ANNALS OF FAMILY MEDICINE

Online Supplementary Material

Dickinson LM, Basu A. Multilevel modeling and practice-based research. *Ann Fam Med.* 2005;3(Suppl 1):S52-S60.

http://www.annfammed.org/cgi/content/full/3/Suppl_1/S52

Supplemental Appendix. Descriptions of Statistical Models

HLM Model 1: Random-Effects Analysis of Variance (ANOVA)

The level-1 (patient-level) model is written as

$$y_{ij} = \beta_{0j} + r_{ij} \qquad r_{ij} \sim N(0, \sigma^2) \qquad (1)$$

Here, we assume that the level-1 error, r_{ij}, is normally distributed with a mean of zero and a constant level-1 variance, σ^2, and β_{0j} is the mean outcome for j^{th} clinic. The level-2 (clinic-level) model is written as:

$$\beta_{0j} = \gamma_{00} + u_{0j} \qquad u_{0j} \sim N(0, \tau_{00}) \qquad (2)$$

where γ_{00} is the grand mean outcome in the population and u_{0j} is a clinic-level random effect that represents the difference between the mean number of alcohol-free weeks for patients in practice j and the grand mean. This random effect is assumed to have a mean of zero and a constant variance, τ_{00}. Substituting (2) in (1) yields the combined (mixed) model

$$y_{ij} = \gamma_{00} + u_{0j} + r_{ij} \qquad (3)$$

which is the 1-way ANOVA model with grand mean γ_{00}, a clinic effect u_{0j}, and a patient effect r_{ij}. Note that there are 2 sources of variation, 1 at the level of the individual patient and 1 at the clinic level. Because there are no patient- or practice-level predictors, this model is also known as the *unconditional model.*

Intraclass Correlation (ICC)

The ICC is the proportion of the total variance that is due to differences among clinics (level 2) and can be expressed

$$\rho = \tau_{00} / (\tau_{00} + \sigma^2)$$

where τ_{00} is the variance due to differences among clinics, σ^2 is the variance among individuals within the same clinic, and $(\tau_{00} + \sigma^2)$ is the total variance in number of alcohol-free weeks for all patients in the study.

REG Model 1: Traditional Linear Regression Model 1

The most commonly used model to study the effect of x_{ij} on y_{ij} can be written as follows:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + r_{ij} \qquad r_{ij} \sim N(0, \sigma^2) \qquad (4)$$

This model is essentially a patient-level (level-1) model. However, one can visualize this model as a 2-level model wherein the clinic-level (level-2) model is specified as:

$$\beta_0 = \gamma_{00}$$
$$\beta_1 = \gamma_{10} \qquad (5)$$

These equations indicate that neither the intercept nor the slope parameter in this model is allowed to vary by clinic. In other words, the mean effect of physician advice on alcohol-free weeks (γ_{10}) and the average number of alcohol-free weeks without any physician advice (intercepts) (γ_{00}) are *assumed to be the same across all clinics*.

HLM Model 2: Random-Intercept Model

In this model, we allow the intercept to be specified as random to allow for variability in mean number of alcohol-free weeks without any physician advice (ie, x_{ij} =0) across clinics. At the patient level (level 1), the model is specified as:

$$y_{ij} = \beta_{0j} + \beta_1 x_{ij} + r_{ij} \qquad r_{ij} \sim N(0, \sigma^2) \qquad (6)$$

Note, in contrast to (4), the intercept in this model is clinic specific as denoted by the subscript *j* in β_{0j}. The clinic-level (level-2) model is

$$\beta_{0j} = \gamma_{00} + u_{0j} \qquad u_{0j} \sim N(0, \tau_{00})$$
$$\beta_1 = \gamma_{10} \qquad (7)$$

Here γ_{00} represents the average clinic mean of number of alcohol-free weeks without any physician advice in the population of clinics, whereas $\beta_{0j} = \gamma_{00} + u_{0j}$ represents the corresponding mean in clinic *j*. Thus u_{0j} is a clinic-level random effect that can be interpreted as the deviation of the intercept for clinic *j* from the population mean intercept. The variance of the random effects u_{0j}, τ_{00}, represents the population variance in the mean number of alcohol-free weeks without physician advice among clinics. Consequently, σ^2 represents the residual variance of the outcome in the population after controlling for clinic-level baseline heterogeneity in the patient population. Also note, similar to (4), the effect of variable x_{ij}, hours of physician advice, is *constrained to be the same for all clinics (ie, the slope is fixed).* This model is often referred to as a random-intercept model, and it can also be thought of as a 1-way ANCOVA with random effects.

HLM Model 3: Random-Coefficients Model

In this model, we allow both the intercept and the slope to be random to account for variability across clinics; therefore, at the patient level (level 1), the model is specified as:

$$y_{ij} = \beta_{0j} + \beta_{1j} x_{ij} + r_{ij} \qquad r_{ij} \sim N(0, \sigma^2) \qquad (8)$$

Note, in contrast to the slope in (4) and (6), the slope in this model is clinic specific as denoted by the subscript *j* in β_{1j}. The clinic-level (level-2) model is specified as:

$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}$$
$$\begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix}\right) \qquad (9)$$

Here β_{0j} and its related components in (9) have the same interpretation as in (7). The random effect u_{1j} is the difference in slope for practice j from the population mean slope and is distributed normally with mean 0 and variance τ_{11}. It is appropriate to test whether there is any significant heterogeneity in the effects of physician advice on alcohol consumption across clinics by testing the hypothesis H_0: $\tau_{11} = 0$ before deciding on a final model. The covariance between the random effects for the intercept and the slope is denoted by τ_{01}.

