In this commentary, structural equation models (SEMs) are discussed as a tool for epidemiologic analysis. Such models are related to and compared with other analytic approaches often used in epidemiology, including regression analysis, causal diagrams, causal mediation analysis, and marginal structural models. Several of these other approaches in fact developed out of the SEM literature. However, SEMs themselves tend to make much stronger assumptions than these other techniques. SEMs estimate more types of effects than do these other techniques, but this comes at the price of additional assumptions. Many of these assumptions have often been ignored and not carefully evaluated when SEMs have been used in practice. In light of the strong assumptions employed by SEMs, the author argues that they should be used principally for the purposes of exploratory analysis and hypothesis generation when a broad range of effects are potentially of interest.

causal inference; causality; causal modeling; confounding factors (epidemiology); epidemiologic methods; regression analysis; structural equation model

Abbreviations: MSM, marginal structural model; SEM structural equation model.

Every so often a paper is published in the epidemiologic literature that employs structural equation models (SEMs). Sometimes the authors of these papers advocate that such methods be employed in epidemiology more broadly. The paper by Arlinghaus et al. (1) in this issue of the Journal is one such paper. In spite of these attempts, the approach remains relatively uncommon in epidemiology. Other methods, of more recent origin than SEMs, seem to have been picked up more readily by epidemiologists (2, 3). Why not SEMs? In this commentary, I briefly discuss the relation of SEMs to other methods in the epidemiologic toolkit and then return to the question of whether, when, and for what purpose epidemiologists should consider the use of SEM methods.

**STRUCTURAL EQUATION MODELING**

**Relation to regression analysis**

Far and away, the most common analytical approach in epidemiology today is the use of regression methods: sometimes linear regression, sometimes logistic, sometimes Poisson, sometimes the proportional hazards model. These models are employed when the relation between a particular exposure and a particular outcome is of interest. Consideration is given to what pre-exposure covariates might confound the relation between the exposure and the outcome. These covariates are generally included in the model as well. We can interpret the coefficient for the exposure in these models as a causal effect if we have adequately controlled for confounding, and provided there is no selection bias or measurement error. With observational data, such control is never perfect; but we attempt to collect data on and adjust for pre-exposure covariates related to both the exposure and the outcome of interest.

If we have a single exposure and a single outcome, how does this differ from structural equation modeling and why might we use it rather than the familiar regression-based approaches? Arlinghaus et al. state that one reason is an increase in statistical power (1). However, this power comes at a price. The price we pay is assumptions. Of course, we make assumptions with traditional regression...
methods as well. We assume that the expected value of the outcome conditional on covariates (or the logit of this expectation, or the log hazard ratio) is linear in the exposure and covariates. SEMs make this assumption as well, but assumptions of this form are made much more extensively with SEMs. If our outcome of interest is injuries, with a traditional logistic regression analysis we would assume that the log odds of the outcome are linear in our exposure (e.g., hours of sleep) and the covariates. However, with SEMs, such as the SEM shown in Arlinghaus et al.’s Figure 1 (adapted here (Figure 1)), we also assume that sleep duration (or the log odds of its categories) is linear in age, body mass index, distress, and working hours, etc.; we assume that working hours are linear in distress, gender, race, education, etc.; and we assume that psychological distress is linear in sadness, nervousness, restlessness, etc. With an SEM, we make linearity assumptions not simply for 1 outcome variable (our primary variable of interest) but for many. However, these are not the only assumptions we make. With SEMs, we also make assumptions about how the covariates are related to each other. In the SEM shown in Arlinghaus et al.’s Figure 1, there are no arrows pointing from gender, race, education, and occupation directly to sleep. The model Arlinghaus et al. employ assumes that the entire effect of these variables on sleep is mediated by weekly working hours. Do we believe these assumptions? Do we need to make them? Our traditional regression models do not make them. With traditional regression methods, we model the relation between the outcome on the one hand and the exposure and covariates on the other, but we do not directly make assumptions about the relations among the covariates themselves. Of course, if we do make such assumptions, as one generally does with an SEM, one can gain power.

**Relation to causal diagrams**

In epidemiology, the use of causal diagrams has become increasingly common over the last 12 years, following the publication of an introduction to the topic by Greenland et al. (2). In these causal diagrams, as in SEMs, we typically make assumptions about the relations the covariates have with one another. How then do these causal diagrams or directed acyclic graphs relate to SEMs?

Causal diagrams, as formalized by Pearl (4, 5), are a graphical representation of very general nonparametric structural equations, causally interpreted. Unlike traditional SEMs, causal diagrams do not make distributional assumptions or assumptions about functional form—no linearity assumptions are made with causal diagrams. These causal diagrams effectively developed out of the SEM literature (5), but the assumptions are far weaker with causal diagrams than with traditional SEMs. Causal diagrams in epidemiology have been used not as a statistical tool but rather...
as a conceptual tool. No assumptions are made about distribution or functional form; no model is fitted to the data. Rather, epidemiologists have used these diagrams to reason about confounding—specifically about whether conditioning on a particular set of covariates suffices to control for confounding. The so-call backdoor path criterion (2, 4) is most often employed, but a complete set of rules is now available, describing when a particular set of covariates suffices to control for confounding (6) or when a causal effect is identified from data (7).

