Published by Oxford University Press on behalf of the International Epidemiological Association
International Journal of Epidemiology 2014;43:264–269

Advance Access publication 23 December 2013
doi:10.1093/ije/dyt221

EDUCATION CORNER

# Fixed effects analysis of repeated measures data

**Fiona Imlach Gunasekara,* Ken Richardson, Kristie Carter and Tony Blakely**

Health Inequalities Research Programme, Department of Public Health, University of Otago, Wellington, New Zealand

*Corresponding author. Department of Public Health, University of Otago, PO Box 7343, Wellington 6242, New Zealand.
E-mail: fiona.imlachgunasekara@plunket.org.nz

**Accepted** 25 September 2013

The analysis of repeated measures or panel data allows control of some of the biases which plague other observational studies, particularly unmeasured confounding. When this bias is suspected, and the research question is: ‘Does a change in an exposure cause a change in the outcome?’, a fixed effects approach can reduce the impact of confounding by time-invariant factors, such as the unmeasured characteristics of individuals. Epidemiologists familiar with using mixed models may initially presume that specifying a random effect (intercept) for every individual in the study is an appropriate method. However, this method uses information from both the within-individual/unit exposure-outcome association and the between-individual/unit exposure-outcome association. Variation between individuals may introduce confounding bias into mixed model estimates, if unmeasured time-invariant factors are associated with both the exposure and the outcome. Fixed effects estimators rely only on variation within individuals and hence are not affected by confounding from unmeasured time-invariant factors. The reduction in bias using a fixed effects model may come at the expense of precision, particularly if there is little change in exposures over time. Neither fixed effects nor mixed models control for unmeasured time-varying confounding or reverse causation.

## The problem with observational studies

Epidemiology is concerned with discovering and understanding the causal relationships between exposures and health outcomes.[1] A lack of causal evidence leads to denial and inaction around harmful exposures (such as tobacco, until sufficient evidence was produced to demonstrate its link to lung cancer and other diseases)[2] or inappropriate action (e.g. recommending postmenopausal women to take hormone replacement therapy to reduce the risk of cardiovascular disease).[3,4] To avoid these pitfalls, epidemiologists seek causal estimates of exposure-outcome associations that are not affected by bias, particularly bias from confounding, selection and measurement error.[5] Observational studies come under particularly severe scrutiny as, unlike ‘gold-standard’ randomized controlled trials (RCTs), participants are not assigned to an exposure or intervention by chance. Therefore, the exposed and non-exposed groups are likely to be different in important ways, due to self-selection and other processes, and this may bias estimation of the exposure-outcome relationship (e.g. by introducing confounding).[1] Some of this difference may be due to observable and measurable factors (measured confounders) and some to unobserved or unknown factors (unmeasured confounders).[6]

Despite the limitations of observational studies, they are widely used in epidemiology where RCTs are not appropriate, ethical, feasible or where they provide evidence only for a highly selected group of individuals which is not widely generalizable to heterogeneous populations.[7–9] Natural experiments, where the 'treatment' (e.g. a lottery win, policy or law change) occurs in a random and often unexpected manner, may provide causal evidence. However, these require that appropriate data is collected during the experiment, including suitable, valid exogenous variables (which are not determined by any other variables of interest)[6] that can be used in instrumental variable analyses. Thus true natural experiments are rare and many give results that are not widely generalizable.[10,11] Longitudinal or panel surveys which gather repeated measures on the same individuals over time are the best observational studies to limit the effects of bias and improve causal estimation, while remaining representative of whole populations.[12–14] In particular, applying fixed effects methods developed in the econometric and multilevel literature can reduce the impact of some types of confounding, in appropriate circumstances.[13,15,16] The central idea of these methods, applied to longitudinal data, is that changes over time comprise within-individual (variation in each individual's own exposures or outcomes) and between-individual (variation across individuals) components. Modelling the within-individual component opens the door to removing all time-invariant confounding, as each individual acts as their own control.[13]

