



COMMENTARY

On Genes, Individuals, Society, and Epidemiology

Ana V. Diez-Roux

Over their long history, and depending on the theory of disease causation prevalent at the time, the fields of epidemiology and public health have focused on different factors as potential causes of disease (1–3). In the late 19th century, miasmatic theories of disease causation, in which ill health resulted from foul emanations of the soil, air, and water, gave way to the germ theory and the doctrine of specific etiology. Subsequently, with the growing importance of chronic diseases, emphasis shifted to the characteristics of individuals, behaviors, and lifestyles.

Today, epidemiology may be at the brink of a new paradigm, the genetic paradigm. The advent of new technologies, and the accompanying interest in identifying genes (and creating genetic screening tests) for particular diseases, has led to explosive growth in research on the genetics of disease and its molecular mechanisms (4–8). This trend has been accompanied by the publication of several textbooks on genetic epidemiology (9–11), and genes have recently been put forward as important factors in the genesis not only of several common diseases but also of behaviors such as physical activity (12) or psychosocial characteristics such as social support (13). Simultaneously with the interest in genes as causes of disease, there appears to have been a resurgence of research and interest in the social origins of disease—as evidenced, for example, by the increase in publications examining

social class differences in health (14), and by several recent commentaries on the role and future of epidemiology which have emphasized the importance of social factors in studying and understanding the distribution of diseases (3, 15–17).

The coexistence of alternate explanations on the origins of disease is, of course, not new. Throughout the history of epidemiology and public health, there has been a steady current of social medicine and social epidemiology intent on investigating and emphasizing the links between how societies are organized and the patterns of death and disease in their populations (18–27). These “social” explanations have often competed with alternate explanations emphasizing biologic and behavioral factors; but perhaps today, the contrast between levels of explanation as diverse as genetic and social has become more evident than ever. The fragmentation of the field is further evidenced by the emergence of different “types” of epidemiology: There is “social” epidemiology, “risk factor” epidemiology, and “genetic” epidemiology, each with its own literature.

As in other scientific fields, research in public health and epidemiology has been strongly influenced by the notion that the whole can be understood by breaking it down and understanding its component parts. This strategy, known as “reductionism,” has been defined as the attempt to “explain the properties of complex wholes—molecules, say, or societies—in terms of the units of which those molecules or societies are composed” (28, p. 5). The idea that understanding the pieces allows us to understand the whole has indeed led to many discoveries in a number of scientific fields (29). In public health and epidemiology, we often apply this type of strategy. For example, in attempting

Received for publication March 31, 1998, and accepted for publication August 19, 1998.

From the Division of General Medicine, College of Physicians and Surgeons, and the Division of Epidemiology, School of Public Health, Columbia University, New York, NY.

Reprint requests to Dr. Ana V. Diez-Roux, Columbia College of Physicians and Surgeons, 622 West 168th Street, PH 9 East, Room 105, New York, NY 10032.

to understand disease in populations, we break populations down into "independent" individuals. In attempting to understand disease in individuals, we tease apart the "independent" contributions of different factors. This approach has yielded much useful epidemiologic information. To cite just two examples, it has contributed to the identification of aspects of lifestyle and biologic factors associated with cardiovascular disease, the leading cause of death in many of today's societies, and has shown that smoking (an eminently preventable behavior) causes cancer. Today, similar strategies help epidemiologists identify the "independent" contributions of genes to disease. Breaking things down is, of course, part of the process of simplification which is often necessary in scientific inquiry. However, as Levins has noted, "the art of research is the sensitivity to decide when a useful and necessary simplification has become an obfuscating simplification" (30, p. 105). To what extent might this approach in public health and epidemiology contribute to the "obfuscation" of important factors influencing health and disease in populations?

