Individual Causal Models and Population System Models in Epidemiology

James S. Koopman, MD, MPH, and John W. Lynch, PhD, MPH

ABSTRACT

A group of individuals behaves as a population system when patterns of connections among individuals influence population health outcomes. Epidemiology usually treats populations as collections of independent individuals rather than as systems of interacting individuals. An appropriate theoretical structure, which includes the determinants of connections among individuals, is needed to develop a "population system epidemiology."

Infection transmission models and sufficient-component cause models provide contrasting templates for the needed theoretical structure. Sufficient-component cause models focus on joint effects of multiple exposures in individuals. They handle time and interactions between individuals in the definition of variables and assume that populations are the sum of their individuals. Transmission models, in contrast, model interactions among individuals over time. Their nonlinear structure means that population risks are not simply the sum of individual risks.

The theoretical base for "population system epidemiology" should integrate both approaches. It should model joint effects of multiple exposures in individuals as time related processes while incorporating the determinants and effects of interactions among individuals. Recent advances in G-estimation and discrete individual transmission model formulation provide opportunities for such integration. (*Am J Public Health*. 1999;89: 1170–1174)

During much of the modern era in epidemiology, the analytic methods and causal models of epidemiology have been directed toward risk factor effects on individuals. Recently, epidemiology has turned again to more broadly addressing population phenomena whose effects on the health of populations cannot be viewed as the sum of effects on the individuals in that population.[1–3] These phenomena include how population patterns of exposure, and not just numbers of exposed individuals, affect the health of populations. For instance, exposure patterns are particularly important determinants of infection levels in populations. Changing the pattern of connections between exposed and unexposed individuals can often affect population infection levels more than changing the exposure status of individuals in that population.[4–7] Similarly, different patterns in the distribution of income in a population can have population health effects beyond the effects attributable to individual incomes.[8–10]

Analysis of how population level characteristics and patterns of exposure affect disease levels could be called "population system epidemiology." In a system, in contrast to a "heap," the arrangement of elements makes a difference. When the pattern of exposures or connections between individuals in a population has the potential to make a difference to disease levels, we are dealing with a population system, not just a heap of individuals.

The thesis of this commentary is that population system epidemiology needs causal models that formulate dynamic interactions between individuals in ways that the sufficient-component cause model[11] does not, while still preserving the virtues of this model. The sufficient-component cause model has been valuable because it abstracts joint effects of multiple exposures in individuals. In population systems modeling, however, models of nonlinear population processes are needed that define how time-varying patterns of connections among individuals affect population level outcomes. Because infection transmission models do this explicitly, they define causal mechanisms that determine how outcomes in some individuals determine outcomes in other individuals.

The sufficient-component cause model, however, has a structure that assumes that populations can be defined by linear combinations of individuals. Rather than modeling the origins of dependencies between individuals that make a population different than the sum of its individuals, the structure of the sufficient-component cause model requires that variables be defined so that the outcome of each individual is independently determined. Schwartz and Carpenter in this issue employ the sufficient-component cause model to analyze effects whose origins lie at a population rather than an individual level.[12] We will attempt to show here, however, the deficiencies of the sufficient-component cause model for modeling population-level effects.

Some Aspects of Causal Models

To contrast sufficient-component cause models with infection transmission models, we should consider first the nature of causal models. Causal models help scientists by abstracting particular elements of a causal phenomenon while discarding details that get in the way of deriving useful theory and insights about the phenomenon under study.

Reality, especially the biological and social reality with which epidemiology deals, is diverse and multifaceted. But theory, by its very nature, must be focused and unifying. A narrow focus on a causal phenomenon that disregards realistic details can help achieve wide applicability of theory. For example, to develop a revolutionizing and productive theory about bodies in motion, Newton had to ignore Aristotle's insistence that the slowing of bodies in motion had to be a central part of theories of motion because the observation of such slowing was so universal. To formulate universal laws of motion, Newton used a model of motion that ignored this universal reality. Newton's laws of motion are validated by their theoretical utility, not their fit to reality.

Sufficient-component cause theory, in a similar fashion to Newton's laws, ignores several aspects of reality but has demonstrated its utility for developing theory about the effects of multiple exposures in individuals. The ability of this model to generate productive new insights is linked to the way it

The authors are with the Department of Epidemiology, University of Michigan, Ann Arbor.

