Unobserved heterogeneity and the comparison of coefficients across nested logistic regression models: how to avoid comparing apples and oranges

Patrick Brzoska · Odile Sauzet · Jürgen Breckenkamp

Introduction

In public health research the focus is often on dichotomous outcomes, such as onset of disease, recovery or death. Logistic regression or other non-linear probability models are commonly used to model such outcomes, for example, for purposes of controlling for confounding variables. Usually, beta coefficients (β) or odds ratios (e^β) are reported as measures of effect size (Hosmer et al. 2013). Often researchers are also interested in comparing effect measures across nested models, for example, to examine whether the association between an exposure (e.g., smoking) and an outcome (e.g., cardiovascular disease) is suppressed or confounded by a third variable (e.g., socioeconomic status) (MacKinnon et al. 2000). In this regard, models are often presented in a step-wise manner, where a crude, i.e., unadjusted, baseline model is subsequently extended by the inclusion of one or more third variables. The difference between the coefficients of the baseline and the subsequent models are often interpreted substantively (Hosmer et al. 2013).

For continuous outcomes modeled using linear models, this comparison is straightforward. In logit and other non-linear probability models, however, the comparison may be biased by unobserved heterogeneity between the models,

i.e., variation of the dependent variable resulting from the influence of unobserved variables. This effect is due to particular assumptions regarding the fixed variance of the residuals that are pertinent to these models. It has been acknowledged in econometrics quite some time ago (Wooldridge 2010) and has recently been also discussed in sociological literature (Mood 2010); however, it is often neglected in public health research. By means of the present article we aim to increase the awareness for this characteristic of the logit model among public health researchers. We also illustrate potential remedies which are available to take unobserved heterogeneity into account.

Logit models and unobserved heterogeneity

A linear model is defined as

$$y_i = \beta_0 + x_{i1}\beta_1 + \ldots + x_{ij}\beta_j + \varepsilon_i, \tag{1}$$

where x_{ij} is the jth independent variable in the model observed for the ith individual, β_j is its coefficient, β_0 is the intercept parameter, and ε_i are the residuals, which are normally distributed with expected value of 0 and variance σ^2. The variance of y_i is composed of the variance explained by the model and the residual variance. Once covariates are entered into the model, the proportion of explained variance increases while the residual variance decreases. The total variance of y_i remains constant.

The situation is different for a logit model, which models the probability for the occurrence of a certain event. The logit model can be conceptualized as a threshold model, according to which the observed dichotomous variable is determined by an underlying latent continuous variable y^* (Long and Freese 2014). This latent variable can be considered to represent the propensity for the observed dichotomous outcome to occur (but does not

P. Brzoska (✉)
Epidemiology Unit, Faculty of Behavioral and Social Sciences, Institute of Sociology, Chemnitz University of Technology, Chemnitz, Germany
e-mail: patrick.brzoska@soziologie.tu-chemnitz.de

O. Sauzet · J. Breckenkamp
Department of Epidemiology & International Public Health, School of Public Health, Bielefeld University, Bielefeld, Germany

necessarily always have a real meaning). Unless a certain threshold is met, y_i equals 0, else y_i equals 1. y^* has a linear relationship with covariates similar to Eq. 1. However, the residuals follow a standard logistic distribution and have a fixed variance of $\pi^2/3$. Because of this constraint implemented in the definition of the logit model, the total variance of y^* changes once covariates are entered into the model. This can be easily demonstrated with a simple simulated dataset ($n = 10,000$) consisting of the three normally distributed variables y, x_1 and x_2 for which the following conditions apply [cf. other, in part more complex, simulations, for example in Mood (2010)]: x_1 and x_2 are moderately correlated with y ($r_{x1, y} = 0.6$; $r_{x2, y} = 0.6$) but are uncorrelated with each other. When y_i is regressed on x_1 in a crude model (Model 1), the effect size (β) of x_1 is—as can be expected—almost equal to the effect size of x_1 in a model in which also x_2 is included, because both independent variables are uncorrelated (Table 1). In a logit model (for illustration purposes we dichotomized y_i at the median), this is not true. The effect size for x_1 in terms of β (or odds ratio) for the crude model (Model 1) is considerably smaller than for the adjusted model (Model 2), in which also x_2 is taken into account, despite both variables being uncorrelated (Table 1). The reason is that unlike in the linear model, the latent variance of the underlying dependent variable (in the logit case y^*) changes once x_2 is added to the model, resulting in a rescaling of the coefficients (the respective Stata script illustrating this simulation can be obtained from the authors). This is similar to a situation of comparing coefficients of two models which both examine weight as the outcome, but where weight is measured in kilograms in one model and in pounds in the other model.

