EPIDEMIOLOGY IN A CHANGING WORLD:

VARIATION, CAUSATION AND UBIQUITOUS RISK FACTORS

Neil Pearce[1,2]

[1] Centre for Public Health Research
Massey University Wellington Campus
Private Box 756
Wellington, New Zealand

[2] Department of Medical Statistics
Faculty of Epidemiology and Population Health
London School of Hygiene and Tropical Medicine
Keppel Street
London WC1E 7HT
United Kingdom

August 2010
(Revised September 2010)
(Revised December 2010)

Address for correspondence:
Professor Neil Pearce
Director, Centre for Public Health Research
Massey University Wellington Campus
Private Box 756
Wellington
Phone: 64-4-380-0606
Fax: 64-4-380-0600
E-mail: n.e.pearce@massey.ac.nz
Website: http://publichealth.massey.ac.nz/

Word count: 4651

Medical Subject Headings: Epidemiology, Methods, Biostatistics, Non-communicable Disease

ABSTRACT

We are all living in the era of globalization and, like it or not, it is going to change the way we practice epidemiology, the kinds of questions we ask, and the methods we use to answer them. However, the methods, and ways of thinking about the health of populations, that will be required for epidemiology in the 21st century are in some instances quite different from the standard epidemiological techniques that are taught in most textbooks and courses today. As we develop epidemiological methods for addressing the scientific and public health problems of the 21st century, it is important that we consider, once again, the distinction between the analysis of variance and the analysis of causes. This has primarily been considered with respect to genetic research, and also with regards to the problems of making comparisons between different populations and environments at the same point in time. It has not been considered in depth with regards to the issues of conducting epidemiologic research in a world that is changing over time. In this paper, I first consider the statistical and scientific issues involved in the distinction between the analysis of variance and the analysis of causes. I then discuss some examples of the implications of this distinction for the theory and practice of epidemiology in a changing world, particularly with regards to risk factors that become ubiquitous over time. Sometimes the most important causes of disease are invisible because they are everywhere.

INTRODUCTION

We are all living in the era of globalization and, like it or not, it is going to change the way we practice epidemiology, the kinds of questions we ask, and the methods we use to answer them[1]. The methods that will be required for epidemiology in the 21st century are likely to be quite different from the standard epidemiological techniques that are taught in most textbooks and courses today[2 3]. As we develop epidemiological methods for addressing the scientific and public health problems of the 21st century, it is important that we consider, once again, the distinction between the analysis of variance and the analysis of causes. This issue, which is important in statistical, aetiological, and public health terms, has received further attention and debate in the *International Journal of Epidemiology*[4-8] as a result of the reprinting of Lewontin's classic 1974 paper[9] (Int J Epidemiol 2006; 35: 520-525). It has primarily been considered with respect to genetic research, although it was also considered indirectly with regards to non-genetic causes of chronic disease in Rose's seminal 1985 paper[10], which was also reprinted in the *International Journal of Epidemiology*[11-15] (Int J Epidemiol 2001; 30: 427-432). These two papers explore the same phenomena from different starting points and perspectives. In both instances, they clearly demonstrate that the major determinants of population variation in a disease, may not be the main causes of the disease. Nevertheless, although this distinction has been explored to some extent with regards to the problems of making comparisons between different populations and environments at the same point in time, it has not been considered in depth with regards to the issues of conducting epidemiologic research into non-genetic factors, or gene-environment interactions, in a world that is changing over time. In particular, if the environment (global, local, social, physical) is changing over time, then the relative population importance of various risk factors may change in a non-intuitive manner, and a risk factor may become relatively more important as a cause of disease, and have an increased population attributable risk, while explaining less of the population variation in the disease.

In this paper, I therefore first briefly consider the statistical and scientific issues involved in the distinction between the analysis of variance and the analysis of causes. I then discuss some specific examples where this distinction is important. Finally, I discuss the implications of this distinction for the theory and practice of epidemiology in a changing world.

THE ANALYSIS OF VARIANCE AND THE ANALYSIS OF CAUSES

The assessment of the relative importance of different risk factors for disease, particularly with regards to genetic and environmental risk factors (nature vs nurture), has been an ongoing source of debate. In his classic 1974 paper[9], Lewontin showed that estimates of the percentage of population variation that is explained by heritability, are not generalizable, since they depend on the distribution of genes and environmental factors in a particular environment. This is one example of a more general phenomenon, in that the percentage of population variation explained is not a valid or generalizable effect measure[16]. Greenland et al[17 18] have examined this issue in depth, and give a number of examples of hypothetical studies in which the relative risk for an exposure-disease association is the same in two populations, but the percentage of population variation explained by the exposure (the square of the correlation coefficient) is different.