Requests for reprints should be sent to James S. Koopman, MD, MPH, Department of Epidemiology, SPH-1, University of Michigan, 109 Observatory St, Ann Arbor, MI 48109-2029 (e-mail: jkoopman@umich.edu).

Editor's Note. Please see related article by Schwartz and Carpenter (p 1175) in this issue.

ignores details on timing of exposure and on population phenomena arising from patterns of connection between individuals that it cannot neatly encompass. Transmission system theory, on the other hand, has employed dynamic models in which timing and connection between individuals are central but joint effects of multiple exposures are ignored. Rather than modeling discrete individuals, transmission system models most commonly model interactions between continuous population segments. This enables transmission system theory to use differential equations to focus on population rather than individual outcomes.

These 2 contrasting approaches to modeling causal issues of central concern to epidemiology focus on such different phenomena as to seem perhaps incommensurate. We propose, however, that causal model formulations unifying the virtues of each approach are needed and possible.

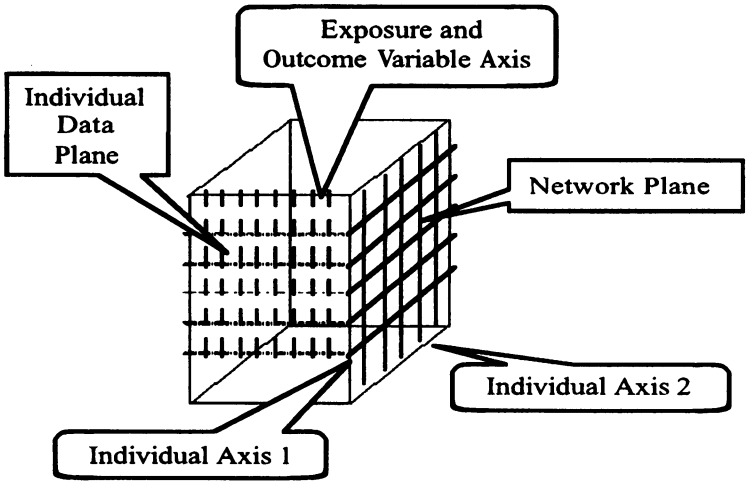

FIGURE 1—The 3-dimensional shape of epidemiologic data.

The Structure of Epidemiologic Data

Perhaps the best way to see the need for integrating sufficient-component cause theory with transmission theory is to begin with data rather than theory. Standard epidemiologic analysis methods array data for separate individuals into rows and data on both outcome and predictor variables for each individual into columns. This is represented in the individual data plane in Figure 1. The fact that most standard analyses do not view individuals as being part of a system is manifest by the assumption that the row an individual is in (i.e., the arrangement of individuals) makes no difference to the results of standard analyses. This 2-dimensional data arrangement at the individual level can be viewed as the face of a 3-dimensional cube, depicted in Figure 1, where the third dimension defines the pattern of connections between individuals that generate population system phenomena.

Both social network analysis[13] and phylogenetic analysis[14] are performed in the network plane that is perpendicular to the individual data plane. For these analytic methods, the data are arranged as a square matrix with individuals along both axes. The values in the matrix represent degrees of connection between individuals. If connections are described dichotomously, there may be a 1 or a 0 in each cell depending on whether a connection between 2 individuals exists, or there may be multiple continuous connection variables. These might have directionality from axis 1 to axis 2, or vice versa. In the case of phylogenetic analysis using DNA sequence data, there might be a variable for each base pair location indicating identity or difference at the site.

Many variables collected in field studies might not directly measure degrees of connection between individuals; they might only reflect chances of having connections with different groups of individuals. One class of variables of this type includes variables that describe an aspect of a contact that both individuals report identically. For example, the type of sex act or the courtship time between meeting and having sex can be used to reflect who is likely to be connected to whom. Geographic or social locations where contacts are made could be used to reflect many different types of contact beyond sexual contact. The intrinsic value of these variables lies in the network plane. If they are analyzed only in the individual data plane, much of their value will be lost because their value to population systems epidemiology lies in describing relationships in the network plane.

Figure 1 is presented for its heuristic value. It is not an exact representation of the shape of epidemiologic data needed for the analysis of any particular model. The number of different connection variables may not correspond to the number of individual variables measured, as the existence of a layer of connection for each individual variable in the figure might imply. Although Figure 1 does not include the time dimension that is central to dynamic analysis of population systems, it captures the essential argument we wish to make: that the dimension of connections between individuals is an integral part of epidemiologic data, even if it is ignored in the data analysis itself.

