Dynamic Path Analysis in Life-Course Epidemiology

Michael Gamborg*, Gorm Boje Jensen, Thorkild I. A. Sørensen, and Per Kragh Andersen

* Correspondence to Michael Gamborg, Institute of Preventive Medicine, Øster Søgade 18, DK-1357 Copenhagen K, Denmark (e-mail: mga@ipm.regionh.dk).

Initially submitted February 22, 2010; accepted for publication December 20, 2010.

Life-course epidemiology seeks to better understand the mechanisms that lead to the development of chronic diseases. An example is the mechanism leading from body size to coronary heart disease (CHD); one way to acquire a better understanding of this mechanism is to investigate to what extent it works through other risk factors. In this paper, the dynamic path analysis model is presented as a tool to analyze these dynamic mechanisms in life-course epidemiology. A key feature of dynamic path analysis is its ability to decompose the total effect of a risk factor into a direct effect (not mediated by other variables) and indirect effects (mediated through other variables). This is illustrated by examining the associations between repeated measurements of body mass index (BMI) and systolic blood pressure (SBP) and the risk of CHD in a sample of Danish men between 1976 and 2006. The effect of baseline BMI on the risk of CHD is decomposed into a direct effect and indirect effects going through later BMI, concurrent SBP, or later SBP. In conclusion, dynamic path analysis is a flexible tool that by the decomposition of effects can be used to increase the understanding of mechanisms that underlie the etiology of chronic disease.

blood pressure; body mass index; epidemiologic methods; heart diseases; path analysis; survival analysis

Abbreviations: BMI, body mass index; CHD, coronary heart disease; SBP, systolic blood pressure.

Life-course epidemiology seeks to better understand the development of chronic diseases by studying long-term mechanisms leading to disease. The statistical methods commonly used in current epidemiology have important recognized limitations in this regard, which the dynamic path analysis by simultaneously modeling several aspects of a dynamic process may overcome.

An example of a mechanism that needs further examination is the process leading from body size to incidence of coronary heart disease (CHD). Both current and earlier body sizes are associated with the risk of CHD (1–4). Body size is also associated with a number of CHD risk factors, such as elevated systolic blood pressure (SBP). Although studies have shown weight reduction to be associated with an improvement in the CHD risk factor profile, including SBP (5–8), the benefit of weight reduction on CHD remains uncertain (9–13). It is therefore important to acquire a better understanding of the complex mechanism leading from body size to incidence of CHD. One way of accomplishing this is first to investigate to what extent the association between body mass index (BMI) and the risk of CHD is mediated through other risk factors, such as SBP, and second to investigate to what extent the association between change in BMI and the risk of CHD is mediated through the level or change in level of other risk factors, such as SBP.

These challenges call for the combination of path analysis and survival analysis. By use of ordinary path analysis (14, 15), it is possible to decompose the total effect of an exposure on a normally distributed outcome into a direct effect and indirect effects mediated through other risk factors as discussed by Baron and Kenny (16). Survival analysis is needed to move beyond the normally distributed outcome to a time-to-event outcome. The recently developed dynamic path analysis model (17) combines path analysis and survival analysis.

Our aim in the present paper is to present the dynamic path analysis model as a tool to analyze and visualize dynamic mechanisms in life-course epidemiology. This is illustrated by examining the association between repeated measurements of BMI and SBP and the risk of CHD in a sample of Danish men. Our main focus is illustration of the dynamic path analysis model, whereas an in-depth analysis...
would need to consider potential confounders and mediators other than just BMI and SBP.

**MATERIALS AND METHODS**

**Sample and study design**

The Copenhagen City Heart Study provided repeated information on BMI and SBP in adulthood. The first examination was performed in 1976–1978 (18), the second was in 1981–1983 (18), and the third was in 1992–1994 (19). The study consists of randomly selected age-stratified samples of subjects who lived in a well-defined area of Copenhagen. Phase 5 SBP was measured on the left upper arm with the subjects in a sitting position after at least 5 minutes’ rest. BMI was calculated from measured height without shoes and weight in light clothes. Measurements were taken at ages 20–80 years (Table 1). To remove the age effect and the possible effect of examination number on levels of BMI and SBP, they were transformed into age- and examination-specific \( z \) scores.