Unless unobserved heterogeneity is taken into account, comparisons of coefficients between nested logit models may therefore be distorted because they are based on different scales—basically leading to the comparison of apples and oranges. Consequently, if differences between coefficients are observed across models, it remains unclear, whether they present substantive effects or are partially or fully reflecting a bias from unobserved heterogeneity.

Considering unobserved heterogeneity in logit models

Different remedies have been discussed in literature to take unobserved heterogeneity in the comparison of nested models into account (see Mood 2010 and Karlson et al. 2012 for an overview), all of which have limitations. Most prominent solutions advocate the use of coefficients other than β and odds ratio. While beta coefficients that are standardized on the latent variance of y^* (y-standardization) have been shown to potentially lead to wrong conclusions if the predicted logit is highly skewed (Karlson et al. 2012; Best and Wolf 2012), measures based on predicted probabilities, such as average marginal effects (AME), are less affected by bias arising from unobserved heterogeneity in most cases unless independent variables are extremely skewed and unobserved heterogeneity is extremely high (Table 1). AMEs tell for each variable in a regression model, how much, averaged across all observations, the probability for an event changes by one unit increase of the independent variable. As can be seen from Table 1, for the simulated data, the AMEs for x_1 for the crude and adjusted models are very similar.

Karlson et al. (2012) recently proposed another type of solution (the Karlson–Holm-Breen [KHB] method), which allows to separate the effects of confounding and rescaling by re-parameterizing the crude model in a way that the scaling remains equal to the scaling of the adjusted model (Table 1) (see also Kohler et al. 2011 for a practical application). This is facilitated by an intermediate step in which the residuals from a regression of the outcome on all covariates are considered as a third variable in the crude

Table 1 Illustration of different approaches to control for unobserved heterogeneity. Results for the coefficients of variable x_1 based on a simulated data set

	Model 1 (crude; only x_1 included)	Model 2 (adjusted; x_1 and x_2 included)
Linear model (linear regression of y_i on x_1)		
Variance of y_i	9.00	9.00
β	0.66	0.67
Logit model (logistic regression of dichotomized y_i on x_1)		
Latent variance of y^*	5.41	24.55
β [odds ratio]	0.48 [1.62][#]	1.10 [3.00][#]
Average marginal effect (AME)	0.087[#]	0.090[#]
β [odds ratio] rescaled using KHB method[a]	1.09 [2.97][#]	1.10 [3.00][#]

[#] $p < 0.01$

[a]   As proposed in Karlson et al. (2012)

model. As is shown by simulation studies (Karlson et al. 2012), this decomposition procedure is robust against bias arising from unobserved heterogeneity and is also not affected by a skewed distribution of the independent variables.

An empirical example: effectiveness of rehabilitation among migrants

In empirical studies, the bias introduced by unobserved heterogeneity can be less large than shown in the simulation study, and frequently, the conventional approach of model comparison using β-coefficients and odds ratios will lead to exactly the same conclusions as the use of alternative measures (Best and Wolf 2012). However, the difference of coefficients between nested models may also be under- or overestimated if unobserved heterogeneity is large and not taken into account. In the following, we illustrate this by means of an empirical example concerning migrant health. A frequent question of social epidemiology is whether differences in terms of the utilization and effectiveness of health services that are observed between migrants and the autochthonous population are caused by a different distribution of demographic and socioeconomic factors between both population groups or whether other factors going beyond the role of social determinants play a role (see for example, Brzoska et al. 2010 and Brzoska et al. 2016 for a substantive discussion of this type of research).

In Table 2 we illustrate how much differences in low occupational performance after rehabilitation (a frequently used measure of rehabilitation effectiveness) between German and Turkish nationals are affected by demographic and socioeconomic factors. We use a random sample ($n = 8839$) of all German and Turkish cases who completed a rehabilitation after diseases of the circulatory system in Germany in the years 2011–2013 granted by the

German Statutory Pension Insurance Scheme (the secondary dataset is available from the German Statutory Pension Insurance Scheme as a public use file; Deutsche Rentenversicherung Bund 2016).