Consider the causal question of whether changes in income lead to changes in health, which may be confounded by measured time-invariant confounders such as sex, ethnicity and education, and unmeasured time-invariant confounders such as intelligence, ability, beliefs and social upbringing. Although it is well established that people with lower socioeconomic position have worse health outcomes, the recommendation that often follows from this observation, that interventions to raise socioeconomic position will improve health, are based on limited evidence.[5,17] All estimators of exposure-outcome associations have to deal explicitly with the fact that these associations are potentially biased by individual, and often unknown, characteristics (between-individual differences), that include time-invariant individual-level confounders. However, within-individual changes in income and health can be used to give an estimate of the exposure (income)–outcome(health) association that is not affected by time-invariant confounding bias—provided the estimator used is unbiased and consistent. One simple regression-based approach that has these properties under certain assumptions, and is well established in the econometric literature, has become known as the 'fixed effects' model. More properly this is an 'unobserved effects model', and to avoid later confusion we note that the term 'fixed effects' is synonymous with regression estimators that control for some types of unmeasured confounders. This usage is quite distinct from that of the statistical literature, where it is often met in the context of random effects or mixed models. In statistical jargon, a fixed effect is a parameter associated with an entire population (to be estimated) and a random effect is a parameter describing the variability of experimental units (e.g. individuals) drawn randomly from the population.[18] This distinction is irrelevant for unobserved ('fixed') effects models, since estimation is unbiased and consistent regardless of whether 'effects' are considered fixed or random provided model assumptions are satisfied (see later).

In what follows we focus on methods for estimating effects of changes over time at an individual level. Where estimates averaged over the population in question are sought, population average models (e.g. generalized estimating equations or GEE) can be useful.[19]

## Fixed effects methods to control for confounding

Kaufman 2008 notes that the econometric fixed effects estimate, which relies solely on within-individual changes, 'eliminates confounding by all [of these] innumerable and unmeasurable influences. This is the really remarkable promise of the fixed effects model, and one that makes it so attractive for social epidemiology, where exposures are often heavily confounded by myriad contextual, behavioural and attitudinal quantities that would be difficult to assess exhaustively.'[16] In this statement, Kaufman highlights the value of the fixed effects model in controlling for time-invariant unmeasured confounding, although glossing over its inability to control for other important biases, such as reverse causation and time-varying unmeasured confounding.

To demonstrate how a fixed effects model controls for time-invariant confounding when applied to longitudinal data, consider a causal linear model where outcome $y_{it}$ for the $i$th of N individuals measured at time $t$ is predicted by time-varying ($x_{it}$) and time-invariant ($Z_i$) exposures. Defining $\varepsilon_{it}$ as the random error term (representing 'disturbances' to the outcome, assumed to be homoscedastic, i.e. have constant variance across time), $\beta_{0t}$ as the intercept and $\alpha_i$ as a time-invariant covariate representing unmeasured time-invariant confounding, then the linear causal model is (Equation 1):

$$y_{it} = \beta_{0t} + \beta_1 x_{it} + \beta_2 Z_i + \alpha_i + \varepsilon_{it}. \qquad (1)$$

For estimation, covariates are random variables (random samples from a population) that are not required to be independent of each other or of $\alpha_i$, so they can genuinely represent confounders which influence both the exposure and outcome. However, there are some restrictive assumptions around their relationship with the disturbances $\varepsilon_{it}$ (see later under Limitations). One approach to estimation in the linear regression context uses the fixed effects 'dummy' method, treating $\alpha_i$ as N unknown parameters,[20] equivalent to a fixed intercept for each individual.[13] But calculating the model in this way is prohibitive for large samples (with many $\alpha_i$ to estimate) and the dummy variables are typically 'nuisance' variables of no inherent interest.

