

ALCOHOL CONSUMPTION AND ISCHEMIC HEART DISEASE MORTALITY: ARE TIME-SERIES CORRELATIONS MEANINGFUL?

By: [Harvey W. Gruchow](#), Alfred A. Rimm, and Raymond G. Hoffmann

Gruchow HW, Rimm AA, and Hoffmann RG: Alcohol consumption and ischemic heart disease mortality: are time-series correlations meaningful? *American Journal of Epidemiology* , 118:641-650, 1983.

Made available courtesy of Oxford University Press: <http://www.oup.com/us/>

***** Note: Figures may be missing from this format of the document**

Abstract:

Recently, time-series correlations of aggregated data have been used to demonstrate the length of latency periods for environmental factors, such as economic conditions and alcohol consumption, in influencing heart disease mortality. Latency periods were specified by lagging mortality rates relative to the economic indicators or rates of alcohol consumption until the highest correlations were achieved. The tendency has been to interpret these correlations without regard to whether the latency periods described are biologically plausible. The authors have identified four models which represent all the possible outcomes of correlational studies of time-series data. Using United States and Canadian mortality rates in relationship to alcohol consumption, they have demonstrated the application of each of these models. For three of the four models, the time-series (lag) correlations are uniform regardless of the number of years mortality is lagged relative to alcohol consumption, and this uniformity does not permit a latency period to be identified. Only the lag correlations between two nonlinear variables show variations over time, depending on the degree of correspondence between the increasing and decreasing line segments of the two curves. Correlations ranging from high positive to high negative are possible, and several peak correlations (positive and negative) can occur. However, the biologic interpretation of multiple peaks with the same or different signs is problematic. The authors conclude that time-series correlations of aggregated data are not useful for the study of latency periods, and that analysis of time-series correlations for this purpose can be at best ambiguous, and at worst, completely misleading.

Keywords: alcoholic beverages; arteriosclerosis; beer; epidemiologic methods; heart diseases; statistics; wine

Article:

Variations in disease patterns, by place and time, are of fundamental interest to epidemiologists in the search for causal factors. Often, the first indications of a potential risk factor for a disease are provided by statistical associations based on ecologic analyses of aggregated data for different regions or different time periods. However, the interpretations which can be made of relationships found in aggregated data are very limited (1-4).

Recently, time-series correlations of aggregated data have been used to try to demonstrate the length of latency periods between environmental factors, such as economic conditions and alcohol consumption, and ischemic heart disease mortality. The lengths of the latency periods were specified by aerially lagging annual mortality rates relative to the annual risk factor rates until the highest correlations were achieved. A latency of zero to five years was suggested between economic downturns and increased ischemic heart disease mortality (5, 6) and five years between increased alcohol consumption and the corresponding decrease in ischemic heart disease mortality (7).

The studies of economic conditions related to heart disease have been criticized on both the legitimacy of the methods used and on the appropriateness of the etiologic models employed (8, 9). Although no commentary has been published on the ischemic heart disease-alcohol time-series study, similar criticisms apply. Also, common to these time-series studies of environmental factors and ischemic heart disease has been the absence of evidence that the latency periods described are biologically plausible.

Here, we present evidence, based on time-series analysis of ischemic heart disease mortality rates in relationship to alcohol consumption, to support these criticisms. Furthermore, we have identified four models which represent all the possible outcomes of correlational studies of time-series data. From these models, we conclude that time-series correlations of aggregated data are not useful for the study of latency periods, and that analysis of time-series correlations for this purpose can be at best ambiguous, and at worst, completely misleading.

SOURCES OF DATA

The data for this study were obtained from United States and Canadian government' published statistics. Per capita alcohol and cigarette consumptions for both countries were estimated from annual sales, based on the adult populations 15 years of age and older (10-12). Alcohol sales volumes were converted to absolute alcohol equivalents using the following proportions: beer, 4.5 per cent; wine, 15 per cent; spirits, 45 per cent.

