Ecological effects in multi-level studies

Tony A Blakely, Alistair J Woodward

Abstract
Multi-level research that attempts to describe ecological effects in themselves (for example, the effect on individual health from living in deprived communities), while also including individual level effects (for example, the effect of personal socioeconomic disadvantage), is now prominent in research on the socioeconomic determinants of health and disease. Such research often involves the application of advanced statistical multi-level methods. It is hypothesised that such research is at risk of reaching beyond an epidemiological understanding of what constitutes an ecological effect, and what sources of error may be influencing any observed ecological effect. This paper aims to present such an epidemiological understanding. Three basic types of ecological effect are described: a direct cross level effect (for example, living in a deprived community directly affects individual personal health), cross level effect modification (for example, living in a deprived community modifies the effect of individual socioeconomic status on individual health), and an indirect cross level effect (for example, living in a deprived community increases the risk of smoking, which in turn affects individual health). Sources of error and weaknesses in study design that may affect estimates of ecological effects include: a lack of variation in the ecological exposure (and health outcome) in the available data; not allowing for intraclass correlation; selection bias; confounding at both the ecological and individual level; misclassification of variables; misclassification of units of analysis and assignment of individuals to those units; model mis-specification; and multicollinearity. Identification of ecological effects requires the minimisation of these sources of error, and a study design that captures sufficient variation in the ecological exposure of interest.
(*J Epidemiol Community Health* 2000;**54**:367–374)

Department of Public Health, Wellington School of Medicine, University of Otago, PO Box 7343, Wellington, New Zealand

Correspondence to:
Dr Blakely (e-mail:
tblakely@wnmeds.ac.nz)

The aim of this paper is to discuss epidemiological issues in the investigation of ecological determinants of health. There has been a resurgence of interest in ecological research, premised on the assumption that social contexts may shape health status as much as traditional individual risk factors.[1–7] In particular, many researchers of the socioeconomic determinants of health are beginning to analyse ecological and individual level exposures simultaneously, often using advanced multi-level statistical methods.[8–11] We are concerned that the application of multi-level statistical methods may have surged ahead of a theoretical framework in which to conduct meaningful and robust analyses. In this paper we describe briefly the historical background to multi-level analysis, review possible categorisations of ecological variables, and present three basic types of ecological effect. Then we explore the possible sources of error in ascribing ecological effects in multi-level analysis that we believe are currently being overlooked in multi-level analyses, and some general strategies to overcome these sources of error.

Background

Research that combines the ecological and individual level has a long history in sociology. Durkheim is credited with the first such attempt when he investigated suicide.[12] During and after the second world war interest in the United States increased,[13] with an ensuing debate about the validity of ecological effects.[14–18] That debate was about whether the effect of an ecological exposure on health is causally valid, independent of explanatory and intervening individual level causes. For example, is it valid to consider ascribing causation to the effect of the ecological exposure "living in a deprived community" on the outcome "individual health status"? Or should we always seek to reduce such observed associations to individual level causal mechanisms like individual socioeconomic status, smoking, and other risk factors? In public health it is accepted that disease causation operates via chains, or webs, of events,[2] and most public health practitioners are comfortable with the notion of proximal and distal causes. For example, we have no difficulty attributing cases of whooping cough to both exposure to the bacteria B pertussis (a proximal cause), and the loss of herd immunity (a distal or population level cause). Indeed, disregarding distal causation may overlook important causal mechanisms; immunisation against pertussis will have less apparent benefit if individual level protection (vaccine efficacy) only is considered, compared with also considering the impact of community level immunisation on the background incidence of whooping cough.

Susser has proposed that links should be made between possible levels of analysis,[7] and uses the analogy of chinese boxes.[6] It is possible to posit an infinite number of levels of organisation, from the individual up (for example, families, neighbourhoods, counties, states), from the individual down (for example, body organs, cellular matrices, DNA), and for overlapping units (for example, area of residence and work environment). This paper considers

Table 1   A classification of ecological variables

Ecological variable	Description	Examples
Aggregate (Morg) Contextual (Susser) Analytical (LM)	Aggregate of attributes measured at the individual level. It is often expressed as a measure of central tendency (eg, mean, median), but may be extended to include measures of variation of individual level variables (eg, standard deviation).	Mean income Median social class Proportion smoking Area based composite indices of need/deprivation Income inequality
Contagion (Susser)	Aggregate of the individual level outcome, rather than exposure(s), that in turn affects the probability of the same outcome in individuals in the same population who are not yet affected.	Prevalence of infectious disease Suicide rate
Environmental (Morg)	Physical characteristics of a place, with an individual level analogue that usually varies between individuals (though it may remain unmeasured at the individual level).	Hours of sunlight Environmental pollutant Latitude and longitude Weather
Structural (LM)	Measure the pattern of relationships and interactions between individuals belonging to the group.	Social networks Weather
Global (Morg) Global (LM) Integral (Susser)	Measure attributes of groups, organisations or places, and are not reducible to the individual level. They are fixed for all, or nearly all, individual group members.	Social (dis)organisation Social capital Legislation or regulation

Susser = Susser, 1994[1]; LM = Lazarfeld and Menzel, 1961[19]; Morg = Morgenstern.[20]

the ecological and individual levels, and for simplicity focuses on just a single ecological and a single individual level; underlying principles can be extended to more levels (or "chinese boxes") if required.

