Qual Quant
DOI 10.1007/s11135-012-9765-5

Adjusting for differential misclassification in multilevel models: the relationship between child exposure to smoke and cognitive development

Maria Eugénia Ferrão · Harvey Goldstein

Abstract Misclassification is found in many of the variables used in social sciences and, in practice, tends to be ignored in statistical analyses, and this can lead to biased results. This paper shows how to correct for differential misclassification in multilevel models and illustrates the extent to which this changes fixed and random parameter estimates. Reliability studies on self-reported behaviour of pregnant women suggest that there may be differential misclassification related to smoking and, thus, to child exposure to smoke. Models are applied to the Millennium Cohort Study data. The response variable is the child cognitive development assessed by the British Ability Scales at 3 years of age and explanatory variables are child exposure to smoke and family income. The proposed method allows a correction for misclassification when the specificity and sensitivity are known, and the assessment of potential biases occurring in the multilevel model parameter estimates if a validation data sample is not available, which is often the case.

Keywords Differential misclassification · Multilevel model · MCMC estimation · Millennium Cohort Study · Pregnancy · Smoking

1 Introduction

Ignoring measurement error and misclassification in the predictor variables of statistical models typically leads to biased parameter estimates and standard errors as well as a loss of power in detecting the impact of explanatory variables.

M. E. Ferrão (✉)
Department of Mathematics, University of Beira Interior, Rua Marquês d' Avila e Bolama, 6200-001 Covilhã, Portugal
e-mail: meferrao@ubi.pt

H. Goldstein
University of Bristol, 2 Priory Road, Bristol BS8 1TX, UK
e-mail: h.goldstein@bristol.ac.uk

Published online: 09 September 2012

Table 1 Sensitivity and Specificity of self-reported smoking during pregnancy

Study	Country	Sample size	Sensitivity P(1|1)	Specificity P(0|0)	P(0|1)	P(1|0)
Shipton et al. (2009)	West of Scotland	3475	0.74	–	0.22; 0.26;0.39	–
Burstyn et al. (2009)	Canada (Edmonton)	377	0.47	0.95	0.53	0.05
England et al. (2007)	US	4289	0.78	–	0.22	–
Pärna et al. (2005)	Estonia (Tallinn)	1360	0.79	–	0.21	–
Lindqvist et al. (2002)	Sweden (Blekinge)	509	0.94	–	0.06; 0.32	–
Klebanoff et al. (2001)	US	105	0.85	0.95	0.15	0.05
Ford et al. (1997)	New Zealand		0.78	–	0.22	–

A large statistical literature on the modelling of such errors exists, mostly dealing with the case of single level models (Carroll et al. 2006; Fuller 2006). However, the effect of measurement error and misclassification in multilevel models has been far less explored. Recent work by Goldstein et al. (2008) has developed Bayesian methods to adjust for both measurement errors and misclassification in multilevel models with a continuous response variable.

In the 1960s the self-report on smoking by pregnant women was assumed to be fairly accurate. Knowledge about the harmful effects of mothers' smoking during pregnancy has increased over the last few decades. Maternal smoking during pregnancy is associated with adverse perinatal outcomes (Butler et al. 1972; Kramer 1987; Simpson 1957; Walsh 1994) and child development (Davie et al. 1972; Goldstein 1971; Julvez et al. 2007). Even recent studies that report no evidence of association between maternal smoking during pregnancy and cognitive performance point out the occurrence of exposure misclassification as a limitation for the results obtained (e.g. Kafouri et al. 2009).

In fact self-reporting of smoking behaviour will be subject to underreporting and the degree of underreporting will also depend on the way data are collected. Studies conducted in different social contexts and countries show that the specificity of self-reporting smoking is close to 1 but the sensitivity varies between 0.47 and 0.96 (Burstyn et al. 2009; England et al. 2007; Ford et al. 1997; Klebanoff et al. 2001; Lindqvist et al. 2002; Pärna et al. 2005; Shipton et al. 2009). Table 1 summarizes the results obtained in seven studies conducted in different countries.

Despite increasing efforts to encourage researchers to be more critical as regards measuring and to promote the best means of reducing measurement errors and misclassification to a minimum, most research papers ignore them. This paper extends the approach proposed by Goldstein et al. (2008) in order to consider differential misclassification in the predictor variables, and evaluates the impact of both types of misclassification in a variance components model. The method can be straightforwardly applied more complex multilevel models. The Millennium Cohort Study (MCS) (Institute of Education/University of London 2009) dataset is used to analyse the relationship between child exposure to smoke and cognitive development. We consider the effect of both non-differential and differential misclassification in the binary predictor variable that represents exposure, on the fixed and random parameter estimates.

