Cross-classified multilevel models improved standard error estimates of covariates in clinical outcomes – a simulation study

Doedens, P.; ter Riet, G.; Boyette, L.-L.; Latour, C.; de Haan, L.; Twisk, J.

DOI
10.1016/j.jclinepi.2022.01.005

Publication date
2022

Document Version
Final published version

Published in
Journal of Clinical Epidemiology

ELSEVIER

Journal of Clinical Epidemiology 145 (2022) 39–46

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

Cross-classified multilevel models improved standard error estimates of covariates in clinical outcomes – a simulation study

Paul Doedens[a,b,*], Gerben ter Riet[b,c], Lindy-Lou Boyette[d], Corine Latour[b], Lieuwe de Haan[a,e], Jos Twisk[f]

[a] *Department of Psychiatry, Amsterdam UMC, location Academic Medical Center, Amsterdam, The Netherlands*
[b] *Urban Vitality – Centre of Expertise, Faculty of Health, Amsterdam University of Applied Sciences, Amsterdam, The Netherlands*
[c] *Department of Cardiology, Amsterdam UMC, location Academic Medical Center, Amsterdam, The Netherlands*
[d] *Department of Clinical Psychology, University of Amsterdam, Amsterdam, The Netherlands*
[e] *Arkin, Amsterdam, The Netherlands*
[f] *Department of Epidemiology and Biostatistics, VU University Medical Center, Amsterdam, The Netherlands*

Accepted 13 January 2022; Available online 19 January 2022

Abstract

Objective: To compare estimates of effect and variability resulting from standard linear regression analysis and hierarchical multilevel analysis with cross-classified multilevel analysis under various scenarios.

Study design and setting: We performed a simulation study based on a data structure from an observational study in clinical mental health care. We used a Markov chain Monte Carlo approach to simulate 18 scenarios, varying sample sizes, cluster sizes, effect sizes and between group variances. For each scenario, we performed standard linear regression, multilevel regression with random intercept on patient level, multilevel regression with random intercept on nursing team level and cross-classified multilevel analysis.

Results: Applying cross-classified multilevel analyses had negligible influence on the effect estimates. However, ignoring cross-classification led to underestimation of the standard errors of the covariates at the two cross-classified levels and to invalidly narrow confidence intervals. This may lead to incorrect statistical inference. Varying sample size, cluster size, effect size and variance had no meaningful influence on these findings.

Conclusion: In case of cross-classified data structures, the use of a cross-classified multilevel model helps estimating valid precision of effects, and thereby, support correct inferences.

Keywords: Cross-classification; Multilevel modelling; Simulation study; Nursing science; Epidemiology

1. Introduction

Rigorous applied clinical research on clinical patient outcomes is essential to enhance quality of care. In clinical settings, such as hospitals, nursing homes or mental health facilities, the quality of nursing staff is associated with adequate quality and safety of patient care [1–4]. The influence of nurses on quality of care in clinical settings has consequences, for instance, for studies on risk factors. When such research targets the effects of nurses on care, clustering of observations is a potential source of bias or incorrect inference in the analysis. Several studies on patient outcomes, also take staff (e.g., nurses) characteristics into account [5–7]. To be valid, the data-analysis of such studies should account for the (oftentimes) clustered data structure. For instance, clustered data structures due to participation of multiple centers or multiple wards within a single center can occur and demand multilevel modelling to account for clustering of observations within centers or wards. However, multilevel modelling may not completely suffice in case of clustered observations in more than one cluster, whose relationship is not strictly hierarchical. When a strictly hierarchical model is not adequate, cross-classified multilevel models may be needed [8–10].

Several authors use "school and neighborhood effects" to clarify the nature of cross-classification [11,12]. School and neighborhood effects describe observations on children's clustered in their schools and in their neighborhoods simultaneously [11,12]. Contextual data are often cross-classified as students living in the same neighborhood may attend different schools and students in the same school

* Corresponding author. Tel.: +31208913500.
E-mail address: p.doedens@amsterdamumc.nl (P. Doedens).

https://doi.org/10.1016/j.jclinepi.2022.01.005
0895-4356/

What is new?