Types of ecological variables
A classification of ecological variables is provided in table 1, including the different terms for the same (or similar) variable used by different authors.[1 19 20] In the epidemiological literature, an "ecological variable" most commonly refers to the first variable in table 1, an "aggregate variable". Measures such as the mean income of a group have a parallel at the individual level, that is an individual's income.[19] Some aggregate variables do not have such direct parallels, for example the standard deviations of an individual level variable. Aggregate variables are used most commonly in epidemiology to infer the association of the parallel individual level variable (for example, individual income) with some individual health state (for example, self reported health). Such "ecological inference" (or "cross level inference") is perceived by some as the only reason for conducting ecological research.[21]

At the other end of the spectrum from aggregate variables are "global variables", which cannot be measured at the individual level, and as such are uniquely ecological variables. In between aggregate and global variables are what Morgenstern (1998) refers to as "environmental variables".[20] These variables are the physical properties of the environment (for example, sunlight hours) that can be measured at either the ecological or individual level, but are usually measured at the ecological level for practical reasons, gaining efficiency but sacrificing a determination of the actual within group variation in exposure or dose (for example, individual exposure to sunlight). The fourth category in table 1 is the "structural variable", defined by Lazarsfeld and Menzel (1961) as being the relationships and interactions between individuals within a group.

While Lazarsfeld and Menzel defined the structural variable as distinct from the global variable, we believe that in most circumstances structural variables could be assigned as global variables—hence the dotted line in table 1.

The "contagion variable" may be defined as the aggregate of the individual level outcomes.[1] It is particularly applicable to infectious disease epidemiology where, for example, the number of infected people affects the risk of the infection for other non-immune individuals in the same population.[22] Wilson and Daly (1997) have proposed a similar "dependent happening" related to socioeconomic factors and health.[23] In a study of Chicago neighbourhoods, they concluded that their results were consistent with life expectancy itself being a determinant of risk taking. People living in a neighbourhood with a low life expectancy may be more likely to indulge in high risk behaviours as there is "less to lose".

What does an ecological effect look like?
The two level model used in this paper includes three types of variable: the ecological exposure(s), X; the individual level exposure(s), x ; and the individual level outcome, y. There are three ways that X can have a cross level effect on y: by directly affecting y (direct cross level effect); by modifying the relation between x and y (cross level effect modification); and by affecting x, which in turn affects y (indirect cross level effect). These ecological cross level effects are presented in figure 1. Effect modification may also occur between ecological variables, but is not shown in figure 1 as it is a step removed from the impact of one ecological exposure on a health outcome—nevertheless it is important when two or more ecological exposures are considered simultaneously.

In a reductionist sense, ecological variables cannot impact "directly" on individuals; instead their effect must be mediated by intermediate variables at the individual level.[24] For example, possible mechanisms linking income distribution to health include: variations in individual's

access to life opportunities and material resources (for example, health care, education); social cohesion, whereby mutual support and cooperation secure better health outcomes; and possible direct psychosocial processes related to relative perceptions of position on the socioeconomic hierarchy.[25] Taking the latter mechanism of socioeconomic hierarchy to a lower level again, animal models have found that experimental manipulation of social status in monkeys affects development of atherosclerosis.[26] Likewise, social ranking of monkeys has been associated with adrenocorticoid profiles.[27] Therefore, it may be argued that neither direct cross level effect nor cross level effect modification are complete causal chains, but require reduction to indirect cross level effects as shown in figure 1. However, to do so would require perfect information on all possible variables. Such reductionism is helpful to understand aetiologically how ecological exposures affect health, but is often unnecessary, and may even be counterproductive, for the identification of intervention points for public health policy and action.[3 4 10 28] The choice of level of causation, and hence the intervening variables to include or exclude, may therefore be a pragmatic one.