An alternative and computationally less demanding way to calculate the linear fixed effects model is the mean-centring approach. In this case the mean (over time) of measurements for each individual is subtracted from all the individual's measurements. The time-invariant terms (which are not independently identifiable) are eliminated in the mean-centring Equation (2), and only parameters associated with time-varying covariates can be estimated by the model.[19] The model becomes:

$$y'_{it} = \beta_{0t} + \beta_1 x'_{it} + \varepsilon'_{it}, \tag{2}$$

where $y'_{it} = y_{it} - \bar{y}_i$, $\bar{y}_i$ is the mean of $y_{it}$ over time (and similarly for $x'_{it}$ and $\varepsilon'_{it}$) for each individual. Using either the dummy variable or the mean-centring approach achieves the objective of removing terms representing measured and unmeasured time-invariant confounding ($\beta_2 Z_i$ and $\alpha_i$) from the model.

Both approaches can be estimated using standard linear regression software, but the mean-centring approach requires a 'degrees-of-freedom' correction to standard errors for time-varying parameters. In fact many software packages (e.g. SAS, STATA) do this automatically, and also provide a variety of adjustments to standard errors for heteroscedasticity and serial correlation to improve inference. The linear fixed effects model has found wide application in the econometrics literature, and we have used it here to illustrate key concepts. However, in epidemiology, outcomes are often categorical, and non-linear models assume greater importance. In general non-linear fixed effects models are more challenging but, for several non-linear models important to epidemiologists, relatively straightforward methods are available. These include conditioning the parameter representing time-invariant confounding out of the likelihood (logistic models) or explicitly modelling within-individual changes in a multilevel group-mean-centred mixed model (ordinal models).[13,21–23] Fixed effects models are therefore available for count and categorical (including binary) outcomes, although the consistency and bias of these estimators are more sensitive to departures from assumptions than is the case for linear fixed effects models. In particular, if heteroscedasticity is suspected, parameter estimates may be biased, and providing robust standard errors (as in the linear case) for a biased estimate makes little sense.[15,24]

## Comparison with mixed models

Mixed models are familiar to epidemiologists for dealing with hierarchical or grouped data, particularly in the context of research on neighbourhoods.[25–27] Mixed models are often loosely referred to as 'random effects' models and include both fixed (in the statistical sense—see above) and random effects. In a simple longitudinal random intercept mixed model, $\alpha_i$ in Equation (1) is assumed to be a random variable (one per individual $i$) with its own probability distribution, typically normal with zero mean. The key distinction between mixed and (econometric) fixed effects models is whether $\alpha_i$ is assumed to be a confounder, i.e. correlated with other covariates in the model.[15] In the simple random intercept mixed model, $\alpha_i$ is assumed to be independent of other covariates in the model and if this assumption is violated, as when $\alpha_i$ represents unmeasured confounding, the random/mixed effects estimator is biased and inconsistent, and confounding bias will not have been removed.[19,28]

Estimates from simple random intercept mixed models combine variation both from within-individuals and between-individuals.[15,18,20,29] However, in many longitudinal data analyses the 'between-individual' differences are likely to include potential confounders—not only observed variables (such as education, age, ethnicity, labour force status and wealth, of importance for the causal question in this example) but also unobserved variables (such as intelligence or genetic variability). If such unobserved variables are important confounders, mixed model estimates will not remove the significant bias introduced by those confounders.

Using our previous example, income is likely to be correlated with time-invariant unmeasured confounders represented by $\alpha_i$ and hence the application of a mixed model to the research question of whether changes in income cause changes in health is unhelpful, since the estimator is biased and inconsistent under these conditions. This applies particularly where the number of individuals (N) is large but the number of data collection points (T) relatively small, as in most longitudinal data analyses. Where T is also large, the mixed model estimate will be dominated by within-individual variation and the difference between estimates from fixed effects and simple random intercept mixed models is reduced.[15]

Figure 1 heuristically demonstrates this via results from a simulation, highlighting two units ('individuals' A and B) with observations on continuous exposure x and continuous outcome y at five time points (solid dot points), with fixed effects (FE—solid lines) and random effects (RE—dotted lines) slopes fitted for each unit. The cloud of data points represent the full simulated longitudinal dataset to which the pooled model is fitted. The pooled (P—dashed line) estimator does not model individual-specific effects (to account for unmeasured confounding) or account for serial correlation (where random disturbances across time for the same individual are correlated). The simulation parameters were chosen to make the difference between the RE and FE estimators clear, with a large N (1000 units), small number of time points (T = 5) and substantial