- Clinical research that accounts for patient and staff characteristics often has a cross-classified data structure, which we refer to as patient and shift effects.
- Ignoring a cross-classified data structure has little effect on fixed effects, but may lead to invalid standard errors.
- Clinical researchers should use cross-classified multilevel models when data structures deviate substantially from strict hierarchy to prevent invalid estimates of the precision of effects.

may live in different neighborhoods. The cross-classified multilevel model (CCMM), also known as cross-classified random-effects model, allows researchers to take into account this particular data structure in one analysis [11].

In the health sciences, the need to distinguish cross-clustering of institutions and neighborhoods may also occur, such as in community mental health care [13,14] and hospital care [15–17]. Application of CCMM in clinical nursing science is, however, slightly different from cross-classification of institution and neighborhood effects. On clinical wards, several nurses care for a patient during a shift. This implies "clustering of nurses within patients". On the other hand, nurses care for more than one patient during their shifts, which implies "clustering of patients within nurses" (Fig. 1A). Thereby, patients and nurses are crossed factors in this data structure (Fig. 1B). In analogy with "school and neighborhood effects", we refer to this structure as "patient and shift effects".

Due to this multiple clustering, CCMM is the appropriate model when performing research on patient outcomes. However, aside from some studies in neonatal intensive care [18–20], few authors in nursing research have used CCMM for their analyses. Unfamiliarity with (recognition of) cross-classified data structures and additional complexity in statistical models might explain this. For school and neighborhood effects, several (simulation) studies are available to assist researchers in choosing between CCMM and other regression techniques [11,21–23]. However, regarding the influence of using CCMM on patient and shift effects under several circumstances, no information is available as far as we are aware.

Given that, theoretically, CCMM is the correct approach to analyze cross-classified data, it is important to evaluate the effect of using different, more commonly used approaches to data with a cross-classified structure due to patient and shift effects in order to assess the magnitude of errors that may result from using theoretically suboptimal approaches. We, therefore, aim to familiarize clinical researchers with cross-classification and assist them in the decision whether the added complexity of CCMM is a price worth paying. Building on the guidance for good quality simulation studies [24], we performed a simulation study to compare the different techniques under various scenarios. We based our simulation data on a real-life observational study [25]. In this study, we analyzed the influence of nursing teams on the frequency of seclusion of patients on a closed psychiatric ward. We found that teams with majority of female nurses and teams with high mean scores of personality trait openness were associated with higher seclusion probabilities for patients. Each shift had a different nursing team composition and the patient population on the ward is changing over time. This is a clear example of patient and shift effects. Therefore, we used this structure as blueprint for our simulations. We added the full STATA code of our analysis as an online supplement for other authors to perform CCMM [25].

2. Method

2.1. Procedures

We performed simulations of two-level cross-classified multilevel models using Markov chain Monte Carlo simulations in STATA, version 14. The full code of our simulations is available upon request including the code used for the CCMM analyses.

For each scenario of interest, we generated 1000 samples to compare the statistical methods using a normally distributed continuous outcome variable at the patient level. We generated two covariates at the level of the patients, namely, sex (dichotomous, 50% male and 50% female) and age (continuous with mean [M] = 50 and standard deviation [SD] = 10). We also created two covariates at the level of nursing teams, namely, team composition (46% of the teams were male only vs. 54% all female or a mixed team of male and female nurses) and the mean number of years of work experience in a team (continuous with M = 5 and SD = 2). We created a categorical variable indicating daily work shifts (day shift, evening shift and night shift) represented by two dummy variables and used day shift as a reference category.

2.2. Scenarios

We performed moderately independent simulations on different scenarios that varied in sample sizes, cluster sizes, effect sizes and between group variances (Table 1). Moderately independent means that we use the same simulated data sets to compare the statistical methods [24].