Age-adjusted United States and Canadian ischemic heart disease mortality rates for the years 1950 through 1976 were excerpted from governmental publications (13, 14). In addition, four other causes of death were selected for comparative study because their rates exhibited different trends during this same time period. These other causes were cerebrovascular disease, lung cancer, rectal cancer, and cirrhosis of the liver. All causes of death were classified according to the Eighth Revision of the *International Classification of Diseases, Injuries and Causes of Death*, 1969. Detailed list codes used for this study were: ischemic heart disease, 410-414; cerebrovascular disease, 430—438; lung cancer, 162; rectal cancer, 154; cirrhosis, 571.

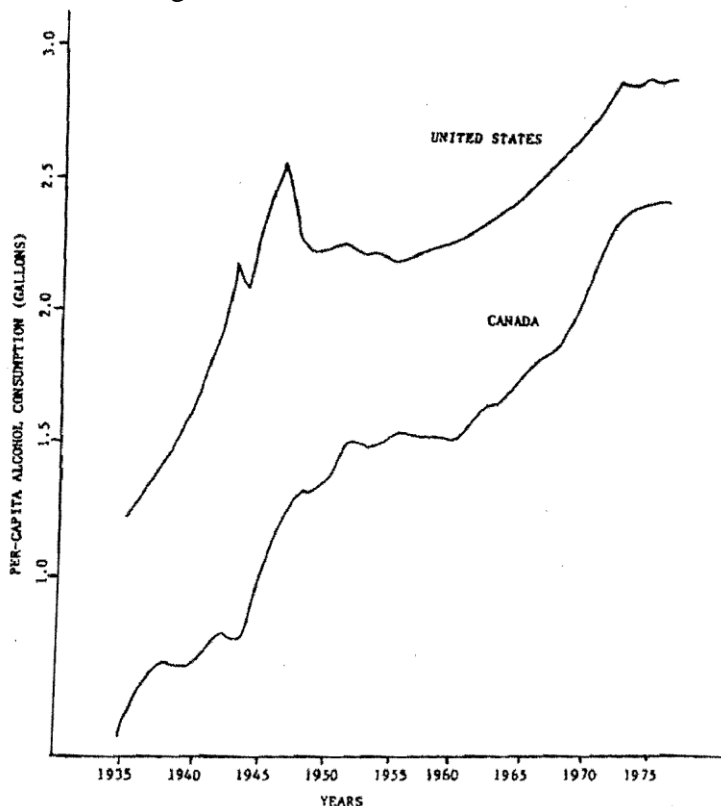


FIGURE 1. Per capita alcohol consumption (ethanol equivalent) for adults 15 years and older, United States and Canada, 1935-1976.

ALCOHOL CONSUMPTION PATTERNS

The annual rates of per capita total alcohol consumption for the United States and Canadian populations are shown in figure 1. In both countries, there has been an overall upward trend in consumption since the mid-1930s, with consumption in the United States substantially higher than that in Canada throughout this period. The only other remarkable difference between the two curves is the pronounced short-term increase in United States rates in 1942-1944, followed by a sharp decrease in 1945-1946. A similar short-term fluctuation was not observed in the Canadian curve. However, in both countries, there was a leveling of consumption rates during the 1950s, followed by an upward surge in the 1960s, and by an apparent further leveling in the 1970s.

ISCHEMIC HEART DISEASE MORTALITY PATTERNS

The rates of ischemic heart disease mortality in the United States and Canada since 1952 are presented in figure 2. The rates in both countries generally increased until the mid-1960s, and decreased thereafter. United States rates were substantially higher than Canadian rates throughout the period 1952-1976, and the slope of the increase in mortality prior to 1965-1967, as well as the slope of the decrease after 1965-1967, was greater for the United States population.