As summarised by Helman (1984), "… the idea of cause has become meaningless other than as a convenient designation for the point in the chain of event sequences at which intervention is most practical."[29]

Fallacies

Diez-Roux (1998) provides an excellent overview of four types of fallacy in multi-level analysis: table 2 is adapted from this paper.[30] The ecological fallacy is well documented in epidemiology,[31–33] being a false inference of the association of individual level variables on the basis of the observed association of the parallel ecological variables. For example, national GDP may be positively associated with motor vehicle fatality rates by country, but within countries the highest death rate from motor vehicle crashes may be for the low income groups. An example of the psychologistic fallacy given by Diez-Roux is where immigrants in a particular study are found to have higher rates of depression, but unbeknown to the researcher this was only true for immigrants living in communities where they represent a minority. Ignoring this contextual effect may wrongly lead to assigning the increased
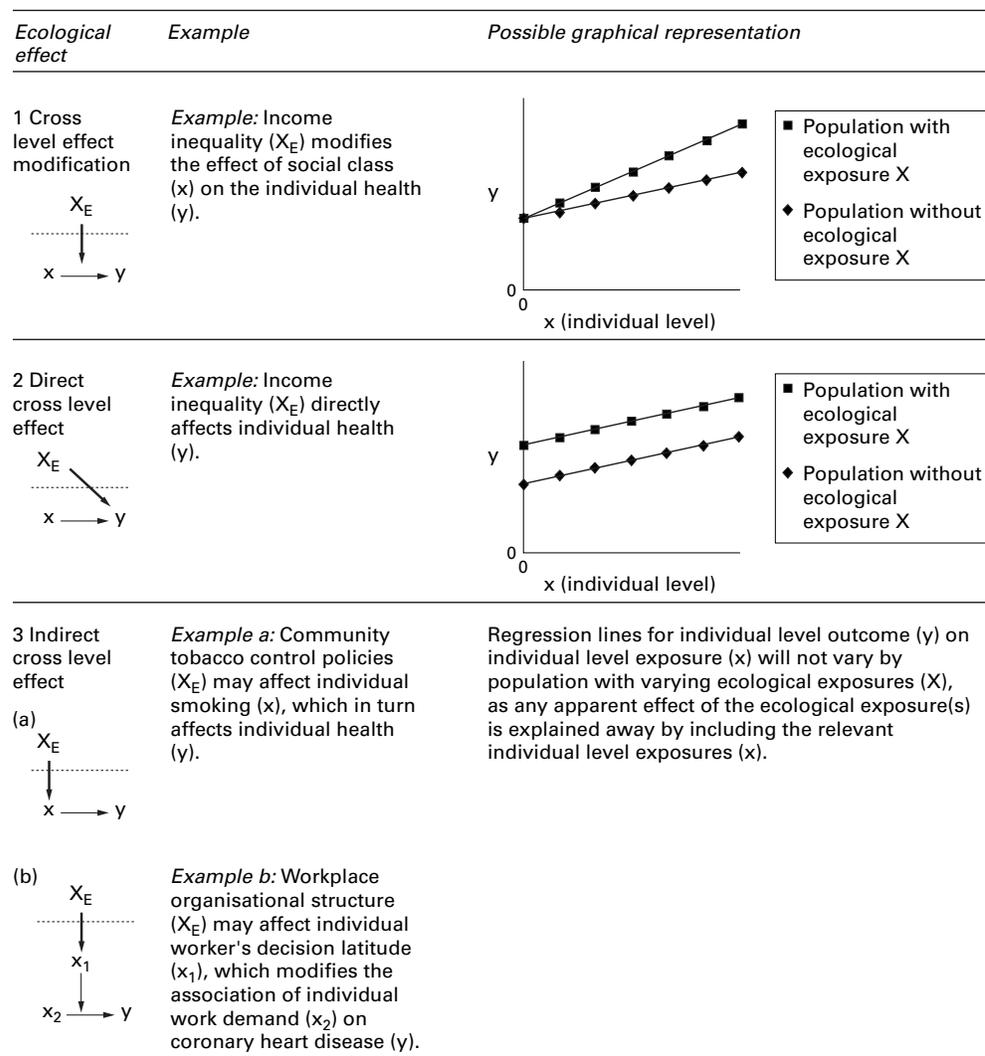
Ecological effect	Example	Possible graphical representation
1 Cross level effect modification X_E ↓ $x \longrightarrow y$	*Example:* Income inequality (X_E) modifies the effect of social class (x) on the individual health (y).	 ■ Population with ecological exposure X ◆ Population without ecological exposure X
2 Direct cross level effect X_E ↘ $x \longrightarrow y$	*Example:* Income inequality (X_E) directly affects individual health (y).	 ■ Population with ecological exposure X ◆ Population without ecological exposure X
3 Indirect cross level effect (a) X_E ↓ $x \longrightarrow y$	*Example a:* Community tobacco control policies (X_E) may affect individual smoking (x), which in turn affects individual health (y).	Regression lines for individual level outcome (y) on individual level exposure (x) will not vary by population with varying ecological exposures (X), as any apparent effect of the ecological exposure(s) is explained away by including the relevant individual level exposures (x).
(b) X_E ↓ x_1 ↓ $x_2 \longrightarrow y$	*Example b:* Workplace organisational structure (X_E) may affect individual worker's decision latitude (x_1), which modifies the association of individual work demand (x_2) on coronary heart disease (y).	

Figure 1   Three types of ecological effect.

