ABSTRACT

Data in health research are frequently structured hierarchically. For example, data may consist of patients nested within physicians, who in turn may be nested in hospitals or geographic regions. Fitting regression models that ignore the hierarchical structure of the data can lead to false inferences being drawn from the data. Implementing a statistical analysis that takes into account the hierarchical structure of the data requires special methodologies.

In this paper, we introduce the concept of hierarchically structured data, and present an introduction to hierarchical regression models. We then compare the performance of a traditional regression model with that of a hierarchical regression model on a dataset relating test utilization at the annual health exam with patient and physician characteristics. In comparing the resultant models, we see that false inferences can be drawn by ignoring the structure of the data.

ABRÉGÉ

Dans le domaine de la recherche en santé, les données sont souvent structurées de façon hiérarchique. Par exemple, des données peuvent regrouper des patients reliés à des médecins, qui à leur tour sont reliés à un hôpital ou une région géographique. L'élaboration de modèles de régression qui négligent cette structure hiérarchique peut mener à des conclusions erronées. La réalisation d'une analyse statistique qui tient compte de la hiérarchie des données requiert des méthodes spécifiques.

Dans notre article, nous présentons le concept des structures hiérarchisées de données et initions le lecteur aux modèles de régression hiérarchiques. Nous comparons ensuite les résultats d'un modèle de régression traditionnel à ceux d'un modèle hiérarchique appliqué à un fichier qui établit des liens entre l'utilisation de tests lors d'examen annuels de santé et les caractéristiques des patients et des médecins en cause. La comparaison entre les deux modèles montre que l'on peut tirer de fausses conclusions si l'on ne tient pas compte de la structure des données.

An Introduction to Multilevel Regression Models

Peter C. Austin, PhD,¹ Vivek Goel, MD, MSc, FRCPC,²
Carl van Walraven, MD, FRCPC, MSc³

Data in health research frequently have a hierarchical structure, with variables measured at each level of the hierarchy. For example, in a study to determine patient and physician factors that influence the use of Pap tests, women would be clustered within physicians. Subjects within the same cluster are often more alike than two randomly chosen subjects, as they will likely have some correlation on important variables. For example, women seen by the same physician may be alike in sociodemographic characteristics. Traditional statistical methods ignore the correlation of outcomes within clusters and tend to underestimate standard errors.¹ This artificially increases the significance of hypothesis tests, increasing the risk of falsely concluding that significant associations exist. Additionally, they do not allow one to incorporate characteristics measured at different levels of the hierarchy. The purpose of this paper is to introduce the reader to hierarchical and

multilevel regression techniques that allow one to explicitly incorporate the hierarchical nature of the data into the analyses, to incorporate variables measured at different levels of the hierarchy, and to examine how regression relationships vary across clusters.

If one is interested simply in making inferences regarding characteristics measured at the lowest level of the hierarchy, then one can analyze the data with sample survey techniques that incorporate the design effect into the analysis.² Alternatively, if the number of groups is small, and one is not interested in the effect of cluster-level characteristics, then one can introduce a categorical group membership variable into the regression models.³ This approach is useful if the groups are distinct entities, and the researcher only wants to draw inferences about those groups, and not to generalize to a larger population of groups. This would occur if the groups denoted ethnic or religious identity.

If one is interested only in characteristics measured at the lowest level of the hierarchy, one wants to assess the variability of regression relationships across clusters, the number of groups is large, the number of subjects in any given group is small, then hierarchical regression analysis techniques – such as random intercept or random coefficient models – can be employed. If one is interested in the effect of characteristics measured at higher levels of the hierarchy, then multilevel analysis is recommended.³⁻¹¹

In this paper we introduce the concept of hierarchical and multilevel regression models, and compare their performance with that of a traditional regression model on a hierarchically structured dataset.