To consider An example that Rose previously used with respect to geographical populations[19], table 1 shows data on smoking and lung cancer in a hypothetical population in two different time periods. In time period 1, 50% are smokers, whereas in time period 2, 90% are smokers. The incidence of lung cancer in smokers (2,000 per 10,000), the incidence of lung cancer in non-smokers (200 per 10,000), and the relative risks (10.0) are identical in the two time periods. However, the proportion of population variation in lung cancer explained is 8.4% in time period 1 and only 2.0% in time period 2,;in contrast, smoking accounts for 82% of lung cancer cases in period 1 and 89% in period 2.

These differences occur because of the differences in smoking prevalence in the two time periods, which produce the non-intuitive finding that the proportion of population variation in lung cancer explained by smoking is much lower in the time period in which more people smoke, and in which smoking accounts for a higher proportion of lung cancer cases. More generally, in a time period in which there is a wide range in smoking frequency (ranging from zero to several packs a day) smoking would explain a high proportion of the variation in lung cancer incidence rates. In another time period with little variation in smoking frequency (e.g. if almost everyone smoked a pack a day), the relative risk for smoking and lung cancer would still be the same (i.e. 10 times), but the proportion of population variation explained would be very low.

This problem of non-generalizability of the proportion of variation explained also applies when the exposure and/or outcome variables are continuous, even if they are jointly normally distributed[18]. The problems of non-generalizability also apply to correlation coefficients and standardized regression coefficients[16]. In contrast, measures such as the incidence rate ratio (for a dichotomous exposure and outcome) or the regression coefficient(s) (for a continuous exposure and outcome) are in principle valid and generalizable effect measures[16]; . These statements should be qualified by noting that all effect measures, including the relative risk, are only generalizable provided that we are prepared to assume that there is no confounding or other bias, and that the distribution of effect modifiers is the same in the two populations or time periods being compared; however, the percentage of population variation explained has the additional problem that it is not generalizable if the prevalence of exposure is different in the two time periods. A further implication is that comparing proportions of population variation explained, for two different risk factors for disease, tells us nothing about which risk factor is the more important cause of disease, in terms of relative risk, absolute risk, or population attributable risk.

A related issue is that it is meaningless to attempt to partition variation into the percentage explained by various risk factors. In particular, it is invalid to partition the population variation of a disease into the percentage explained by genetics (nature) and the percentage explained by the environment (nurture) because their effects are non-additive (in fact, their joint effect is more than additive):

> "For example, if two men lay bricks to build a wall, we may quite fairly measure their contribution by counting the number laid by each; but if one mixes the mortar and the other lays the bricks, it would be absurd to measure their relative quantitative contributions by measuring the volume of bricks and of mortar. It is obviously even more absurd to … ascribe so many inches of a man's height to his genes and so many to his environment." [9]

EXAMPLE: PHENYLKETONURIA (PKU)

These issues have been discussed in the epidemiological context in Rothman's theory of sufficient and component causes, using the example of phenylketonuria (PKU)[20]. Although Rothman's approach is simplistic (in particular, it assumes that the exposures and outcome are all dichotomous), it can be used to illustrate some basic concepts about causation, and to distinguish it from variation. Briefly, PKU is a classic "genetic" disease which can be detected with a simple blood test at birth. PKU occurs due to a combination of the presence of the PKU gene, and a diet high in phenylalanine. Each of these factors (the PKU gene and a high-phenylalanine diet) are necessary but not sufficient causes of the disease, but together they are sufficient causes, i.e. the causal constellation which involves both of these 'exposures' is a sufficient causal constellation.

We live on a planet where almost everyone has a high phenylalanine diet, whereas only a small proportion of people have the PKU gene (figure 2). Thus, the disease appears to be almost

entirely genetic, since almost everyone who has the gene develops the disease. Thus, virtually 100% of the population variation is explained by genetics, and almost none of it is explained by diet. Nevertheless, changing the gene is currently impractical, and the solution (once the problem has been identified) is to change the diet for the first few years of life. Thus, PKU is regarded as a classically genetic disease, but the intervention is environmental. This is because PKU is *caused* by the *joint effect* of the gene and the high-phenylalanine diet. In fact, 100% of cases of PKU are caused by the gene (and essentially its heritability is 100%), since 100% of cases could in theory be prevented by eliminating the gene from the population. However, 100% of cases can also be prevented by reducing phenylalanine in the diet. Thus, the *variation* is almost 100% genetic and almost 0% environmental (dietary) - these two figures, by definition, sum to 100%. However, the *causation* is 100% genetic and 100% environmental, and these two figures necessarily sum to more than 100%.

In contrast, figure 3 shows a hypothetical planet in which everyone has the PKU gene, but there is considerable variation in dietary phenylalamine intake. On this planet, any infant with a high phenylalamine diet would develop PKU, whereas infants with a low phenylalamine diet would not; thus, PKU would appear to be a nutritional disease with no genetic component (its heritability would be zero). Now, the *variation* would be almost 100% environmental and almost 0% genetic. Nevertheless, the *causation* would still be 100% genetic and 100% environmental.

