# Supplemental materials (1553 words)

## Statistical details

### Data

The study data contains the following variables:

* stratum: stratum identifier ()
* individual: individual identifier()
* copd: COPD status (1: new diagnosis of COPD; 0: otherwise)()
* old: dummy variable for age (1: old; 0: young)()
* female: dummy variable for gender (1: female; 0: male)()
* medinc, lowinc: dummy variables for medium and low income (the reference category is high income)(,)
* lowed: dummy variable for education (1: low; 0: high) ()
* alone: dummy variable for living alone (1: yes; 0: no) ()
* immigrant: dummy variable for immigrant (1: yes; 0: no) ()

where 96 strata have been defined by forming the combinations between age, gender, income, education, alone and immigrant ().

### Model 1

The two-level variance-components logistic regression model for can then be written as

where denotes the probability of a new COPD diagnosis for individuals in stratum , denotes the intercept, and denotes the stratum random effect. The model includes no covariates and so the stratum random effect captures both the main effects of the variables used to define the strata and their two- and higher-way interactions. The stratum random effects are assumed normally distributed with zero mean and constant between-stratum variance , a parameter to be estimated.

The latent response formulation of the above model can be written as

where denotes the continuous latent response or propensity of receiving a new COPD diagnosis and is the individual level residual assumed to follow the standard logistic distribution with mean zero and variance 3.29 ( where is the mathematical constant ).

The degree of clustering in the latent responses is typically summarized by one of two statistics: the intraclass correlation (ICC) or the variance partition coefficient (VPC). In the current and subsequent models, the formula for the ICC and VPC coincide.

The ICC is interpreted as the expected correlation between the latent responses of two individuals from the same stratum. The VPC is interpreted as the proportion of the latent response variation which lies between strata. The proportion of the latent response variation which lies within strata is given by . These stratum and individual level VPCs are typically multiplied by 100 and reported as percentages.

The probability of receiving a new COPD diagnosis in stratum is calculated as

The relative risk (RR) comparing the probability in stratum to that in stratum is therefore calculated as

The odds of receiving a new COPD diagnosis in stratum are calculated as

The odds ratio (OR) comparing the odds in stratum to those in stratum is therefore calculated as

The average stratum has and so the odds ratio comparing stratum to the average stratum is therefore calculated as .

### Model 2

Model 2 extends model 1 by entering one covariate at a time. For example, the model which enters old, the dummy variable for older individuals, can be written as

Adding stratum level covariates to the model will explain away the between-stratum variance. The degree to which the between-stratum variance reduces as we move from model 1 to 2 can be expressed by the proportional change in variance (PCV) statistic, calculated as

where and denote the between-stratum variance from models 1 and 2 respectively. PCVs are typically multiplied by 100 and reported as percentages.

The ICC and VPCs for this model are calculated as before, but are now interpreted as conditional ICC and VPCs that summarize the degree of clustering in the adjusted latent responses.

### Model 3

Model 3 extends models 1 and 2 by entering all the covariates: old, female, medinc, lowinc, lowed, alone and immigrant. The model can be written as

Here the seven fixed-effects covariates capture the main effects of the explanatory variables on the log-odds of COPD, while the two-way and all higher-way interaction effects between these variables are captured by the random stratum effect .

Model 3 therefore decomposes the log-odds of a new COPD diagnosis into two parts. That part due to the main effects of the explanatory variables (the fixed-part of the model) and that part due to interactions between the explanatory variables (the random-part of the model: the random stratum effect). Thus, the interpretation of the stratum effect is that it quantifies the difference between the log-odds in stratum when we acknowledge the interaction effects and the log-odds in stratum when we ignore the interaction effects.

To aid interpretation, we can work on the probability scale. The probability of receiving a new COPD diagnosis in stratum is calculated as

This probability can be decomposed into that part due to the main effects of the explanatory variables and that part due to the interactions between the explanatory variables

We calculate the former as

We calculate the latter by subtraction.

