International Journal of Epidemiology

Vol. 18, No. 1
Printed in Great Britain

Ecological Bias, Confounding, and Effect Modification

SANDER GREENLAND AND HAL MORGENSTERN

Greenland S (Division of Epidemiology, UCLA School of Public Health, Los Angeles, California 90024, USA) and Morgenstern H. Ecological bias, confounding, and effect modification. *International Journal of Epidemiology* 1989, **18**: 269–274.

Ecological bias is sometimes attributed to confounding by the group variable (ie the variable used to define the ecological groups), or to risk factors associated with the group variable. We show that the group variable need not be a confounder (in the strict epidemiological sense) for ecological bias to occur: effect modification can lead to profound ecological bias, whether or not the group variable or the effect modifier are independent risk factors. Furthermore, an extraneous risk factor need not be associated with the study variable at the individual level in order to produce ecological bias. Thus the conditions for the production of ecological bias by a covariate are much broader than the conditions for the production of individual-level confounding by a covariate. We also show that standardization or ecological control of variables responsible for ecological bias are generally insufficient to remove such bias.

Ecological bias (also known as aggregation bias or cross-level bias) refers to the failure of ecological- (aggregate-) level associations to properly reflect individual-level associations.[1–3] Epidemiological confounding refers to the failure of a crude (or partially adjusted) association to properly reflect the magnitude of an exposure effect, due to differences in the distribution of extraneous risk factors among exposed and unexposed individuals.[4–6] It is well known (cf[2,3]) that regional differences in background rates can produce large amounts of ecological bias. Perhaps as a result, it is not uncommon to see ecological bias equated with confounding or attributed to the presence of confounding factors that vary across ecological groups. We often encounter this view when authors argue that their ecological results are unbiased because certain other risk factors are not confounding at an individual level or not associated with the exposure.

Effect modification refers to variation in the magnitude of an effect measure across levels of a third variable.[6,7] We can consider group as the third variable (where 'group' is a nominal variable indicating the ecological group to which an individual belongs). The key to understanding the connection between ecological bias, confounding, and effect modification is to evaluate group as a confounder and an effect modifier *at the individual level*. We will show that ecological bias may result from group acting as a confounder *or* a modifier of the exposure effect.

Ecological bias analogous to confounding occurs when the background rate of disease (ie the rate in the unexposed) varies across groups. Such variation arises from the differential distribution of extraneous risk factors across groups. These other risk factors need not be confounders or modifiers of the exposure effect at the individual level (within groups). Ecological bias can also occur when group appears as an effect modifier, ie when the exposure effect varies across groups.[6,7] Such variation can arise from the differential distribution of individual-level effect modifiers across groups. Such ecological bias can be severe even when the effect modification is relatively weak and there is no confounding by group. Consequently, it can be misleading to claim that a particular ecological analysis is unbiased simply because no confounding is present.

We will also show that ecological control of confounders and other covariates responsible for ecological bias cannot be expected to completely remove the biases such covariates produce, and may even worsen bias. Our observations may have some potential importance, in that implausibility of strong confounding is sometimes used to argue against the presence of ecological bias, and yet there are rarely any *a priori* reasons for claiming absence of effect modification. In fact, when comparing regional impacts of the introduction of health services or technologies (such as fetal monitoring[8]), inter-regional differences in administration and training should lead one to expect regional variation in effectiveness, ie effect modification by region.

Mathematical derivations of the results we discuss are given in the appendix. There we show that the ecological linear regression coefficient[1–3,9] can be decomposed

Division of Epidemiology, UCLA School of Public Health, Los Angeles, CA 90024 USA.

into three components: the average rate difference across groups, a bias factor due to confounding by group, and a bias factor due to effect modification by group.

Our first two illustrations provide simple numeric counterexamples to the notion that ecological bias requires or always originates from individual-level confounding. For simplicity, we will assume that all rates and prevalences are age-adjusted and apply to white males only, although none of our findings depend on this assumption.

