

# Latent Variable Modeling in Epidemiology

BENGT O. MUTHÉN, PH.D.

*Results of epidemiologic research can be misinterpreted because of imprecision in the measurement of the factors, or variables, involved. Latent variable modeling is a useful technique for avoiding such misinterpretations.*

An important function of epidemiologic research is to investigate the factors influencing the occurrence of disease. For example, one might study the effect of blood pressure on the risk of coronary heart disease, the effect of diet on the risk of breast cancer, or the effect of alcohol on the risk of liver disease. Results of such studies are essential for planning disease prevention strategies.

However, the results of such research are often misinterpreted because of imprecision in the measurement or estimation of the factors, or variables, involved. For example, results of a population survey might be analyzed to determine whether heavy alcohol consumption increases the probability of developing liver cirrhosis. The *predictor variable*, in this case the level of alcohol consumption, is subject to measurement error, because survey respondents may underreport their level of drinking.

A similar problem might arise when attempting to measure alcohol dependence (alcoholism). This variable cannot be observed directly, but must be estimated through observation of related vari-

ables. These might include any number of diagnostic criteria for alcoholism, such as giving up important activities in favor of drinking (see below).

In the above examples, the "true" value of the predictor variable is said to be latent, or hidden. The problem of latent variables may occur when a given variable cannot be observed directly, or contains measurement error. If latent variables are not carefully dealt with, they can distort conclusions from epidemiologic studies. *Latent variable modeling* is emerging as a useful technique for avoiding such distortions.

Important contributions to latent variable modeling have come out of the behavioral and social sciences, and such techniques are just beginning to be appreciated in epidemiologic and clinical studies (see, for example, Breslow 1988). Special uses of latent variable modeling are now emerging in fields as diverse as exposure assessment in air pollution epidemiology (Tosteson et al. 1989), nutritional epidemiology (Plummer and Clayton in press), developmental toxicity studies in laboratory animals (Catalano

and Ryan 1992), and twin studies of alcoholism, depression, and phobias (Kendler et al. 1992*a,b,c*).

This article discusses some uses of latent variable modeling in alcohol epidemiology. Problems associated with measurement error will be illustrated using computer-simulated data. Examples demonstrate how latent variable modeling can be used to represent indirectly observed variables. Examples of latent variable modeling for alcohol abuse and dependence will be discussed in detail using data from the Alcohol Supplement of the 1988 National Health Interview Survey (NHIS88; Massey et al. 1979).

## AN ILLUSTRATION OF THE EFFECTS OF MEASUREMENT ERROR

Measurement error may affect the results of studies that attempt to identify the risk factors for a disease. For example, such a study might be testing the hypothesis that high blood pressure increases the risk for coronary heart disease (CHD). Blood pressure in humans varies throughout the