The application of strategies emphasizing the breaking down and dissecting of risk to epidemiologic research has contributed to (and has in turn been reinforced by) the notions that 1) the true causes of illness are to be found in the biologic—i.e., once we know details of the biologic (and now genetic) mechanisms involved we will know the causes of disease—and 2) the causes of disease are to be found exclusively at the level of individuals. The challenge to epidemiology today is to "put things back together" again, after they have been examined in pieces, and to reconsider dimensions of populations or societies which we may have lost in our examination of the individual-level causes of disease. Moving beyond biologic and individual-based explanations does not imply denying biology, but rather involves viewing biologic phenomena within their social contexts and examining the tight interrelations between the social and the biologic at multiple levels (17). Neither does it imply denying individual-level explanations, but rather entails integrating them into broader models incorporating interactions between individuals, as well as group-level or society-level determinants (which may modify or interact with individual-level properties).

Putting the parts back together into wholes, and understanding the dynamic interrelations between them, is no easy task. The tension between holistic and individualistic explanations of social phenomena (and their correlates, methodological holism and methodological individualism) has long been present (and is far from being resolved) in the social sciences, includ-

ing sociology (31) and economics (32). But this issue is not exclusive to the social sciences; in fact, it is one that permeates science generally, as evidenced, for example, by the ongoing and lively debate on this topic in the field of biology (28, 33, 34). In its efforts to integrate social and biologic factors and individual- and group-level factors in the study of health, epidemiology faces many challenges, some of which will be briefly addressed below.

THEORIES AND MODELS OF DISEASE CAUSATION

A key challenge to today's epidemiology is the development of theories and accompanying conceptual models of how multiple factors, often operating at different levels, are related to the causation of disease in societies (1, 15, 35). Over 60 years ago, in an often-cited quote, Wade Hampton Frost wrote that "epidemiology at any given time is something more than the total of its established facts. It includes their orderly arrangement into chains of inference which extend more or less beyond the bounds of direct observation" (36, p. ix). As Cassel noted, it is the model of disease causation we espouse that guides us in developing these chains of inference (37). The implicit model of disease causation in much current epidemiologic research is the multicausal or "web of causation" model initially proposed by MacMahon et al. (38), in which several different individual-level factors are presumed to be independently related to the outcome. The multicausal model has been undeniably useful in epidemiology, allowing researchers to account for the possibility that several different factors may be involved in shaping patterns of health and disease. However, the widespread application of this model has sometimes resulted in the reduction of epidemiologic research to the investigation of associations between a factor and a disease, after adjustment for multiple additional factors. Chains of causation and the different levels at which factors operate are often ignored: Hierarchies are collapsed, and interest centers on estimating "independent" effects. In doing this, the model implicitly tends to favor more proximate (and therefore biologic and individual-level) determinants over more distal and society-level ones (17, 39).

Models for a new epidemiology may need to go beyond the incorporation of yet another factor into the "web of causation," and there has been a growing discussion of alternate types of models in the epidemiologic literature (15–17, 35, 40). These models will need to take into account the role of multiple levels (e.g., molecular, individual, and societal) in shaping

health outcomes, as well as dynamic interactions within and between levels. Along these lines, Koopman has recently argued that "epidemiology is in transition from a science that identifies risk factors for disease to one that analyzes the systems that generate patterns of disease in populations" (40, p. 630), and Loomis and Wing have held that "cause in epidemiology is not a property of agents but one of complex systems in which the population phenomena of health and disease occur" (41, p. 2).