AN EXAMPLE OF ECOLOGICAL BIAS FROM PURE EFFECT MODIFICATION

Suppose one studies the relationship between smoking and regional variation in oesophageal cancer incidence. If there is some small variation in background prevalence of a cofactor (eg a nutritional deficiency) which is itself not a risk factor in the absence of the study factor (smoking), the background (non-smoker) rates need not vary, but the study factor effect will vary.[6] Table 1 illustrates this sort of phenomenon. Note that since the non-smoker rates do not vary by region, region cannot be a confounder in an individual-level analysis of this table.[4,5] Thus we should get identical results from an analysis of exposure effect that standardizes for region and a 'crude' analysis that simply combines numbers across the regions (groups), as in the last panel of Table 1. This can be verified directly by comparing the crude rate ratio (obtained by combining all regions) with the standardized morbidity ratio (SMR).[4,5] The total number of smoking cases observed in all the regions in a year, O, would (on average) be [12(100) + 15(80) + 20(60)]/100 = 36, corresponding to a rate of 36/240 000 = 15 cases per 100 000 people per year. The total number expected among the smokers if they had experienced the same rates as the non-smokers, E, would be 3(240)/100 = 7.2, corresponding to a rate of 7.2/240 000 = 3 cases per 100 000 people per year. This expected rate is exactly equal to the rate among the non-smokers in the combined population. As a result, SMR = O/E = 36/7.2 = 5.0 is exactly equal to the crude rate ratio obtained by ignoring region, 15/3 = 5.0. Similarly, the standardized morbidity and crude rate differences are both equal to 12 per 100 000 people per year. Thus, there is no confounding by region. (This result is little changed by choice of standard; for example, using the total population as the standard yields a standardized rate ratio of 5.2).

An ecological analysis would make use only of the data in the last two lines of Table 1. An ecological linear regression[1-3,9] of y = regional (cancer) rate on x = smoking prevalence, y = a + bx, would find a = 9 and b = −3 cases per 100 000 people per year. Thus b, which corresponds to the ecological estimate of the rate difference, is in the wrong direction. The ecological estimate of the rate ratio[9] is 1 + (−3)/9 = 0.67, incorrectly showing a protective effect of smoking. This is similar to the inverse ecological association of smoking and oesophageal cancer found in Richardson *et al* (ref 3, Figure 6).

One can obtain an alternative ecological estimate of the rate ratio by using a linear regression of the natural log of the cancer rate y on smoking prevalence x, ie fitting the log-linear model $\log_e y = a + bx$. Under this model, exp(b) will then be the rate ratio estimate.[3] For the data in Table 1, this model yields a rate-ratio estimate of exp(−0.385) = 0.68, essentially just as biased as the linear-model estimate.

Note that both the effect modification and the ecological bias in this example do *not* qualitatively depend on the parameter or scale chosen to measure effects: Because the baseline rate in Table 1 is constant across groups, any reasonable measure of effect will vary across the groups. Similarly, it is obvious, no matter what measure of association is used, that the ecological association of smoking prevalence and cancer is in the opposite direction of the individual-level associations. Thus the bias cannot be ascribed to the scale-dependent[6,7] properties of effect modification. A key feature in this example is that no confounding by region or factors associated with region is present: a crude (unadjusted) analysis of the combined individual data yields an unbiased estimate of effect. Thus the profound ecological bias cannot be attributed to the effect of some strong extraneous risk factor that varies across regions. In fact, the net effect of the unmeasured cofactors responsible for effect modification in the above example could be smaller than the effect of the study exposure: note that the observed modification adds only 8 more cases per 100 000 people per year to the exposed rate in region C, compared to the 9 cases per 100 000 people per year excess rate seen among the exposed in region A. In general, the effect of a cofactor may be too weak to conceivably 'confound away' the exposure effect in a non-ecological analysis, and yet, because of its modification of exposure effect, it may still produce enough ecological bias to completely reverse the estimate of exposure effect.

ECOLOGICAL BIAS AND EXTRANEOUS RISK FACTORS

One phenomenon noted in the literature,[2] but sometimes overlooked by researchers, is the possibility for a risk factor to introduce ecological bias without having any individual-level association with the study variable within groups. More generally, a risk factor need not

TABLE 1 *Numerical example of ecological bias induced by effect modification in the absence of confounding by region*

	Region A		Region B		Region C		All regions combined	
Smoking	Yes	No	Yes	No	Yes	No	Yes	No
Oesophageal cancer rate*	12	3	15	3	20	3	15	3
Population size	100 000	100 000	80 000	120 000	60 000	140 000	240 000	360 000
Rate ratio	4.0		5.0		6.7		5.0	
Rate difference	9		12		17		12	
Regional rate of:								
Smoking (x)	0.50		0.40		0.30			
Oesophageal cancer* (y)	7.5		7.8		8.1			

* Mortality per 100 000 people per year.
Ecological regression of y (cancer rate) on x (smoking prevalence): $y = 9.0 - 3.0x$.

be a modifier or individual-level confounder within groups in order to produce ecological bias.