On the other hand, figure 4 shows a planet in which there is variation in both the gene and the diet. Now, the *variation* would be about 50% environmental and about 50% genetic. Nevertheless, the *causation* would still be 100% genetic and 100% environmental.

These three examples, one of which approximates the real situation on planet earth and two of which are hypothetical, illustrate that the amount of population variation due to a particular factor can vary wildly between populations (or planets), even though it is an equally important cause in each situation. Once again, the proportion of population variation explained is not generalizable, and is not a valid effect measure[16]. In fact, almost every disease is 100% genetic and 100% environmental, but the proportion of population variation due to genetics or environment (which necessarily add up to 100%) will vary widely between populations.

We live on a planet which is changing over time, and in which genetic factors are changing slowly, whereas environmental changes can be more rapid. This can mean that we can change from one situation (e.g. figure 4) to another (e.g. figure 2) fairly rapidly, and this can produce dramatic non-intuitive changes in the apparent relative importance of genetic and environmental risk factors.

EXAMPLE: GENES, EXERCISE, ENERGY INTAKE AND OBESITY

I will now consider a more complex example, which has multiple risk factors, both genetic and non-genetic, which are changing over time. Much has been written about the causes of the apparent global epidemic of obesity[21], which seems to even be occurring in our pets[22], with considerable dispute as to whether there is[23], or is not[24], an epidemic, and what the major causes (if any) may be[25]. I do not intend to replicate these debates here. Rather, I will use a hypothetical example, loosely based on the current debates about obesity, to illustrate the potential difficulties of determining the major causes of a disease/condition in a world which is changing over time.

Briefly, and very simplistically, people become obese because they consume more energy (in the form of food/drink) than they expend (in the form of exercise)[25]. This process is mediated by individual metabolism which is strongly affected by genetics. There has been considerable debate about the relative importance of these factors. Clearly, genetic factors play a role, but they cannot alone account for

the dramatic increases in obesity in recent decades because these have occurred too rapidly to be due to genetic changes. Thus, the *changes* in the prevalence of obesity over time must primarily be due to environmental changes. There has been considerable debate as to whether changes in diet or exercise are more important in this regard. Clearly they are both important and they interact with each other (a decrease in energy intake may be offset by a decrease in energy expenditure). One argument, among many, has been that energy intake has increased relatively little, whereas energy expenditure has decreased markedly because of changes in the urban environment; these include urban design, safety concerns (which discourage walking), the rise of the car, and the near demise of public transport, as well as the many labour saving devices that now permeate daily life[26]. We have (largely unintentionally) produced a society in which exercise is no longer an integral part of daily life, either at work[27] or outside of work[26]. Now, exercise is something that we choose to do by going to the gym (figure 5), or by engaging in other physical activity. Obviously, whether we choose to do this is strongly affected by a large number of societal and individual factors including the availability of facilities, income, having the necessary time, and having the ability and motivation to exercise in this way[26].

Figure 6 (simplistically) illustrates the former situation in which exercise was, for most people, an integral part of daily life, and there was a wide variation in exercise levels in the population. If we assume simplistically that having (unknown) "obesity genes" and "not exercising" are each necessary causes of obesity, and that their combination is a sufficient cause, then obesity was 100% genetic and 100% environmental, but the population variation was (hypothetically) about 50% for each.

Figure 7 (once again very simplistically) illustrates the situation that we have moved towards, in which there is much less population variation in exercise levels. In this scenario, the main change has been in exercise levels, but *because* of this change, exercise explains less of the population variation, and a greater percentage than previously is explained by genetics. The ultimate example of this hypothetical scenario is twin studies. Twins are perfectly matched on birth cohort, so experience the same changes in environment over time. It follows that environmental factors account for little of the variation in disease risk between twins, whereas genetic factors account for about two-thirds of the variation[28 29]. Thus twin studies show very high heritability for obesity, even though time trend studies show that environmental factors are of overwhelming importance[28]. If everyone lives in the same obesogenic environment, then it is individual genetic susceptibility that determines which individuals become obese and which don't. Thus, factors at the individual level (i.e. genetics) appear to be more important than factors at the environmental level since there is so little variation at the environmental level.

Figures 8 and 9 illustrate similar issues with regards to different individual-level risk factors, namely exercise and energy intake. In this scenario, we have moved from a situation where there was substantial variation in both factors (figure 8) to one in which exercise levels have reduced, and there is correspondingly less population variation. In both situations, obesity occurs because of an imbalance between energy intake and energy expenditure (in genetically susceptible people). Thus, each factor is necessary, but not sufficient, for obesity to occur, e.g. lack of exercise is necessary, but is not sufficient, because it may be countered by low energy intake. It is the combination of the two factors that results in obesity. Thus, in both situations obesity is 100% attributable to low exercise and 100% attributable to high energy intake. However, the reductions in exercise over time, which under this scenario are the main reason for the obesity epidemic, result in the paradoxical situation where exercise accounts for a smaller proportion of the population variation, and energy intake appears to have become relatively more important. It is therefore crucial that cross-sectional data on population variation in disease are interpreted carefully and appropriately in order to avoid incorrect policy decisions. In particular, if changes over time are ignored, and the environment at a particular point in time is regarded as 'fixed', then macro-level measures such as the provision of public transport, safe cities, etc[26] will not be

considered, and there will be an inappropriate focus on changing the behaviour of individuals[2], rather than a more balanced approach which involves interventions both at the population and individual levels.