In addition, new epidemiologic models may require the consideration of alternate types of "causation" or "determination." For example, in his classical text on causation in modern science, Bunge (42) defines "determination" (or "lawful production") as including more than traditional causation (or causal determination). He describes a spectrum of categories of determination applicable in science which are irreducible to one another but which are also interconnected. These include not only causal determination (determination of the effect by an external cause, as in "among susceptible individuals, smoking causes lung cancer") and statistical determination (as in "x percent of persons with high cholesterol will develop a myocardial infarction"), which are the types of determination commonly implicit in epidemiologic research (43), but also other types of determination such as reciprocal causation and structural or holistic determination, which may also have applications in public health and epidemiology. Reciprocal causation (determination of the consequent by mutual action) would be present if, for example, a person's consumption of "unhealthy" foods is influenced by the types of foods available where he or she lives, and if in turn food availability is influenced by consumption in the area. Holistic determination (determination of the parts by the whole) would be present if a person's risk of adopting a certain behavior were influenced by the prevalence of that behavior in the social group to which he or she belonged, or if a person's risk of disease depended on the degree of social inequality in his or her society. As suggested by Almeida Filho (44), different types of "determination" (such as those described above, or others) may be appropriate for different scientific questions and for different levels of analysis (i.e., along the continuum from molecules to society). Of course, it is easier to enunciate these models in general terms than to elaborate them in specific and empirically testable formulations. The development of new theories and models of disease causation that can be operationalized and tested presents enormous challenges to epidemiology, but the process is also ripe with new possibilities.

POPULATIONS AS MORE THAN COLLECTIONS OF INDIVIDUALS

An important factor which has in part limited epidemiology's ability to examine the causes of disease in populations is the "individualization" of epidemiology. By "individualization" I mean the notion that the risk of disease depends exclusively on individual-level characteristics. This notion has been reflected in the behavioral model of disease (in which disease stems from the choices and behaviors of individuals, isolated from their social contexts), and it reappears today in some aspects of the genetic model (in which disease is strongly influenced by an individual's unique genetic makeup). Although epidemiology has often been referred to as the study of the distribution of disease in populations, much of today's epidemiology conceptualizes populations merely as aggregates of individuals (useful from a statistical point of view), rather than as groups of interacting individuals with social relationships and social organizations and with group-level properties that may partly influence risk of disease (16, 41, 44). Although much has been written in epidemiologic journals on the "ecological fallacy" (the fallacy of inferring individual-level associations from group-level data), there has been comparatively little mention of the fallacy inherent in focusing exclusively on the individual level without taking group-level factors into account (the psychologistic or individualistic fallacy) (45). Types of group-level variables which have recently appeared in epidemiologic analyses include, for example, income inequality (46, 47), neighborhood characteristics (48, 49), the prevalence of infection in a community (50), and contact patterns between individuals (51).

A major challenge to epidemiology today is the development of models and methods that integrate individuals within their groups or social contexts, that examine the interacting effects of both individual-level and group-level variables, and that take into account the role of interactions between individuals in shaping the distribution of health and disease. Recent publications on multilevel analysis (45, 52, 53)—as one analytical strategy for including variables operating at multiple levels in epidemiologic analyses—and systems analysis (40, 54)—an approach that analyzes the dynamic systems which generate patterns of health and disease in populations, allowing for interactions between components and processes—reflect a growing theoretical and methodological interest in these areas.

BEYOND INDEPENDENT EFFECTS

Many of the analytical methods used in epidemiology today focus on the need to isolate the "independ-

dent" contributions of different variables to an outcome. The enormous development of multivariate statistical methods over the past few decades, together with the availability of the necessary computing resources, has disseminated these approaches and facilitated their use. These multivariate methods (which are the methodological correlates of the multivariate model of disease causation described above) have allowed the incorporation of multiple variables into epidemiologic analyses. However, although the statistical assumptions of these techniques are usually discussed, their implicit causal-model assumptions often are not. Even the types of questions asked may be partly driven by the analytical method available (rather than the other way around). For example, the research questions themselves are sometimes framed in terms of estimating independent effects (e.g., determining the independent contribution of race/ethnicity to an outcome) rather than in terms of examining and explaining the processes involved (the processes leading to the observed racial/ethnic differences).