For every scenario (eighteen in total), we performed four different analyses, namely 1) standard linear regression ignoring clustering at both patient and nursing team level.; 2) multilevel regression with random intercept at patient level, ignoring clustering at nursing team level; 3) multilevel regression with random intercept at nursing team

Figure 1a: Presence of patients and nursing teams

Shift	Team 1	Team 2	Team 3	Team 4	Team 5	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5
January 1, day shift	X					X	X			
January 1, evening shift		X				X	X	X		
January 1, night shift			X			X	X	X	X	
January 2, day shift	X					X	X	X	X	
January 2, evening shift				X		X	X	X	X	X
January 2, night shift			X				X	X	X	X
January 3, day shift					X		X	X	X	X
January 3, evening shift				X			X	X	X	X
January 3, night shift			X				X	X		X

Figure 1b: Example of patient and shift effects based on Figure 1a

	Patient 1	Patient 2	Patient 3	Patient 4	Patient 5
Team 1	X X	X X	X	X	
Team 2	X	X	X		
Team 3	X	X X X	X X X	X X	X X
Team 4	X	X X	X X	X X	X X
Team 5		X	X	X	X

Fig. 1. It is an example of patient and shift effects, we explore patients and nursing teams of a ward in three consecutive days (i.e., nine unique shifts). Patients one and two were already admitted before the first shift. Patient one was discharged during the evening shift of January 2, patient two was still admitted during the last shift. Patient three was admitted during the evening shift on January 1 and stayed until the last shift. Patient four was admitted during night shift on January 1 and discharged in the evening shift of January 3. Patient five was admitted in the evening shift of January 2 and stayed until the last shift. Each nursing team has a unique combination of individual nurses. Nursing team one worked two consecutive day shifts (January 1 and 2), team two only worked during the evening shift of January 2. Team three worked the three-night shifts. Team four worked the last two evening shifts and team five worked the day shift on January 3. We see that patients form clusters within nursing teams, but nursing teams also form clusters within patients. The clustering is not strictly hierarchical. This is a cross-classified data structure.

Table 1. The components of simulated scenarios

Sample size (2 options)	**Cluster size (2 options)**	**Effect size (2 options)**	**Variance (3 options)**
Larger groups N = 50 patients; N = 100 teams	Larger clusters N = 25 shifts (patients); N = 10-15 shifts (teams)	Stronger effect $\beta = 2$ (team composition, all male teams); $\beta = 1$ (work experience, years); $\beta = 2$ (sex); $\beta = 0.2$ (age, years); $\beta = -1$ (shift, two dummy variables)	Large variance $\sigma^2 = 0.3$
Smaller groups N = 10 patients; N = 20 teams	Smaller clusters N = 5 shifts (patients); N = 2-3 shifts (teams)	Weaker effect $\beta = 0.2$ (team composition, all male teams); $\beta = 0.1$ (work experience); $\beta = 0.2$ (sex); $\beta = 0.02$ (age, years); $\beta = -0.1$ (shift, two dummy variables)	Intermediate variance $\sigma^2 = 0.2$
			Small variance $\sigma^2 = 0.1$

Box 1. STATA code to run the different analyses

```
1 mixed y sex age maleteams experience
  eveningshift nightshift
2 mixed y sex age maleteams experience
  eveningshift nightshift || patient:
3 mixed y sex age maleteams experience
  eveningshift nightshift || team:
4 mixed y sex age maleteams experience
  eveningshift nightshift || _all: R.patient || team:
5 mixed y sex age maleteams experience
  eveningshift nightshift || _all: R.team || patient:
```

1, Linear mixed model without random intercepts; 2, Linear mixed model with a random intercept on patient level; 3, Linear mixed model with a random intercept on nursing team level; 4, CCMM taking into account both levels of clustering; 5, Alternative for CCMM taking into account both levels of clustering.

level, ignoring clustering at patient level; and 4) CCMM, which takes into account clustering both at nursing team level and patient level (Box 1).

Our basic scenario consisted of 50 unique patients who stayed on the ward for a duration of 25 shifts (approximately eight days), making up 1250 shifts. 100 unique nursing teams covered all 1250 shifts with each team attending 10–15 shifts. Every shift has a unique outcome measure; therefore, the number of shifts that a patient stayed at the ward is equivalent to the number of observations. In other words, the basic scenario had samples with 50 patients with 25 observations and 100 unique nursing teams with 10–15 observations. We varied the basic scenario by lowering the number of patients (sample size), resulting in samples with 10 patients with 25 observations and 20 nursing teams with 10–15 observations. We also varied by lowering the number of observations (cluster size), with samples consisting of 50 patients with 5 observations and 100 nursing teams with 2–3 observations.