The outline of the paper is as follows. Section 2 presents the MCS description, the sensitivity and specificity of self-reported smoking by pregnant women, and the specification and estimation for the multilevel model with the adjustment for differential misclassification in the predictor (exposure) variable. The results obtained for the illustrative example are analyzed in Section 3. Section 4 presents a discussion.

2 Data and methods

2.1 MCS dataset

The MCS (Institute of Education/University of London 2009) sample consists of children born between 1 September 2000 and 31 August 2001 (in England and Wales), and between 23 November 2000 and 11 January 2002 (in Scotland and Northern Ireland), alive and living in the UK at age 9 months, eligible to receive Child Benefit at that age, and for as long as they remain living in the UK at the time of sampling (Plewis 2007). Four waves have been conducted to date: MCS1 (age 9 months), MCS2 (age 3 years), MCS3 (age 5 years) and MCS4 (age 7 years). The total sample size of MCS1 is 18,818, including 246 sets of twins and ten sets of triplets (*op.cit.*, p.34). The potential longitudinal sample declines over time to the extent that children die or permanently emigrate from the UK. The sample size of MCS2 is 15 590 (Plewis and Ketende 2006). In this paper we use data from the first and second sweeps (Hansen 2008). The units excluded from analyses are: those units where the relationship between the main respondent at sweep one and the cohort member is not natural mother; and twins (2nd cohort member) and triplets (2nd and 3rd cohort members). After merging MCS1 and MCS2 files and doing these exclusions the total number of cases is 14,867 grouped in 398 wards. Missing data in variables such as BAS score and family income are 929 (6 %) and 1,194 (8 %), respectively. For present purposes, missing data are assumed to be completely at random, although some caution needs to be adopted over this assumption.

Child exposure to smoke is a derived variable that represents both mother's smoking during and after pregnancy, and/or whether someone smokes in the same room where the child is. On average women declare a reduction of smoking during pregnancy, mainly over the first months. However, 20 % of pregnant women self-report smoking. After pregnancy, that percentage is 26. At least 11 % of children are passive smokers in the early childhood (0–3 years).

2.2 Sensitivity and specificity

Several studies have shown misclassification of self-reported smoking, possibly because the knowledge about the harmful effects of mothers' smoking during pregnancy has increased and, thus this has became a sensitive item of information. Several studies based on cotinine validated smoking also report such misclassification. Some of them are included in Table 1. Since X=1 means that the child is exposed to smoke, the sensitivity is defined as the probability that a true exposure is correctly classified, $P(Xobs = 1|X = 1)$, while the specificity is the probability that a true non-exposure is correctly classified, $P(Xobs = 0|X = 0)$. Therefore, $P(Xobs = 1|X = 0)$ and $P(Xobs = 0|X = 1)$ are probabilities of misclassification. Differential misclassification occurs when the probability of being misclassified differs across groups of study (Porta 2008). According to Shipton et al. (2009), for example, the probability of misclassification is higher among women who live in least deprived areas compared with women in most deprived areas.

The values of sensitivity reported by studies in Table 1 vary a lot. The variability depends on the context where the study took place, on the setting in which the questions were asked, the methods used for cotinine validated smoking, the chosen cotinine cut-off value, etc. Among these studies there are two that report differential misclassification, depending on the family socioeconomic status (Shipton et al. 2009) or depending on the amount of tobacco smoked (Lindqvist et al. 2002).

2.3 The multilevel model

We consider a two-level variance components model (Goldstein 2011) with children (indexed by i) at level 1 and wards (indexed by j) at level 2. The model we wish to estimate, based on true values, is written as

$$\begin{aligned} y_{ij} &= \beta_{0j} + \beta_1 x_{ij} + \beta_2 z_{ij} + e_{ij} \\ \beta_{0j} &= \beta_0 + u_{0j} \\ u_{0j} &\sim N(0, \sigma_{u0}^2), \quad e_{ij} \sim N(0, \sigma_e^2) \end{aligned} \tag{1}$$

where the response variable y_{ij} is the BAS naming vocabulary, age adjusted and standardised score. The variable x_{ij} represents the exposure to smoke of child *ij* and z_{ij} is family income. Income is a proxy for family socioeconomic status, the most common confounder of maternal cigarette smoking during pregnancy.