1. Scientist, Institute for Clinical Evaluative Sciences (ICES) and Assistant Professor, Department of Public Health Sciences, University of Toronto, Toronto, ON
2. Associate Professor, Departments of Health Administration and Public Health Sciences, University of Toronto
3. Assistant Professor of Medicine, University of Ottawa; Associate Investigator, Clinical Epidemiology Unit; and Adjunct Scientist, ICES

Correspondence and reprint requests: Peter Austin, Institute for Clinical Evaluative Sciences, G-160, 2075 Bayview Avenue, North York, ON, M4N 3M5, Tel: 416-480-6131, Fax: 416-480-6048, E-mail: peter.austin@ices.on.ca
Dr. Goel is supported in part by a National Health Scholar Award from Health Canada. Dr. van Walraven was an R. Samuel McLaughlin Foundation research fellow at ICES when part of this study was conducted and is currently an Arthur Bond Scholar of the Physicians Services Incorporated Foundation. The views expressed herein are solely those of the authors and do not represent the views of any of the sponsoring organizations.

Hierarchical models

Formal, mathematical descriptions of each hierarchical and multilevel regression model are contained in Table I. Throughout this paper, we will assume the use of unweighted data. If the data arose through multistage cluster sampling or other complex sampling technique, then sampling weights can be incorporated into hierarchical and multilevel regression analyses.

Consider data consisting of patients nested within physicians. Assume that we have measured each patient's age, and the outcome is the cost of treating a certain medical condition. In order to simplify the interpretation of the regression coefficients, patient age is centred on the cohort average (the subject's age minus the average age).

A *traditional linear regression model* can be fit to the data (Figure 1). The intercept term is the average cost of treatment for a patient of average age (age=0, since age is centred). The slope is the amount by which the cost of treatment increases for every year increase in age. In this model, the only source of variation is variation around the regression line. A traditional regression model assumes that the subjects are independent of each other. However, this model ignores the fact that patients treated by the same physician may not be independent of each other, since the outcome – cost of treatment – is heavily dependent on physician behaviour. Additionally, a skilled physician may attract sicker patients, thus increasing his/her cost of treatment. Therefore, greater homogeneity may occur within a given practice than one would expect due to chance alone. This may result in incorrect estimates for the standard errors of the coefficients. Traditional regression models also assume that the effect of patient age on cost of treatment is the same for all physicians. In reality, the effect of patient age on the outcome may vary across physicians. For these reasons, the traditional linear model is of limited utility and may give very deceptive results when used in a hierarchical situation. Furthermore, traditional regression models do not allow one to incorporate physician-level characteristics.

Our next model, a *random intercept model* (Figure 2) starts to incorporate the hierarchical nature of the data. In this

Model Name	Formula	Distributional Assumptions
Traditional Regression Model	$y_i = \beta_0 + \beta_1 \text{age}_i + \varepsilon_i$	$\varepsilon_i \sim N(0, \sigma^2)$
Random Intercept Model	$y_{ij} = \beta_{0j} + \beta_1 \text{age}_{ij} + \varepsilon_{ij}$	$\varepsilon_{ij} \sim N(0, \sigma^2)$ $\beta_{0j} \sim N(\beta_0, \sigma_0^2)$
Random Coefficients Model	$y_{ij} = \beta_{0j} + \beta_{1j} \text{age}_{ij} + \varepsilon_{ij}$	$\varepsilon_{ij} \sim N(0, \sigma^2)$ $\beta_{0j} \sim N(\beta_0, \sigma_0^2)$ $\beta_{1j} \sim N(\beta_1, \sigma_1^2)$ $\text{corr}(\beta_{0j}, \beta_{1j}) = \rho$
Full Hierarchical Model	$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + \varepsilon_{ij}$ $\beta_{0j} = \alpha_{00} + \alpha_{01} w_j + \delta_{0j}$ $\beta_{1j} = \alpha_{10} + \alpha_{11} w_j + \delta_{1j}$	$\varepsilon_{ij} \sim N(0, \sigma^2)$ $\delta_{0j} \sim N(0, \sigma_0^2)$ $\delta_{1j} \sim N(0, \sigma_1^2)$

Note: age_i is the age of the i^{th} patient centred on the cohort average. Similarly, age_{ij} denotes the centred-age of the i^{th} patient treated by the j^{th} physician. The speciality of the j^{th} physician is denoted by w_j .