We varied the effect sizes (i.e., regression coefficient β [beta]) of the covariates between stronger and weak effects. Stronger effect size meant $\beta = 2$ for the effect of patients' sex and the effect of team composition (all male teams). $\beta = 1$ for mean work experience of nursing teams, $\beta = 0.2$ for the effect of patient age (effect per year older) and $\beta = -1$ for shift covariates (evening shift and night shift). Weak effect size meant $\beta = 0.2$ for both dichotomous covariates (patients' sex and only male nurses present), $\beta = 0.1$ for mean work experience of nursing teams, $\beta = 0.02$ for patients' age and $\beta = -0.1$ for shift covariates (evening shift and night shift). We based the ratio between the magnitudes of β on the findings in a real-life study on which we based our data structure. We analyzed all scenarios with large between group variance ($\sigma^2 = 0.3$), intermediate between group variance ($\sigma^2 = 0.2$) and small between group variance ($\sigma^2 = 0.1$).

For each scenario, we estimated β and standard error (SE) of fixed parameters (i.e., covariates on both patient and nursing team level) as well as variances at the different levels and reported coverage and bias to assess model performance. Bias represents the deviation from the true value of the effect size in the simulation [24]. Coverage represents the proportion of times that the confidence interval of the simulations contains the true value of the regression coefficient. Coverage should be close to the chosen confidence interval, in our case 95% [24] (Table 2).

3. Results

We summarize full results of our simulation in Online supplement 1. Effect size estimations (i.e., regression coefficients) were stable with all regression approaches used. Changes in sample size, cluster sizes and between-cluster variances had no major influence on the estimated effects.

We observed an effect on the SEs of the covariates. When applying a multilevel model with a random intercept at the patient level and ignoring clustering at the nursing

Table 2. Definition of reported criteria and abbreviations

	Definition
M	Mean
SD	Standard deviation
B (beta)	Regression coefficient of fixed effects (or: covariates)
SE	Standard error of fixed effects (or: covariates)
Bias	Relative deviation of estimate β compared to true β $(\frac{True\ \beta - \beta}{\beta})*100\%$
Coverage	Proportion of times that the simulated confidence interval contains the true regression coefficient β, coverage should be around 95%

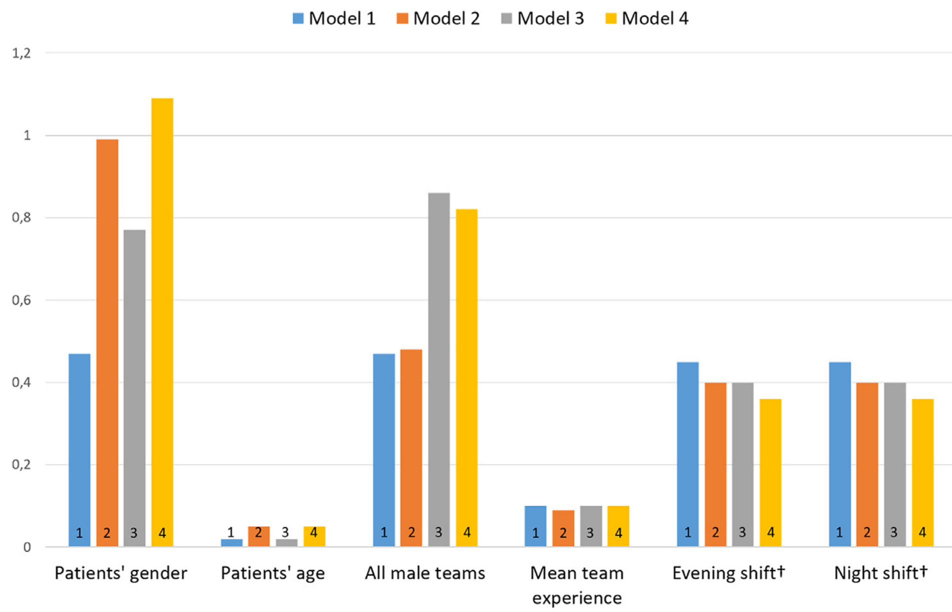

Fig. 2. Standard error in basic scenario. † = dichotomous dummy variable, compared to day shift. 1 = Linear mixed model without random intercepts; 2 = Linear mixed model with a random intercept on patient level; 3 = Linear mixed model with a random intercept on nursing team level; 4 = CCMM taking into account both levels of clustering. “(For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)”