The slope β_1 quantifies the difference between cognitive developments of children that are exposed to smoke comparing to those who are not exposed, controlled for family income. This is the main parameter of concern.

The term u_{0j} is the ward level residual whilst the e_{ij} is a child level residual. Level one and two residuals are assumed to follow independent normal distributions. For further details see Goldstein (2011).

2.4 Adjusting for differential misclassification

The model (1) is based on the assumption that we observe true values for (Y, X, Z). In fact we rather observe (Y, X_{obs}, Z) as a result of misclassification. We follow the MCMC algorithm proposed by Goldstein et al. (2008). Regarding the misclassification adjustments, those authors use Eqs. (2) and (3) to calculate the posterior probabilities $P(X = 0|X_{obs} = 0)$ and $P(X = 1|X_{obs} = 1)$,

$$P\left(X = 0|X_{obs} = 0\right) = \frac{L_{00} P\left(X = 0\right)}{L_{00} P\left(X = 0\right) + L_{01} P\left(X = 1\right)} \tag{2}$$

$$P\left(X = 1|X_{obs} = 1\right) = \frac{L_{11} P\left(X = 1\right)}{L_{11} P\left(X = 1\right) + L_{10} P\left(X = 0\right)} \tag{3}$$

where $L_{00,} L_{01,} L_{10,} L_{11}$ are given by

$$L_{00} = exp\left(-\frac{\tilde{y}^2}{2\sigma_\varepsilon^2}\right) P\left(X_{obs} = 0|X = 0\right) \tag{4}$$

$$L_{01} = exp\left(-\frac{\left(\tilde{y} - \beta_1\right)^2}{2\sigma_\varepsilon^2}\right) P\left(X_{obs} = 0|X = 1\right). \tag{5}$$

The prior distribution of X is given by

$$P\left(X = 0\right) = \frac{P\left(X_{obs} = 1|X = 1\right) - P\left(X_{obs} = 1\right)}{P\left(X_{obs} = 0|X = 0\right) + P\left(X_{obs} = 1|X = 1\right) - 1} \tag{6}$$

$$P\left(X = 1\right) = 1 - P\left(X = 0\right) \tag{7}$$

where $P(X_{obs} = 1|X = 1) = 1 - P(X_{obs} = 0|X = 1)$ are assumed known. After sampling the new set of true values X, standard steps of the MCMC algorithm are applied.

3 Results

The results obtained by assuming non-differential misclassification are presented in Table 2. Table 3 includes the results obtained by assuming differential misclassification. The variance component model parameter estimates with no misclassification (column [1]) shows a negative (statistically significant) effect of exposure to smoke on child cognitive development and a positive association between socioeconomic status and the response variable. This remains true for every set of assumptions, but it can be observed that the effect of exposure increases with the severity of misclassification. Comparing estimates presented in column [1] to those presented in column [5] one can see that the estimate more than doubles if we consider no misclassification (column [1]) or consider the strongest P(0|1) value such as reported by Burstyn et al. (2009) (column [5]). The respective standard errors become larger as the probability of misclassification increases. In addition, the estimates presented also show that around 10 % of the variance of cognitive development is due to differences across wards, after controlling for exposure and family income.

The differential misclassification probabilities used in the models are presented in Table 3, and are those probabilities reported by Lindqvist et al. (2002) and Shipton et al. (2009).

The estimates in column [3] were obtained from models fitted with the adjustment of differential misclassification according to following: "Of 407 women, reporting to be non-smokers, 6 % were most likely smokers…Of 60 women, reporting smoking 1–10 cigarettes per day, 32 % were likely to smoke more.". Column [7] presents a more extreme case of misclassification when the number of cigarettes per day is more than ten. Only 6 % of pregnant women in the MCS sample declare smoking more than ten cigarettes per day. Finally, results in column [8] were obtained using a similar assumption according to the evidence that "a greater proportion of smokers in the least deprived areas did not report their smoking (39 %)