The relative risk (RR) comparing the probability in stratum to that in stratum when we only acknowledge the main effects of the explanatory variables, that is, ignoring the interaction effects, is calculated as

Alternatively, we can re-express the stratum effects as odds-ratios by exponentiating them . Thus, if, for example, stratum has an predicted odds-ratio of 2 then the actual odds of a new COPD diagnosis in stratum would be two times higher than the odds we would predict when we naively ignore the presence of interaction effects between the covariates.

### A note on comparing predicted stratum effects across models

Conceptually the stratum probabilities (based on including all terms in the model) will be identical across the three models, but in practice the corresponding predicted stratum probabilities will typically exhibit small differences across the three models. In this section we explain the intuition for this difference.

Each model simply decomposes the logit of the stratum probabilities into two parts: the fixed- and random-parts of the model. The models differ, however, in how they define their fixed-parts and the definition of their random-parts therefore change to accommodate these differences leading to no change in the resulting stratum probabilities. The stratum probabilities are identical across models. However, when we attempt to assign values to the stratum probabilities by predicting them post-estimation, small discrepancies arise across the models. These discrepancies relate to the standard application of shrinkage in the prediction of the stratum random effects from which the predicted stratum probabilities are formed. The more reliable the observed stratum proportions (due to higher observed proportions and larger cell sizes) the less severe the shrinkage and the smaller any observed discrepancies between models. Taken to the extreme, when the predicted stratum effects exhibit no shrinkage (or where we simply ignore the shrinkage when making the calculations) the predicted stratum probabilities are once again identical across the three models.

### Estimation

We fit all models by Markov chain Monte Carlo (MCMC) methods as implemented in MLwiN 3.01 (Browne, 2017; Charlton et al., 2017). We call MLwiN from Stata using the runmlwin command (Leckie and Charlton, 2013).

We specify ‘diffuse’ (vague, flat, or minimally informative) prior distributions for all parameters. We fit all models with a burn-in period of 5,000 iterations and a monitoring period of 50,000 iterations. Informal visual assessments of the parameter chains and standard MCMC convergence diagnostics suggest that these periods are sufficiently long.

The parameter point-estimates and 95% credible intervals (CIs) are the means and the 2.5th and 97.5th percentiles of the respective MCMC chains for each parameter. We use the deviance information criterion (DIC) to compare the fit of models (Spiegelhalter, Best, Carlin, & van der Linde, 2002): Models with smaller DIC values are preferred to those with larger values, with differences of five or more considered substantial (Lunn, Jackson, Best, Thomas, Spiegelhalter, 2012).

### A note on calculating 95% credible intervals for functions of the model parameters

An advantage of MCMC methods is that it is easy to calculate 95% CIs around complex non-linear function of the model parameters. One simply calculates the function of interest at every iteration of the MCMC chains. This gives an MCMC chain for the function of interest. The mean of this chain is then reported as the point estimate of the function while the 2.5th and 97.5th percentiles of the chain are presented at the 95% CI. We apply this general approach when calculating the 95% CIs for the predicted stratum effects presented in Table 2 and the predicted stratum effects and their decomposed parts in Table 3.

### References

Browne, W.J. (2017). MCMC Estimation in MLwiN v3.00. Centre for Multilevel Modelling, University of Bristol.

Charlton, C., Rasbash, J., Browne, W.J., Healy, M. and Cameron, B. (2017). MLwiN Version 3.00. Centre for Multilevel Modelling, University of Bristol.

Leckie, G., & Charlton, C. (2013). Runmlwin: A program to run the MLwiN multilevel modelling software from within Stata. Journal of Statistical Software, 52, 1-40.

Lunn, D., Jackson, C., Best, N., Thomas, A., & Spiegelhalter, D. (2012). The BUGS book: A practical introduction to Bayesian analysis. Boca Raton, FL: Chapman & Hall/CRC Texts in Statistical Science.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B, 64, 583–639.