Table 2 gives a hypothetical example of this phenomenon in an ecological study of household radon levels and lung cancer mortality in older males. For simplicity, smoking and radon level have been dichotomized, and only three regions are considered. Household surveys indicate that within several regions in the US, smoking and radon level are unassociated (Cohen, unpublished data) and Table 2 is constructed to reflect this: high radon and smoking are unassociated in all regions. Note that the true (individual) rate ratios for radon effect and smoking effect are 2 and 10; both are constant across regions. The rate ratio for radon effect from the crude individual data (ignoring both smoking and region) is 1.9, only slightly confounded by smoking.

Despite the lack of association of smoking and radon within regions, the association of smoking with region results in an association of region with lung cancer rates that is independent of radon level. Consequently, the ecological estimate of radon effect is biased: A simple linear regression of regional cancer rates on high-radon prevalence would yield an ecological rate ratio estimate of $1 + b/a = 1 + (-59)/84 = 0.3$, the reverse of the true association; similarly, a simple log-linear regression yields a rate-ratio estimate of 0.4. Control for smoking in the ecological regression removes only a portion of the bias: the rate-ratio estimate for radon effect from a multiple linear regression including smoking is $[29 + 7 + 76(0.40)]/[29 + 76(0.40)] = 1.1$ when evaluated at the mean smoking prevalence of 0.40; a multiple log-linear regression yields an estimate of 1.0. Thus the ecological estimates of radon effect are severely biased, despite the lack of confounding *within* regions and the miniscule amount of confounding in the crude individual analysis.

As explained in the Appendix, confounding of within-region (individual-level) associations by a covariate will generally contribute to bias in the ecological estimates of effect, even if the covariate is identically distributed across groups or populations (although such a contribution may of course be cancelled by other sources of bias). A more distressing property of ecological analysis is that ecological control of a confounder cannot be expected to completely control confounding, and may yield little or no bias reduction.

Table 3 presents a numerical example of this phenomenon. Here, the effect of alcohol use on oesophageal cancer mortality is to be estimated, and smoking is a known confounder: at the individual level, alcohol use and smoking are strongly associated, as reflected in

TABLE 2 *Numerical example of ecological bias induced by a covariate with no individual-level association with study factor*

	Region A				Region B				Region C			
	Smokers		Non-smokers		Smokers		Non-smokers		Smokers		Non-smokers	
Radon level	High	Low	High	Low	High	Low	High	Low	High	Low	High	Low
Lung cancer rates*	200	100	20	10	200	100	20	10	200	100	20	10
Population size	26 000	74 000	26 000	74 000	28 000	52 000	42 000	78 000	30 000	30 000	70 000	70 000
Regional rate of:												
High radon (x_1)	0.26				0.35				0.50			
Smoking (x_2)	0.50				0.40				0.30			
Lung cancer* (y)	69.3				62.1				55.5			

* Mortality per 100 000 people per year.
Ecological regressions of y (cancer rate) on
x_1 (high-radon prevalence) : $y = 84 - 59x_1$
x_1 and x_2 (smoking prevalence) : $y = 29 + 7x_1 + 76x_2$.

TABLE 3 *Numerical example in which ecological confounder control increases bias*

	Region A				Region B				Region C			
	Smokers		Non-smokers		Smokers		Non-smokers		Smokers		Non-smokers	
Alcohol use	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No
Oesophageal cancer rate*	25	5	5	1	25	5	5	1	25	5	5	1
Population size	50 000	50 000	30 000	70 000	58 750	21 250	21 250	98 750	60 000	12 000	38 000	90 000
Regional rate of:												
Alcohol use (x_1)	0.40				0.40				0.49			
Smoking (x_2)	0.50				0.40				0.36			
Oesophageal cancer* (y)	8.6				8.9				9.2			

* Mortality per 100 000 people per year.
Ecological regressions of y (cancer rate) on
x_1 (alcohol-use prevalence) : $6.8 + 5.0x_1$
x_1 and x_2 (smoking prevalence) : $y = 9.3 + 2.0x_1 - 3.0x_2$.