Standard epidemiologic analyses assume that causal events in the dimension connecting individuals are irrelevant. Even models that take into account dependence between individuals generated by clustered sampling assume that the outcome in one individual is independent of the outcome in other individuals. This assumption is also inherent to any use of the sufficient-component cause model to represent populations. This assumption is violated whenever transmission of infection generates new sources of infectious agent or whenever the level of infection in one segment of a population affects the risk of infection in other population segments. Because this violation is so readily apparent for infectious diseases, we use infectious disease examples in this commentary. But the existence of social network connections is also likely to have other health effects that violate this standard assumption of epidemiologic analysis. Examples include social support, social stress, transmission of behavior norms, transmission of knowledge that influences behavior, and transmission of power relationships imposing order or generating exploitation.

Because standard epidemiologic analyses and the sufficient-component cause model ignore network connections between individuals, they also ignore the larger political, cultural, and economic forces that determine different patterns of network connections. Models that incorporate network structure could help highlight these important determinants of disease and bring them into the realm of epidemiologic investigation.

Network Connections, Individual Risk, and Population Risk

We can think of infection as flowing in the network plane that is ignored by standard analytic methods in epidemiology. Transmission

system models are constructed mainly so that their contact rates and transmission probabilities describe causal phenomena in this plane. Transmission models must incorporate the interactions of all individuals in a population.

Epidemiologic data, in contrast, usually deal only with a sample of individuals from the population. Consequently, the most common network data in epidemiology indicate only the class of individuals contacted rather than the exact individuals who were contacted. Data gathered from individuals about their contacts are said to be "egocentric": they describe the contacts around individuals but not the overall population pattern of contacts. The examples we present in Figures 2 and 3 use complete network data. In our examples, circles represent individuals and lines represent the existence of a connection between individuals. Time relationships are ignored.

The example in Figure 2 demonstrates that egocentric data are critically incomplete even when all the individuals in the population are identical. Egocentric data describe the connections of individuals but not how those connections are linked into a population network. Such data from either population A or population B in Figure 2 would indicate that each individual is connected to 2 other individuals. But egocentric information does not establish whether those individuals form chains that sustain transmission. In A, they would; in B, they would not. Although all individuals in populations with either pattern A or B would appear to be the same, the population with pattern A would have higher levels of infection. This illustrates the fact that the 2 populations are not just the sum of the individuals therein as would be assumed by analysis restricted to the individual data plane. The 2 populations need to be defined by the pattern of connections between individuals in the network plane as well.

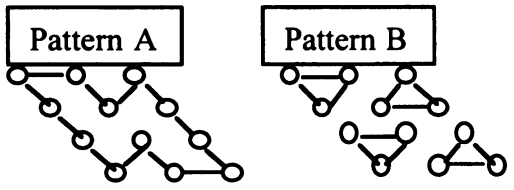

FIGURE 2—Distinct patterns of connection between identical individuals with 2 connections.

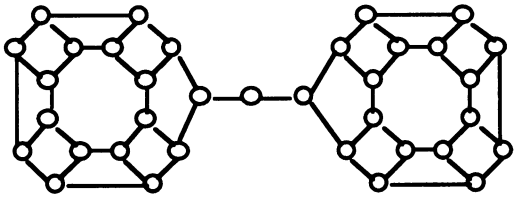

FIGURE 3—A network in which the most important individual determining population levels of infection has the lowest risk of infection.

Different Effects of Network Roles and Individual Risk on Population Risk

In Figure 2, all individuals in a network play the same role in their network. Now we consider an example in which individuals play different roles in their network. In Figure 3, individuals with 3 contacts can be distinguished by their proximity to a connecting link between 2 groups. Moreover, most individuals have 3 connections, but 1 individual has only 2.

Consider the situation in which transmission across each link occurs with some specified probability and there is random introduction of infection into the network. A range of transmission probability values exists in which the individual with only 2 links has the lowest chance of becoming infected after random introduction of infection to the population. Thus, from an individual risk view, this individual has the lowest risk of infection. The contribution of that individual to infection levels in the population system can, however, at certain transmission probabilities, be greater than that of any other individual. Eliminating one connection to this individual can do more to lower average infection levels in the population after introduction of infection than eliminating a connection to any other individual. Any risk analysis assuming independence of outcomes in individuals at risk would miss this fact.