team level (model 2), the estimations of the SE of patient level covariates increased, while the estimations of the SE at the nursing team level covariates were stable. Similarly, in model 3 (with a random intercept at the nursing team level and ignoring clustering at the patient level), we observed that the SE of nursing team level covariates increased, while the estimations of the SE for patient level covariates did not. In the CCMM model, we observed increased SEs for both patient level covariates and nursing team level covariates.

For example, when performing standard linear regression (model 1) in our basic scenario with stronger effect size and large variance, SE was 0.47 for patients’ sex and 0.47 for team composition. Adding a random intercept at the patient level (model 2), SE was 0.99 for patients’ sex and 0.48 for team composition. However, adding a random intercept at the nursing team level (model 3) yielded a SE of 0.77 for patients’ sex 0.86 for all male teams. Finally, in the CCMM (model 4), SE was 1.09 for patients’ sex and 0.82 for team composition. Fig. 2 summarizes the findings on SE across the different simulations for the basic scenario.

Furthermore, we observed that the random effect variance per level decreased when adding more levels. Random effect variance parameters in CCMM (compared to linear mixed model with random intercept on patient or nursing team level) indicate that it is important to account for clustering at both levels, otherwise variance is misattributed either to the level that was included or unaccounted for and remains as residual variance.

CCMM models had better model performance, especially concerning the coverage of confidence intervals. We observed more bias in simulations with weak effect sizes compared to those simulating stronger effect sizes. In addition, effect sizes of continuous covariates were more stable than the effect sizes for dichotomous covariates. Furthermore, scenarios with smaller sample sizes showed some under-coverage (<95%). Under CCMM, both clusters of patient and nursing covariates showed least biased estimations with acceptable coverage (between 93% and 96%).

The shift covariates (measured at the lowest level) were remarkably stable in terms of their effect size, and coverage. However, the more variance taken into account, the lower the estimated SEs, with the CCMM models showing the smallest SEs for these covariates (see Table 3).

4. Discussion

We investigated the effect of using different statistical techniques on data with a cross-classified structure, specifically on effect estimates and SEs of the covariates. We found that standard and multilevel models caused little bias in the estimates of effect of fixed covariates at the level of patients and nursing teams, but underestimated the true SE of these covariates. CCMM resulted in better coverage compared to hierarchical multilevel models for the covariates related to the ignored crossed level.

Patient and shift effects are common when taking into account both patient level and nursing team level covariates in a statistical model. It is unlikely that ignoring the cross-classified data structure will lead to opposite conclu-

Table 3. Results of the basic scenario simulation

	Model 1: Standard linear model			Model 2: Multilevel model (patients)			Model 3: Multilevel model (nurses)			Model 4: CCMM		
	β (SE)	Bias	Coverage	β (SE)	Bias	Coverage	β (SE)	Bias	Coverage	β (SE)	Bias	Coverage
Patients												
Sex	1.94 (0.47)	-3%	56%	1.97 (0.99)	-2%	88%	1.96 (0.77)	-2%	79%	1.98 (1.09)	-1%	94%
Age	0.20 (0.02)	0%	55%	0.20 (0.05)	0%	94%	0.20 (0.02)	0%	53%	0.20 (0.05)	0%	93%
Nursing team												
All male teams	1.97 (0.47)	-2%	63%	2.01 (0.48)	0%	69%	1.97 (0.86)	-2%	92%	2.01 (0.82)	0%	95%
Experience	1.00 (0.09)	0%	76%	1.00 (0.09)	0%	82%	1.00 (0.10)	0%	82%	1.00 (0.10)	0%	94%
Shift[a]												
Evening shift	-1.0 (0.45)	0%	96%	-1.0 (0.40)	0%	95%	-1.0 (0.40)	0%	94%	-1.0 (0.37)	0%	94%
Night shift	-1.0 (0.45)	0%	93%	-1.0 (0.40)	0%	95%	-1.0 (0.40)	-1%	95%	-1.0 (0.36)	0%	96%
Variance												
Residual	6.46			5.68			5.52			4.99		
Patient				3.08						2.88		
Nursing team							3.37			2.95		