Table 3. An individual-level analysis of alcohol use and cancer would find a smoking-adjusted rate ratio of 5.0 (whether or not region was controlled; region is not a confounder conditional on smoking). The crude rate ratio from the individual-level analysis is 9.2, a result biased by the strong alcohol-smoking association at the individual level.

In contrast, a simple linear regression of regional cancer rates on alcohol-use prevalence (without controlling smoking) would yield a rate-ratio estimate of $1 + 5.0/6.85 = 1.7$, a result biased by the variation in the alcohol-smoking association across regions. (This variation results in differential confounding by smoking of the alcohol-cancer association across regions; this in turn results in 'region' appearing as a confounder and effect modifier in an individual-level analysis when smoking is ignored.) The same rate-ratio estimate (1.7) is obtained from a simple log-linear regression. Nevertheless, an ecological linear regression of cancer rates on alcohol and smoking prevalence would yield a rate ratio estimate of $[9.3 + 2.0 - 3.0(0.42)]/[9.3 - 3.0(0.42)] = 1.2$ when evaluated at the mean smoking prevalence of 0.42. Again, the same rate-ratio estimate (1.2) is obtained using a multiple log-linear regression. Thus the effect estimate from multiple regression analysis is more biased than the simple regression result, ie ecological control of smoking (the confounder) increased bias.

While it can be helpful, rate standardization will not in general remove all confounding by a covariate in an ecological analysis. Consider the alcohol-use and cancer-mortality rates from Table 3, directly standardized for smoking using the total (combined) population of the regions as the standard:

Region	A	B	C
Alcohol use, x_{1S}	0.174	0.411	0.522
Cancer (per 100 000/yr), y_S	7.58	9.26	10.37

The ecological linear regression of y_S, the standardized cancer rate, on x_{1S}, the standardized alcohol-use prevalence, is $y_S = 6.0 + 8.3x_{1S}$. This yields a rate-ratio estimate of $1 + 8.3/6.0 = 2.4$, better than the crude estimate of 1.7 but still much less than the correct value of 5.0. A log-linear regression of y_S on x_{1S} yields a rate-ratio estimate of 2.6, only a slight improvement. The results are also little changed by using any other reasonable standard.

Rosenbaum and Rubin have shown that if (as is often the case) only the outcome rate is standardized, the result may be to increase bias.[10] For example, if in the preceding illustration one does a linear regression of the standardized cancer rates (y_S above) on the unstandardized alcohol-use rates (x_1 in Table 3), the resulting equation is $y_S = -0.61 + 22.4x_1$. This yields a *negative* (and hence meaningless) rate-ratio estimate of $1 - 22.4/0.61 = -36$. If instead one does a log-linear regression of y_S on x_1, the resulting rate-ratio estimate is 11.8, which is still more biased than the earlier results.

DISCUSSION

In addition to the sources of bias in results of individual-level analyses,[6,7] ecological estimates of effect can be biased from two other sources: confounding by group (the unit of analysis) and effect modification by group. Since it is not possible to identify empirically these sources of ecological bias from aggregate data alone, the investigator must rely on prior knowledge of intergroup variation in the distribution of other risk factors and effect modifiers. Unfortunately, as we have shown, extraneous risk factors responsible for ecological bias may not be confounders or effect modifiers at the individual level. The problem of identifying ecological bias is made even more difficult by the fact that factors responsible for ecological bias may not even *appear* to be confounders or modifiers at the ecological level.[2] Yet in all these cases the amount of ecological bias may

be substantial (even reversing the direction of an observed association), especially when the observed range of average exposure level across groups is small or the exposure under study is not a strong risk factor.[3]

As shown in the Appendix, there will be no ecological bias if both the background (unexposed) rate of disease and the exposure effect do not vary across groups, and there is no confounding within groups. Unfortunately, simultaneous fulfillment of these conditions is rather unlikely, and small departures from them may result in substantial bias. Of course, there may be little or no ecological bias in certain situations in which one or more of these conditions are not met, but such situations would be impossible to identify without individual-level data.