Teasing apart independent effects is part of the process of breakdown and simplification which is often necessary in scientific inquiry. However, as we test more complex models involving multiple interacting levels, separating "independent effects" may not always be the most appropriate strategy. In epidemiology, the strong emphasis on separating independent effects or "dissection of risk" has perpetuated the notion that effects are always separable, when in reality they often are not. Stallones already alluded to this issue of "separating out the inseparable" in 1973 when he wrote: "[T]he burden of disease on a human population is part of an environmental system and the interrelatedness of the components of the system cannot be understood by pursuing research whose rationale is to divide and isolate the components in ever greater detail" (55, p. 29). In the extreme case, as a result of the emphasis on separating out independent effects, even the "multicausal model" is forgotten, as each factor is statistically abstracted from the web or pattern of factors of which it is a part (39), and the multicausal model itself is reduced to a collection of uncausal relations (44). New analytical strategies will undeniably flow from the need to investigate hypotheses based on new models of disease causation. The examination of joint and interacting (rather than merely independent) effects operating across levels will need to be a part of these strategies.

PUTTING EPIDEMIOLOGY BACK TOGETHER AGAIN

The epidemiologic approach focused on identifying individual-level risk factors for diseases has been fruit-

ful in increasing our understanding of many factors influencing the distribution of disease in populations, including behaviors, biologic factors, and today, genes. These methods have also been successfully used to identify social factors involved in disease causation—as illustrated, for example, by the many studies documenting social class differences in cardiovascular disease (56–58). Breaking things down is a useful scientific tactic with which to analyze many problems. The difficulty arises when the method itself is reified into an ontologic stance, a "true" and "complete" representation of reality (33) (i.e., "the world is like the method" rather than "the method helps us understand some aspects of the world"). An exclusive focus on individualizing and dissecting risk may hamper our ability to test more sophisticated (and realistic) models of disease causation.

In a sense, today's increasing emphasis on genes as fundamental causes of disease exemplifies the biologization and individualization of epidemiology in its maximum expression (i.e., genes as the fundamental biologic substrate and genetic makeup as a unique characteristic of individuals). To a greater or lesser extent, genes will be involved in every disease. However, for most diseases, gene expression, and indeed the degree to which genetic differences are important in understanding the distribution of disease, will depend on the broader context, and this context will be largely influenced by how we live with each other in society (59). Genetic epidemiology and traditional risk factor epidemiology hold potential for enhancing our understanding of the causes of disease, but in isolation the picture they give us is incomplete. Analogously, social explanations of the causes of disease presuppose biologic explanations (and must take them into account), although they are not reducible to them. The challenge is to develop theories (and the methodologies needed to test them) that integrate genes (or other biologic variables) within their broader behavioral, cultural, and social contexts. This is necessary for understanding not only the whole, but also the parts; not only populations, but also the persons within them.

Today we need an epidemiology that moves beyond the dichotomies of social/biologic and groups/individuals. The reason for this is scientific: Our object of study demands it. And of course, good science is one (although by far not the only) requirement for good policy. This new epidemiology will require new ways of thinking about our theories and our methods. It may also require recovering some of the descriptive and population-based aspects of traditional epidemiology (16). It will not replace current approaches but rather will complement, enrich, and transform them. The challenges are enormous—but an epidemiology

capable of moving beyond exclusively biologic and individualistic explanations is a *sine qua non* for both good epidemiologic science and good public health policy.

ACKNOWLEDGMENTS

Dr. Joan Benach contributed greatly to many of the ideas developed in this paper. The author also thanks Drs. José Tapia, Javier Nieto, and Steven Shea for their helpful comments on an earlier version of this paper.