Information on CHD events was obtained by linking to the National Cause of Death Register and the National Hospital Discharge Register. Discharge diagnoses and causes of death were classified according to the International Classification of Diseases, Eighth Revision, before 1994 and the Tenth Revision thereafter. CHD was defined as codes 410.0–414.9 in the former and codes I20.0–I25.9 (chronic and acute ischemic disease including angina pectoris and myocardial infarction) in the latter. Primary and secondary diagnoses from in- and outpatients were considered.

For simplicity, the analysis was restricted to males. The follow-up period was restricted to ages 50–80 years. Since the methods used in this study estimate cumulative effects, inclusion of the few earlier events could cause problems. Follow-up of subjects began at the first examination or at the subjects’ 50th birthday, whichever came later. Follow-up ended on the date of the first CHD event; date of death, emigration, or loss to follow-up; the subjects’ 80th birthday; or December 31, 2006, the date of the last available update from the Cause of Death Register, whichever came first. Refer to the Web Appendix, which is posted on the Journal’s website (http://aje.oxfordjournals.org/), for further details. In total, 6,152 male subjects were at risk of having their first CHD at ages 50–80 years. The subjects were followed for 85,476 person-years, and 1,754 subjects had their first CHD event during follow-up.

**Statistical analysis**

Dynamic path analysis, as described by Fosen et al. (17), is a way of modeling time-dependent processes leading to an event. Dynamic path analysis allows the mutual associations between repeatedly measured covariates and the associations between these covariates and the rate of event to vary over time. In this study, associations between repeatedly measured BMI \( z \) scores and SBP \( z \) scores and the rate of a CHD event were examined. Dynamic path analysis allows the decomposition of an effect into a direct effect and indirect effects going through other covariates in the model.

The dynamic path analysis model is a combination of the Aalen additive hazards model and ordinary path analysis. The Aalen model is used to model the rate of CHD as a function of BMI and SBP, and the ordinary path analysis is used to estimate the association between BMI and SBP.

To decompose the total effect of BMI on the rate of CHD into indirect and direct effects, the indirect effect of BMI through SBP on the rate of CHD is estimated by combining the estimate of the effect of BMI on SBP from the ordinary path analysis with the estimate of the effect of SBP on the rate of CHD. This decomposition is the key feature of dynamic path analysis.

In dynamic path analysis, the Aalen model is chosen over the standard Cox proportional hazards model, because the Aalen model is a linear model that is essential for the decomposition of total effects into direct and indirect effects and because the Aalen model allows the effect of covariates to vary with age.

A path diagram is used to describe the assumed causal relations between the variables in a dynamic path analysis model. An arrow indicates that one variable is regressed on

the other variable. Although it is not possible to demonstrate causality on the basis of observational data, it is, for the simplicity of exposition, assumed that the path diagram of Figure 1 represents a full accounting of the causal factors involved in CHD. The path diagram in Figure 1 is a directed acyclic graph (20), directed because it shows the direction of associations and acyclic because it does not include feedback loops.

The Aalen additive hazards model. The Aalen additive hazards model was introduced by Aalen in 1980 (21) and further described in 1989 (22). It is a very flexible nonparametric survival model that is easy to fit (e.g., using the timereg package (23)) for the R statistical software package (24). It models the hazard rate for an event at each age \( t \) as a linear function of the covariates. This fundamental assumption of linearity must be checked. The Aalen model allows both the covariates themselves and the effect of these variables to change with age. The hazard rate for CHD is modeled as a linear function of baseline BMI and SBP and current BMI and SBP:

\[
\alpha(t) = \beta_0(t) + \beta_{B0}(t) \text{BMI}_b + \beta_{SB0}(t) \text{SBP}_b + \beta_{B1}(t) \text{BMI}_c(t) + \beta_{SB1}(t) \text{SBP}_c(t).
\]  