Examples similar to those given here can be constructed using various other magnitudes for the background rates, exposure prevalences, and sizes of exposure effect. Analogous examples can also be constructed using a polytomous or continuous exposure variable. However, the more extreme the range in exposure distribution across regions, the more extreme the effect modification must be in order to produce ecological bias without confounding. In the limit in which each region is either 100% exposed or 100% unexposed, the ecological estimates cannot be distorted (as an estimate of exposure effect) by the phenomena described here. This is easily seen from the fact that in this limit an ecological analysis is equivalent to a non-ecological crude analysis: the average rate of the 100% exposed regions will equal the exposed rate of all regions combined, and the average rate of the remaining (100% unexposed) regions will equal the unexposed rate of all regions combined. Similarly the potential severity of the biases illustrated in Tables 2 and 3 is inversely proportional to the between-group variation in the distributions of exposure, extraneous risk factors, and disease. These observations complement Richardson *et al's*[3] findings that ecological bias due to model misspecification is inversely proportional to the between-group variation in disease rates.

A special problem of ecological studies is that ecological control of a covariate contributing to ecological bias will generally be inadequate to remove the bias produced by the covariate, even in the absence of measurement error. This is in sharp contrast to individual-level analyses, in which detailed control of an accurately measured covariate will remove any confounding by the covariate.[5–7] While we do not wish to discourage the use of ecological studies for exploratory purposes, because of the problems highlighted here we recommend that ecological analyses should be accompanied by thorough consideration of biases unique to such analyses, as well as biases common to all studies.

APPENDIX

Derivation of Results Cited in Main Text

Let the ecological groups (eg regions) be indexed by k, and let p_k, r_{0k}, and r_{1k} denote the exposure prevalence, rate in unexposed, and rate in exposed in group k. The crude rate in group k is then

$$r_{+k} = p_k r_{1k} + q_k r_{0k} = r_{0k} + p_k D_k = r_{0k}(p_k R_k + q_k) \quad (1)$$

where $D_k = r_{1k} - r_{0k}$ and $R_k = r_{1k}/r_{0k}$ are the individual-level rate difference and ratio for group k, and $q_k = 1 - p_k$. In the linear (ecological) regression of r_{+k} on p_k, we have that $r_{+k} = a + bp_k$, $1 + b/a$ is the ecological rate-ratio estimate,[9] and $a + b(1) - [a + b(0)] = b$ is the ecological rate-difference estimate. From standard linear regression theory,[11]

$$\begin{aligned} b &= \mathrm{cov}(p_k, r_{+k})/v \\ &= [\mathrm{cov}(p_k, r_{0k}) + \mathrm{cov}(p_k, p_k D_k)]/v \end{aligned} \quad (2)$$

where $v = \mathrm{var}(p_k)$ and the covariances and variances are over k (ie are based on averages over k, with weighting factors as appropriate). Since $v = 0$ (and hence b is undefined) only if the exposure prevalence does not vary across populations, it will be assumed that $v > 0$.

Let $E(D_k)$ and $E(p_k)$ be the averages (weighted as appropriate) of the D_k and p_k over k. Upon substituting the relation

$$\mathrm{cov}(p_k, p_k D_k) = E(D_k)v + \mathrm{cov}([p_k - E(p_k)]p_k, D_k)$$

into expression 2 and simplifying, b reduces to

$$E(D_k) + \mathrm{cov}(p_k, r_{0k})/v + \mathrm{cov}([p_k - E(p_k)]p_k, D_k)/v \quad (3)$$

The ecological rate difference b thus equals the average individual-level difference $E(D_k)$ plus two bias factors. The second term in expression 3 may be viewed as a bias factor due to confounding by group, for only if r_{0k} varies across k will there be confounding by group at the individual level, and will $\mathrm{cov}(p_k, r_{0k}) > 0$. The third term in the right-hand sum may be viewed as a bias factor due to effect modification, for only if D_k varies across k will $\mathrm{cov}([p_k - E(p_k)]p_k, D_k) > 0$. A similar but approximate partition can be derived for the ecological rate-ratio estimate from a log-linear regression by noting that $\log r_{+k} = \log r_{0k} + \log(p_k R_k + q_k)$.

Even if both bias factors in expression 3 are zero, b may still be a biased measure of effect: note that the D_k appearing in expression 3, although referring to individual-level associations, are unadjusted measures of effect. Thus the presence of confounding within ecological groups on the individual level implies that the D_k will themselves be biased measures. Confounding in the D_k will, in turn, contribute [via $E(D_k)$] to

the bias in the ecological rate difference b, except in the unlikely situation that the biases in the D_k fortuitously cancel upon averaging. If in addition the true differences are uniform, differential confounding will produce spurious heterogeneity of the D_k, and hence cause the third term in expression 3 to be non-zero.