Table 2 Non-differential misclassification adjustment

	Estimate (SE) [1] P(0|1)=0 P(1|0)=0	Estimate (SE) [2] P(0|1)=0.14 P(1|0)=0.05	Estimate (SE) [3] P(0|1)=0.22 P(1|0)=0.05	Estimate (SE) [4] P(0|1)=0.32 P(1|0)=0.05	Estimate (SE) [5] P(0|1)=0.53 P(1|0)=0.05
Constant	0.092	0.098	0.102	0.111	0.165
	(0.019)	(0.019)	(0.020)	(0.021)	(0.036)
Exposure to smoke	−0.060	−0.072	−0.075	−0.085	−0.130
	(0.018)	(0.022)	(0.023)	(0.025)	(0.042)
Income	0.199	0.199	0.199	0.199	0.201
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
Level 2 variance	0.089	0.090	0.090	0.090	0.089
	(0.009)	(0.008)	(0.008)	(0.008)	(0.008)
Level 1 variance	0.792	0.792	0.792	0.791	0.790
	(0.010)	(0.010)	(0.010)	(0.010)	(0.010)

Table 3 Differential misclassification adjustment

	Estimate (SE) [6] Number of cigarettes [1,10]: P(0|1)=0.32 > 10: P(0|1)=0.06 P(1|0)=0.05	Estimate (SE) [7] Number of cigarettes [1,10]: P(0|1)=0.32 > 10: P(0|1)=0.53 P(1|0) =0.05	Estimate (SE) [8] Family Income Deprived: P(0|1)=0.22 Normal: P(0|1)=0.26 Affluent: P(0|1)=0.39 P(1|0)=0.05
Constant	0.111	0.108	0.107
	(0.021)	(0.021)	(0.021)
Exposure to smoke	−0.087	−0.097	−0.081
	(0.025)	(0.025)	(0.024)
Income	0.199	0.199	0.203
	(0.009)	(0.009)	(0.009)
Level 2 variance	0.089	0.089	0.089
	(0.008)	(0.008)	(0.008)
Level 1 variance	0.791	0.791	0.790
	(0.010)	(0.010)	(0.010)

compared with women in the most deprived areas (22 %)…”. In our study we considered as most deprived families those whose income is less than £10,400 per year (17.6 %) and less deprived families those which income is more than £31,000 (24.7 %). It can be observed that estimates change, but not much when compared with those presented in Table 2. This is explained by the low proportion of cases in the sample with high misclassification probabilities.

4 Discussion

In this paper, we present an extension of the MCMC algorithm proposed by Goldstein et al. (2008) to address the issue of differential misclassification in multilevel models with continuous response variable. The relationship between child exposure to smoke (pre-natal and in the early childhood) and cognitive development is analyzed assuming that reliance on self-reported smoking status tends to result in misclassification of child exposure. Some authors report the differential misclassification depending on the amount of cigarettes pregnant women smoke per day, and also depending on the family socioeconomic status. Based on validated data, those studies report the sensitivity and specificity of self-reported attitudes towards smoking. Thus, assuming that the probabilities of misclassification are known, the procedure was applied under various different assumptions. The variance component models fitted suggest that the magnitude of the negative effect of child exposure on cognitive development more than doubles in the worst situation of misclassification. In fact, as the probability that a reported non-smoker is in fact a smoker increases, so does the estimates of the effect of smoking after adjustment for this misclassification probability. In addition, the impact of differential misclassification in the fixed parameter estimates depends not only on the severity of misclassification but also on the marginal distribution of the variable that differentiates the probability of misclassification. It was also shown that standard errors tend to become larger. This suggests that data analysts should pay particular attention to such misclassifications and incorporate them in their analyses.

Further work is planned in order to correct for income measurement error. Finally, we have relied upon published values for misclassification probabilities, but these are difficult to obtain, and hence the sensitivity analyses we have carried out should be treated with caution. Additional research devoted to obtaining good estimates of misclassification probabilities would be useful.

Acknowledgements We wish to acknowledge support from the Foundation for Science and Technology, Portugal, and the Economic and Social Research Council, UK. The data set was supplied by the UK Data Archive.

References

Burstyn, I., Kapur, N., Shalapay, C., Bamforth, F., Wild, T.C., Liu, J., LeGatt, D.: Evaluation of the accuracy of self-reported smoking in pregnancy when the biomarker level in an active smoker is uncertain. Nicotine Tob. Res. **11**(6), 670–678 (2009). Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19395685

Butler, N.R., Goldstein, H., Ross, E.M.: Cigarette smoking in pregnancy: its influence on birth weight and perinatal mortality. BMJ **2**, 127–130 (1972)

Carroll, R.J., Ruppert, D., Stefanski , L.A., Crainiceanu, C.: Measurement Error in Nonlinear Models: A Modern Perspective. pp. 488 Chapman & Hall/CRC, London (2006)