DISCUSSION: VARIATION, CAUSATION AND UBIQUITOUS RISK FACTORS

Are these scenarios realistic? Of course this is debatable, but they do offer simplistic hypothetical examples of how changes in a risk factor over time (e.g. exercise) can cause an epidemic of disease, but can also paradoxically make the risk factor appear less important, in terms of the population *variation*, even though it remains of crucial importance, in terms of *causation*. Rose[10] lists other similar examples of ubiquitous risk factors, including softness of the public water supply in Scotland and cardiovascular disease rates, dietary fat and coronary heart disease, and diet and blood pressure and overweight; further examples are added by Khaw and Marmot in the recently published new edition of Rose's book[19]. Rose, Khaw and Marmot[19] thus note that entire populations may be exposed to a particular risk factor and there is usually a continuum of disease risk (rather than a clear distinction between the "sick" and the "healthy") across the population. Small improvements in the health of a "sick population" may be more effective than attempts to treat or prevent illness in "sick individuals"[10 19]. However, before such public health measures are adopted, the important near-ubiquitous risk factors must first be identified.

In this context, a distinction should be drawn between Rose's[10] formulation of this problem, and that presented here. Rose[10] argues that when a risk factor is ubiquitous in a population, it may strongly influence the population incidence of a disease, but may not identify high-risk individuals within a population. This is technically correct, e.g. in a society where everyone smokes, smoking will not identify high risk individuals for lung cancer. However, Rose goes further than this to argue that "the determinants of individual cases, and the determinants of incidence rate" may be different[10]. He thus moves from an argument about methods of *identifying* high risk cases, to an argument about the *causes* of high risk cases. This has been quoted by most commentators on Rose's paper[11 12 14 15] to argue that the causes of incidence are not necessarily the same as the causes of cases. Some commentators[14] have taken this distinction further and also argued that the causes of incidence are necessarily "upstream" in Rose's formulation whereas the causes of cases are more "proximal". Although the distinction between "upstream" and "downstream", "proximal" or "distal" causes or various levels of causation is important[10 30-32], it is not essential to the issues discussed in this paper. In fact, the examples discussed in this paper, mostly involving relative "proximal" causes, illustrate that the causes of incidence *are* also the causes of individual cases. For example, in a population where everyone smokes, smoking is a cause both of the population incidence of lung cancer and of individual cases. The important distinction is not between the causes of incidence and the causes of cases (all of the ubiquitous risk factors considered here are causes of individual cases); rather it is between the causes of incidence and the causes of population variation. In particular, a near-ubiquitous risk factor (e.g. smoking, high cholesterol) can be an important cause of incidence, *and* of individual cases, but may explain little of the population variation and may not be a useful means of identifying high risk individuals within a particular population (this is essentially the point that Rose was making, but the distinction was perhaps less clear than it could have been due to his use of the word "determinant" in the latter context).

So can the role of near-ubiquitous risk factors be studied and their importance for disease causation quantified? Yes they can, but it's not easy, and they can't always be studied with our "standard" methods such as randomized controlled trials, cohort studies or case-control studies.

Firstly, it may be possible to identify important risk factors by comparisons between populations or comparisons over time. For example, there are substantial variations in levels of exercise even between various "Western" countries, with the lowest levels being observed in the USA, and the highest levels in countries such as The Netherlands[26]. Many of the recent discoveries on the causes of cancer (including dietary factors and colon cancer, HBV and liver cancer, aflatoxins and liver cancer, human papilloma virus and cervical cancer) have their origins, directly or indirectly, in the systematic international comparisons of cancer incidence conducted in the 1950s and 1960s[33]. These suggested hypotheses concerning the possible causes of the international patterns, which were investigated in more depth in further studies. In some instances these hypotheses were consistent with biological knowledge at the time, but in other instances they were new and striking, and might not have been proposed, or investigated further, if the population level analyses had not been done.

More recently, a huge amount of funding has been spent on studying the "known" causes of asthma in affluent countries (e.g. air pollution, allergen exposure), and it is only now that standardized studies are revealing major international differences in asthma prevalence that are not explained by these "established" risk factors[34][35]. This has cast doubt on the "orthodox" theory in which allergen exposure in infancy produces atopic sensitization and that continued exposure results in asthma. The striking time trends in asthma prevalence also cast doubt in claims that asthma is a "genetic disease" since it is strongly heritable[36]. Instead, attention is increasingly focusing on the possibility that other factors in utero and in infancy may "programme" the immune system in a manner which increases the risk of developing asthma and allergy[36][37]. Thus, the increases in asthma prevalence may be unrelated to allergen exposure, but rather may reflect changes in immune function as a result of the "cleaner" environment that occurs with Westernization. This evidence from epidemiological studies is now supported by clinical and immunological studies, and is producing a major shift in etiologic thinking and new research initiatives at the population, individual and micro levels. These developments may have occurred anyway, but have been greatly enhanced by the epidemiological evidence from international comparisons[38].