Table 2   Types of fallacy in multi-level research (taken from Diez-Roux[30])

Unit of analysis	Level of inference	Type of fallacy
Group	Individual	Ecological
Individual	Group	Atomistic*
Individual: relevant group level variables excluded	Individual	Psychologistic*
Group: relevant individual level variables excluded	Group	Sociologistic

*Also called individualistic by some authors.

rates of depression to an individual factor such as race, rather than the context. The sociologistic fallacy is opposite to the psychologistic, whereby both analysis and inference occur at the group level, but relevant individual level variables are excluded. For example, ecological studies may find an association of income inequality with health status that is actually attributable to confounding by individual factors such as smoking. Multi-level research tries to avoid all four of these fallacies, but the ascription of ecological effects (as in figure 1) is particularly at risk of the sociologistic fallacy.

Estimating ecological effects: study design, sources of error and strategies to correctly identify ecological effects

In the remainder of this paper we consider limitations in study design and sources of error affecting the estimation of ecological effects, and strategies to correctly identify ecological effects. The framework is organised under six subheadings: ensuring variation of the ecological exposure; precision and multi-level statistical methods; selection bias; confounding; information bias; model specification and multicollinearity. Other authors have considered sources of error giving rise to cross level bias, or the ecological fallacy, in ecological inference of an individual level association from the observed aggregate level association.[20][31–34] These sources of error are not directly transferable to multi-level studies, where ecological effects in themselves are estimated—however, there is some overlap.

ENSURING VARIATION OF THE ECOLOGICAL EXPOSURE

It is a "sine qua non" of epidemiology that to detect any effect there must be variation in the exposure (and outcome) under study. This essential prerequisite may be problematic for ecological exposures. Often macro-level socioeconomic exposures (for example, income inequality) do not vary within the eligible study population (for example, state or country, or more pragmatically the available dataset) at one point in time. The identification of small ecological effects in a study may, therefore, actually be just the tip of the iceberg, and should not dismissed as inconsequential. When there is insufficient variation in the ecological exposure in the eligible population at one point in time, extension of the study design across time or populations may provide the necessary variation. Firstly, additional populations with different levels of the ecological exposure may be added to the analysis (for example, cross national studies). A likely drawback, however, is a lack of comparability of unmeasured covariates between populations/datasets. For

example, "culture" may vary between countries and be independently associated with health. Secondly, a times series study of one population may capture variation in the ecological exposure, but controlling for secular trends is difficult. Thirdly, data for both multiple populations/datasets and different time periods may be combined in a mixed study design,[33] thus combining the two former study designs. This mixed study design allows a simultaneous analysis of within group changes over time in ecological exposure and outcome, and between group variation in ecological exposure and outcome. Unfortunately, datasets of this richness are likely to be rare.

PRECISION AND MULTI-LEVEL STATISTICAL METHODS

Multi-level studies entail hierarchically clustered units of analysis, for example individuals within census tracts within counties. Such study designs are subject to intraclass correlation, whereby individuals within groups are more alike than individuals across groups. Statistical analysis that ignores the multi-level nature of the dataset may underestimate the standard error of ecological effects. To more conservatively estimate the standard error of ecological effects, separate random error terms may be specified for each level of analysis (that is, randomly varying intercepts between ecological units for the regression equation). Random error terms may also be included for the individual level coefficients (that is, randomly varying slopes between ecological units). Overviews of multi-level statistical methods can be found elsewhere.[8][9][35][36] Given a fixed number of individuals, the balance of the number of ecological units to the number of individuals in each ecological unit that maximises the precision of estimated ecological effects is a complex function of the intraclass correlation and covariances. As a general study design rule in social epidemiology, more precise estimates of ecological effects will usually be obtained by increasing the number of ecological units compared with just increasing the number of individuals within each ecological unit.

SELECTION BIAS

Selection bias is a potential source of error, both from systematic bias in the selection of individuals within ecological units, and the possible selection of ecological units themselves.