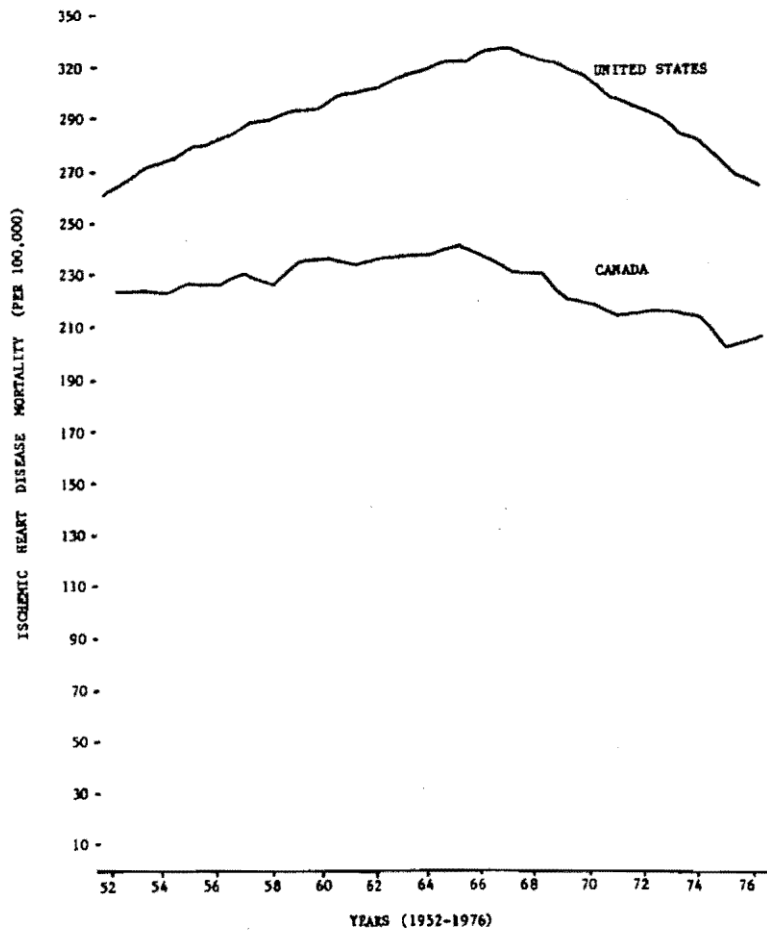


FIGURE 2. Age-standardized annual mortality rates (per 100,000) from ischemic heart disease, United States and Canada, 1952-1976.

TIME-SERIES CORRELATIONS

The time-series correlations between alcohol consumption rates and ischemic heart disease mortality rates for both countries are shown in figure 3. For the United States population, the correlations are extremely variable. Mortality is only moderately correlated (negatively) with consumption of the same year (zero year's lag). However, when mortality is correlated with consumption rates of preceding years (lag correlations), the coefficients are negative for one- to 10- year lags, and positive for 10- to 20-year lags. The strongest negative correlation is for a lag of six years ($r = -0.75$), and the strongest positive correlation is for a lag of 16 years ($r = 0.50$). In other words, alcohol consumption rates are negatively correlated with ischemic heart disease mortality up to 10 years later, but are positively correlated with ischemic heart disease mortality rates 10 to 20 years later.

For the Canadian population, the correlations were negative throughout the lag period, but there was a marked decrease in the strength of the negative correlation between five and 20 years' lag, with the weakest negative correlation at 13 years ($r = 0.32, p > 0.10$).

The ambiguity of the United States ischemic heart disease mortality-alcohol lag correlation pattern raises several questions concerning the appropriate interpretation of these correlations. LaPorte et al. (7) based their

conclusion of a five-year latent period for the protective effect of alcohol on ischemic heart disease using the largest negative lag correlation for beer consumption rates in the United States ($r = -0.94$ at five years).