Caption: Our basic scenario had larger groups and clusters (50 patients with 25 shifts and 100 teams with 10–15 shifts), large variance ($\sigma 2 = 0.3$) and stronger effects. True β was 2.0 for patients' sex, 0.2 for patients' age, 2.0 for all male teams, 1.0 for mean nurses' experience in the team, -1.0 for evening shift and -1.0 for night shift. The full overview of our results can be found in Online supplement 1.

[a] compared to day shift

sions about the direction and magnitude of effect sizes, as our study showed that taking into account clustering at the patient level and/or nursing team level had no major influence on the estimation of the covariates. However, we found that ignoring cross-classification could lead to underestimation of SEs. Underestimation of the SE results in too small confidence intervals and may result in incorrect inference based on statistical significance, although we would advise against the rigid use of statistical significance for inference [26]. In contrast, the SEs of the indicators for the daily work shifts showed overestimation in the analysis ignoring (either part) of the clustering (i.e., estimation of the SEs are larger than their true value) when not using CCMM. This is a phenomenon often observed for covariates measured at the lowest level [27].

To our knowledge, our study is the first to examine patient and shift effects in a simulation study that compares hierarchical multilevel models with CCMM. Several other authors performed simulations of CCMM in other applications, such as cross-sectional studies [22,28], longitudinal studies [8,29] and meta-analyses [30]. Our finding that the effect size estimation of the covariates showed little bias is in line with previous simulation studies [22,28–30]. Consequently, if researchers omit correction for cross-classification, the risk of an incorrect conclusion about the magnitude or direction of an effect seems limited. However, the underestimation of SEs of covariates may well lead to incorrect inferences. Several other authors report similar consequences on SEs. For instance, Meyers and Beretvas [28] reported that when the model ignores clustering of a factor (e.g., students within schools), SEs associated with that factor were highly underestimated. Other authors reported comparable findings in studies with both simulated and real-world data [11,22,29,30].

Interpretation of our findings should take into account the following uncertainties. First, we performed simulations with a normally distributed continuous outcome variable. Secondly, we assessed several scenarios with different sample sizes and cluster sizes, but we did not evaluate uneven distribution of the size of samples and clusters. In clinical practice, it is plausible that some nursing team compositions are much more prevalent than others are. Milliren, Evans [23] investigated this uneven sample size distribution across levels in an example of the school-neighborhood effect and found no systematic bias because of this phenomenon. Thirdly, the distribution of between-level variance in our CCMM is (roughly) equal between the two crossed levels. In real world data, this is not necessarily the case. Dunn, Richmond [11] used the correct model for cross-classified data in a real-world example with school and neighborhood effects and compared this to the hierarchal models with part of the clustering (school or neighborhood) ignored. The between-level variance in both hierarchical models was roughly equal.

However, in the CCMM model, schools caused almost all between-level variance while neighborhood was no important factor [11]. Fourthly, we did not simulate various levels of (partial) cross-classification, but we suggest this as an interesting subject for future simulation studies. Finally, we simulated a relatively simple model between patients and nursing teams. In real clinical research, more complex data structures are common. For instance, in a multicenter study, a partial cross-classified structure could be the care with patients and nursing teams as crossed factors, both hierarchically clustered within wards or hospitals. However, Luo and Kwok [29] simulated cross-classified longitudinal data with three levels and found similar results about the fixed effects and their SEs for the covariates associated with the ignored crossed level.