Davie, R., Butler, N.R., Goldstein, H.: From Birth to Seven. Longmans, London (1972)

England, L.J., Grauman, A., Qian, C., Wilkins, D.G., Schisterman, E.F., Yu, K.F., Levine, R.J.: Misclassification of maternal smoking status and its effects on an epidemiologic study of pregnancy outcomes. Nicotine Tob. Res. **9**(10), 1005–1013 (2007). Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17852766

Ford, R.P., Tappin, D.M., Schluter, P.J., Wild, C.J.: Smoking during pregnancy: how reliable are maternal self reports in New Zealand?. J. Epidemiol. Community Health **51**(3), 246–251 (1997)

Fuller, W.A.: Error Models. pp. 440 Wiley, New York (2006)

Goldstein, H.: Factors influencing the height of seven year old children. Results from the national child development study. Hum Biol **43**, 92–111 (1971)

Goldstein, H.: Multilevel Statistical Models, 4th edn. Wiley, Chichester (2011)

Goldstein, H., Kounali, D., Robinson, A.: Modelling measurement errors and category misclassifications in multilevel models. Stat. Model. **8**(3), 243–261 (2008). SAGE Publications. doi:10.1177/1471082X0800800302

Hansen, K.: Millennium Cohort Study: First, Second and Third Surveys. A Guide to the Datasets, 4th edn. Centre for Longitudinal Studies, London (2008)

Institute of Education/University of London. Millennium Cohort Study: First Survey, 2001–2003, 8th edn. Colchester, Essex. (2009)

Julvez, J., Ribas-Fitó, N., Torrent, M., Forns, M., Garcia-Esteban, R., Sunyer, J.: Maternal smoking habits and cognitive development of children at age 4 years in a population-based birth cohort. Int. J. Epidemiol. **36**(4), 825–832 (2007). Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17550944

Kafouri, S., Leonard, G., Perron, M., Richer, L., Séguin, J.R., Veillette, S., Pausova, Z., et al.: Maternal cigarette smoking during pregnancy and cognitive performance in adolescence. Int. J. Epidemiol. **38**(1), 158–172 (2009). Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19039007

Klebanoff, M.A., Levine, R.J., Morris, C.D., Hauth, J.C., Sibai, B.M., Catalano, P., Wilkins, D.G.: Accuracy of self-reported cigarette smoking among pregnant women in the 1990s. Paediatr. Perinatal Epidemiol. **15**(2), 140–143 (2001). doi:10.1046/j.1365-3016.2001.00321.x

Kramer, M.S.: Determinants of low birth weight: methodological assessment and meta-analysis. Bull. World Health Organ **65**(5), 663–737 (1987)

Lindqvist, R., Lendahls, L., Tollbom, O., Aberg, H., Håkansson, A.: Smoking during pregnancy: comparison of self-reports and cotinine levels in 496 women. Acta Obstet Gynecol Scand **81**(3), 240–244 (2002)

Plewis, I.: Millennium Cohort Study: Technical Report on Sampling, 4th edn. Centre for Longitudinal Studies, Institute of Education, London (2007)

Plewis, I., Ketende, S.: Millennium Cohort Study: Technical Report on Response, 1st edn. Centre for Longitudinal Studies, University of London, London (2006)

Porta, M.: A Dictionary of Epidemiology, 5th edn. Oxford University Press, Oxford (2008)

Pärna, K., Rahu, M., Youngman, L.D., Rahu, K., Nygård-Kibur, M., Koupil, I.: Self-reported and serum cotinine-validated smoking in pregnant women in Estonia. Matern. Child Health J. **9**(4), 385–392 (2005)

Shipton, D., Tappin, D.M., Vadiveloo, T., Crossley, J.A., Aitken, D.A., Chalmers, J.: Reliability of self reported smoking status by pregnant women for estimating smoking prevalence: a retrospective, cross sectional study. Br. Med. J. **339**, b4347 (2009). BMJ Publishing Group Ltd. Retrieved from http://eprints.gla.ac.uk/18454/

Simpson, W.J.: A preliminary report on cigarette smoking and the incidence of prematurity. Am. J. Obstet. Gynecol. **73**, 808–815 (1957)

Walsh, R.A.: Effects of maternal smoking on adverse pregnancy outcomes: examination of the criteria of causation. Hum. Biol. **66**(6), 1059–1092 (1994)