CONFOUNDING

In general, there are two types of confounding of ecological exposures: within ecological level confounding by ecological covariates, and cross level confounding by individual level covariates (figure 2). Within ecological level confounding is conceptually the same as confounding in single level epidemiology—both the exposure and confounders are at the same level of analysis. Cross level confounding may be more conceptually challenging. A commonly cited example is individual level income as a confounder of the association of income inequality with health.[37] As the association of individual
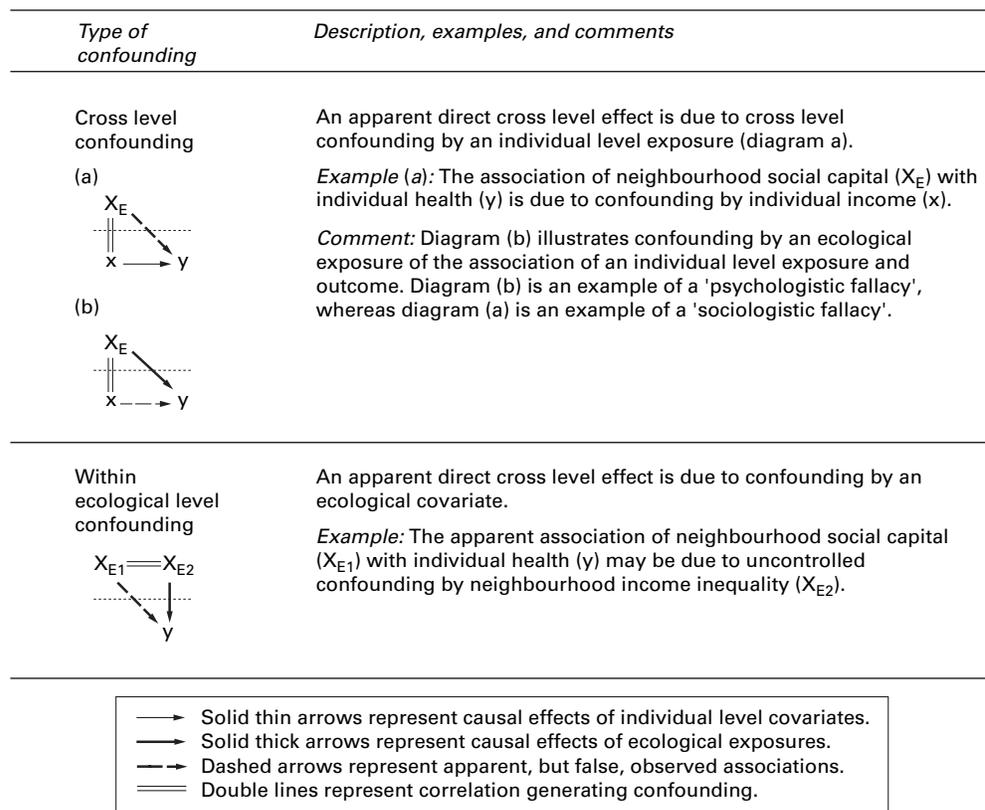
Type of confounding	Description, examples, and comments
Cross level confounding (a) X_E ⤏ ‖ x → y (b) X_E ‖ x ⤏ y	An apparent direct cross level effect is due to cross level confounding by an individual level exposure (diagram a). *Example (a):* The association of neighbourhood social capital (X_E) with individual health (y) is due to confounding by individual income (x). *Comment:* Diagram (b) illustrates confounding by an ecological exposure of the association of an individual level exposure and outcome. Diagram (b) is an example of a 'psychologistic fallacy', whereas diagram (a) is an example of a 'sociologistic fallacy'.
Within ecological level confounding X_{E1}═X_{E2} ⤏ ↓ y	An apparent direct cross level effect is due to confounding by an ecological covariate. *Example:* The apparent association of neighbourhood social capital (X_{E1}) with individual health (y) may be due to uncontrolled confounding by neighbourhood income inequality (X_{E2}).

> → Solid thin arrows represent causal effects of individual level covariates.
> → Solid thick arrows represent causal effects of ecological exposures.
> ⤏ Dashed arrows represent apparent, but false, observed associations.
> ═ Double lines represent correlation generating confounding.

Figure 2   Confounding as a source of error for estimated effects.

income with health is non-linear,[38] it is possible that the average income by ecological unit is not associated with income inequality by ecological unit, yet individual income could still be confounding the association of income inequality with health. To control for this cross level confounding, individual income must be included in the model and specified as a categorical variable or some appropriate transformation of absolute income (for example, the natural logarithm).

Note that confounding purely within the individual level cannot bias an ecological effect, unless one of the individual level confounders is also associated with the ecological exposure—which is cross level confounding.

An important issue in multi-level research is that it may be difficult to differentiate between individual level covariates as confounders or intermediary variables. If the latter, then "controlling" for the individual level covariate will lead to overlooking indirect cross level effects. For example, work in progress by one of us (TB) suggests that the association of state level income inequality with self rated health in the United States is reduced when education is included at the individual level. Should education here be considered a confounder or an intervening variable between income inequality and health? The answer is not clear. It is suggested that that less egalitarian states (that is, states with high income inequality) tend to under invest in education,[39] thus placing individual education, in part at least, as an intermediary variable. Analyses with and without the individual level covariate should be presented to give an upper and lower bound within which the reader may judge the "true" ecological effect.

It is possible to describe a third type of confounding of an ecological effect that arises not from the association of one covariate with the ecological exposure, but the association of the joint distribution of two or more covariates with the ecological exposure. This possibility is analogous to the demonstration by Greenland and Morgenstern that effect modification within the individual level can result in cross level bias in ecological inference.[32] For example, if smoking and alcohol consumption interact at individual level in their association with health, and the percentage of heavy alcohol drinkers that were smokers varied by ecological unit, and the variation in this latter joint distribution was correlated with both the ecological exposure and health outcome of interest, error may occur in the measurement of an effect for the ecological exposure of interest. Such variation of the joint distribution of individual level variables, over and above variation in their singular distribution, is probably unlikely.