Thus, comparisons between populations, and comparisons over time, have played a major role in generating hypotheses and identifying important population risk factors for disease. They play a particularly important role when important risk factors are ubiquitous, or nearly ubiquitous, in particular countries, or in particular time periods. Of course, such studies are fraught with difficulties because of problems of "the ecologic fallacy"[39]. Thus, ecologic studies provide virtually no evidence as to causation. On the other hand, they may be the best way (or the only way) to identify the major population-level determinants of disease, which can then be studied in depth using other study designs. Thus, while ecologic studies are not sufficient in themselves for identifying risk factors or establishing causation, they are an essential part of a wider scientific process[40]. In this context, it is important to emphasize that the appropriateness of any research methodology depends on the phenomenon under study: its magnitude, the setting, the current state of theory and knowledge, the availability of valid measurement tools, and the proposed uses of the information to be gathered. The appropriateness of a research method in epidemiology is determined by the nature of the problem under consideration, the community resources and skills available and the prevailing norms and values at the national, regional or local level[41]. If we want to discover (or at least generate hypotheses about) the major population-level determinants of disease, then ecologic studies will play an essential role in those process, despite their methodological limitations.

Secondly, once a particular hypothesis has been proposed, studies can be conducted within particular countries, even if there is little variation in exposure to a particular risk factor. For example, suppose we have two countries, in one of which 50% of the population are smokers, and in the other 95% are smokers. The percentage of population variation explained will be different in the two countries, and will in fact be greater in the country in which only 50% smoke.

However, the incidence rate ratio comparing smokers and non-smokers, and the dose-response curve, will be about the same in the two countries (i.e. about a 10-fold risk for smoking a pack a day for many years). Thus, it is still possible in principle to identify the major risk factors for disease, provided that there is at least some population variation. However, the problem is that too often the major population risk factors will not be hypothesized (and therefore will not be studied) when they are ubiquitous. In particular, lack of exercise has become such a pervasive aspect of daily life that it requires a historical perspective to realize how much daily life has changed, and that what appears 'normal' now would have been far from normal 50 years ago[26].

So it is possible to study near-ubiquitous risk factors, some of which may be major causes of disease at the population level. However, as illustrated in this paper, a factor that discourages the investigation of near-ubiquitous risk factors is the continued misuse of measures such as the percentage of population variation explained to identify the 'major' risk factors for disease, genetic or non-genetic. As Stephen Rose notes[6] "the practical relevance of claiming that some character in the environment is 80% heritable provides no guidance for how to respond – except in a purely ideological way". However "in the years since 1974 many researchers have not heeded Lewontin's suggestion to 'stop the endless search for better methods of estimating useless quantities'… heritability estimates have continued to fuel policy and popular debates"[7]. In fact, Lewontin argues that to some extent things have got worse with the revolution in studies at the level of DNA and RNA which have provided 'a powerful reinforcement of the erroneous notion that variation in phenotype is entirely the consequence of genetic variation… If the ANOVA did nothing else it created a mindset that was much closer to the truth than the naïve current prejudice that DNA has in it all the information necessary to specify the organism."[5] Unfortunately, confusion between the analysis of variance and the analysis of causes is not unique to genetic research, and has also influenced other areas of epidemiology. These issues are likely to become, if anything, more important in the 21st century, as we increasingly see global changes in the environment, with major effects on human health[42 43].

If we are to move beyond being "prisoners of the proximate"[32], and to move "upstream"[31 41] to address the major population-determinants of health, this requires us to move beyond studying variation to studying causation. The most important population-level determinants of disease may vary little within a particular population or time period, and therefore may not be easily identifiable with our "standard" epidemiological methods. "Inferior" methods such as ecologic studies may play an important role in the cycle of hypothesis generation and testing[40] and more complex methods which do not fit neatly into the randomized controlled trial paradigm may also be valuable[44]. Thus, it is unscientific to claim that some study designs are inherently better than others, as if any single study could produce "the answer" to any scientific or public health question. It's important that, instead, epidemiological research is viewed as a process, in which all of the different study designs may play a role, depending on the scientific and public health questions under study. In turn, it is important that this research process takes into account the changes in populations over time, which may result in important risk factors becoming ubiquitous. Sometimes the most important causes of disease are invisible because they are everywhere.