Two general points have received considerable emphasis in this literature on causal diagrams that are also relevant to traditional SEMs but are, unfortunately, often neglected when SEMs are employed. First, every absence of an arrow on the diagram is making a strong assumption (5). If we show an arrow on the diagram where a connection is absent, we are being conservative in our inferences about confounding control; but if we leave an arrow off of the diagram when a connection is in fact present, we can draw the conclusion that we have controlled for confounding when in fact we have not. There are many arrows missing in Arlinghaus et al.’s Figure 1. I have already commented on the absence of arrows from gender, race, education, and occupation to sleep; but none of these variables have an arrow pointing to psychological distress either, or to any of the variables related to psychological distress (sadness, nervousness, restlessness, etc.). Nor is there an arrow pointing from body mass index to psychological distress (or vice versa), from age to psychological distress, from age to body mass index, or from gender or race to occupation or education. Do we really believe all of these arrows are absent? Arlinghaus et al. do not comment. The arrows that are present seem plausible; the arrows that are absent less so. With SEMs, as in causal diagrams, the absence of any arrow needs to be justified on substantive grounds. Too often with SEM applications, the model that is used corresponds to what the investigators feel are the most important arrows, with little attention given to what arrows are assumed to be absent.

The literature on causal diagrams has also emphasized that in order to employ the rules about confounding and interpreting estimates as causal, the diagram must be such that any common cause of 2 variables on the graph must also be on the graph (5). Otherwise, confounding structures may be present that are unrepresented by the diagram, and consequently reasoning from the diagram may well be invalid. Likewise, with SEMs, causal interpretation of the effect estimates requires that any common cause of 2 variables on the graph must also be on the graph. Is it likely that we have no unmeasured common causes for the model shown in Arlinghaus et al.’s Figure 1? Might “number of children” be an unmeasured common cause of both working hours and hours of sleep? Might “exercise” or “physical fitness” affect both hours of sleep and injury?

Unmeasured confounding is, of course, pervasive in epidemiologic studies. It is generally unavoidable, but we can at least attempt to assess its importance. This is typically done through sensitivity analysis techniques. When we have a single exposure and a single outcome in view, the implementation of such sensitivity analysis techniques is quite straightforward (8–10). With an SEM, however, in which each effect estimate essentially depends on the entire structure of the model, it becomes less clear how such techniques can be implemented.

Unmeasured common causes are often represented on causal diagrams; such unmeasured common causes can, and should, also be represented on an SEM. In fact, under assumptions of linearity and normality, statistical inference using SEMs can sometimes still be done even with these unmeasured common causes (11)—but not always. The effects of interest are not always identified when there are unmeasured common causes. Some sufficient conditions for identification have been given in the SEM literature (11), but even for linear systems, a complete characterization of which effects are identified when there are unmeasured variables in the SEM is in fact still an open problem (12). We arguably do not understand all of the intricacies of these models as the simple graphical picture with nodes and arrows might suggest. In any case, careful thought needs to be given to unmeasured common causes when formulating an SEM, just as needs to be done with other causal diagrams. This too, unfortunately, is often a step that is neglected when SEMs are used in practice.

Relation to causal mediation analysis

Arlinghaus et al. make another argument for using SEMs. They state that the SEM methods will allow them to identify the direct and indirect effects of weekly working hours on injury as mediated through sleep versus other pathways (1). Methods from the causal inference literature have also recently been developed to answer these questions of mediation and direct and indirect effects (13–22), and in fact, these methods arguably also developed out of the SEM literature (22, 23). However, unlike the SEM literature, these techniques from causal inference, by relying on counterfactual-based definitions of direct and indirect effects, can much more easily accommodate interactions between the exposure and mediator and/or other variables and nonlinearities. These methods from causal mediation analysis are also more akin to the traditional regression-based approaches for the effect of a single exposure insofar as they focus on a particular exposure-mediated-outcome relation and also focus specifically on the assumptions needed to identify the particular direct and indirect effects of interest. SEMs, as traditionally employed, focus on a host of different effects and as a result make a whole host of assumptions across all of the variables on the diagram; but they do, in turn, deliver every possible mediated or path-specific effect in which one might be interested.