Risk Factors for Transmission

The failure of individual risk analyses to identify key individuals determining transmission at the population level is particularly notable when individuals are distinguished by risk factor status. Consider 2 populations in which for each individual in the first population, there is a corresponding individual in the second population with exactly the same risk factors and exactly the same history of contacts. Any description of individuals in the individual data plane will find these 2 populations to be identical: they differ only in the way individuals are connected. By changing contact patterns, however, we can change the level of infection from zero to complete.

Consider the transmission of an infection that does not induce immunity in a population with individuals who do and do not have a risk factor. Suppose for the sake of exposition that this risk factor is a gene that increases susceptibility to infection. Individuals in the group with the gene might make all of their contacts with individuals who do not have this gene. If these latter individuals are incapable of sustaining chains of transmission on their own, then the population will not sustain transmission. If individuals with the gene make all of their contacts with each other, infection might flow quickly between them but not at all to the group without the gene. If individuals in the group with the gene make just enough contacts with each other to sustain circulation between them, they might make the rest of their contacts with the group without the gene and infect many or most of that group's members.

Even if data are available on whether each contact of an individual does or does not have a susceptibility gene, analysis of this network data in the individual data plane cannot determine the population level of infection. Thus, individual risks for infection cannot be determined. This is true regardless of whether the population system is at equilibrium. Given such data, a transmission model analysis capturing phenomena occurring in the network plane can determine expected infection levels.

The need for network analysis is further emphasized when the gene under consideration does not increase the chances that an individual will be infected but does increase the level of infectious agent in the event of infection and therefore increases contagiousness. Gene effects on susceptibility and contagiousness can be quite distinct, because the survival of a transmitted agent can depend on very different things than the proliferation of the agent in the host. Standard risk factor

analysis has a strong bias to detect risk factors for susceptibility rather than risk factors for contagiousness. By defining exposure as contact with someone who has the gene under consideration, a standard analysis could detect a risk factor affecting contagiousness. Given difficulties in determining the identity and gene status of individuals contacted, however, the ability to detect contagiousness effects will be limited. Just as in the susceptibility case, even if perfect measurements are available, an individual-level analysis cannot determine population levels of infection or individual infection risks in a transmission system.

An important observation that does not even fall into the realm of individual-level analysis is that if the gene we are considering increases contagiousness, it will affect population levels of infection more than if it affects susceptibility. A gene might make an individual x times as susceptible to infection or x times as contagious if infected. An increase in contagiousness will raise infection to higher levels than will a comparable increase in susceptibility. Not only do contagiousness risk factors have greater effects than susceptibility risk factors, they are also likely to be more frequent. There are more opportunities to affect agent replication once it is underway than there are opportunities to keep replication from beginning. Standard methods in epidemiology do not focus on contagiousness risk factors, whereas transmission system analyses more naturally focus on such risk factors. Transmission system analysis may thus not only move us beyond a risk factor focus but also enhance the focus on a class of risk factors that has been ignored.

The Sufficient-Component Cause Model

The sufficient-component cause model has been and should continue to be a productive causal model. It has focused and unified epidemiologic thinking,[11] clarified concepts of confounding and effect modification, and helped to clarify the mathematical relationships for assessing the public health impact of exposure to risk factors.[15] It has served as the basis for defining observations that can distinguish whether 2 risk factors act in distinct causal pathways, have distinct roles in the same causal pathway, or play the same role in pathogenesis.[16] Its limitations for analysis of infectious disease effects, however, have long been noted.[17]

Two such limitations deserve discussion. The first is that the sufficient-component cause model does not treat time in a fashion that allows for dynamic analysis. For a dynamic analysis, dynamic processes must be explicitly formulated in time. In the sufficient-component cause model, time is invoked in the definition of whether a factor is present or absent but dynamic processes are not formulated.

The second limitation of the sufficient-component cause model is that effects in the network plane are not incorporated in an analyzable fashion. The examples just provided are meant to make that clear. The problem is not just that defining exposure as a function of the contacts made is unwieldy in the context of the sufficient-component cause model. Rather, the problem is that individual characteristics alone do not determine the population risks. The outcome of the sufficient-component cause model is a dichotomous disease classification of individuals. The way that the outcome in one individual depends on the outcomes in other individuals, though, makes it impossible for the sufficient-component cause model to define individual infection risks in a transmission system before infection spreads in the population. In nonlinear population systems, such as infectious agent transmission systems, population risks are not the sum of individual risks as the sufficient-component model assumes.