Here, \( \alpha(t) \) is the hazard rate of CHD at age \( t \), \( \beta_0(t) \) is the baseline hazard rate, here the hazard rate of an average individual (with all \( z \) scores equal to 0), and \( \beta_{B0}(t) \), \( \beta_{SB0}(t) \), \( \beta_{B1}(t) \), \( \beta_{SB1}(t) \) are the regression functions describing the effect of the covariates as a function of age \( t \). Thus, \( \beta(t) \) is the excess hazard rate at age \( t \) corresponding to a 1 \( z \)-score difference in the corresponding covariate holding all other covariates fixed. BMI\( b \) and SBP\( b \) are the baseline BMI and SBP (measurements are taken at the first examination and therefore do not change with age), and BMI\( c(t) \) and SBP\( c(t) \) are the current BMI and SBP as a function of age. When fitting the model, the most recent values, which are updated for each new examination, are used as proxies for the current value. Each of these 4 regression functions is represented by an arrow in the path diagram (Figure 1).

The cumulative regression function \( B(a, b) = \int_a^b \beta(t) dt \) (the area under the regression function from age \( a \) to age \( b \)) measures the cumulated excess rate from age \( a \) to age \( b \). If the cumulative regression function is small and competing risks are negligible (here competing risk is death from any cause other than CHD), then the cumulative regression function can be interpreted as the excess fraction of a population that will experience an event between age \( a \) and age \( b \) compared with another population if the populations have a 1-unit difference in the corresponding covariate and all other covariates fixed.

Figure 2 shows an example of a cumulative regression function in which the excess rate caused by a 1 \( z \)-score difference in baseline BMI is estimated. Because the function increases with age, it implies that a high baseline BMI causes an elevated risk of CHD. If the function had decreased, it would have meant that a high baseline BMI...
BMI is fairly constant across age, which means that the excess rate of CHD caused by baseline estimated cumulative regression function is fairly constant, causes a reduced risk of CHD. In the figure, the slope of the estimated cumulative regression function is fairly constant, which means that the excess rate of CHD caused by baseline BMI is fairly constant across age.

Ordinary path analysis. Ordinary path analysis is an extension of linear regression analysis. In this study, it is a joint analysis of the associations between BMI and SBP at different time points, with simultaneous estimation of 3 regression equations—one for baseline SBP, one for current BMI, and one for current SBP. The path model was based on the time structure of the data and on biologic knowledge. Thus, baseline SBP was assumed to depend on baseline BMI, current BMI was assumed to depend on baseline BMI, and current SBP was assumed to depend on baseline BMI, baseline SBP, and current BMI. To allow the effects to change with age, the path analysis model was fitted for each event age, including all subjects at risk for CHD at the given age:

\[ SBP_t = \alpha_{S_b}(t) + \alpha_{S_b,b}(t) \text{BMI}_b + \epsilon_1(t) \]  

\[ \text{BMI}_t = \alpha_{S_b}(t) + \alpha_{S_b,b}(t) \text{BMI}_b + \epsilon_2(t) \]  

\[ \text{SBP}_t = \alpha_{S_b}(t) + \alpha_{S_b,b}(t) \text{BMI}_b + \alpha_{S_s,b}(t) \text{SBP}_b + \alpha_{S_s,b}(t) \text{BMI}_b + \epsilon_3(t) \]

Here \( \alpha_{S_b,b}(t), \alpha_{S_b,b}(t), \alpha_{S_s,b}(t), \alpha_{S_s,b}(t), \) and \( \alpha_{S_s,b}(t) \) are regression functions or path coefficients as functions of

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Direct(^a) Cumulated Effect(^a)</th>
<th>Direct(^a) Cumulated Effect(^a)</th>
<th>Total Cumulated Effect(^a)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Body mass index(^d) (z score)</td>
<td>0.038 0.020, 0.055</td>
<td>0.008 0.005, 0.012</td>
<td>0.046 0.029, 0.063</td>
</tr>
<tr>
<td>Systolic blood pressure(^d) (z score)</td>
<td>0.038 0.020, 0.055</td>
<td>0.038 0.020, 0.055</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviation: CI, confidence interval.