Acknowledgements

REFERENCES

1. Pearce N. The globalization of epidemiology. *International Journal of Epidemiology.* 2004;33:1127-31.
2. Pearce N. Traditional epidemiology, modern epidemiology, and public health.[comment]. *American Journal of Public Health.* 1996;86:678-83.
3. Pearce N. Epidemiology as a population science. *International Journal of Epidemiology.* 1999;28:S1015-8.
4. Freese J. Commentary: The analysis of variance and the social complexities of genetic causation. *International Journal of Epidemiology* 2006;35:534-536.
5. Lewontin RC. Commentary: Statistical analysis or biological analysis as tools for understanding biological causes. *International Journal of Epidemiology* 2006;35:536-537.
6. Rose SPR. Commentary: Heritability estimates - long past their sell-by date. *International Journal of Epidemiology* 2006;35:525-527.
7. Taylor P. Commentary: The analysis of variance is an analysis of causes (of a very circumscribed kind). *International Journal of Epidemiology* 2006;35:527-531.
8. Vreeke GJ. Commentary: The attainability of causal knowledge of genetic effects in complex human traits. *International Journal of Epidemiology* 2006;35:531-534.
9. Lewontin R. The analysis of variance and the analysis of causes. *American Journal of Human Genetics* 1974;26:400-411.
10. Rose G. Sick individuals and sick populations. *International Journal of Epidemiology* 1985;14:32-38.
11. Ebrahim S, Lau E. Commentary: Sick populations and sick individuals. *International Journal of Epidemiology* 2001;30:433-434.
12. Hunt K, Emslie C. Commentary: The prevention paradox in lay epidemiology - Rose revisited. *International Journal of Epidemiology* 2001;30:442-446.
13. McCormick J. Commentary: Reflections on sick individuals and sick populations. *International Journal of Epidemiology* 2001;30:434-435.
14. Schwartz S, Diez-Roux R. Commentary: Causes of incidence and causes of cases - a Durkheimian perspective on Rose. *International Journal of Epidemiology* 2001;30:435-439.
15. Weed DL. Commentary: A radical future for public health. *International Journal of Epidemiology* 2001;30:440-441.
16. Rothman KJ, Greenland S, Lash TL. *Modern epidemiology.* 3rd ed. Philadelphia: Lippincott-Williams, 2008.
17. Greenland S, Maclure M, Schlesselman JJ, Poole C, Morgenstern H. Standardized Regression-Coefficients - a Further Critique and Review of Some Alternatives. *Epidemiology* 1991;2:387-392.
18. Greenland S, Schlesselman JJ, Criqui MH. The Fallacy of Employing Standardized Regression-Coefficients and Correlations as Measures of Effect. *American Journal of Epidemiology* 1986;123:203-208.
19. Rose G, Khaw K-T, Marmot M. *Rose's strategy of preventative medicine.* Oxford: Oxford University Press, 2008.
20. Rothman KJ. Causes. *American Journal of Epidemiology* 1976;104:587-592.
21. Ebrahim S. Obesity, fat, and public health. *International Journal of Epidemiology* 2006;35:1-2.
22. Nijland ML, Stam F, Seidell JC. Overweight in dogs, but not in cats, is related to overweight in their owners. *Public Health Nutrition* 2010;13:102-106.
23. Kim S, Popkin BM. Commentary: Understanding the epidemiology of overweight and obesity - a real global public health concern. *International Journal of Epidemiology* 2006;35:60-67.
24. Campos P, Saguy A, Ernsberger P, Oliver E, Gaesser G. Response: lifestyle not weight should be the primary target. *International Journal of Epidemiology* 2006;35:81-82.

25. Cummins S, Macintyre S. Food environments and obesity - neighbourhood or nation? *International Journal of Epidemiology* 2006;35:100-104.

26. Frumkin H, Frank L, Jackson R. *Urban sprawl and public health: designing, planning and building for healthy communities*. London: Island Press, 2004.

27. Bain C. Commentary: What's past is prologue. *International Journal of Epidemiology* 2006;35:16-17.

28. Davey Smith G, Ebrahim S. Epidemiology - is it time to call it a day? *International Journal of Epidemiology* 2001;301:1-11.

29. Maes HHM, Neale MC, Eaves LJ. Genetic and environmental factors in relative body weight and human adiposity. *Behavior Genetics* 1997;27:325-351.

30. Krieger N. Proximal, distal, and the politics of causation: What's level got to do with it? *American Journal of Public Health* 2008;98:221-230.

31. McKinlay JB. The promotion of health through planned sociopolitical change: challenges for research and policy. *Soc Sci Med* 1993;36:109-17.

32. McMichael AJ. Prisoners of the proximate: loosening the constraints on epidemiology in an age of change. *Am J Epidemiol* 1999;149:887-97.

33. Doll R, Payne P, Waterhouse J. Cancer Incidence in Five Continents: A Technical Report. Berlin: Springer-Verlag (for UICC), 1966.

34. Asher MI, Keil U, Anderson HR, Beasley R, Crane J, Martinez F, et al. International Study of Asthma and Allergies in Childhood (ISAAC): rationale and methods. *European Respiratory Journal.* 1995;8:483-91.