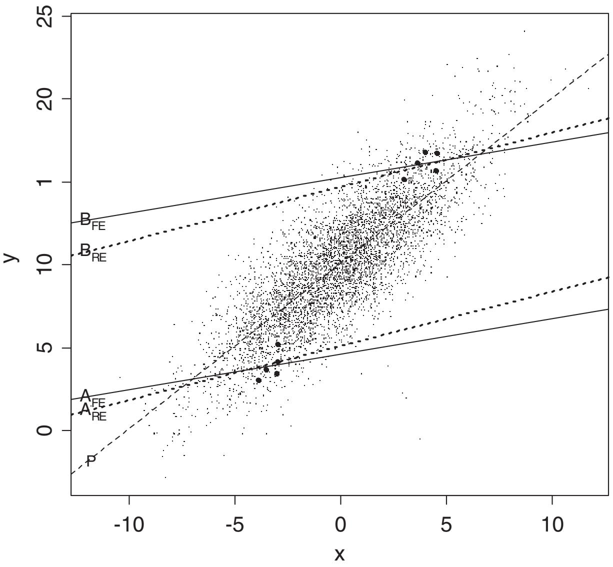

**Figure 1** Results of simulation showing the difference between random effects(RE, dotted line) and fixed effects (FE, stepped line) models using five observations each on two individuals: A (dots) and B (bold dots) and relationship to estimator from a pooled model (P, dashed line) using data from all simulated points (cloud of small dots)

unmeasured confounding (correlation between 'unmeasured' confounder and exposure was 0.8). All of these factors increase the error in the RE estimator.[15] The Appendix contains more detailed information about the simulation (available as Supplementary data at *IJE* online).

In this simulation, the mixed random intercept regression model (RE) correctly treats longitudinal data as grouped (at the individual level), but does not control for unmeasured confounders and therefore does not provide a reliable estimate of causal interest (the average of the within-individual slopes, given by FE). Technically, mixed model random intercepts are 'shrunk' towards the overall intercept from the pooled model so that the slope estimate from the mixed model lies between the fixed effects and pooled estimates.[18]

One way to formally test whether the orthogonality assumption (no unmeasured time-invariant confounding) required by the linear random intercept mixed model estimator holds is to use the Hausman test statistic.[30] If the null hypothesis (no statistically significant difference between the fixed and random effects estimates) is rejected, this is interpreted as evidence against the orthogonality assumption and the linear fixed effects model is preferred.

## Limitations of fixed effects models

The main advantage of the fixed effects model is that it only uses within-individual variation, but this can lead to lack of precision (mixed models are potentially more efficient, with narrower confidence intervals). Another major disadvantage is that parameters for time-invariant variables, such as sex and ethnicity, are not estimated (since they do not change in individuals over time). However, time-invariant covariates may be interacted with time-varying exposures of interest, e.g. to investigate whether the effect of income on health varies by sex, poverty status or educational attainment.[13] Similarly, fixed effects models are not useful for investigating the exposure-outcome association in respondents who do not change their exposure levels (e.g. the effect of persistent low income on health), or appear in just one time period, because only observations where the exposure varies contribute to the fixed effects estimate (see Appendix: Figure A1, available as Supplementary data at *IJE* online).