Transmission Models

Transmission system models explicitly model phenomena in the network plane that most epidemiologic analysis ignores. These models have a long and exponentially growing tradition of development in epidemiology. Many useful concepts coming out of this tradition have been presented by Anderson and May.[18] The differential equation models commonly used in this tradition handle time explicitly and therefore offer a basis for dynamic analysis that the sufficient-component cause model lacks. Perhaps the characteristic that most distinguishes transmission models from the sufficient-component cause model is that they model nonlinear population effects.[4] Nonlinearity of effects at the population system level has 2 important implications: first, individual effects will not sum to population effects, and, second, patterns of connection between individuals will have effects at the population level.

The dominant tradition in transmission system modeling ignores many aspects of reality so that, as with Newton's model of motion, the essence of the phenomenon under study can be illuminated, insights can be gained, and new analytic tools can be developed. The major simplification made by transmission models is that populations are treated as continuous entities and individuals are ignored. This simplification is intrinsic to the use of differential equation models, and it imposes further simplifications with regard to contact patterns. The type of differential equations commonly used to construct transmission system models cannot capture details of individual connection patterns like those seen in Figures 2 and 3. Moreover, in differential equation transmission models, contact is an instantaneous event. Differential equations can be defined that incorporate continuous population segments in which the basic units are not individuals but pairs of individuals.[19] In this type of differential equation model, contact can have duration and need not be instantaneous. But contact models of this type still cannot define individual networks of the type in Figure 2 or 3.

Recently there has been better acceptance of discrete individual models in the transmission system modeling tradition. Discrete individual models of transmission have been especially useful in the examination of sexually transmitted infections.[20–24] The reason for this acceptance is that individual network patterns in which contacts occur in ongoing relationships are now seen to be important determinants of population infection levels of sexually transmitted infections. Individual models can capture these determinants better than differential equation models.

Blended Transmission and Sufficient-Component Cause Models

One reason for promoting the development of transmission models with discrete individuals who connect to each other in population patterns is that by including individuals in transmission models, one can meld the theoretical insights of transmission models with the theoretical insights of the sufficient-component cause model. Such blended models could provide a basis for fully using all the dimensions of epidemiologic data. A new approach we have taken to discrete individual models of transmission systems may offer some advantages in this endeavor.[25] One advantage of this new approach is that it allows for theoretical model analysis that is difficult or impossible with other approaches. Another advantage is that it explicitly incorporates data on the social and geographic setting of contact; such data are readily collectible in epidemiologic studies. Most other discrete individual models incorporate only data from the network plane relevant to connections between specific individuals, but such data are not readily collectible in epidemiologic studies.

The sufficient-component cause model cannot be integrated directly into transmission

models because it lacks an explicit integration of time. Graph theoretic models may have an advantage over the sufficient-component cause model in this regard.[26] Graph theoretic models, like the sufficient-component cause model, are designed to address the joint effects in individuals of multiple variables. By including time relationships in arrows instead of only in variable definitions, graph theoretic models clarify some aspects of confounding that were not so readily apparent from analysis of the sufficient-component cause model.[26] In their current form, however, graph theoretic models do not incorporate data relevant to the network plane. They model only individual effects, and they assume that there are no dependencies between individuals. In the same manner that G-estimation methods have been used to incorporate time data,[27] however, graph theoretic models might be elaborated to incorporate network data.

The point of this commentary is not to advocate any particular approach to developing integrative causal models. Rather, we argue that the theoretical foundation for population system epidemiology cannot depend wholly on individual-level models. Dynamic models of population processes are needed that incorporate exposure patterns, time, and contact networks in the manner of transmission models while also incorporating the joint effects of multiple exposures in individuals in the manner of the sufficient-component cause model. Bringing both of these traditions together could enhance the pursuit of a population systems approach to epidemiology. □