\(^a\) The effects are estimated by using the dynamic path model described in Equations 3 and 5 and cumulated from 50 to 65 years of age. Because the cumulated effects are small and the risk of death from any cause other than coronary heart disease is negligible, the cumulated effects can be interpreted as the excess fraction of a population that will experience a coronary event between the ages of 50 and 65 years compared with another population if the populations have a 1 z-score difference in the covariate.

\(^b\) The direct effect of body mass index is adjusted for systolic blood pressure and vice versa.

\(^c\) The indirect effect of body mass index is the part of the effect that is mediated through systolic blood pressure.

\(^d\) Baseline measurements from 20 to 65 years of age.
age, and $\varepsilon_2$, $\varepsilon_3$, and $\varepsilon_4$ are mutually independent, normally distributed error terms with mean 0. Each of these 5 path coefficient functions is represented by an arrow in the path diagram (Figure 1). Note that it was assumed that current BMI did not depend on baseline SBP.

**Dynamic path analysis.** Performing dynamic path analysis includes simultaneous fitting of the Aalen model (Equation 1) and an ordinary path analysis model (Equations 2–4) for each event age by using data from all individuals at risk of a CHD event at the given age. By allowing all effects to be estimated for each age, the dynamic path analysis is very flexible. Dynamic path analysis even allows effects to disappear or appear at given ages. In this study, however, it was assumed that all the fitted ordinary path analysis models had the structure, described in the path diagram (Figure 1).

**Direct, indirect, and total effect.** When looking at the path diagram to acquire a better understanding of the mechanism leading from BMI to incidence of CHD, 2 questions come to mind. First, to what extent is the effect of BMI on the rate of CHD mediated through SBP, and second, to what extent is the effect of change in BMI (current BMI given baseline BMI) on the rate of CHD mediated through SBP?

The direct effect of a covariate is the part of the effect that is not mediated through other covariates in the model; the direct effects can be estimated by using the Aalen model (Equation 1).

An indirect effect (mediated through a potential mediator) is estimated at each event age by multiplying the appropriate regression functions. A simple example of a dynamic path analysis model is given by Equations 2 and 5 of the associations between baseline BMI and SBP and the rate of CHD.
As a 1-unit difference in baseline BMI causes a $S_b$; difference in baseline SBP, and a $S_b$; difference in baseline SBP causes a $S_b$; times $S_b$ (from Equation 5) excess rate of CHD, the indirect effect of BMI going through SBP at age $t$ is estimated by the following:

$$\alpha(t) = \beta_0(t) + \beta_{B_B}(t) \text{BMI}_t + \beta_{S_B}(t) \text{SBP}_t.$$  (5)

These age-specific indirect effects are then added by age to get the indirect cumulative regression function. In more complex models, Equation 6 is extended; if there is more than one mediator in the path from exposure to outcome, then all the regression functions along the path are multiplied to obtain the indirect effect along this path. If there is more than one possible path from exposure to outcome, then indirect effects along each path are added to obtain the overall indirect effect.

The total effect of a covariate on the rate can be estimated by using the Aalen model without adjustment for any potential mediating variables. Because of the linearity of the dynamic path model, the indirect effects and the direct effect sum to the total effect.

**RESULTS**

**Baseline BMI and CHD**

A natural starting point for the investigation of the effect of BMI on the rate of CHD is to examine the total effect of baseline BMI on the rate of CHD by using the Aalen model:

$$\alpha(t) = \beta_0(t) + \beta_{B_B}(t) \text{BMI}_t.$$  (7)

The cumulative regression function generally increased with age almost linearly (Figure 2), which demonstrates that a high baseline BMI caused a constant elevated rate of CHD at all ages.