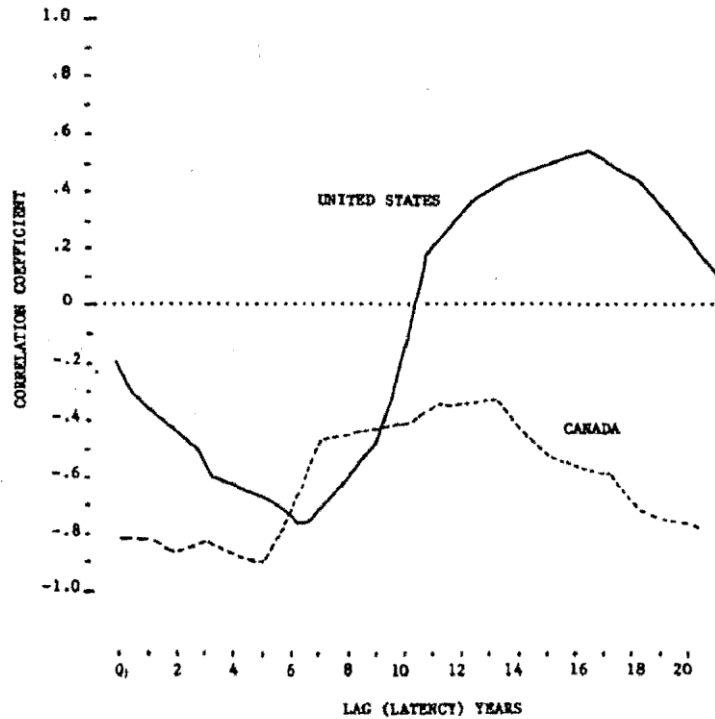


FIGURE 3. Time-series (lag) correlations for ischemic heart disease mortality with per capita alcohol consumption, for the United States (solid line) and Canada (broken line). Mortality rates between 1952 and 1976, and alcohol consumption rates between 1935 and 1976 were used. Lag years, used in computing the correlation coefficients, represent the number of years between mortality and previous consumption.

However, by studying various lag periods, we observed that the correlations between ischemic heart disease mortality and both beer and total alcohol consumption for lag periods longer than 10 years were positive, with the highest positive correlation for beer consumption at 13 years ($r = 0.70$, data not shown). Since this high positive correlation was also significantly different from zero ($p < 0.01$), any etiologic interpretation of the time-series between ischemic heart disease mortality and alcohol consumption would have to account for positive correlations after 10 years' latency as well as negative correlations for shorter latency periods.

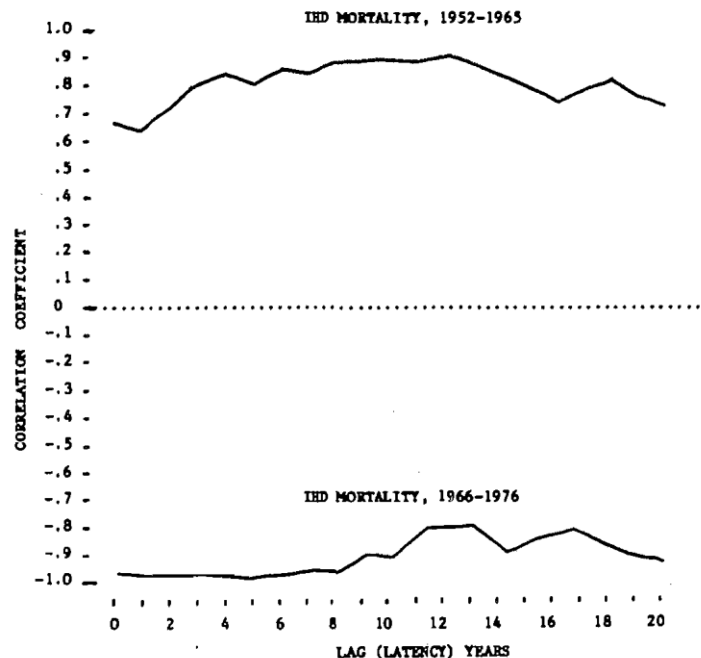


FIGURE 4. Time-series (lag) correlations for ischemic heart disease (IHD) mortality with per capita alcohol consumption, for Canada, for two different periods of mortality: 1952-1965 (top curve) and 1966-1976 (bottom curve).

The purpose of computing lag correlations is to determine whether a trend in one variable (i.e., alcohol consumption) can be related to a trend in another variable (i.e., ischemic heart disease mortality) at a later time, and to determine the length of the latent period, which is identified by the lag period producing the highest correlation. However, despite the apparent logic in this approach, it is severely limited in its usefulness because the strength and the sign of the lag correlation coefficient are determined entirely by the relative trends in the variables being correlated, but not by their relative rates of change. To illustrate this limitation, we computed separately the lag correlation for pre- and post-1965 ischemic heart disease mortality rates for Canada with alcohol consumption (figure 4). The lag correlations between post-1965 ischemic heart disease mortality and alcohol consumption were strongly negative ($p < 0.01$), while the lag correlations for pre-1965 ischemic heart disease mortality were strongly positive ($p < 0.01$).