INFORMATION BIAS
We broadly differentiate information bias here into misclassification or mismeasurement of the ecological exposure and covariates, and incorrect assignment of individuals to groups or ecological units of analysis.

Non-differential misclassification bias of exposure nearly always causes a bias to the null in single level epidemiology,[40] but may cause bias in either direction in multi-level research dependent upon the nature of the exposure (binary, or continuous) and the level of

measurement (ecological or individual level).[41] Consider a binary individual level exposure (home ownership as a proxy for wealth) non-differentially misclassified during measurement at the individual level, and then represented as an aggregate ecological variable. Assume that the "unexposed" regions have 85% home ownership, the "exposed" regions 15% home ownership, and that there is a direct cross level effect of home ownership on health. If home ownership was non-differentially misclassified at the individual level, then those regions with 85% home ownership would have a lower observed home ownership: if 10% of all home ownership was recorded incorrectly by individuals then $((85\% \times 0.90) + (15\% \times 0.10)) = 78\%$ (rather than 85%) will be observed as home owners in the "unexposed" regions. The reverse will happen for the exposed region: 22% of individuals will be observed as homeowners. If one then extrapolates any direct cross level effect for home ownership to the hypothetical instance of regions with full home ownership versus those with none, the ecological effect will be overestimated by $(1/(0.78{-}0.22)) / (1/(0.85{-}0.15)) = 1.25$, a bias away from the null.

Secondly, consider a continuous individual level variable randomly mismeasured at the individual level, and then represented as a mean aggregate ecological exposure. Here, there is no bias in the estimated ecological effect: the random mismeasurements for all individuals within groups should sum to zero, meaning that there is no bias in the summary mean for the group. Thirdly, consider random misclassification and mismeasurement of ecological exposures measured directly at the ecological level (for example, global and environmental ecological exposures): here measurement is at the same level as representation of the exposure, and effect measures will be biased to the null as for single level epidemiology generally.

The lag time between an ecological exposure and individual level health outcome is a form of misclassification bias that deserves specific mention. Many multi-level studies that consider ecological socioeconomic exposures have used cross sectional survey data.[10 42 43] Not only does this introduce the possibility of reverse causation (health status affecting the ecological exposure), but it also implies a zero lag time between exposure and outcome. It is usually implausible for the effect of an exposure to be instantaneous, particularly in social epidemiology. If the ecological exposure is stable over time, then specification of a lag time may not be necessary—otherwise incorrect specification of lag time is another source of misclassification bias. Investigation of lag times between socioeconomic ecological exposures and individual outcomes is required.

Regarding non-differential misclassification of confounders, misclassification of individual level confounders and ecological level confounders (measured directly at the ecological level) will generally reduce the ability to control for confounding. However, for ecological confounders that are first measured at the indi-

vidual level and then aggregated up, non-differential misclassification during measurement at the individual level may not reduce the ability to control for confounding.[44]

An important issue is the grouping of individuals into ecological units, yet the implications of grouping strategies are often overlooked.[43 45] The level of aggregation is considered under the next subheading (theory and model specification); here the incorrect assignment of individuals to groups is considered as an information or misclassification bias. As an example, consider an individual assigned to the wrong neighbourhood in a study of the association between neighbourhood cohesiveness and individual health. A first bias is that the level of cohesiveness for the assigned neighbourhood may not be the same as the individual's true neighbourhood, resulting in misclassification of the ecological exposure for that individual. Such misclassification might be expected to be non-differential, biasing the observed association of cohesiveness and health to the null. A second bias may arise if the measurement of cohesiveness was based on aggregated individual level responses including the incorrectly assigned individual, thus biasing the observed level of cohesiveness for the given neighbourhood. These two sources of bias are magnified when grouping is not conducted specifically for the given study, but instead existent administrative groups (for example, census tracts) are used with likely incorrect assignment of both individuals and group "boundaries".[42 43] The likely effect of using convenient rather than theoretically predetermined ecological units is a reduced ability to detect any ecological effect.

THEORY AND MODEL SPECIFICATION

There are numerous theoretical and model specification issues that confront anyone doing multi-level studies. We consider just a few examples.

An ecological effect may vary with the level of aggregation. Soobader and LeClere (1999) found a stronger association between income inequality and morbidity at the county level,

compared with the census tract level, in the United States.[46] The authors concluded that the level of aggregation was important, such that at lower levels of aggregation (census tract) the effect of income inequality was "mediated through neighborhood consequences of income inequality and individual processes". Thus, not only may the strength of the observed association vary by level of aggregation, but so too may the mechanisms. Extending Soobaders and LeCleres' example, the possible underlying mechanism of income inequality affecting health at the census tract may be relative perceptions of social hierarchy, at the county level may be via segregation of neighbourhoods, and at the state or national level may be via policies that affect individual's access to life opportunities and material resources (for example, health care, education). Thus, both the quantitative and qualitative association of an ecological exposure with an individual level outcome may vary by level of aggregation. The correct level of aggregation for socioeconomic ecological variables is the subject of ongoing research.