Basic fixed effects models work under the assumption of strict exogeneity, which prohibits some types of feedback from past outcomes to current covariates and current outcome to future covariates. Under this assumption, having controlled for a given set of (possibly lagged) covariates at each time point, past values of the selected covariates cannot independently modify the current outcome and past outcomes cannot independently modify future values of those covariates. However, this assumption may be problematic in some situations of interest to epidemiologist. Figure 2 presents a simple directed acyclic graph of one exposure and outcome over two time periods, highlighting departures from the strict exogeneity assumption (in dashed lines). These include unobserved time-varying confounding (Figure 2, pathway B, e.g. where unmeasured shifts in societal attitude to a health issue affect the future reporting of the health outcome), reverse causation (Figure 2, pathway A, e.g. where health status impacts on income level, as well as income affecting health) and the presence of measurement error (not shown in Figure 2).[15] Dynamic fixed effects (single equation) models have been developed with relaxed exogeneity assumptions that allow the inclusion of a time-lagged outcome (also known as state dependence, Figure 2, pathway C).[15] The initial conditions problem, which may be important in dynamic fixed effects models, is a particular example of unmeasured time-varying confounding because the first observation of a study is rarely the 'true' initial state[31] and covariates prior to the initial time period may affect the current exposure and outcomes. Accounting for this bias by including (endogenous) observed initial outcomes requires additional assumptions about the static nature of processes occurring before the study genesis, which may not be sound. Similar remarks apply to conditioning unobserved initial outcomes out of the model likelihood.[15]

More complex effects such as reverse causation require multiple equation methods: cross-lagged fixed effects structural equation models (SEMs) have been used in this context,[13,32] but have significant limitations in the presence of time-dependent confounding.[33–35] In circumstances where there are complex dynamics of

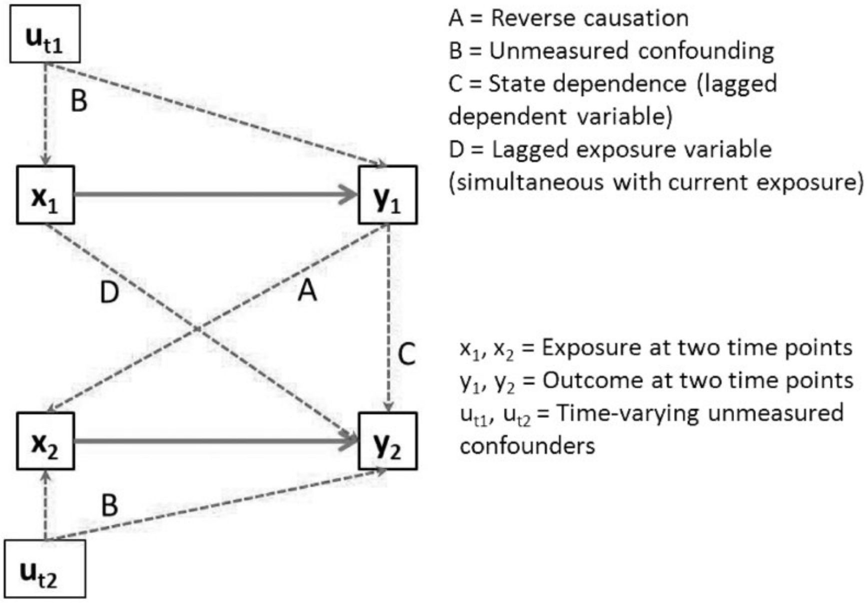
$u_{t1}$
B
$x_1$
$y_1$
D
A
C
$x_2$
$y_2$
B
$u_{t2}$
A = Reverse causation
B = Unmeasured confounding
C = State dependence (lagged dependent variable)
D = Lagged exposure variable (simultaneous with current exposure)
$x_1$, $x_2$ = Exposure at two time points
$y_1$, $y_2$ = Outcome at two time points
$u_{t1}$, $u_{t2}$ = Time-varying unmeasured confounders

**Figure 2** Directed acyclic graph (DAG) depicting potential correlations (in dashed lines) between repeated measures of an exposure ($x_1$ and $x_2$) and outcome ($y_1$ and $y_2$) that violate the assumption of strict exogeneity

evolving exposures and outcomes, standard regression models (including simple random intercept mixed models and econometric fixed effects models) are limited and more advanced causal models (e.g. g-method causal estimators) may be required to provide unbiased results.[34,36,37] Even then, validation studies or sensitivity analyses may be needed to test for bias from unmeasured confounding.[38]