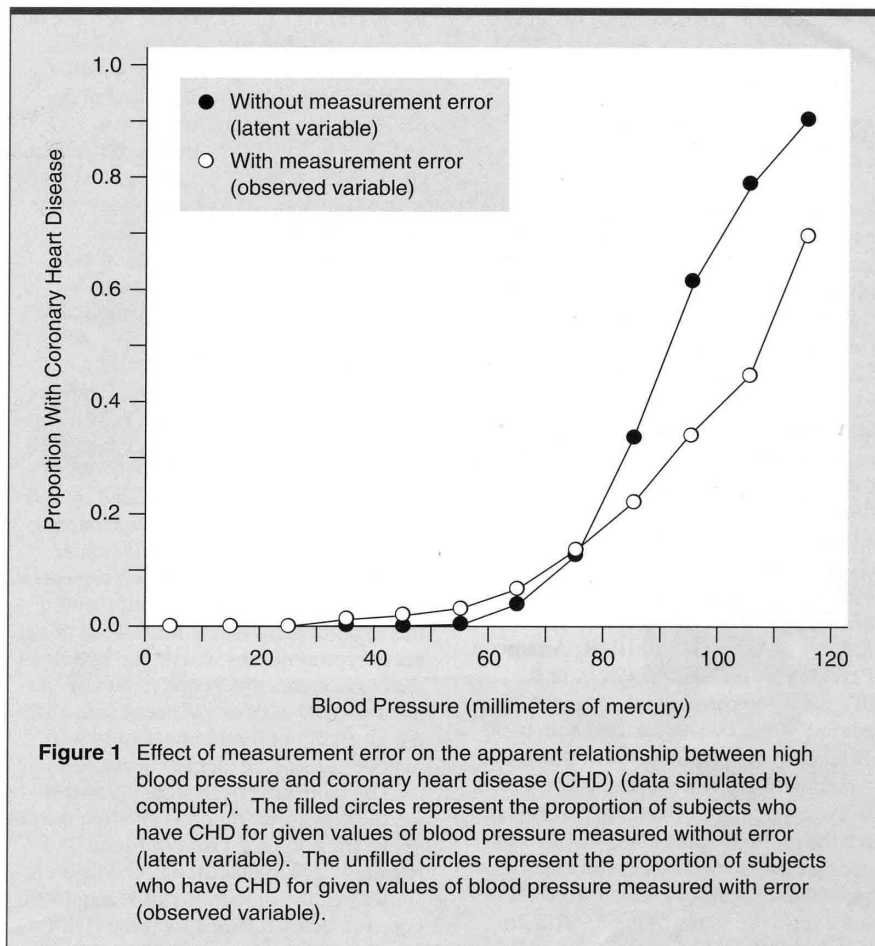


day and from day to day. Therefore, a given blood pressure measurement reflects just one point in time, and thereby contains measurement error. The true risk factor for CHD is the long-term average blood pressure; this is the latent variable in the following example.

To illustrate the effect of measurement error, a computer was used to generate simulated blood pressure data corresponding to a random sample of 10,000 people. Two variables representing blood pressure were generated, one containing no measurement error and one containing measurement error corresponding to a reliability of 0.67. The reliability of a variable describes the precision with which the variable is measured. Reliability values range from 0.0 to 1.0, with higher values representing greater precision. A reliability of 0.67 was chosen for this illustration, because it is typical of variables in many epidemiologic studies (see Stefanski and Carroll 1985; Willet et al. 1985).

Figure 1 shows two sets of points from the computer-simulated data. The filled circles represent the proportion of subjects who have CHD for given values of blood pressure measured without error (latent variable). The unfilled circles represent the proportion of subjects who have CHD for given values of blood pressure measured with error (observed variable). The observed values in figure 1 produce a flatter curve than do the true (latent) values; this represents a weakening of the apparent relationship between blood pressure and CHD. Thus, the risk for CHD in this simulation is grossly underestimated for individuals with high observed blood pressure. This is because some of the individuals with high observed blood pressure actually have lower true blood pressure, for which the risk of CHD is lower. In addition, the risk for CHD is slightly overestimated for individuals with low observed blood pressure, because some of these individuals have higher true blood pressure.

The amount of attenuation, or weakening, of the relationship seen in figure 1 depends on the reliability of the predictor variable (blood pressure): the lower the



**Figure 1** Effect of measurement error on the apparent relationship between high blood pressure and coronary heart disease (CHD) (data simulated by computer). The filled circles represent the proportion of subjects who have CHD for given values of blood pressure measured without error (latent variable). The unfilled circles represent the proportion of subjects who have CHD for given values of blood pressure measured with error (observed variable).

reliability, the greater the attenuation. However, measurement error does not always give rise to attenuation, but may inflate relationships when there is more than one predictor. For example, measurement error in one predictor may give rise to overestimation of the effect of a second predictor (Fuller 1991).

There are two types of measurement error—random error and systematic error. By definition, a random error is equally likely to add to or subtract from the true value. A systematic error is one that is more likely to either add to or subtract from the true value.

Blood pressure measurement is subject to random error. Recorded blood pressure may be either higher or lower than true average blood pressure because of day-to-day variation and because of random errors in the measuring instrument. True blood pressure is the latent variable in this case, and can be approximated by the long-term average of several followup blood pressure measurements (see MacMahon et al. 1990).

Examples of systematic error can be found in studies of the hypotheses that saturated fat intake influences the risk for breast cancer, and that alcohol consumption increases the risk for liver disease. Systematic error in these studies occurs when subjects who eat or drink heavily underreport, respectively, their dietary intake or alcohol consumption.

Systematic measurement error is handled by collecting additional information from a validation sample taken from the main study. For example, in the Nurses' Health Study (Rosner et al. 1989), a self-administered food frequency questionnaire was used to measure dietary fat intake in 89,538 women. Of these women, 173 were also asked to weigh and record their food intake for four 1-week periods at 3-month intervals over 1 year. When such data are available, the relationship between the two measures can be determined. This relationship was used to estimate the actual dietary fat intake from the food frequency questionnaire for subjects in the main study. Similar modeling

*BENGT O. MUTHÉN, PH.D., is a professor in the Graduate School of Education, University of California, Los Angeles.*

*This research was supported by grant AA08651 from the National Institute on Alcohol Abuse and Alcoholism.*



is described in Tosteson and colleagues (1989) and Carroll and Stefanski (1990).