Often only direct cross level effects are considered explicitly; cross level effect modification and indirect cross level effects are implicitly overlooked. For example, Boyle and Willms (1999) found little association between "place" variation in health status in a multi-level study using the Ontario Health Survey, having included individual level covariates.[43] The authors interpreted this as suggesting little or no ecological effect. Such conclusions may be valid for the allocation of health resources at a given point in time for a given society, but they are not necessarily aetiologically valid (as acknowledged by Boyle and Willms). By default, the conclusion of little or no ecological effect pertains only to little or no direct cross level effect: cross level effect modification was not considered, and individual level covariates of education and income were assumed to be confounders rather than components, in part at least, of any indirect cross level effect. More generally, the study by Boyle and Willms, and others (for example, Duncan *et al*[10 42]), highlights that a lack of variation in health status by place may not reflect a lack of ecological effect, but more a lack of variation in ecological exposures within the given population, time period, or dataset—the subject of the first subheading in this section. Regarding cross level effect modification, the researcher must be explicit whether an underlying additive or multiplicative model is assumed. Rothman and Greenland argue that an additive model is the causally relevant one, requiring modelling strategies other than just including interaction products in multiplicative regression models.[40]

Two final sources of error in multi-level analysis require mentioning: model misspecification and multicollinearity. As an example of the former, the association of individual income with health is non-linear,[38] and will be a source of error if not modelled as either a categorical variable or some non-linear function of income. Little is known about the form of the relation of ecological exposures with health. It would be prudent, therefore, to model ecological variables as categorical variables in the first instance. Multicollinearity is more likely for ecological variables than for individual level variables,[33] and may make it impossible to estimate independent effects for more than one ecological exposure simultaneously.

Conclusion

Multi-level studies, and the accompanying statistics, are complex. The motivation for this paper was our own difficulty grappling with the complexity of multi-level studies, in particular the actual implementation of the call to incorporate the ecological level into epidemiological practice.[6 7] We anticipate that as researchers move beyond the initial exhilaration of applying the "magic" of multi-level statistical methods to data, there will be an increasing and necessary focus on theory, study design, and sources of error. For example, it is likely that studies will suggest that a range of ecological exposures are related to health, but multicollinearity between these ecological exposures will beg the question "'which ones are the important ones", and "what is the causal web"? In this paper we have attempted to step back from the statistics and clearly define the nature of an ecological effect, the sources of error that may be incurred ascribing an ecological effect, and research strategies to enhance (correct) identification of an ecological effect. Specific recommendations we make to other researchers conducting a multi-level study include:

- consider each of the three types of ecological effect (direct cross level effect, cross level effect modification, and indirect cross level effects)
- consider whether there is sufficient variation in the ecological exposure in the available dataset
- assess possible sources of error (selection bias, confounding, and information biases)
- present results both with and without an individual level covariate in the model when it is possible that the individual level covariate is an intervening variable between the ecological exposure and individual health outcome
- consider time lags between exposure and outcome
- consider the limitations of the ecological units available on administrative datasets
- conduct sensitivity analyses with different models and datasets.

1 Susser M. The logic in ecological: I. The logic of analysis. *Am J Public Health* 1994;**84**:825–9.
2 Krieger N. Epidemiology and the web of causation: has anyone seen the spider? *Soc Sci Med* 1994;**39**:887–903.
3 Mackenbach J. Public health epidemiology. *J Epidemiol Community Health* 1995;**49**:333–4.
4 Pearce N. Traditional epidemiology, modern epidemiology, and public health. *Am J Public Health* 1996;**86**:678–83.

5 Woodward A. What makes populations vulnerable to ill health? *NZ Med J* 1996;**109**:265–7.

6 Susser M, Susser E. Choosing a future for epidemiology: II. From black box to chinese boxes and eco-epidemiology. *Am J Public Health* 1996;**86**:674–7.

7 Susser M. Does risk-factor epidemiology put epidemiology at risk? Peering into the future. *J Epidemiol Community Health* 1998;**52**:608–11.

8 Bryk A, Radenbush S. *Hierarchical linear models: applications and data analysis methods.* Newbury Park: Sage, 1992.

9 Goldstein H. *Multilevel statistical models.* London: Edward Arnold, 1995.

10 Duncan C, Jones K, Moon G. Health-related behaviour in context: a multilevel modelling approach. *Soc Sci Med* 1996;**42**:817–30.

11 Diez-Roux A, Nieto F, Muntaner C, *et al*. Neighourhood environment and coronary heart disease: a multilevel analysis. *Am J Epidemiol* 1997;**146**:48–63.