## Conclusion

When analysing longitudinal survey data, econometric fixed effects models provide a method for assessing exposure-outcome associations adjusted for all time-invariant confounding and measured time-varying confounding. These models are useful for assessing exposure-outcome associations when there is a large number of respondents, low dropout and regular, detailed data collection over time (typical of longitudinal data), when the strict exogeneity conditions are defensible, and where exposures change over time for at least some respondents. For analyses of persons within clusters (e.g. neighbourhoods), where the number of units can be small but the number of observations in each unit is large, simple random intercept mixed models are often preferred to fixed effects models as they can be more efficient and have greater flexibility in dealing with individual observations at multiple levels.[39] Fixed effects models are a useful and easily applied exploratory tool where time-invariant confounding is likely to cause significant bias in causal estimates. More complex models may be required when reverse causation, state dependence or unmeasured time-varying confounding are likely to violate the fixed effects model assumptions.

## Supplementary Data

Supplementary data are available at *IJE* online.

## Funding

This work was done within the Health Inequalities Research Programme, which is primarily funded by the Health Research Council of New Zealand.

**Conflict of interest:** None declared.

**KEY MESSAGES**

- Fixed effects models are a useful exploratory tool when applied to longitudinal data because they control for all time-invariant confounding, both measured and unmeasured, by using only the changes in exposure occurring within individuals to estimate the outcome.
- The major limitation of fixed effects models is the inability to control for other biases which may be important in longitudinal data analysis, including reverse causation and time-varying confounding.

## References

[1] Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd edn. Philadelphia, PA: Wolters Kluwer, 2008.

[2] Terry L, Woodruff S. *Smoking and Health: Report of the Advisory Committee to the Surgeon General of the United States*. Washington, DC: U-23 Department of Health, Education, and Welfare, 1964.

[3] Bath PMW, Gray LJ. Association between hormone replacement therapy and subsequent stroke: a meta-analysis. *BMJ* 2005;**330:**342.

[4] Prentice RL, Langer R, Stefanick ML *et al*. Combined Postmenopausal Hormone Therapy and Cardiovascular Disease: Toward Resolving the Discrepancy between Observational Studies and the Women's Health Initiative Clinical Trial. *Am J Epidemiol* 2005;**162:**404–14.

[5] Harper S, Strumpf EC. Commentary: Social Epidemiology: Questionable Answers and Answerable Questions. *Epidemiology* 2012;**23:**795–98.

[6] Imlach Gunasekara F, Carter K, Blakely T. Glossary for econometrics and epidemiology. *JEpidemiol Community Health* 2008;**62:** 858–61.

[7] Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 2000; **342:**1878–86.

[8] Vandenbroucke JP, von Elm E, Altman DG *et al*. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. *PLoS Med* 2007;**4:**e297.

[9] Kaufman JS, Kaufman S, Poole C. Causal inference from randomized trials in social epidemiology. *Soc Sci Med* 2003; **57:**2397–409.

[10] Moffit R. Remarks on the analysis of causal relationships in population research. *Demography* 2005;**42:**91–108.

[11] Glymour MM. Natural experiments and instrumental variable analyses in social epidemiology. In: Oakes JM, Kaufman JS (eds). *Methods in Social Epidemiology*. San Francisco, CA: Jossey-Bass, 2006.

[12] Singer JD, Willett JB. *Applied Longitudinal Data Analysis*. Oxford, UK: Oxford University Press, 2003.

[13] Allison PD. *Fixed Effects Regression Analysis for Longitudinal Data Using SAS*. Cary, NC: SAS Institute, 2005.

[14] Wagner GG, Frick JR, Schupp Jr. *The German Socio-Economic Panel Study (SOEP) – Scope, Evolution and Enhancements. SOEPpapers on Multidisciplinary Panel Data Research at DIW Berlin*. Berlin: DIW Berlin, 2007.

[15] Wooldridge JM. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press, 2002.

[16] Kaufman JS. Commentary: Why are we biased against bias? *Int J Epidemiol* 2008;**37:**624–26.

[17] Ludbrook A, Porter K. Do interventions to increase income improve the health of the poor in developed economies and are such policies cost effective? *Appl Health Econ Health Policy* 2004;**3:**115–20.