### MULTIPLE INDICATORS OF LATENT VARIABLES

Latent variable modeling can be used in any situation where variables of interest are not directly observed, including situations where the variable must be estimated from a number of related variables (indicators). (For a classic reference, see Lord and Novick 1968.) The following examples include latent variable models with multiple indicators.

Consider a set of diagnostic criteria for measuring alcohol dependence, such as those formulated for the International Classification of Diseases, 10th revision (ICD-10; World Health Organization 1992), the *Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised* (DSM-III-R; American Psychiatric Association 1987), or the diagnostic options proposed for the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-IV; American Psychiatric Association 1992). In the following example, these criteria are the multiple indicators, and alcohol dependence is the unobserved latent variable, the status of which we want to infer from the status of the criteria. In order to formulate a trustworthy statistical model of how well the indicators measure the latent variable, it is necessary to translate theory into a statistical measurement model and then test how well the statistical model fits the observed data.

Such a statistical measurement model could, for example, draw on Edwards' (1986) conceptual model of the alcohol dependence syndrome, traditionally designated "alcoholism." The following discussion adapts Edwards' concepts to our illustration. According to Edwards, the syndrome "occurs with graded intensity"—that is, the manifestations of alcoholism can be ranked according to increasing severity. Thus, there is a single underlying continuum, or dimension, along which alcohol dependence becomes more severe. The syndrome may be recognized by the clustering of certain "elements." These elements can be interpreted here to represent the diagnostic criteria. Not all of the criteria need be present, or present in the same degree, to establish the diagnosis. However, the syndrome tends to manifest itself with greater clarity

as the criteria that are present increase in number and severity.

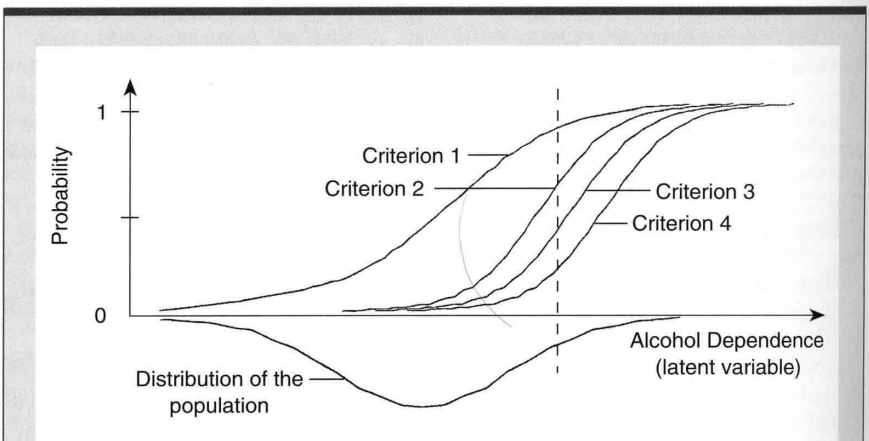
Edwards' concept of a single underlying dimension along which alcohol dependence becomes more severe is reflected in the DSM-III-R definition of alcohol dependence, which requires that at least three out of nine diagnostic criteria be fulfilled. This notion is also incorporated into DSM-III-R and DSM-IV, where severity modifiers of "mild," "moderate," and "severe" are applied to diagnoses of alcohol dependence, based on the number of criteria fulfilled.

One possible statistical translation of this theory is shown in figure 2. The horizontal axis represents the continuous variable of alcohol dependence, which corresponds to Edwards' "graded intensity." This is a latent variable, and cannot be measured directly. The bell-shaped curve below the horizontal axis represents the distribution of the population along this axis. In general, the far left tail of the curve represents abstainers, the middle of the curve represents people who use alcohol to a lesser or greater extent, and the far right tail represents people who may be considered alcohol dependent.

The indicators of the latent variable are the diagnostic criteria. The four curves above the horizontal axis in figure 2 represent four such criteria. The figure shows that the probability of a diagnostic criterion being fulfilled increases with increasing intensity. For an intensity value in the left tail or at the center of the

population distribution, the probability of a person fulfilling any criterion is very low. For values in the right tail, the probability of fulfilling all criteria is very high. This reflects Edwards' concept of clustering of elements, as well as his notion that the alcohol dependence syndrome shows increasing clarity with mounting intensity of alcohol dependence. The figure shows that even at a high value of the latent intensity variable, not all criteria are necessarily fulfilled. Subjects with a latent variable value at the dotted line would have a probability of less than 0.5 (50 percent) of fulfilling two of the four criteria. The fact that some criteria are not fulfilled at high intensity values may reflect pure measurement error (such as individual misunderstandings in the interview), recording errors, or idiosyncrasies of individual alcohol problems.