**Baseline BMI and CHD mediated by baseline SBP**

To investigate if baseline SBP is a mediator of the effect of baseline BMI on the rate of CHD, we performed the dynamic path analysis in Equations 2 and 5. The direct
The effect of change in BMI and SBP on CHD

To investigate the potential effect of change in BMI and change in SBP on the rate of CHD, the current measurements of BMI and SBP were included in the dynamic path analysis model; this model (Equations 1–4) is illustrated by using the path diagram (Figure 1). Note that replacing the current measurements of BMI or SBP with the difference between current and baseline measurements does not change the regression functions. So the parameter $\beta_8$ (Equation 1) can be interpreted as both the effect of current BMI for a given baseline BMI and also as the effect of change in BMI between the baseline measurement and the current measurement for a given baseline measurement. The estimated cumulative regression functions of the Aalen model (Equation 1) are shown in Figure 4A–D. For each event age, the parameters in the ordinary path analysis model in Equations 2–4 were estimated, by using data from all subjects at risk for CHD at the given age. All effects in these regressions were significant at all ages (not shown).

The direct effect and the total effect of current SBP on the rate of CHD (Figure 4A) are identical, as there are no potential mediating variables between current SBP and CHD in the path diagram (Figure 1). This cumulative regression function generally increased with age, indicating that a high current SBP for a given baseline SBP, baseline BMI, and concurrent BMI (i.e., an increase in SBP for a given baseline SBP, baseline BMI, and concurrent BMI) caused an elevated rate of CHD. The cumulated direct effect of recent SBP for a given BMI and baseline SBP from age 50 to 65 years is estimated to be 0.030 (95% confidence interval: 0.007, 0.052).

The direct cumulative regression function of current BMI (Figure 4B, dotted line) decreased at first but then later increased (although not significantly), suggesting that an increase in weight for a given baseline BMI, baseline SBP, and concurrent SBP might cause a lower rate of CHD at younger ages and an elevated rate at later ages. The cumulated direct effect of recent BMI for a given SBP and baseline BMI from age 50 to 65 years is estimated to be $-0.019$ (95% confidence interval: $-0.058, 0.020$). The total effect of change in BMI for a given baseline BMI and baseline SBP (Figure 4B, solid line) includes the indirect effect going through a concurrent change in SBP. The difference between the direct and the total effect was small, indicating that the positive indirect effect of a change in BMI through change in SBP was small.

The total cumulative regression function of baseline BMI generally increased with age (Figure 4C, solid line). It also increased at a higher rate than the direct cumulative regression function (Figure 4C, dotted line). This indicated that a high baseline SBP for a given baseline BMI caused an elevated rate of CHD and that there was a positive indirect effect going through later SBP.

The total cumulative regression function of baseline BMI generally increased with age (Figure 4D, solid line), indicating that a high baseline BMI caused an elevated rate of CHD. The total effect can be decomposed into the positive direct effect illustrated by the direct cumulative regression function (Figure 4D, dotted line) and an indirect effect, which was small, as the direct and the total effect were similar in size. The indirect effect is a sum of the 5 indirect effects: a non-significantly negative indirect effect going through current BMI (Figure 5, medium dotted line; Table 3); and 4 paths going through baseline and/or current SBP. The sum of these indirect effects was positive (Figure 5, thick dotted line; Table 3).
DISCUSSION

This study presented an application of the recently developed dynamic path analysis as a tool in life-course epidemiology. This analytical method can assist in investigating the mechanisms behind the association between risk factors, especially changes in risk factors, and later morbidity or mortality. In this context, dynamic path analysis was a flexible method that allowed the effect of a risk factor or a change in the risk factor to be decomposed into indirect effects mediated through other variables in the model and a direct effect, not mediated by the other variables in the model. This separation has the potential to enhance our understanding of the mechanisms by which risk factors operate.

In dynamic path analysis, the Aalen additive hazards model for survival analysis was chosen over the more conventional Cox proportional hazards model. When one performs standard Cox regression, the assumption of proportionality is made; that is, the hazards of 2 subjects with different exposure status are assumed to be proportional with the same proportionality factor at all ages in the follow-up period. When one performs Aalen’s additive hazards regression, there is no proportionality-like assumption, as the effect of the exposure is allowed to vary with age. In standard Cox regression, the effect of a covariate on the rate is assumed to be linear on the logarithmic scale, whereas in Aalen regression, the effect of a covariate on the rate is assumed to be linear on the original scale. These assumptions are not identical because they assume linearity on different scales. Nonetheless, both are restrictive and require that an examination of linearity is performed, for example, by using splines (25), for results to be trustworthy.