12 Durkheim E. *Suicide.* Glencoe, IL: Free Press, 1951. [Original work published in 1898].

13 van den Eeden P, Huttner H. Multi-level research. *The Journal of the International Sociological Association* 1982;**30**:1–182.

14 Blau P. Structural effects. *American Sociological Review* 1960;**25**:178–93.

15 Hauser R. Context and consex: a cautionary tale. *American Journal of Sociology* 1970;**75**:645–64.

16 Hauser R. Contextual analysis revisited. *Sociological Methods and Research* 1974;**2**:365–75.

17 Farkas G. Specification, residuals and contextual effects. *Sociological Methods and Research* 1974;**2**:333–63.

18 Hauser R. Reply to Barton. *American Journal of Sociology* 1970;**76**:517–20.

19 Lazarfeld P, Menzel H. On the relation between individual and collective properties. In: Etzioni A, ed. *Complex organization.* New York: Holt, Reinhart and Winston, 1961:422–40.

20 Morgenstern H. Ecologic studies. In: Rothman K, Greenland S, eds. *Modern epidemiology.* 2nd ed. Philadelphia: Lippincott-Raven, 1998:459–80.

21 Piantadosi S, Byar D, Green S. The ecological fallacy. *Am J Epidemiol* 1988;**127**:893–904.

22 Halloran M, Struchiner C. Study designs for dependent happenings. *Epidemiology* 1991;**2**:331–8.

23 Wilson M, Daly M. Life expectancy, economic inequality, homicide, and reproductive timing in Chicago neighbourhoods. *BMJ* 1997;**314**:1271–4.

24 Susser M. *Causal thinking in the health sciences: concepts and strategies in the health sciences.* New York: Oxford University Press, 1973.

25 Kawachi I, Kennedy B. Income inequality and health: pathways and mechanisms. *Health Serv Res* 1999;**34**:215–7.

26 Shivley C, Clarkson T. Social status and coronary artery atherosclerosis in female monkeys. *Arterioscler Thromb* 1994;**14**:721–6.

27 Sapolsky R, Alberts S, Altmann J. Hyperchortisolism associated with social subordinance or social isolation among wild baboons. *Arch Gen Psychiatry* 1997;**54**:1137–43.

28 Rose G. *The strategy of preventive medicine.* Oxford: Oxford University Press, 1992.

29 Helman C. *Culture, health and illness.* Bristol: Wright, 1984.

30 Diez-Roux A. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *Am J Public Health* 1998;**88**:216–22.

31 Greenland S. Divergent biases in ecologic and individual-level studies. *Stat Med* 1992;**11**:1209–23.

32 Greenland S, Morgenstern H. Ecological bias, confounding, and effect modification. *Int J Epidemiol* 1989;**18**:269–74.

33 Morgenstern H. Ecologic studies in epidemiology: concepts, principles, and methods. *Annu Rev Public Health* 1995;**16**:61–81.

34 Greenland S, Robins J. Invited commentary: ecologic studies - biases, misconceptions, and counterexamples. *Am J Epidemiol* 1994;**139**:747–60.

35 Hox J. *Applied multilevel analysis.* Amsterdam: TT-Publikaties, 1995.

36 Burton P, Gurrin L, Sly P. Extending the simple linear regression model for correlated responses: an introduction to generalized estimating equations and multi-level mixed modeling. *Stat Med* 1998;**17**:1261–91.

37 Gravelle H. How much of the relation between population mortality and unequal distribution of income is a statistical artefact? *BMJ* 1998;**316**:382–5.

38 Backlund E, Sorlie PD, Johnson NJ. The shape of the relationship between income and mortality in the United States: Evidence from the National Longitudinal Mortality Study. *Ann Epidemiol* 1996;**6**:12–20.

39 Kaplan G, Pamuk E, Lynch J, *et al*. Inequality in income and mortality in the United States: analysis of mortality and potential pathways. *BMJ* 1996;**312**:999–1003.

40 Rothman K, Greenland S. *Modern epidemiology.* 2nd ed. Philadelphia: Lippincott-Raven, 1998.

41 Brenner H, Greenland S, Jockel K-H, *et al*. Effects of non-differential exposure misclassification in ecologic studies. *Am J Epidemiol* 1992;**135**:85–95.

42 Duncan C, Jones K, Moon G. Do places matter? A multi-level analysis of regional variations in health-related behaviour in Britain. *Soc Sci Med* 1993;**37**:725–33.

43 Boyle M, Willms J. Place effects for areas defined by administrative boundaries. *Am J Epidemiol* 1999;**149**:577–85.

44 Brenner H, Greenland S, Savitz D. The effects of non-differential confounder misclassification in ecologic studies. *Epidemiology* 1992;**3**:456–9.

45 Iversen G. *Contextual analysis.* Newbury Park, CA: Sage, 1991.

46 Soobader M-J, LeClere F. Aggregation and the measurement of income inequality: effects on morbidity. *Soc Sci Med* 1999;**48**:733–44.