The criteria act as indicators of the latent variable in that the value of the latent variable is likely to be high as more criteria are fulfilled. Generally speaking, the measurement of the latent variable is improved when more criteria are used and when the relationship between each criterion and the latent variable is strong. The strength of the relationship can be described from figure 2—the steeper the curve, the stronger the relationship. Analysis of such relationships is the topic of factor analysis, a statistical technique frequently used in psychological studies and proposed for psychiatric epidemiology (Duncan-Jones et al. 1986).



**Figure 2** Computer-generated model illustrating the use of multiple indicators for estimating a latent variable. The horizontal axis represents the continuous variable of alcohol dependence. Below is the distribution of the hypothetical population for that variable. It is a latent variable, and cannot be measured directly. The indicators of the latent variable are the diagnostic criteria, four of which are shown.

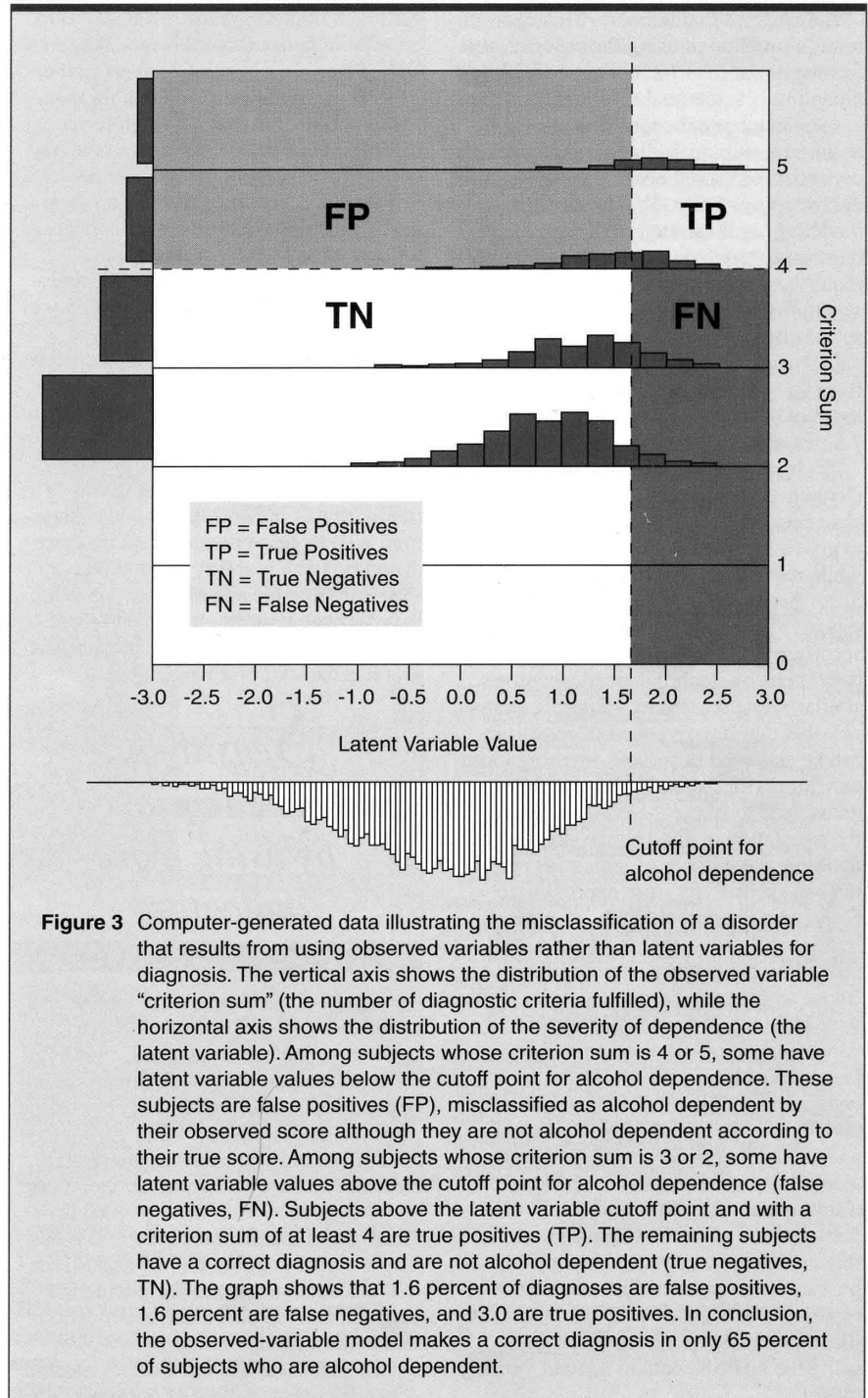


## MISCLASSIFICATION ERRORS IN DIAGNOSIS

Latent variable models such as the one just described can be used to study misclassification errors in diagnosis. A similar method is often used to evaluate medical tests. This is illustrated below using computer-simulated data representing 10,000 subjects generated according to a model such as the one illustrated in figure 2. Each subject has a latent variable value and a score of 0 or 1 on each of 11 criteria. Zero means that the criterion is not fulfilled, and 1 means that it is fulfilled. Although the data are artificial, the parameters determining the relationships between the criteria and the latent variable have been given realistic values. These values were obtained from a study of the Alcohol Supplement of the NHIS88, sponsored by the National Institute on Alcohol Abuse and Alcoholism (NIAAA; Muthén et al. in press). A more detailed simulation study was carried out by Muthén (in submission).

In one option proposed for DSM-IV, a diagnosis of alcohol dependence is made if at least 4 of 11 criteria are fulfilled. In the simulated data, 4.6 percent of the subjects are classified as alcohol dependent due to their criterion sum being 4 or more. These subjects are not, however, the same as those who represent the top 4.6 percent on the latent variable of alcohol dependence. Figure 3 shows the misclassification that results from using observed variables rather than latent variables for diagnosis. The vertical axis of figure 3 shows the distribution of the criterion sum for scores of 2 through 5, while the horizontal axis shows the distribution of the latent variable, reflecting severity of dependence.