Model 1 presents different types of crude coefficients for Turkish nationals. In Model 2 these coefficients are adjusted for demographic and socioeconomic factors. German nationals are the reference category in both models. Model 1 shows that Turkish nationals are at a 2.8-times higher chance (odds ratio = 2.80) of having an only limited occupational performance at rehabilitation discharge. Once demographic and socioeconomic factors are controlled for, the odds ratio decreases to 2.09 (Model 2). Obviously, social determinants play a role in explaining differences between German and Turkish nationals in terms of rehabilitation effectiveness. The question is how large this role really is. Based on the underlying β-coefficients, the reduction in effect size corresponds to 28.5%. As outlined previously, this type of comparison could be biased by differences in the scaling of the two models resulting from unobserved heterogeneity. Comparisons of rescaled coefficients or of AMEs can therefore provide a more accurate picture of the true difference between the crude and adjusted coefficients. As these coefficients in Table 2 show, the proportion of the difference between Turkish and German nationals regarding the effectiveness of rehabilitation which is explained by demographic and socioeconomic factors is in fact considerably larger than initially assumed using a conventional comparison of odds ratios (between 39.9 and 43.1%, depending on the method used).

Conclusion

Taking unobserved heterogeneity into account in the comparison of coefficients across logistic regression models is a complex issue for which awareness in public health

Table 2 Limited occupational performance following rehabilitation after diseases of the circulatory system in German and Turkish nationals residing in Germany (random sample of all cases who completed medical rehabilitation in the years 2011–2013 granted by

the German Statutory Pension Insurance Scheme; logistic regression models adjusted for demographic and socioeconomic factors; $n = 8839$; Deutsche Rentenversicherung Bund 2016)

	Model 1 (crude)	Model 2 (adjusted for demographic and socioeconomic factors[a])	Reduction of effect size between Model 1 and 2 (%)
Coefficient for Turkish nationals (reference group: German nationals)			
β [odds ratio]	1.03 [2.80][#]	0.74 [2.09][#]	28.5
Average marginal effect (AME)	0.153[#]	0.092[#]	39.9
β [odds ratio] rescaled using KHB method[b]	1.30 [3.67][#]	0.74 [2.09][#]	43.1

[a] Sex, age, marital status, educational level, occupational position, occupation and area of residence

[#] $p < 0.01$

[b] As proposed in Karlson et al. (2012)

research must be increased. Researchers have different adjustment methods at hand, which are also subject of a growing amount of method research on that topic. Although there is no consensus on which method is the best solution, the decomposition procedure suggested by Karlson et al. (2012) has been shown to be robust against bias arising from unobserved heterogeneity. To the best of our knowledge, currently only Stata allows a user-friendly application of this procedure through the user-written 'khb' program (Kohler et al. 2011). Alternatively, the use of predicated probabilities, for example in the form of AMEs, has been shown to be an easy-to-apply strategy which is less affected by unobserved heterogeneity and is available in most statistical packages.

Compliance with ethical standards

Ethical approval The use of the secondary data presented in this article follows the requirements as defined by the German Social Code VI, IX and X. Since the data are fully anonymized, no additional ethical approval for the present analysis was necessary.

Conflict of interest The authors declare that they have no conflict of interest.

References

Best H, Wolf C (2012) Modellvergleich und Ergebnisinterpretation in Logit-und Probit-Regressionen. Kölner Z Soziol Soz 64:377–395

Brzoska P, Voigtländer S, Spallek J, Razum O (2010) Utilization and effectiveness of medical rehabilitation in foreign nationals residing in Germany. Eur J Epidemiol 25:651–660

Brzoska P, Sauzet O, Yilmaz-Aslan Y, Widera T, Razum O (2016) Self-rated treatment outcomes in medical rehabilitation among German and non-German nationals residing in Germany: an exploratory cross-sectional study. BMC Health Serv Res 16:105

Deutsche Rentenversicherung Bund (2016) Public Use File Abgeschlossene Rehabilitation im Versicherungsverlauf 2006 bis 2013 (PUFRSDLV13B). Forschungsdatenzentrum der Deutschen Rentenversicherung Bund, Berlin

Hosmer DW, Lemeshow S, Sturdivant RX (2013) Applied logistic regression. Wiley, Hoboken

Karlson KB, Holm A, Breen R (2012) Comparing regression coefficients between same-sample nested models using logit and probit: a new method. Sociol Methodol 42:286–313

Kohler U, Karlson KB, Holm A (2011) Comparing coefficients of nested nonlinear probability models. Stata J 11:420–438

Long JS, Freese J (2014) Regression models for categorical dependent variables using Stata. Stata press, College Station

MacKinnon DP, Krull JL, Lockwood CM (2000) Equivalence of the mediation, confounding and suppression effect. Prev Sci 1:173–181

Mood C (2010) Logistic regression: why we cannot do what we think we can do, and what we can do about it. Eur Sociol Rev 26:67–82

Wooldridge JM (2010) Econometric analysis of cross section and panel data. MIT press, Cambridge