HLM Model 4: Intercept as Outcomes Model

Here, the level-1 model is exactly the same as in the random-intercept model in (6) or as in the random-coefficients model in (8). The corresponding level-2 models are described below.

The level-2 model corresponding to a random-intercept level-1 model is given by

$$\beta_{0j} = \gamma_{00} + \gamma_{01} w_j + u_{0j} \quad u_{0j} \sim N(0, \tau_{00}) \qquad (10)$$
$$\beta_1 = \gamma_{10}$$

This model suggests that the heterogeneity in the mean number of alcohol-free weeks without any physician advice (intercepts) across clinics may be explained by clinic-level characteristics such as urbanicity of the clinic (w_j). Whereas γ_{00} represents the average clinic mean number of alcohol-

free weeks without any physician advice in the population of clinics in rural areas, γ_{01} represents the difference in average clinic mean of number of alcohol-free weeks between the population of clinics in urban and rural areas. This model still assumes, however, that there is no heterogeneity in the effect of physician advice on alcohol consumption across clinics (slope). As in (7), u_{0j} is a clinic-level random effect that can be interpreted as the deviation of the intercept for clinic *j* from the population mean intercept either in a rural or in an urban area. The proportion of variance in β_{0j} that is explained by introducing a level-2 covariate is given by comparing the estimates of τ_{00} in (7) and (10):

$$\% \text{ of Var}(\beta_{0j}) \text{ explained by } w_j = \frac{\hat{\tau}_{00}(7) - \hat{\tau}_{00}(10)}{\hat{\tau}_{00}(7)} \qquad (11)$$

HLM Model 5: Intercept and Slope as Outcomes Model

This model is very similar to the intercept as outcomes model (above). Here we try to explain heterogeneities in slopes as well as intercepts with level-2 characteristics. Again, the level-1 model is the same as in the random-intercept model (6). The level-2 model corresponding to a level-1 random-coefficients model is given as

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01} w_j + u_{0j} \\ \beta_{1j} &= \gamma_{10} + \gamma_{11} w_j + u_{1j} \end{aligned} \qquad \begin{pmatrix} u_{0j} \\ u_{1j} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_{00} & \tau_{01} \\ \tau_{01} & \tau_{11} \end{pmatrix} \right) \qquad (12)$$

This model suggests that both the heterogeneity in the mean number of alcohol-free weeks without any physician advice (intercept) across clinics and the heterogeneity in the effect of physician advice on alcohol consumption (slope) across clinics *may be explained by clinic-level characteristics such as urbanicity of the clinic*. The interpretations of β_{0j} and its related components in (12) are the same as those in (10). In the context of the alcohol data, γ_{10} represents the average physician advice slope for the population of clinics in rural areas, and γ_{11} is the difference in average physician advice slope between the population of clinics in urban and rural areas.

If after adjusting for urbanicity of clinics, the residual variance for slope parameter β_{1j} is not statistically significant at the 5% level (although it is significant at the 10% level), the researcher may want to retain the random slope or may choose to specify the slope as nonrandomly varying such as:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01} w_j + u_{0j} \quad u_{0j} \sim \text{N}(0, \tau_{00}) \\ \beta_1 &= \gamma_{10} + \gamma_{11} w_j \end{aligned} \qquad (13)$$

REG Model 2: Traditional Regression Model 2

Traditionally, an interaction model is used to study how w_j moderates the effect of x_{ij} on y_{ij} as follows:

$$y_{ij} = \gamma_{00} + \gamma_{01} w_j + \gamma_{10} x_{ij} + \gamma_{11} x_{ij} * w_j + r_{ij} \qquad r_{ij} \sim \text{N}(0, \sigma^2) \qquad (14)$$

This model is essentially a patient level (level-1) model. However, one can visualize this model as an HLM with nonrandomly varying slopes and intercept. Here the level-1 model is the same as that in (8) and the level-2 model is specified as:

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01} w_j \\ \beta_{1j} &= \gamma_{10} + \gamma_{11} w_j \end{aligned} \qquad (15)$$

This equation indicates that both the intercept and the slope parameter in this model are allowed to vary by the urbanicity of the clinic but not otherwise. This model can lead to inefficient estimation of parameters *if the residual variances for the slope and intercept are significant even after controlling for* w_j. For example, we would expect that the traditional model 2 will generate this type of inefficiencies from the alcohol data set because Table 2 reveals that the residual variance of

the intercept is significant even after controlling for urbanicity (w_j).

HGLM Model: Adaptation of Logistic Regression Model to Hierarchical Structure

Level-1 link function. For binary outcomes, the level-1 model will use a logit link

$$\eta_{ij} = \log\ (\varphi_{ij}/(1 - \varphi_{ij})), \qquad (16)$$

where φ_{ij} is the probability of success and η_{ij} is the log of the odds of success (eg, patient i in clinic j is screened).

Level-1 structural model. In the level-1 model, individual outcomes, in terms of log(odds), will be modeled as a function of p patient characteristics (fixed effects), such as social and demographic variables, and comorbid conditions.

$$\eta_{ij} = \beta_{0j} + \beta_{1j} X_{1ij} + \beta_{2j} X_{2ij} + \ldots + \beta_{pj} X_{pij} \qquad (17)$$

Level-2 model. As in the case of continuous outcomes, the clinic-level models specify the relationship between the clinic-level predictor(s) and the coefficients in the level-1 model. For example, we would be interested in testing whether the intervention increased the likelihood that a diabetic patient would be screened, after adjusting for patient-level covariates. The level-2 model would be specified in the same manner as for continuous outcomes, as in (12) above.

HLM = hierarchical linear model; REG = regression; ANCOVA = analysis of covariance; HGLM = hierarchical generalized linear model.