[18] Pinheiro J, Bates D. *Mixed-effects Models in S and S-PLUS*. New York: Springer, 2009.

[19] Gardiner JC, Luo Z, Roman LA. Fixed effects, random effects and GEE: What are the differences? *Stat Med* 2009;**28:**221–39.

[20] Verbeek M. *A Guide to Modern Econometrics*. 3rd edn. Chichester, UK: John Wiley, 2008.

[21] Mukherjee B, Ahn J, Liu I, Rathouz PJ, Sánchez BN. Fitting stratified proportional odds models by amalgamating conditional likelihoods. *Stat Med* 2008;**27:**4950–71.

[22] Allison PD. *Logistic Regression Using SAS*. Cary, NC: SAS Institute, 2008.

[23] Lancaster T. The incidental parameter problem since 1948. *J Econometrics* 2000;**95:**391–413.

[24] Greene WH. *Econometric Analysis*. Upper Saddle River, NJ: Prentice Hall, 2012.

[25] Diez Roux AV. Multilevel analysis in public health research. *Annu Rev Public Health* 2000;**21:**171–92.

[26] Merlo J, Chaix B, Yang M, Lynch J, Rastam L. A brief conceptual tutorial on multilevel analysis in social epidemiology: interpreting neighbourhood differences and the effect of neighbourhood characteristics on individual health. *J Epidemiol Community Health* 2005;**59:** 1022–29.

[27] Blakely T, Subramanian SV. Multilevel studies. In: Oakes J, Kaufman JS (eds). *Methods in Social Epidemiology*. San Francisco, CA: Jossey Bass, 2006.

[28] Heagerty PJ, Kurland BF. Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika* 2001;**88:**973–85.

[29] Trivedi P, Cameron A. Linear panel models: basics. *Microeconomics: Methods and Applications*. Cambridge, UK: Cambridge University Press, 2005.

[30] Hausman JA, Taylor WE. Panel data and unobservable individual effects. *Econometrica* 1981;**49:**1377–98.

[31] Baltagi BH. *Econometric Analysis of Panel Data*. 3rd edn. Chicester, UK: John Wiley, 2005.

[32] Allison PD. Causal inference with panel data. Paper presented at the Annual Meeting of the American Sociol ogy Association. Philadephia, PA: ASA, 2005.

[33] Pearl J. *Causality: Models, Reasoning, andIinference*. 2nd edn. New York: Cambridge University Press, 2009.

[34] Robins J, Hernán M. Estimation of the causal effects of time-varying exposures. In: Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G (eds). *Longitudinal Data Analysis*. Boca Raton, FL: CR Press Taylor and Francis Group, 2009.

[35] VanderWeele TJ. Invited commentary: structural equation models and epidemiologic analysis. *Am J Epidemiol* 2012; **176:**608–12.

[36] Greenland S, Robins J. Identifiability, exchangeability and confounding revisited. *Epidemiol Perspect Innov* 2009; **6:**4.

[37] Pearl J. Statistics and Causal Inference: A Review. *Sociedad de Estad'ıstica e Investigaci'on Operativa* 2003;**12:** 101–65.

[38] VanderWeele TJ, Onyebuchi A. Bias formulas for sensitivity analysis of unmeasured confounding for general outcomes, treatments and confounders. *Epidemiology* 2011;**22:**42–52.

[39] Moerbeek M, van Breukelen GJP, Berger MPF. A comparison between traditional methods and multilevel regression for the analysis of multicenter intervention studies. *J Clin Epidemiol* 2003;**56:**341–50.

[40] Agresti A. *Categorical Data Analysis*. Hoboken, NJ: John Wiley, 2002.

[41] Press WH, Teukolsky SA, Vetterling WT, Flannerly BP. *Numerical Recipes: The Art of Scientific Computing*. 3rd edn. Cambridge, UK: Cambridge University Press, 2007.

[42] Petersen T. Analyzing panel data: Fixed-and random-effects models. In: Hardy MA, Bryman A (eds). *Handbook of Data Analysis*. London: Sage, 2004.