Another reason (and the main reason why the Aalen model was chosen) was because it is a linear model as opposed to the Cox model, which is multiplicative. The choice of a linear survival model is essential to decompose the total effect of an exposure into direct and indirect effects, as it is not possible to obtain a similar decomposition when using nonlinear survival models like the Cox model (17).

The dynamic path analysis model illustrated in Figure 1 is obviously a simplistic model of the mechanism leading from BMI to later CHD. First, the model does not include any potential confounders of the association between BMI and the risk of CHD, such as diet, genes, physical activity, smoking, or social position. If the total effect of BMI on the rate of CHD is to be considered causal, one of several assumptions is that there are no unmeasured covariates that are a cause of both BMI and CHD. Second, the model does not include any potential confounders of the association between SBP and the rate of CHD. If unmeasured confounders of this association exist, the decomposition of the total effect into a direct effect and an indirect effect through SBP could be biased (26). Third, SBP acts as a marker for the general CHD risk profile and not just as SBP per se, as SBP is correlated with a number of other risk factors for CHD. This implies that a part of the estimated indirect effect of BMI on the rate of CHD going through SBP might actually be going through other risk factors for CHD. To address these problems, potential confounders and other risk factors for CHD could be included in the model. Fourth, it is assumed that there is no interaction among BMI, SBP, and the occurrence of CHD. These issues are beyond the scope of this paper, as the aim here is to illustrate the usefulness of the dynamic path analysis model in life-course epidemiology and, to achieve this, a simple model was deliberately chosen.

The broad definition of CHD used in this study increases statistical power, but if the associations between BMI and the different conditions included in CHD are not the same, it causes problems when interpreting the estimates. The most recent measurements of BMI and SBP are used to approximate the current values, which could cause bias. Refer to the Web Appendix for a discussion of this topic.

The data example analyzed showed that baseline BMI is positively associated with the rate of CHD. The decomposition of this effect showed that, even though both baseline and later SBP were influenced by the baseline BMI, the part of the effect of baseline BMI that was mediated through baseline and later SBP was, although statistically significant, small compared with the total effect of baseline BMI. This suggests that SBP mediates only a minor part of the excess rate of CHD associated with a high BMI and that the negative effect of a high BMI cannot be prevented by only monitoring and treating hypertension.

The change in BMI over a 5–15 year period was not associated with the rate of CHD. This is counterintuitive, as baseline BMI is associated with the rate of CHD independent of at what age it was measured. It may suggest that preventing overweight and obesity is more beneficial than treating it. A change in BMI is positively associated with SBP, which in turn is positively associated with the rate of CHD. Therefore, the indirect effect of change in BMI is positive as expected from the positive association between baseline BMI and the rate of CHD.

In conclusion, dynamic path analysis is a flexible tool that can be used for the analysis of dynamic processes. It can, by decomposition of the effect of a risk factor or the effect of change in a risk factor into indirect and direct effects, increase the understanding of the very complex social, behavioral, and biologic mechanisms that underlie the etiology of chronic disease, which eventually may improve the evidence base for management of conditions such as obesity and hypertension.

ACKNOWLEDGMENTS

Author affiliations: Institute of Preventive Medicine, Copenhagen University Hospital, University of Copenhagen, Copenhagen, Denmark (Michael Gamborg, Thorkild I. A. Sørensen); Department of Occupational and Environmental Medicine, Bispebjerg University Hospital, Copenhagen, Denmark (Gorm Boje Jensen); Department of Cardiology, Hvidovre University Hospital, Hvidovre, Denmark (Gorm Boje Jensen); and Department of Biostatistics, University of Copenhagen, Copenhagen, Denmark (Per Kragh Andersen).

This work was supported by the Danish Heart Foundation (grant 05-10-B86-A795-22240F), the Danish Graduate
School in Biostatistics, and the Faculty of Health Sciences of the University of Copenhagen.

The authors thank Dr. Jennifer L Baker for her scientific contributions to this manuscript.

Conflict of interest: none declared.

REFERENCES