REFERENCES

- Cassel J. Potentialities and limitations of epidemiology. In: Health and the community: readings in the philosophy and sciences of public health. Katz A, Felton J, eds. New York, NY: The Free Press, 1965:432–45.
- Susser M, Susser E. Choosing a future for epidemiology. I. Eras and paradigms. *Am J Public Health* 1996;86:668–73.
- McMichael AJ. The health of persons, populations, and planets: epidemiology comes full circle. *Epidemiology* 1995; 6:633–6.
- Ambrosone CB, Kadlubar FF. Toward an integrated approach to molecular epidemiology. *Am J Epidemiol* 1997;146:912–18.
- Ottman R. Genetic epidemiology of epilepsy. *Epidemiol Rev* 1997;19:120–8.
- Duffy DL. Genetic epidemiology of asthma. *Epidemiol Rev* 1997;19:129–43.
- Peyster PA. Genetic epidemiology of coronary artery disease. *Epidemiol Rev* 1997;19:80–90.
- Slooter AJ, van Duijn CM. Genetic epidemiology of Alzheimer disease. *Epidemiol Rev* 1997;19:107–19.
- Morton NE. Outline of genetic epidemiology. Basel, Switzerland: S Karger, AG, 1982.
- Khoury MJ, Beaty TH, Cohen BH. Fundamentals of genetic epidemiology. New York, NY: Oxford University Press, 1993.
- Schulte PA, Perera FP, eds. Molecular epidemiology: principles and practice. New York, NY: Academic Press, Inc, 1993.
- Perusse L, Tremblay A, Leblanc C, et al. Genetic and environmental influences on level of habitual physical activity and exercise participation. *Am J Epidemiol* 1989;129:1012–22.
- Kendler KS. Social support: a genetic-epidemiologic analysis. *Am J Psychiatry* 1997;154:1398–404.
- Kaplan GA, Lynch JW. Whither studies on the socioeconomic foundations of population health? (Editorial). *Am J Public Health* 1997;87:1409–11.
- 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.
- Pearce N. Traditional epidemiology, modern epidemiology and public health. *Am J Public Health* 1996;86:678–83.
- Krieger N. Epidemiology and the web of causation: Has anyone seen the spider? *Soc Sci Med* 1994;39:887–903.
- Rosen G. A history of public health. New York, NY: MD Publications, Inc, 1958. [Expanded edition: The Johns Hopkins University Press (Baltimore, Maryland), 1993].
- Rosen G. What is social medicine? *Bull Hist Med* 1947;21: 674–731.
- Taylor R, Rieger A. Medicine as social science: Rudolf Virchow on the typhus epidemic in Upper Silesia. *Int J Health Serv* 1985;15:547–59.
- Waitzkin H. The social origins of illness: a neglected history. *Int J Health Serv* 1981;11:77–103.
- Sydenstricker E. Health and environment. New York, NY: McGraw-Hill Book Company, 1933. [Reprinted by Ayer Company Publishers (Salem, New Hampshire), 1972].
- Syme SL. Social determinants of disease. *Ann Clin Res* 1987; 19:44–52.
- Susser M, Watson W, Hopper K. Social class and disorders of health. In: *Sociology in medicine*. 3rd ed. New York, NY: Oxford University Press, 1985:213–75.
- Laurell AC. Social analysis of collective health in Latin America. *Soc Sci Med* 1989;28:1183–191.
- Blane D, Brunner E, Wilkinson R. Health and social organization. New York, NY: Routledge, Chapman and Hall, Inc, 1996.
- Wilkinson RG. Unhealthy societies: the afflictions of inequality. New York, NY: Routledge, Chapman and Hall, Inc, 1996.
- Lewontin RC, Rose S, Kamin LJ. Not in our genes: biology, ideology, and human nature. New York, NY: Pantheon Books, Inc, 1984.
- Russell B. The scientific outlook. 2nd ed. Chapter 2. London, England: Allen and Unwin, 1949. [Reprinted by W W Norton Company (New York, New York), 1962].
- Levins R. Ten propositions on science and antiscience. *Soc Text* 1996;14:101–11.
- Lukes S. Methodological individualism reconsidered. In: MacIntyre A, Emmet D, eds. *Sociological theory and philosophical analysis*. New York, NY: Macmillan Publishing Company, Inc, 1970:76–88.
- Schumpeter JA. History of economic analysis. 2nd ed. London, England: Oxford University Press, 1954:888–9.
- Levins R, Lewontin RC. The dialectical biologist. Boston, MA: Harvard University Press, 1985.
- Wilson EO. Consilience: the unity of knowledge. New York, NY: Alfred A Knopf, Inc, 1998.
- Krieger N, Zierler S. What explains the public's health?—A call for epidemiologic theory. *Epidemiology* 1996;7:107–9.
- Frost WH. Introduction to Snow on cholera. New York, NY: The Commonwealth Fund, 1936.
- Cassell J. The contribution of the social environment to host resistance: the Fourth Wade Hampton Frost Lecture. *Am J Epidemiol* 1976;104:107–23.
- MacMahon B, Pugh TF, Ipsen J. Epidemiologic methods. Boston, MA: Little, Brown and Company, 1960.
- Tesh SN. Hidden arguments: political ideology and disease prevention policy. New York, NY: Rutgers University Press, 1990.
- Koopman JS. Emerging objectives and methods in epidemiology. *Am J Public Health* 1996;86:630–2.
- Loomis D, Wing S. Is molecular epidemiology a germ theory for the end of the twentieth century? *Int J Epidemiol* 1990; 19:1–3.
- Bunge M. Causation and determination, causalism and determinism. In: *Causality and modern science*. New York, NY: Dover Publications, Inc, 1979:3–30.
- Rothman KJ, Greenland S. Causation and causal inference. In: Rothman KJ, Greenland S. *Modern epidemiology*. 2nd ed. Philadelphia, PA: Lippincott-Raven, Inc, 1998.
- Almeida Filho N. Epidemiologia sin números. Washington, DC: Pan American Health Organization, 1992. (Serie Paltes no. 28).
- Diez-Roux AV. Bringing context back into epidemiology: variables and fallacies in multilevel analysis. *Am J Public Health* 1998;88:216–22.
- Kennedy BP, Kawachi I, Prothrow-Stith D. Income distribution and mortality: cross sectional ecological study of the Robin Hood index in the United States. *BMJ* 1996;312: 1004–7.
- Kaplan GA, Pamuk ER, Lynch JW, et al. Inequality in income and mortality in the United States: analysis of mortality and potential pathways. *BMJ* 1996;312:999–1003.
- O'Campo P, Gielen AC, Faden RR, et al. Violence by male partners against women during the childbearing year: a contextual analysis. *Am J Public Health* 1995;85:1092–7.