Another serious problem with using the method of relating lags of one time-series with another time-series (this is equivalent to examining the cross-correlation function between the two series) is that if the series do not have the trend removed, then by chance alone there is a very high probability of a statistically significant lag or lead correlation. This problem was noted by Box and Newbold (15) in a discussion of the relationships of the Financial Times Ordinary Share Index and the United Kingdom Car Production.

This situation occurs whenever the two series have memory. For example, alcohol consumption this year depends on alcohol consumption last year, since most of the same people are still consuming alcohol. Similarly, mortality due to heart disease this year depends on the number of people who are ill from the disease—a quantity that is fairly constant since only a small percentage of those with ischemic heart disease die in a given year. A simple way to model a process with a reasonable amount of memory is to use a first-order regressive process (16), $Y(t) = B*Y(t-1) + e(t)$, where B represents the memory factor (0-1.0), $Y(t)$ is this year's alcohol consumption, $Y(t-1)$ is last year's alcohol consumption, and $e(t)$ is a random change. Memory is represented by the correlation from one year to the next; the memory is 0.88 for ischemic heart disease mortality and 0.96 for alcohol consumption.

To demonstrate this effect of memory in a time-series, 20 totally unrelated time-series (of length 26) were generated using the autoregressive model and specifying B as 0.90. Some had an increasing trend, some had a decreasing trend, and some had a trend that increased and decreased. Examining the cross-correlations among these series shows that 95 per cent had a maximum lag correlation greater than 0.4, 75 per cent had a maximum lag correlation greater than 0.5, and 25 per cent had a maximum lag correlation above 0.6. In other words, whether unrelated series have similar trends or not, if they have memory, there will be a significant lag correlation.

The dependency of lag correlations on the relative trends being compared results in a finite set of outcomes for such comparisons. We have devised four models which represent all possible combinations of trends and resulting out-comes. Each model consists of three components: the trends of each of the two variables, and time.

MODEL I: SIMILAR LINEAR TRENDS

In Model I (figure 5A), both variables have similar trends over time. As long as both trends are always increasing (or decreasing), the lag correlations between these variables will be strongly positive, regardless of the individual slopes of the curves, or the length of time between the curves. For example, two variables simultaneously increasing (or decreasing) at the same rate would be as highly correlated as two variables simultaneously increase (or decreasing) at different rates, and since the trends are constant, the same degree of correlation would be found between the variables regardless of the length of time between the two curves. This is illustrated by the trends of cirrhosis mortality and lung cancer mortality in relationship to alcohol consumption. Although lung cancer mortality increased at a much faster rate between 1952 and 1976 than cirrhosis mortality, there is virtually no difference between these two forms of mortality in their lag correlations with alcohol consumption. For neither form of mortality can a probable latent period be specified, based on these consistently strong lag correlations. Time-series correlations of lung cancer mortality with per capita cigarette

consumption show similarly high correlation coefficients which can also preclude identifying a latent period, although cigarette smoking is a well established risk factor for lung cancer.