Figure 3 shows that for the simulated subjects whose criterion sum is 4, some have latent variable values below the cutoff point for alcohol dependence represented by the vertical broken line. These subjects are "false positives"; that is, they are diagnosed as being alcohol dependent by their observed score although they are not alcohol dependent according to their true score. Similarly, among the subjects whose criterion sum is 3, some have latent variable values above the cutoff point for alcohol dependence, and thus constitute "false negatives." Subjects above the latent variable cutoff point and with a criterion sum of at least 4 are "true positives." In total for these data, 1.6 percent of subjects are false positives, 1.6



**Figure 3** Computer-generated data illustrating the misclassification of a disorder that results from using observed variables rather than latent variables for diagnosis. The vertical axis shows the distribution of the observed variable "criterion sum" (the number of diagnostic criteria fulfilled), while the horizontal axis shows the distribution of the severity of dependence (the latent variable). Among subjects whose criterion sum is 4 or 5, some have latent variable values below the cutoff point for alcohol dependence. These subjects are false positives (FP), misclassified as alcohol dependent by their observed score although they are not alcohol dependent according to their true score. Among subjects whose criterion sum is 3 or 2, some have latent variable values above the cutoff point for alcohol dependence (false negatives, FN). Subjects above the latent variable cutoff point and with a criterion sum of at least 4 are true positives (TP). The remaining subjects have a correct diagnosis and are not alcohol dependent (true negatives, TN). The graph shows that 1.6 percent of diagnoses are false positives, 1.6 percent are false negatives, and 3.0 percent are true positives. In conclusion, the observed-variable model makes a correct diagnosis in only 65 percent of subjects who are alcohol dependent.

percent are false negatives, and 3.0 percent are true positives. The observed variable model diagnoses only 65 percent of those who are alcohol dependent. The conclusion of figure 3 is that the sensitivity (the percentage of cases diagnosed correctly) for this hypothetical DSM-IV diagnosis can be as low as 65 percent in a

general population survey. In other words, 35 percent of the cases may be missed.

Figure 3 demonstrates that people classified as alcohol dependent using the criterion sum form a heterogeneous group. Individuals with a criterion sum of 4 vary widely in their values on the latent variable, in fact covering about one-half



of the range of that variable. This illustrates a problem in using the criterion sum for diagnosis, for which latent variable modeling is a potential solution.

The number of criteria that should be required to support a diagnosis of alcohol dependence is a subject of debate among the framers of DSM-IV. Latent variable modeling, as illustrated in figure 2, is relevant for such decisions. For example, requiring a criterion sum of 3 increases sensitivity from 65 to 79 percent, but specificity (the percentage of nondependent subjects correctly diagnosed) drops from 98 to 96 percent, and the true prevalence of 4.6 percent is overestimated as 7.8 percent.