35. Burney PGJ, Luczynska C, Chinn S, Jarvis D. The European Community Respiratory Health Survey. *European Respiratory Journal* 1994;7:954-960.

36. Douwes J, Pearce N. Asthma and the westernization 'package'. *International Journal of Epidemiology.* 2002;31:1098-102.

37. Douwes J, Pearce N. The end of the hygiene hypothesis? *International Journal of Epidemiology* 2008;37:570-572.

38. Enarson D. Fostering a spirit of critical thinking: the ISAAC story. *International Journal of Tuberculosis and Lung Disease* 2005;9:1.

39. Morgenstern H. Ecologic studies. In: Rothman KJ, Greenland S, Lash TL, editors. *Modern epidemiology. 3rd ed*. Philadelphia: Lippincott Williams & Wilkins, 2008.

40. Pearce N. The ecological fallacy strikes back.[comment]. *Journal of Epidemiology & Community Health.* 2000;54:326-7.

41. Pearce N, McKinlay JB. Dissent: Back to the future in epidemiology and public health: response to Dr. Gori.[comment]. *Journal of Clinical Epidemiology.* 1998;51:643-6; discussion 647-9.

42. McMichael A. *Planetary overload: global environmental change and the health of the human species*. Cambridge: Cambridge University Press, 1993.

43. McMichael T. *Human frontiers, environments and disease*. Cambridge: Cambridge University Press, 2001.

44. Pearce N, Merletti F. Complexity, simplicity and epidemiology. *International Journal of Epidemiology* 2006;35:515-519.

Table 1: Hypothetical example of the effect(s) of smoking on lung cancer in two different time periods

		Time period 1			Time period 2		
		Smoking			Smoking		
		Yes	No		Yes	No	
Lung	Yes	1000	100		1800	20	
Cancer	No	4000	4900		7200	980	
	Total	5000	5000		9000	1000	
Incidence rate (per 10,000)		2000	200		2000	200	
Risk ratio			10.0			10.0	
Correlation			0.29			0.14	
% variance explained			8.4%			2.0%	
Population Attributable Fraction (PAF)			82%			89%	

Figure 1: Measures of variation and measures of causation for smoking and lung cancer, by prevalence of smoking
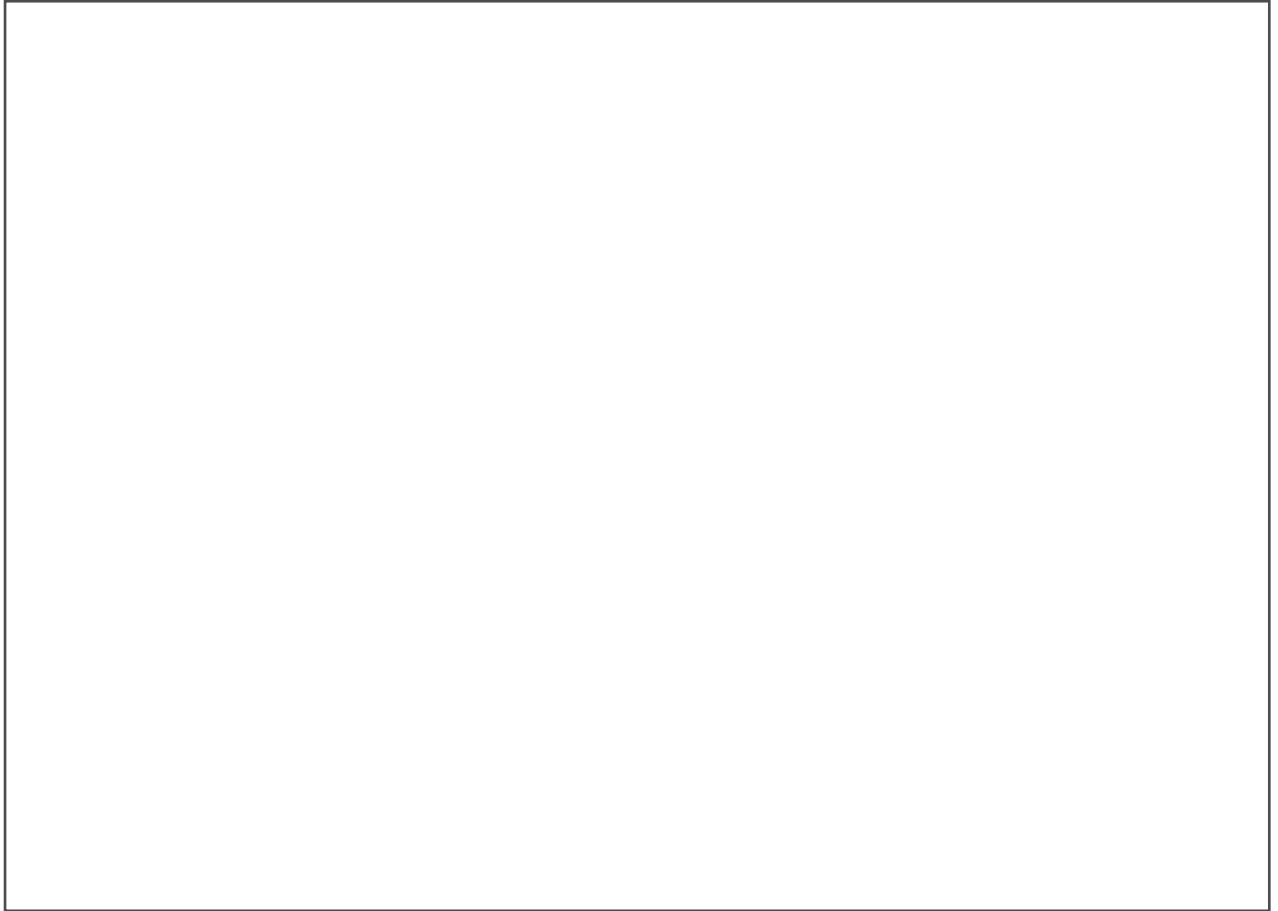
Figure 2: Causation and variation in phenylketonuria (PKU) on planet earth: the roles of genes and diet