49. Diez-Roux AV, Nieto FJ, Muntaner C, et al. Neighborhood environments and coronary heart disease: a multilevel analysis. *Am J Epidemiol* 1997;146:48-63.
50. Koopman JS, Longini IM Jr. The ecological effects of individual exposures and nonlinear disease dynamics in populations. *Am J Public Health* 1994;84:836-42.
51. Koopman JS, Longini IM Jr, Jacquez JA, et al. Assessing risk factors for transmission of infection. *Am J Epidemiol* 1991;133:1199-209.
52. von Korff M, Koepsell T, Curry S, et al. Multi-level analysis in epidemiologic research on health behaviors and outcomes. *Am J Epidemiol* 1992;135:1077-82.
53. Duncan C, Jones K, Moon G. Context, composition and heterogeneity: using multilevel models in health research. *Soc Sci Med* 1998;46:97-117.
54. Weed DL. Beyond black box epidemiology. *Am J Public Health* 1998;88:12-14.
55. Stallones RA. The epidemiologist as environmentalist. *Int J Health Serv* 1973;3:29-33.
56. Marmot MG, Smith GD, Stansfeld S, et al. Health inequalities among British civil servants: the Whitehall II study. *Lancet* 1991;337:1387-93.
57. Kaplan GA, Keil JE. Socioeconomic factors and cardiovascular disease: a review of the literature. *Circulation* 1993;88:1973-98.
58. Marmot MG, Bosma H, Hemingway H, et al. Contribution of job control and other risk factors to social variations in coronary heart disease incidence. *Lancet* 1997;350:235-9.
59. Penchaszadeh VB. Genética, individuo y sociedad: desafíos para la medicina social. [Genetics, individuals, and society: challenges to social medicine]. *Bol Oficina Sanit Panam* 1995;118:254-63.