If, instead, one makes the definition of alcohol dependence more inclusive by changing the latent variable cutoff point to give a true prevalence of 7.8 percent while requiring a criterion sum of at least 3, correct prevalence is obtained, but sensitivity is again low (64 percent), and specificity is 97 percent. Using stricter definitions of alcohol dependence gives similarly low sensitivity values. It seems possible that improvement of sensitivity can be achieved in surveys by using more indicators (for example, the symptom items used to create the criteria). Using the simulation method above suggests that doubling the number of indicators could raise the sensitivity from approximately 65 percent to approximately 85 percent.

Figure 3 shows that most individuals missed by the criterion sum definition (those who have a criterion sum less than 4, or less than 3) have latent variable values just below the cutoff point on the latent dimension. This illustrates the consequence of the strict separation of diagnoses into "not alcohol dependent" and "alcohol dependent." If, however, alcohol dependence can be represented as a continuum, as in figures 2 and 3, then there may be no natural cutoff point on this continuum. In fact, for many purposes there is no need for such a strict separation of diagnoses. These points are illustrated by some real data analyses using the NHIS88 data in a latent variable model.

### LATENT VARIABLE ANALYSIS OF REAL DATA

The appropriateness of a latent variable model should be tested against real data. For example, one may ask if the single latent variable dimension of figure 2 is

sufficient to describe the responses to the criteria. In fact, testing this model against NHIS88 data for current drinkers in the general population showed that this was inappropriate, but that a two-dimensional model fit the data well (Muthén et al. in press; Muthén in submission). This indicates that alcohol dependence is not just a more severe form of alcohol abuse, but reflects a separate phenomenon, or dimension.

In a study by Muthén and colleagues (in press), the first, milder dimension was interpreted as alcohol abuse and was measured well by alcohol-related criteria with relatively high prevalences in the population: drinking more or longer than intended, and drinking in situations in which it is physically hazardous. The second, more severe, dimension was interpreted as dependence and was measured well by less prevalent criteria corresponding to an inability to cut down or stop drinking, the abandonment or reduction of other activities in favor of drinking, and drinking despite the recognition that it is causing problems in one's life.

## *Sensitivity can be improved by using more indicators.*

The analyses showed that the less severe alcohol-related problems of the first dimension and the more severe problems of the second dimension do not represent opposite ends of the same axis, but define phenomena of a distinct kind.

This model represents a different structure of the concepts of alcohol abuse and dependence from that assumed in DSM-III-R and the proposed DSM-IV. The analyses suggest that the definitions in these documents blur the distinction between alcohol abuse and dependence due to the overlap in their criteria. For example, these documents define alcohol dependence using some of the same criteria that measure alcohol abuse.

The values of the latent variables representing alcohol dependence and alcohol abuse can be estimated from subjects' responses to questions related to diagnostic criteria. These responses can be plotted using the continuum of abuse as one axis and the continuum of dependence as the other axis. Such a plot, based on 18,244

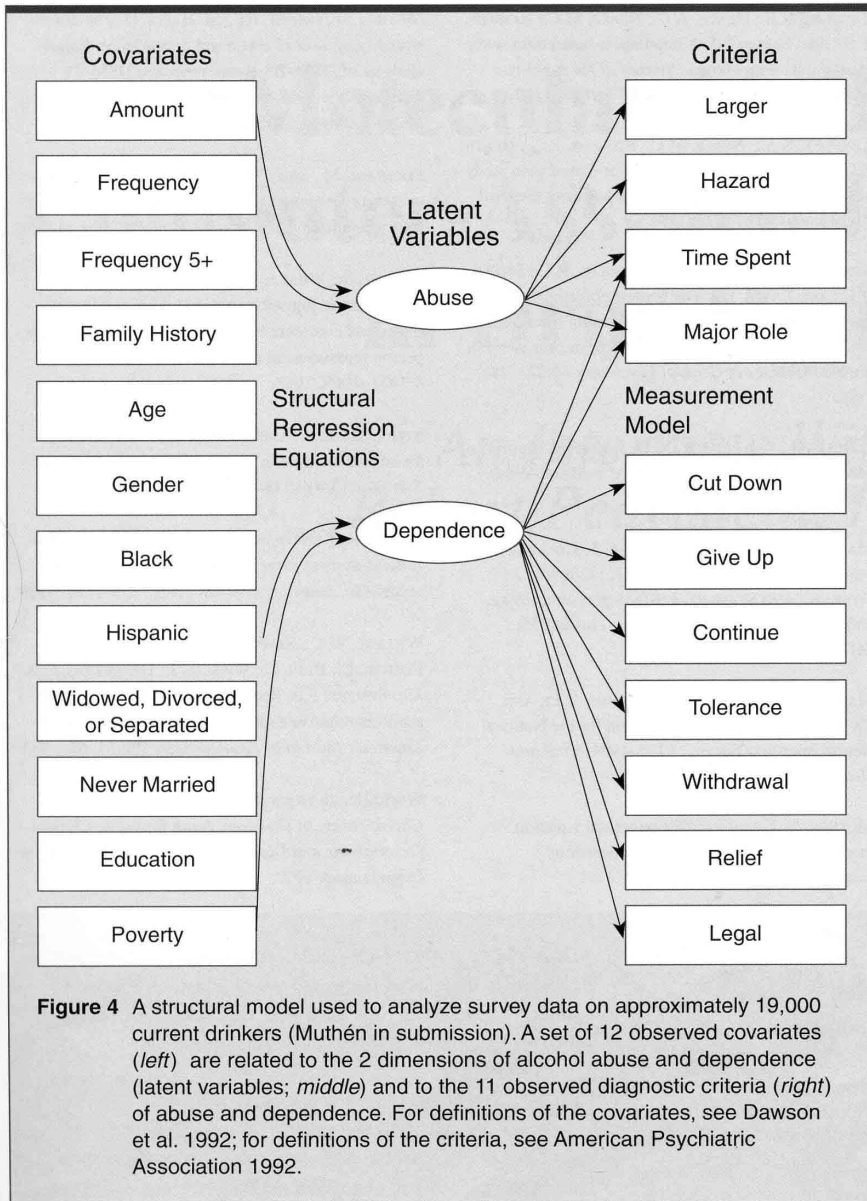
white, current drinkers in the NHIS88, showed no clear cutoff point for either dimension that can be used to define alcohol abuse or dependence.