Note also that the D_k, and hence b, may be biased by the covariate even if the covariate distribution is constant across the ecological groups. For example, sex might confound the D_k within every geographical region and thus produce bias in the ecological difference b, even though the sex distribution is the same across regions.[2] Parallel comments apply when considering the rate ratio or any other effect measure.

Even in the absence of confounding, ecological bias is theoretically unbounded, since v can be made arbitrarily small relative to the second covariance in expression 3, and this covariance may be of opposite sign of the true average effect (as in the example in Table 1). In practical terms the bias in b will be constrained by the fixed (observed) value of $\text{var}(p_k)$, and by plausible limits on the variation of effect [eg note that D_k is bounded by r_{+k}/p_k and $(r_{+k}-r_{0k})/p_k$]. As demonstrated by Table 1, however, these constraints may admit enormous bias for quite reasonable ranges of the p_k and D_k.

Our results are indirectly related to those of Firebaugh,[12] who showed that ecological bias in the regression of y on x occurred if and only if the coefficient b_2 in the individual-level regression model $y = a + b_1x + b_2\bar{x} + \epsilon$ is non-zero, where $\bar{x}$ is the group (region) mean of x, and ϵ is an independent error term; $\bar{x}$ corresponds to the p_k in the earlier notation. Firebaugh assumed the validity of the preceding linear model; in our setting this is equivalent to assuming a constant rate difference across region (which, coding x = 1 = exposed and x = 0 = unexposed, would equal b_1). In a similar fashion, others have assumed a no-interaction model when discussing ecological bias,[1,2] although Richardson *et al*[3] did note that non-linearities in the relationship of exposure to risk could lead to bias in ecological estimates of the rate ratio. The individual-level models underlying Tables 1–3 are non-linear; nevertheless, fitting Firebaugh's linear model to Tables 1–3 yields a non-zero coefficient for p_k, so that his criterion is fulfilled by these examples.

Under Firebaugh's linear model, b_2 is non-zero only if $\bar{x}$ is associated with extraneous determinants of y (other than x, which it is necessarily associated with); hence the view of ecological bias as arising from confounding by group has some basis under this model. Nevertheless, as our first example shows, this interpretation breaks down under only moderate departures from linearity.

REFERENCES

[1] Langbein L I, Lichtman A J. *Ecological Inference.* Beverly Hills, CA, Sage, 1978.

[2] Morgenstern H. Uses of ecologic analysis in epidemiologic research. *Am J Public Health* 1982; **72:** 1336–44.

[3] Richardson S, Stucker I, Hemon D. Comparison of relative risks obtained in ecological and individual studies. *Int J Epidemiol* 1987; **16:** 111–20.

[4] Miettinen O S, Cook E F. Confounding: essence and detection. *Am J Epidemiol* 1981; **114:** 593–603.

[5] Greenland S, Robins J M. Identifiability, exchangeability, and epidemiologic confounding. *Int J Epidemiol* 1986; **15:** 412–8.

[6] Rothman K J. *Modern Epidemiology.* Boston, Little, Brown, 1986.

[7] Kleinbaum D G, Kupper L L, Morgenstern H. *Epidemiologic Research: Principles and Quantitative Methods.* Belmont, CA, Lifetime Learning Publs, 1982.

[8] Neutra R R, Greenland S. Continuous intrapartum fetal monitoring. In: Wald N J (ed). *Antenatal and Neonatal Screening.* New York: Oxford, 1984.

[9] Beral V, Chilvers C, Fraser P. On the estimation of relative risk from vital statistical data. *J Epidemiol Comm Health* 1979; **33:** 159–62.

[10] Rosenbaum P R, Rubin D B. Difficulties with regression analyses of age-adjusted rates. *Biometrics* 1984; **40:** 437–43.

[11] Graybill F A. *Theory and Application of the Linear Model.* North Scituate, MA, Duxbury, 1976.

[12] Firebaugh G. A rule for inferring individual-level relationships from aggregate data. *Am Sociol Rev* 1978; **43:** 557–72.

(Revised version received December 1987)