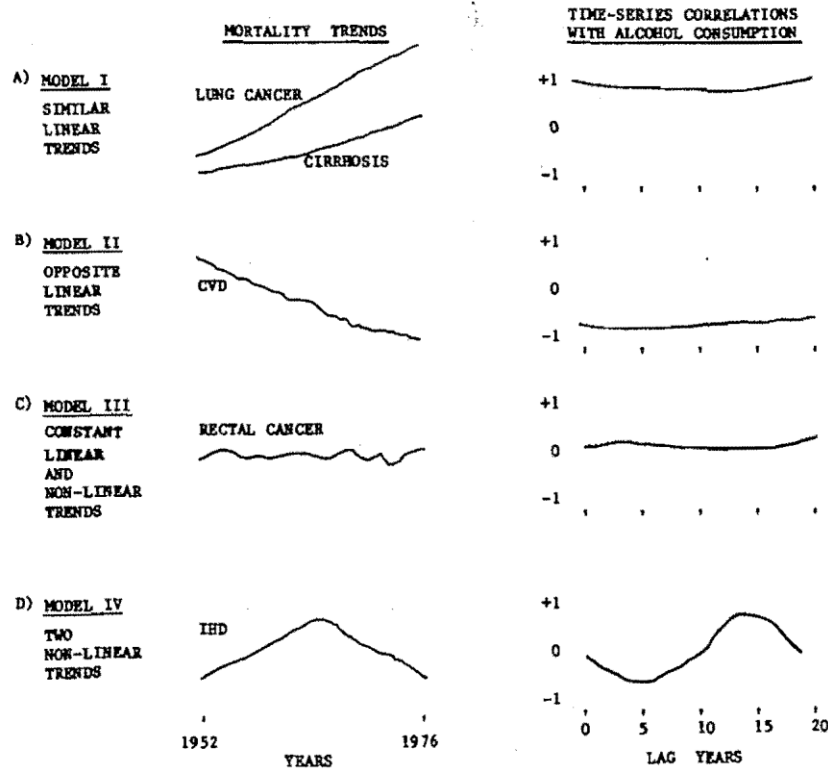


FIGURE 5. Four models representing all possible outcomes of time-series correlations, with examples of each model based on correlations between cause-specific mortality rates and per capita alcohol consumption or per capita cigarette consumption. *Model I*: similar linear trends; cirrhosis mortality and lung cancer mortality with alcohol consumption, for Canada, 1952–1976. *Model II*: opposite linear trends; cerebrovascular disease (CVD) mortality with alcohol consumption, for Canada, 1952–1976. *Model III*: constant linear and nonlinear trends; rectal cancer mortality and per capita constant-tar cigarette consumption, for Canada, 1952–1976. (See text for explanation of “constant-tar”.) *Model IV*: two nonlinear trends; ischemic heart disease (IHD) mortality and alcohol consumption, for United States, 1952–1976.

MODEL II: OPPOSING LINEAR TRENDS

In Model II (figure 5B), the two variables have opposing time trends (increasing vs. decreasing) and the correlations would be strongly negative regardless of the latency period. Cerebrovascular disease mortality and alcohol consumption illustrate the negative correlations between two variables with opposite time trends. Again, it is not possible to detect a latent period from the uniformly strong negative lag correlations.

MODEL III: CONSTANT LINEAR VS. NONLINEAR TRENDS

Model III (figure 5C) represents the situation where one variable is constant (neither increasing nor decreasing) over time in relationship to either a linear or fluctuating trend in the other variable. Because there is no change in the constant, the correlations would be zero, regardless of the slope(s) of the second variable, or of the lag period. The closest approximation to this model among disease entities is the relationship between rectal cancer mortality and constant-tar cigarette consumption. Rectal cancer mortality changed little between 1952 and 1976, but cigarette consumption, corrected for the reduced tar content in recent years (“constant-tar”), increased until about 1966, then decreased. None of the lag correlations is significantly different from zero, and no latent period is detectable.

MODEL IV: TWO NONLINEAR TRENDS

The relationship of two nonlinear variables is shown in Model IV (figure 5D). The correlation between variables of this type depends on the degree of correspondence between the increasing and decreasing line segments of the two curves. If increasing (or decreasing) segments of the two curves correspond directly, a high

positive correlation will result, whereas if increasing segments in one correspond to decreasing segments in the other, a high negative correlation will result. Intermediate degrees of correspondence will result in lesser correlations. Therefore, lag correlations for variables of this type will vary from positive through zero to negative, depending on the lag period used. The lag correlations between ischemic heart disease mortality and alcohol consumption illustrate this type of relationship (cf. figure 3). Although it is possible to identify peak negative and positive lag correlations between two variables of this type, the determination of latency period is problematic. Since both negative and positive correlations occur, either an interpretation which can reconcile both prevention and promotion of a disease by the same etiologic factor must be constructed, or additional information must be used to distinguish between these two possible opposing effects. Therefore, in spite of the fact that the time-series correlations between two variables that conform to this model will have the highest degree of variation between different lag periods, the correlations by themselves add nothing unique to our knowledge of the possible etiologic relationship.