Parameter	Traditional Model Estimate (p-value)	Hierarchical Model Estimate (p-value*)
Intercept	1.6258	$N(1.7460, \sigma^2=0.2106)$
Patient Age	0.0080 ($p \leq 0.0001$)	$N(0.0139, \sigma^2=0.0001)$ ($p \leq 0.0001$)
Female Patient	-0.3832 ($p \leq 0.0001$)	$N(-0.2600, \sigma^2=0.0307)$ ($p \leq 0.0001$)
Male Physician	0.0606 ($p = 0.0413$)	-0.5460 ($p = 0.0312$)
Physician Practice Volume	0.0271 ($p \leq 0.0001$)	0.0022 ($p = 0.8411$)

Physician-specific regression coefficients for the intercept, patient age and patient gender follow a Normal distribution with given mean and variance for the multilevel model. Hence, the effects of patient age and gender on lab test utilization vary across physicians. Similarly, the physician-specific intercepts follow a Normal distribution with given mean and variance. Hence, the mean number of tests, for an average patient, varies across physicians.

* p-value is for testing whether mean of Normal distribution is different from zero, or for testing whether fixed regression coefficient is different from zero.

model, a separate regression line is fit to each physician's patients, with the regression lines constrained to have the same slopes. Hence, the relationships between the predictor variables and the outcomes are forced to be the same for each physician. In Figure 2, the fitted regression lines vary in their intercepts, indicating that the cost of treatment for an average-aged patient varies between doctors. However, the physician-specific regression lines are parallel, indicating that the increase in the cost of treatment with age is the same for each physician's patients. A random intercept model would be fit if only patient-level characteristics were of interest, and if one believed that the cost of treatment for the average patient varied across physicians, whereas the effect of each patient-level characteristic was the same for each physician's patients.

The next level of complexity is a *random coefficients model* (Figure 3). This family of

models incorporates the hierarchical structure of the data, and allows one to gain a deeper understanding of underlying relationships in the data. In a random coefficients model, a separate regression model is fit to each physician's patients. In contrast to the random intercept model, all the regression coefficients are allowed to vary across physician-specific models. As in the random intercept model, each model is assumed to be similar to the other models. This results in each physician's regression line being shrunk towards the average regression model. The effect of this is that each physician's regression line is a compromise between the regression line fit through that physician's patients and the average regression relationship over all physicians. This allows one to include physicians who treat few patients. In fitting a random coefficients model, one can also examine the correlation between estimated regression coefficients. One would fit a

Figure 1. Simple Linear Regression Model

Figure 2. Random Intercept Model

Note: Figures 1 to 3 relate cost of treatment to patient age. In each figure, the horizontal axis denotes patient age, centred on the cohort average. The vertical axis denotes cost of treatment in thousands of dollars.

random coefficients model if one was only interested in patient-level characteristics, and if one wanted to allow the effect of at least some of the characteristics to vary across physicians.

The model, illustrated in Figure 3, reveals important observations regarding the data. First, the intercept varies, indicating that the cost of treatment for the average patient varies across physicians. Second, the slope of the regression line varies. Therefore, the change in the cost of treatment per year increase in age varies across physicians. Finally, there is a negative correlation between intercept and slope. Those regression lines with lower intercepts tend to have larger slopes. Therefore, those physicians with a relatively low cost of treatment for the average patient tend to have a relatively high increase in cost of treatment with increasing patient age. In contrast, those physicians with relatively high costs of treatment

for the average patient tend to have a relatively low increase in cost of treatment with increasing patient age.