NB: The shaded area is where most people "live"; the upper right hand internal box is the area in which people get PKU

Figure 3: Causation and variation in phenylketonuria (PKU) on a planet far away: the roles of genes and diet

NB: The shaded area is where most people "live"; the upper right hand internal box is the area in which people get PKU

Figure 4: Causation and variation in phenylketonuria (PKU) on another planet far away: the roles of genes and diet

NB: The shaded area is where most people "live"; the upper right hand internal box is the area in which people get PKU

Figure 5: Exercise has become a matter of choice rather than a part of daily life
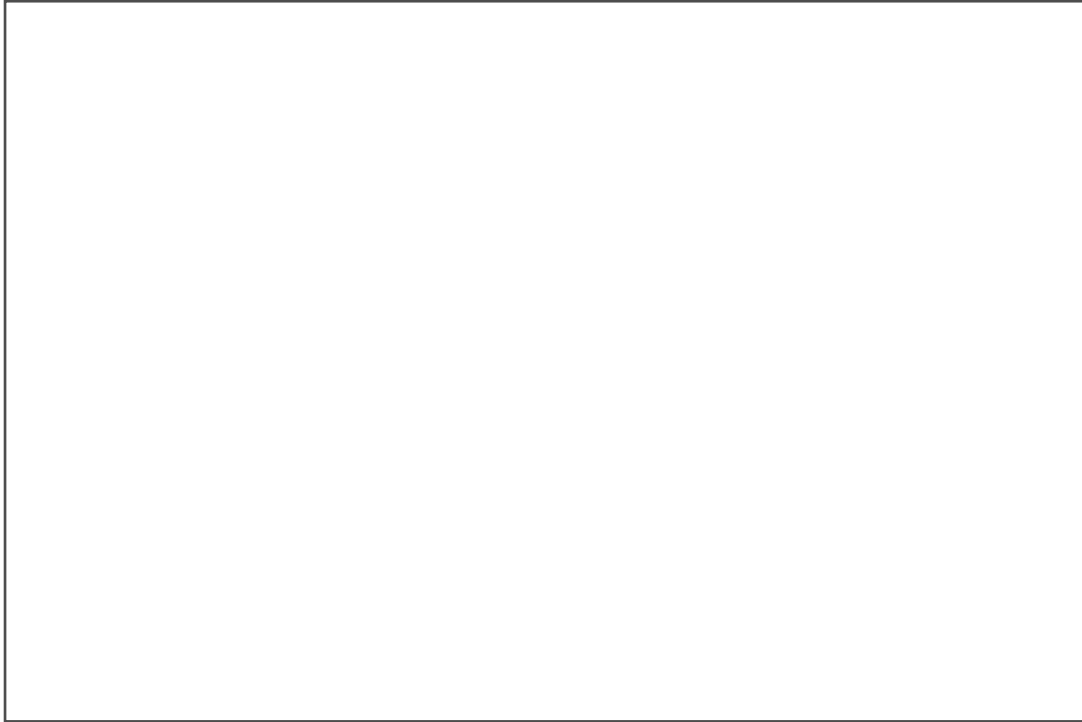
Figure 6: Causation and variance in obesity in the *past*: the roles of genes and exercise

NB: The shaded area is where most people "live"; the upper right hand internal box is the area in which people develop obesity

Figure 7: Causation and variance in obesity in the *present*: the roles of genes and exercise

NB: The shaded area is where most people "live"; the upper right hand internal box is the area in which people develop obesity

Figure 8: Causation and variance in obesity in the *past*: the roles of energy intake and exercise

NB: The shaded area is where most people "live"; the upper right hand internal box is the area in which people develop obesity

Figure 9: Causation and variance in obesity in the *present*: the roles of energy intake and exercise

NB: The shaded area is where most people "live"; the upper right hand internal box is the area in which people develop obesity