### STRUCTURAL MODELING WITH LATENT VARIABLES

The scores of the two dimensions can be related to variables such as the respondent's alcohol consumption, age, and gender. This type of analysis can be carried out by extending the two-dimensional latent variable model to include covariates, observed variables that are assumed to be related to the criteria and their latent variables. This latent variable approach has the advantage of not forcing a choice of cutoff point on the sum of criteria and classifying all subjects as either "nondependent" or "dependent" individuals. Thereby, the misclassification problems demonstrated earlier are avoided. Such statistical modeling is referred to as "structural equation modeling" or "covariance structure modeling." This is a general framework for modeling, encompassing the special cases we have considered so far. For an overview, see Joreskog and Sorbom (1979), Muthén (1983), and Bollen (1989).

For example, figure 4 shows a structural model used to analyze NHIS88 data on nearly 19,000 current drinkers (Muthén in submission). Here a set of 12 covariates were related to the 2 dimensions and their 11 criterion indicators. The observed covariates are shown on the left, including alcohol consumption variables, family history of alcoholism (for definitions, see Dawson et al. 1992), age, gender, ethnicity, and various other sociodemographic characteristics of the respondents. To the right in the figure are the 11 observed diagnostic criteria. In the middle are the two latent variables, designated "abuse" and "dependence." The relationships between the criteria and the latent variables define the measurement model and are described in terms similar to those used for figure 2. If this part of the model alone were present, the statistical approach would correspond to factor analysis. The relationships between the latent variables and the covariates are, however, the ones of primary interest. These relationships are called structural equations.

The structural analysis results of Muthén (in submission) show that the two latent variables have different relation-



**Figure 4** A structural model used to analyze survey data on approximately 19,000 current drinkers (Muthén in submission). A set of 12 observed covariates (*left*) are related to the 2 dimensions of alcohol abuse and dependence (latent variables; *middle*) and to the 11 observed diagnostic criteria (*right*) of abuse and dependence. For definitions of the covariates, see Dawson et al. 1992; for definitions of the criteria, see American Psychiatric Association 1992.

ships to the covariates. Positive family history of alcoholism among first-, second-, or third-degree relatives is strongly related to the more severe dimension of dependence, even when controlling for other variables such as alcohol consumption. This effect is considerably stronger for dependence than for abuse. Furthermore, age and being Hispanic are found to have different relationships with the two dimensions. Increasing age reduces the value of abuse more than dependence. Being Hispanic is associated with lower values on the abuse dimension but higher values on the dependence dimension.

### UNTAPPED POTENTIAL OF LATENT VARIABLE MODELING

The structural model shown in figure 4 exemplifies a general multivariate modeling framework that encompasses many other statistical formulations not discussed here. Latent variable models can be used to study the classification and causes of alcohol disorders, to analyze the progression of alcohol problems, to study the co-occurrence of alcohol dependence and depression, and to study the genetic susceptibility to alcohol dependence. Latent variable modeling has great potential also in genetic model-

ing, especially since the observed variables that indicate the presence of alcohol dependence are based on multiple indicators from interviews.