Complex statistical techniques such as CCMM can be a challenge for (clinical) researchers to comprehend and most statistical literature on this matter focusses on a specialized (statistical) audience. The use of CCMM is possible in several statistical packages. We used the mixed command in STATA to perform our simulations (Box 1). The lme4 package in R also implements CCMM [31,32]. In case of non-continuous endpoints (e.g. binary or count), more specialized software is necessary, such as MLwiN [33]. In the observational study, on which data the simulations were based, we used the runmlwin command in STATA to perform cross-classified logistic regression analysis, of which the code is available elsewhere [25]. Leckie and Charlton [33] provided a comprehensive description of the runmlwin command to benefit from the best of both packages. In order to make adequate decisions, researchers need to recognize cross-classification in de structure of their data. There is no formal test to analyze whether cross-classification is present in data. Domain experts need to argue theoretically whether this type of non-nested clustering is present. In our case, this was the result of extensive debate (with multiple drawings of the data structure) between clinical researchers and epidemiologists. We believe that this paper could assist both clinical researchers and consulting statisticians with this decision-making process. Ultimately, when investigating the influence of nurses on patient outcomes, the use of CCMM could lead to estimations of the precision of effect sizes that are more accurate, which contributes to the further development of nursing care in clinical settings.

Author contributions

PD, GtR, JT: Made substantial contributions to conception and design, or acquisition of data, or analysis and interpretation of data; PD, GtR, LLB, CL, LdH, JT: Involved in drafting the manuscript or revising it critically for important intellectual content; Given final approval of the version to be published. Each author should have participated sufficiently in the work to take public responsibility for appropriate portions of the content; Agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Conflict of Interest

None.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jclinepi.2022.01.005.

References

[1] Butler M, Schultz TJ, Halligan P, Sheridan A, Kinsman L, Rotter T, et al. Hospital nurse-staffing models and patient- and staff-related outcomes. Cochrane Database Syst Rev 2019;4:Cd007019.
[2] Wunderlich GS, Sloan F, Davis CK. Nursing Staff in Hospitals and Nursing Homes: Is It Adequate?. Washington DC: National Academy Press; 1996.
[3] Lee SE, Vincent C, Dahinten VS, Scott LD, Park CG, Dunn Lopez K. Effects of individual nurse and hospital characteristics on patient adverse events and quality of care: a multilevel analysis. J Nurs Scholarsh 2018;50:432–40.
[4] Wynendaele H, Willems R, Trybou J. Systematic review: Association between the patient-nurse ratio and nurse outcomes in acute care hospitals. J Nurs Manag 2019;27:896–917.
[5] Aiken LH, Sloane DM, Bruyneel L, Van den Heede K, Griffiths P, Busse R, et al. Nurse staffing and education and hospital mortality in nine European countries: a retrospective observational study. Lancet 2014;383:1824–30.
[6] Doedens P, Vermeulen J, Boyette LL, Latour C, de Haan L. Influence of nursing staff attitudes and characteristics on the use of coercive measures in acute mental health services-A systematic review. J Psychiatr Ment Health Nurs 2020;27:446–59.
[7] Duffield CM, Roche MA, Dimitrelis S, Homer C, Buchan J. Instability in patient and nurse characteristics, unit complexity and patient and system outcomes. J Adv Nurs 2015;71:1288–98.
[8] Cafri G, Hedeker D, Aarons GA. An introduction and integration of cross-classified, multiple membership, and dynamic group random–effects models. Psychol Methods 2015;20:407–21.
[9] Fielding A, Goldstein H. Cross-classified and multiple membership structures in multilevel models: an introduction and review. University of Birmingham; 2006.
[10] Ross BH, Murphy GL. Food for thought: cross-classification and category organization in a complex real-world domain. Cogn Psychol 1999;38:495–553.
[11] Dunn EC, Richmond TK, Milliren CE, Subramanian SV. Using cross-classified multilevel models to disentangle school and neighborhood effects: an example focusing on smoking behaviors among adolescents in the United States. Health Place 2015;31:224–32.
[12] Raudenbush SW. A crossed random effects model for unbalanced data with applications in cross-sectional and longitudinal research. J Educ Stat 1993;18:321–49.
[13] Weich S, McBride O, Twigg L, Duncan C, Keown P, Crepaz-Keay D, et al. Variation in compulsory psychiatric inpatient admission in England: a cross-classified, multilevel analysis. Lancet Psychiatry 2017;4:619–26.
[14] Traoré M, Vuillermoz C, Chauvin P, Deguen S. Influence of individual and contextual perceptions and of multiple neighborhoods on depression. Int J Environ Res Public Health 2020;17(6):1958.