Both the random intercept model and the random coefficients model possess multiple sources of variation. At the patient level, there is variation around the physician-specific regression line. However, there is also variation at the physician level, since the fitted regression lines vary across physicians. Both the random intercept and the random coefficients model are appropriate if one is interested only in patient-level characteristics, and if one wants to examine the variability of the regression models across clusters. These two families of models assume that the clusters are a random sample of clusters, drawn from a larger population of clusters. Alternatively, the traditional regression model could be fit to the data, and “sandwich” estimators (or Huber variance estimators)³ used to correct the standard

errors. However, such methods do not allow the regression relationship to vary across clusters, but assume a uniform regression relationship across clusters. We would not be able to draw the inference that costs of treatment for the average patient varied across physicians.

Our final regression model, a *full multi-level model*, incorporates explanatory variables measured at different levels of the hierarchy. As before, patient age is measured at the lowest level of the hierarchy. However, we now add a variable denoting physician speciality that is measured at the second level of the hierarchy. As in the random coefficients model, we fit a separate model to each physician’s patients, relating cost of treatment to patient age. We now use the physician-level variable to explain the variation in the physician-specific regression models. In a second set of regression models, we relate the variability in the physician-specific intercepts and

slopes to physician training. By examining these two models, we can determine the association of speciality with both treatment costs for the average patient, and changes in cost with patient age. Multilevel models enable an examination of effects of characteristics measured at different levels of the hierarchy, so that valid inferences may be drawn from the data. Traditional regression models and sample survey methods applied to such data will tend to produce misleading results.

DATA SOURCES

For an example of hierarchical modelling, we will use a random sample of data from a study¹² conducted to determine the effects of patient and physician characteristics on diagnostic testing at the Periodic Health Exam (PHE). We identified laboratory investigations ordered for adult Ontarians at their PHE in 1996, using the Ontario Health Insurance Plan database. This documents all health services for more than 95% percent of Ontarians. The Canadian Task Force on the Periodic Health Exam (CTFPHE) recommends few laboratory tests be ordered routinely for healthy adults at their PHE. In the study the outcome variable was the number of lab tests ordered at the PHE not recommended by the CTFPHE.

The data consisted of 14,949 patients nested within 64 physicians. Patient-level variables were age and gender. The physician-level variables used in the model were physician gender and the mean number of non-PHE outpatient assessments per clinically active day.

METHODS

We determined the association of each patient and physician covariate with the number of discretionary tests using two separate models. First, we fit a traditional Poisson regression model,¹³ which ignores the hierarchical structure of the data. This model was fit using SAS version 6.12.¹⁴ Second, we fit a hierarchical Poisson regression model, in which the nesting of patients within physicians is taken into account. The intercepts and the coefficients for patient age and gender were allowed to vary

across regression models (with each coefficient being shrunk towards the average coefficient for that variable). The variation in the physician-specific intercept terms was modelled as a function of physician gender and physician practice volume. The hierarchical regression model was fit using the MLwiN software package.¹⁵

RESULTS

Table II presents the results from the two different models. For the hierarchical regression model, the physician-specific intercepts and the physician-specific effects of patient age and gender are assumed to follow a Normal distribution across physicians. Hence, these distributions rather than a single measure of effect are presented. Both models show agreement on

patient-level predictors. Patient age was seen to be a significant predictor of test utilization, with discretionary test utilization increasing with increasing patient age. In both models, females received significantly fewer discretionary tests than males.

However, the two models produced divergent results for physician-level predictors. The traditional regression model inferred that male physicians ordered significantly more tests than their female colleagues, while the hierarchical model found the converse to be true. The traditional regression model found that test utilization increased with increasing physician practice volume, whereas the hierarchical regression model found no significant effect due to this variable.

The hierarchical regression model found evidence that the mean number of tests for

an average patient varied significantly across physicians, as did the effect of patient age and gender ($p \leq 0.001$ in all cases). The traditional regression model assumed a uniform effect of patient age and gender, and was unable to examine the variation of these effects across physicians.

DISCUSSION

Data in health research are frequently hierarchical in structure, a characteristic increasingly acknowledged in current statistical analyses. Traditional regression models ignore the possible homogeneity within clusters, or the possible influence that a cluster can exert on the outcome, potentially resulting in false conclusions being drawn from the data. Instances of this can be found in the literature. For example, Bennett¹⁶ presented a traditional analysis that concluded that style of teaching had a significant impact on students' educational progress. Aitkin et al.,¹⁷ upon re-analyzing the data and accounting for the clustering of students in classes and schools, and that variables were measured at different levels of the hierarchy, concluded that the association was no longer significant.