References

1. Susser M, Susser E. Choosing a future for epidemiology, I: eras and paradigms. *Am J Public Health.* 1996;86:668–673.
2. Susser M, Susser E. Choosing a future for epidemiology, I: from black box to Chinese boxes and eco-epidemiology. *Am J Public Health.* 1996;86:674–677.
3. Koopman JS. Emerging objectives and methods in epidemiology. *Am J Public Health.* 1996;86: 630–632.
4. Koopman JS, Longini IM. Ecological effects of individual exposures and nonlinear disease dynamics in populations. *Am J Public Health.* 1994;84:836–842.
5. Koopman JS, Longini IM, Jacquez JA, Simon CP, Martin W, Woodcock D. Assessing risk factors for transmission. *Am J Epidemiol.* 1991; 133:1199–1209.
6. Koopman JS, Jacquez JA, Simon CP, et al. The role of primary HIV infection in the spread of HIV through populations. *J Acquir Immune Defic Syndr Hum Retrovirol.* 1997;14:249–258.
7. Jacquez JA, Koopman JS, Simon C, Sattenspiel L, Perry T. Modeling and the analysis of HIV transmission: the effect of contact patterns. *Math Biosci.* 1988;92:119–199.
8. Kaplan GA, Pamuk ER, Lynch JW, Cohen RD, Balfour JL. Inequality in income and mortality in the United States: analysis of mortality and potential pathways. *BMJ.* 1996;312:999–1003.
9. Wolfson M, Kaplan GA, Lynch JW, Ross N, Backlund E. The relationship between income inequality and mortality is not a statistical artefact. *BMJ.* In press.
10. Lynch JW, Kaplan GA, Pamuk ER, et al. Income inequality and mortality in metropolitan areas of the United States. *Am J Public Health.* 1998;88:1074–1080.
11. Rothman KJ, Greenland S. Causation and causal inference. In: Rothman KJ, Greenland S, eds. *Modern Epidemiology.* Philadelphia, Pa: Lippincott Raven Press; 1998:7–28.
12. Schwartz S, Carpenter KM. The right answer for the wrong question: consequences of type III error for public health research. *Am J Public Health.* 1999;89:1175–1180.
13. Wasserman S, Faust K. *Social Network Analysis: Methods and Applications.* Cambridge, UK: Cambridge University Press; 1994.
14. Swofford DL, Olsen GJ, Waddell PJ, Hillis DM. Phylogenetic inference. In: Hillis DM, Moritz C, Mable BK. *Molecular Systematics.* Sunderland, Mass: Sinauer Associates Inc Publishers; 1996.
15. Koopman JS. Interaction between discrete causes. *Am J Epidemiol.* 1981;113:716–724.
16. Koopman JS, Weed DL. Epigenesis theory: a mathematical model relating causal concepts of pathogenesis in individuals to disease patterns in populations. *Am J Epidemiol.* 1990;132:366–390.
17. Koopman JS. Causal models and sources of interaction. *Am J Epidemiol.* 1977;106:439–443.
18. Anderson RM, May RM. *Infectious Diseases of Humans: Dynamics and Control.* Oxford, England: Oxford University Press; 1992.
19. Dietz K, Hadeler KP. Epidemiological models for sexually transmitted diseases. *J Math Biol.* 1988;26:1–25.
20. van der Ploeg CBP, van Vliet C, de Vlas SJ, et al. STDSIM: a microsimulation model for decision support in STD control. *Interfaces.* 1998; 28:84–100.
21. Ghani AC, Swinton J, Garnett GP. The role of sexual partnership networks in the epidemiology of gonorrhea. *Sex Transm Dis.* 1997;24: 45–56.
22. Kretzschmar M, Morris M. Measures of concurrency in networks and the spread of infectious disease. *Math Biosci.* 1996;133:165–195.
23. Adams A, Barth-Jones D, Chick SE, Koopman JS. Simulations to evaluate HIV vaccine trial designs. *Simulation.* 1998;71:228–241.
24. Welch G, Chick SE, Koopman JS. Effect of concurrent partnerships and sex-act rate on gonorrhea prevalence. *Simulation.* 1998;71:242–249.
25. Koopman JS, Chick SE. Models for evaluating network statistics and surveillance procedures: preliminary results for gonorrhea. Paper presented at: Santa Fe Conference on Sexually Transmitted Disease Models; April 28, 1999; Santa Fe, NM.
26. Greenland S, Pearl J, Robins JM. Causal diagrams for epidemiologic research. *Epidemiology.* 1999;10:37–48.
27. Witteman JC, D'Agostino RB, Stijnen T, et al. G-estimation of causal effects: isolated systolic hypertension and cardiovascular death in the Framingham Heart Study. *Am J Epidemiol.* 1998;148:390–401.