multilevel modeling and policy development: guidelines and applications to medical travel," *Frontiers in Psychology*, vol. 7, pp. 752–813, 2016.
- [58] T. N. Gomes, F. K. Dos Santos, D. Santos et al., "Correlates of sedentary time in children: a multilevel modelling approach," *Bone Marrow Concentrate Public Health*, vol. 14, pp. 890–899, 2014.
- [59] A. Debie, G. Amare, S. Handebo, M. E. Mekonnen, and G. A. Tesema, "Individual-and community-level determinants for complete vaccination among children aged 12-23 Months in Ethiopia: a multilevel analysis," *BioMedical Research International*, vol. 2020, Article ID 6907395, 10 pages, 2020.
- [60] E. E. Chaka and P. Bao, "Examining individual- and community-level factors affecting skilled delivery care among women who received adequate antenatal care in Ethiopia: using multilevel analysis," *BioMedical Research International*, vol. 2020, Article ID 2130585, 8 pages, 2020.
- [61] G. Fazio and D. Piacentino, "A spatial multilevel analysis of Italian SMEs' productivity," *Spatial Economic Analysis*, vol. 5, no. 3, pp. 299–316, 2010.
- [62] J. Thiele and B. Markussen, "Review: modelling invasion probability of giant hogweed (heracleum mantegazzianum) with logistic GLMM," *Centre for Agriculture and Bioscience International Reviews*, vol. 7, pp. 1–12, 2012.
- [63] P. Clarke, "Theory and methods: when can group level clustering be ignored? Multilevel models versus single-level models with sparse data," *Journal of Epidemiology and Community Health*, vol. 62, no. 8, pp. 752–758, 2008.
- [64] D. M. McNeish, "Modeling sparsely clustered data: designbased, model-based, and single-level methods," *Psychological Methods*, vol. 19, no. 4, pp. 552–563, 2014.
- [65] I. G. G. Kreft, Are Multilevel Techniques Necessary? an Overview, Including Simulation Studies, Unpubl. Manuscript, Calif. State Univ. Los Angeles, California, LA, USA, 1996.
- [66] C. J. M. Maas and J. J. Hox, "The influence of violations of assumptions on multilevel parameter estimates and their standard errors," *Computational Statistics and Data Analysis*, vol. 46, no. 3, pp. 427–440, 2004.
- [67] P. Clarke and B. Wheaton, "Addressing data sparseness in contextual population research: using cluster analysis to create synthetic neighborhoods," *Sociological Methods and Research*, vol. 35, no. 3, pp. 311–351, 2007.
- [68] T. A. B. Snijders and R. J. Bosker, "Modeled variance in twolevel models," *Sociological Methods and Research*, vol. 22, no. 3, pp. 342–363, 1994.

- [69] G. Vallejo, E. Tuero-Herrero, J. C. Núñez, P. Rosário, and P. Rosário, "Performance evaluation of recent information criteria for selecting multilevel models in Behavioral and Social Sciences," *International Journal of Clinical and Health Psychology*, vol. 14, no. 1, pp. 48–57, 2014.
- [70] M. Černe, M. Jaklič, and M. Škerlavaj, "Authentic leadership, creativity, and innovation: a multilevel perspective," *Leadership*, vol. 9, no. 1, pp. 63–85, 2013.
- [71] C. M. Vauclair and R. Fischer, "Do cultural values predict individuals' moral attitudes? A cross-cultural multilevel approach," *European Journal of Social Psychology*, vol. 41, no. 5, pp. 645–657, 2011.
- [72] D. Holman and A. Walker, "Social quality and health: examining individual and neighbourhood contextual effects using a multilevel modelling approach," *Social Indicators Research*, vol. 138, no. 1, pp. 245–270, 2018.
- [73] J. Merlo, P. Wagner, and G. Leckie, "A simple multilevel approach for analysing geographical inequalities in public health reports: the case of municipality differences in obesity," *Health*, vol. 58, Article ID 102145, 2019.
- [74] A. F. Hayes, "A primer on multilevel modeling," *Human Communication Research*, vol. 32, no. 4, pp. 385–410, 2006.
- [75] A. Debie, G. Amare, S. Handebo, M. E. Mekonnen, and G. A. Tesema, "Individual- and community-level determinants for complete vaccination among children aged 12-23 Months in Ethiopia: a multilevel analysis," *BioMedical Research International*, vol. 2020, Article ID 6907395, 10 pages, 2020.
- [76] E. Cobo, F. Cardellach, B. Kostov et al., "Effect of using reporting guidelines during peer review on quality of final manuscripts submitted to a biomedical," *Journal: Masked Randomised Trial*, vol. 6783, pp. 1–11, 2011.