[15] Colais P, Bontempi K, Pinnarelli L, Piscicelli C, Mappa I, Fusco D, et al. Vaginal birth after caesarean birth in Italy: variations among areas of residence and hospitals. BMC Pregnancy Childbirth 2018;18:383.

[16] Pruitt SL, Leonard T, Zhang S, Schootman M, Halm EA, Gupta S. Physicians, clinics, and neighborhoods: multiple levels of influence on colorectal cancer screening. Cancer Epidemiol Biomarkers prev 2014;23:1346–55.

[17] Ratnapradipa KL, Lian M, Jeffe DB, Davidson NO, Eberth JM, Pruitt SL, et al. Patient, hospital, and geographic disparities in laparoscopic surgery use among surveillance, epidemiology, and end results-medicare patients with colon cancer. Dis Colon Rectum 2017;60:905–13.

[18] Fortney CA, Pratt M, Dunnells ZDO, Rausch JR, Clark OE, Baughcum AE, et al. Perceived infant well-being and self-reported distress in neonatal nurses. Nurs Res 2020;69:127–32.

[19] Sink DW, Hope SA, Hagadorn JI. Nurse:patient ratio and achievement of oxygen saturation goals in premature infants. Arch Dis Child Fetal Neonatal Ed 2011;96:F93–8.

[20] Tubbs-Cooley HL, Mara CA, Carle AC, Mark BA, Pickler RH. Association of nurse workload with missed nursing care in the neonatal intensive care unit. JAMA Pediatr 2019;173:44–51.

[21] Im MH, Kim ES, Kwok OM, Yoon M, Willson VL. Impact of not addressing partially cross-classified multilevel structure in testing measurement invariance: a monte carlo study. Front Psychol 2016;7:328.

[22] Luo W, Kwok OM. The impacts of ignoring a crossed factor in analyzing cross-classified data. Multivar Behav Res 2009; 44:182–212.

[23] Milliren CE, Evans CR, Richmond TK, Dunn EC. Does an uneven sample size distribution across settings matter in cross-classified multilevel modeling? Results of a simulation study. Health Place 2018;52:121–6.

[24] Burton A, Altman DG, Royston P, Holder RL. The design of simulation studies in medical statistics. Stat Med 2006; 25:4279–92.

[25] Doedens P, ter Riet G, Vermeulen JM, Boyette LLNJ, Latour CHM, de Haan L. Influence of nursing staff characteristics on seclusion in acute mental health care - A prospective two-year follow-up study. Arch Psychiatr Nurs 2021;35:491–8.

[26] Wasserstein RL, Lazar NA. The ASA statement on p-values: context, process, and purpose. Am Stat 2016;70:129–33.

[27] Diez-Roux AV. Multilevel analysis in public health research. Annu Rev Public Health 2000;21:171–92.

[28] Meyers JL, Beretvas SN. The impact of inappropriate modeling of cross-classified data structures. Multivar Behav Res 2006; 41:473–97.

[29] Luo W, Kwok O-m. The consequences of ignoring individuals' mobility in multilevel growth models: a monte carlo study. J Educ Behav Stat 2012;37:31–56.

[30] Fernandez-Castilla B, Maes M, Declercq L, Jamshidi L, Beretvas SN, Onghena P, et al. A demonstration and evaluation of the use of cross-classified random-effects models for meta-analysis. Behav Res Methods 2018;51(3):1286–304.

[31] mixed — Multilevel mixed-effects linear regression. STATA Manuals.

[32] Bates D, Maechler M, Bolker B, Walker S, Christensen RHB, Singmann H, et al. Linear Mixed-Effects Models using 'Eigen' and S4. CRAN; 2021.

[33] Leckie G, Charlton C. runmlwin: A program to run the MLwiN multilevel modeling software from within Stata. J Stat Softw 2013;1:1–40.