An emphasis of the causal inference literature on mediation has been in order to interpret direct and indirect effects estimates causally, one needs to condition on a set of covariates that suffices to control for not just exposure-outcome confounding but also mediator-outcome confounding and exposure-mediator confounding. Sensitivity analysis techniques have been developed in the causal inference literature for assessing the extent to which inferences about direct and indirect effects are influenced by violations of these assumptions (18, 19). Biases, sometimes
quite severe, can arise when control has not been made for such exposure-mediator or mediator-outcome confounders (24–25). Such biases are relevant to SEMs as well. In Arlinghaus et al.’s Figure 1, working hours is taken as the exposure, hours of sleep as the mediator, and injury as the outcome (1). I noted above that number of children might confound the relation between working hours and hours of sleep (an exposure-mediator confounder). Likewise, exercise or physical fitness might affect both hours of sleep and injury (a mediator-outcome confounder). Such unmeasured confounders could substantially change both direct and indirect effect estimates. This problem of unmeasured confounding in the assessment of direct and indirect effects is certainly not unique to the SEM approach, but it needs to be considered and taken seriously, as it does in the counterfactual-based approaches. However, the potential for these types of assumptions to be violated grows as the size of the SEM, and the number of possible mediated effects, increases. These problems thus become more severe, not less severe, with SEMs; and unfortunately, these problems of confounding for direct and indirect effects are often ignored when SEMs are used in practice.

**Relation to MSMs**

There has likewise been growing interest in estimating the effects of time-varying exposures in epidemiology. New techniques, such as marginal structural models (MSMs) (3), have been developed to address the challenges that can arise in these settings, such as time-dependent confounding. SEMs can also be used to estimate the effects of time-varying exposures in some, but not all, of the settings in which MSMs can be employed (26–28). However, even when SEMs can be used, they make assumptions about more variables than do MSM techniques (28). This is essentially because MSMs target a particular set of effects of interest and make only the assumptions needed to identify and estimate those effects. SEMs allow for the estimation of more effects than do MSMs, but they also make more assumptions than MSMs. This point is discussed further elsewhere (28).

**DISCUSSION**

Additional challenges with SEMs arise when using cross-sectional data. We are no longer sure about temporal ordering and the direction of causality. As Arlinghaus et al. noted (1), a bidirectional influence between sleep and body mass index seems likely (29, 30). However, the arrows between hours of sleep and injury or between working hours and injury might also plausibly run in the opposite direction. Might we not expect persons with injury to be unable to work for some time or to have more trouble sleeping? And these potentially bidirectional sets of influences now concern the variables that are most central to the questions of interest to the investigators. With gender or race or maybe even education, we may be quite sure of the direction of causality on the grounds of temporality; but with other variables and relations, directionality is less clear, and mutual influence is likely. Without longitudinal data, we cannot sort out these questions. Similar questions about directionality also arise in an SEM’s latent-variables models, such as the psychological distress latent variable in Arlinghaus et al.’s Figure 1. An arrow in this diagram goes from psychological distress to feelings of worthlessness; but if someone were to be repeatedly told that he or she was worthless, would we not expect psychological distress to increase? Might the arrow then point in the opposite direction? And might we also expect an arrow from worthlessness to feelings of sadness and to restlessness? The causal interpretation of these aspects of an SEM is often not very clear.

Structural equation modeling constitutes a powerful tool; it allows for estimation of numerous effects—direct, indirect, total, path-specific—all across the diagram. But the power of these models comes at the cost of assumptions. The linearity, distributional, and no-confounding assumptions are made not simply for 1 specific outcome and 1 specific relation, but for all variables on the diagram, across the entire SEM. Of course, we need to make assumptions in our epidemiologic analyses; we cannot avoid it. However, we usually would opt for fewer assumptions rather than more, and then use sensitivity analysis techniques (8–10) for assumptions about which we were especially concerned. SEMs opt for more assumptions rather than fewer, but they deliver more as a result.

So when should SEMs be used in epidemiology? I would argue that, in light of the strong assumptions made, they should be used only when 1) we truly are interested in a wide range of different effects and pathways across an entire set of variables for several different outcomes and 2) we are using them principally for exploratory and hypothesis-generating purposes. If we are interested in only the effect of a single fixed exposure, then traditional regression-based approaches deliver what is desired with fewer assumptions. If we are interested in simply 1 specific exposure-mediator-outcome relation (e.g., working hours–sleep–injury), then methods from the causal mediation analysis literature allow for greater flexibility, focus on the relevant no-confounding assumptions, and are readily amenable to sensitivity analysis techniques. If we want to evaluate a host of mediated effects for different outcomes all at once, then in this case we may want to consider an SEM; but we should only do so if we are willing to accept that the results will be accompanied by a host of assumptions and, on these grounds, in most cases are going to be very tentative and easily subject to criticism. For exploratory and hypothesis-generating purposes, such methods can still be useful. However, if we want to carry out the most rigorous possible analysis that our data and the epidemiologic toolkit allow, we should probably turn elsewhere.

**ACKNOWLEDGEMENTS**

Author affiliation: Departments of Epidemiology and Biostatistics, Harvard School of Public Health, Boston, Massachusetts (Tyler J. VanderWeele).
The author was supported by National Institutes of Health grant HD060696.
Conflict of interest: none declared.

REFERENCES