DISCUSSION

Limitations on the interpretation of ecologic correlations have been previously mentioned (1-4). These limitations stem from the aggregated nature of the data, and the fact that it is not possible to demonstrate a statistical association between the behavior of the individual units in a population and the behavior of the aggregated units. We have demonstrated further that time-series correlations of ecologic data have a very limited number of outcomes, determined by the relative time trends of the variables being compared, and are not useful for determining latency between supposed cause and effect.

The use of time-series correlations to identify latency periods between the action of a risk factor and manifestation of a disease results in additional problems of interpretation. Are there underlying biologic principles which can explain the correlations obtained, and the length of the latency period indicated? When there is indication of two or more latency periods (i.e., the lag correlation curve has more than one peak), how are they to be distinguished, and what etiologic interpretation can be made of several peak correlations of the same or opposite *signs*? It is clear that if one refers specifically to the time-series correlations between ischemic heart disease mortality and alcohol consumption for the United States (cf. *figure 3*), the peak lag correlations at six years (negative) and at 16 years (positive) make biologically plausible interpretations difficult. A period of six years as the latency period for a protective effect of alcohol appears too short a time for alcohol consumption to influence so strongly the long-term atherogenic processes, especially when the higher levels of alcohol consumption are found in the younger segments of the population. The peak positive correlation at 16 years is equally difficult to assign a biologic explanation, since it indicates an inducing or promoting effect of alcohol in ischemic heart disease mortality, which is not consistent with most of the published studies on the relationship of alcohol to heart disease. Finally, we are unable to identify a biologic reason or reasons to accommodate both the apparent protective effects of alcohol with a latency period of six years, and the apparent promoting effect with a latency of 16 years.

In sum, we have identified four models that encompass the possible time-series ecologic relationships. It should be apparent that the support or lack of support which these correlations might provide for a hypothesis can be entirely fortuitous. Although we recognize the contributions which informed observations of ecologic associations and trends can make toward the formulation of etiologic hypotheses, the use of these associations to quantify relationships or test hypotheses is invalid.

REFERENCES

1. Robinson WS. Ecological correlations and the behavior of individuals. *Am Sociol Rev* 1950;15:351-7.
2. Stavrakis KM. The role of ecologic analysis in studies of the etiology of *disease*: a discussion with reference to large bowel cancer. *J Chronic Dis* 1976;29:436-44.
3. Selvin S, Merrill D, Sacks ST. An alternative to ecologic regression analysis of mortality rates. *Am J Epidemiol* 1982;115:617-23.
4. Morgenstern H. Uses of ecologic analysis in epidemiologic research. *Am J Public Health* 1982;72:1336-4.4.
5. Brenner MH. Economic changes in heart disease mortality. *Am J Public Health* 1971; 61:606-11.

6. Bunn AR. Ischaemic heart disease mortality and the business cycle in Australia. *Am J Public Health* 1979;69:772-81.
7. LaPorte RE, Cresanta JL, Kuller LH. The relationship of alcohol consumption to atherosclerotic heart disease. *Prev Med* 1980;9:22-40.
8. Kul SV. Mortality and the business cycle: some questions about *research* strategies when utilizing macro-social and ecological data. *Am J Public Health* 1979;69:784-8.
9. Lew EA. Mortality and the business cycle: how far can we push an association. *Am J Public Health* 1979;69:782-3.
10. Statistics Canada. The control and sale of alcoholic beverages in Canada. Catalogue 63-202, Ottawa, 1918-1976.
11. Statistics Canada. Canadian statistical review. Catalogue 1.1-003, Ottawa, 1926-1976.
12. Nobel ED, ed. Third special report to the US Congress on alcohol and health, June 1978. Rockville, MD: National Institute on Alcohol Abuse and Alcoholism.
13. Statistics Canada. Vital statistics. Vol III, deaths. Catalogue 84-206, 1952-1976.
14. US Department of Health, Education, and Welfare. Report of working group on heart disease epidemiology. June, 1979. (NTH publication no. 79-1667).
15. Box GEP, Newbold P. Some comments on a paper of Coen, Comma, and Kendall. *J R Stat Soc, Series A* 1971;133:229-40.
16. Box GEP, Jenkins GM. Time series analysis forecasting and control. San Francisco: Holden-Day, 1970.