Hierarchical regression models allow one to obtain more accurate estimates of regression coefficients and associated p-values, resulting in more valid inferences being drawn from the data. Traditional regression models do not allow the regression relationship to vary across clusters, but assume that the regression model is uni-

form across clusters. If one is only interested in the effect of variables measured at the lowest level of the hierarchy, then sample survey techniques that incorporate the design effect allow one to analyze hierarchical data. There are also methods for correcting the standard errors obtained from traditional regression methods. However, hierarchical regression models allow one to examine how the regression model varies across clusters, and to assess the sources of this variation. Furthermore, only multilevel models allow one to incorporate characteristics measured at higher levels of the hierarchy.

Multilevel regression models can also be used to analyze panel or longitudinal data. In longitudinal data, the same response variable is measured at several different time points on each subject. Such data have a hierarchical structure, with measurement occasions nested within subjects.

In conclusion, hierarchical regression models are a powerful tool to uncover underlying relationships in data that are hierarchically structured. Multilevel models allow one to incorporate variables measured at different levels of the hierarchy. It results in more accurate estimates of regression coefficients and statistical tests of significance, and allows one to draw richer inferences from the data.

REFERENCES

1. Rice N, Leyland A. Multilevel models: Applications to health data. *J Health Services Research and Policy* 1996;1:154-64.
2. Groves RM. *Survey Errors and Survey Costs*. New York, NY: John Wiley & Sons, 1989.
3. Snijders TAB, Bosker RJ. *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling*. London: Sage Publications, 1999.
4. Leyland F, Boddy FA. League tables and acute myocardial infarction. *Lancet* 1998;351:555-58.
5. Duncan C, Jones K, Moon G. Context, composition and heterogeneity: Using multilevel models in health research. *Soc Sci Med* 1998;46:97-117.
6. Normand ST, Glickman ME, Gatsonis CA. Statistical methods for profiling providers of medical care: Issues and applications. *J Am Statistical Association* 1997;92:803-14.
7. Gatsonis CA, Epstein AM, Newhouse JP, et al. Variations in the utilization of coronary angiography for elderly patients with an acute myocardial infarction. *Med Care* 1995;33:625-42.
8. Christiansen CL, Morris CN. Improving the statistical approach to health care provider profiling. *Ann Intern Med* 1997;127:764-68.
9. Langford IH, Bentham G, McDonald A. Multilevel modelling of geographically aggregated health data: A case study on malignant melanoma mortality and UV exposure in the European community. *Statistics in Medicine* 1998;17:41-57.
10. Bryk AS, Raudenbush SW. *Hierarchical Linear Models*. Newbury Park, CA: Sage Publications, 1992.
11. Goldstein H. *Multilevel Statistical Models*, Second Edition. New York, NY: Edward Arnold, 1995.
12. Van Walraven C, Goel V, Austin P. Why are investigations not recommended by practice guidelines ordered at the periodic health examination? *J Evaluation in Clinical Practice* 2000;6:215-24.
13. Cameron AC, Trivedi PK. *Regression Analysis of Count Data*. New York, NY: Cambridge University Press, 1998.
14. SAS Institute Inc., *SAS/STAT Software: Changes and Enhancements through Release 6.12*. Cary, NC: SAS Institute Inc., 1997.
15. Goldstein H, Rasbash J, Plewis I, et al. *A User's Guide to MLwinN. Multilevel Models Project*. London: Institute of Education, University of London, 1998.
16. Bennett N. *Teaching Styles and Pupil Progress*. London: Open Books, 1976.
17. Aitkin M, Anderson D, Hinde J. Statistical modelling of data on teaching styles (with discussion). *J Royal Statistical Society, Series A*, 1981;144:148-61.

Received: December 14, 1999

Accepted: October 16, 2000