Latent variable modeling has important applications in two important data sets of NIAAA: the ongoing National Longitudinal Alcohol Epidemiological Survey (NLAES), and the National Collaborative Studies on Genetics of Alcoholism (COGA). In the NLAES, about 45,000 subjects are being studied over a period of 2 years. The survey is collecting extensive information regarding family history of alcoholism, alcohol consumption levels, and alcohol-related problems. In the COGA, about 600 alcoholics and their family members are being studied (see Holden 1991). COGA is also collecting information for genetic analyses.

In addition to providing tools for these complex epidemiologic tasks, the latent variable modeling framework can provide better descriptive statistics for alcohol problems. It might improve the estimation of alcoholism prevalence from survey data, particularly when differences in prevalence are being estimated among population subgroups; subgroup prevalence can be estimated not only from criterion information but also from covariate information (such as alcohol consumption; see figure 4). This type of approach is well established in other types of national surveys, for example in the education field.

In summary, latent variable modeling is useful in situations where random or systematic measurement error is a problem, where phenomena under study are not directly observed, and where multiple indicators are needed to describe various aspects of a phenomenon. ■

### REFERENCES

- American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised*. Washington DC: the Association, 1987.
- American Psychiatric Association. *DSM-IV Options Book*. Washington, DC: the Association, 1992.
- BOLLEN, K. *Structural Equations With Latent Variables*. New York: Wiley and Sons, 1989.
- BRESLOW, N. Chairman's introduction to the special issue: Proceedings of the workshop on methods for longitudinal data analysis in epidemiological and clinical studies. *Statistics in Medicine* 7:7, 1988.



CARROLL, R.J., AND STEFANSKI, L.A. Approximate quasi-likelihood estimation in models with surrogate predictors. *Journal of the American Statistical Association* 85(411):652-663, 1990.

CATALANO, P.J., AND RYAN, L.M. Bivariate latent variable models for clustered discrete and continuous outcomes. *Journal of the American Statistical Association* 87(419):651-658, 1992.

DAWSON, D.A.; HARFORD, T.C.; AND GRANT, B. Family history as a predictor of alcohol dependence. *Alcoholism: Clinical and Experimental Research* 16(3):572-575, 1992.

DUNCAN-JONES, P.; GRAYSON, D.A.; AND MORAN, P.A.P. The utility of latent trait models in psychiatric epidemiology. *Psychological Medicine* 16:391-405 1986.

EDWARDS, G. The alcohol dependence syndrome: A concept as stimulus to enquiry. *British Journal of Addiction* 81:171-183, 1986.

FULLER, W.A. Regression estimation in the presence of measurement error. In: Biemer, P.P.; Groves, R.M.; Lyberg, L.E.; and Mathiowetz, N.A., eds. *Measurement Errors in Surveys*. New York: Wiley and Sons, 1991. pp. 617-635.

HOLDEN, C. Probing the complex genetics of alcoholism. *Science* 251(4990):163-164, 1991.

JORESOG, K.G., AND SORBOM, D. *Advances in Factor Analysis and Structural Equation Models*. Cambridge, MA: Abt Books, 1979.

KENDLER, K.S.; HEATH, A.C.; NEALE, M.C.; KESSLER, R.C.; AND EAVES, L.J. A population-based twin study of alcoholism in women. *Journal of the American Medical Association* 268(14):1877-1882, 1992a.

KENDLER, K.S.; NEALE, M.C.; KESSLER, R.C.; HEATH, A.C.; AND EAVES, L.J. A population-based twin study of major depression in women. *Archives of General Psychiatry* 49:267-272, 1992b.

KENDLER, K.S.; NEALE, M.C.; KESSLER, R.C.; HEATH, A.C.; AND EAVES, L.J. The genetic epidemiology of phobias in women. The interrelationship of agoraphobia, social phobia, situational phobia, and simple phobia. *Archives of General Psychiatry* 49:273-281, 1992c.

LORD, F.M., AND NOVICK, M.R. *Statistical Theories of Mental Test Scores*. Reading, MA: Addison-Wesley, 1968.

MACMAHON, S.; PETO, R.; CUTLER, J.; COLLINS, R.; SORLIE, P.; NEATON, J.; ABBOTT, R.; GODWIN, J.; DYER, A.; AND STAMLER, J. Blood pressure, stroke, and coronary heart disease. Part 1. *Lancet* 335: 765-774, 1990.

MASSEY, J.T.; MOORE, T.F.; PARSONS, V.L.; AND TADORS, W. Design and estimation for the National Health Interview Survey, 1985-1994. *Vital and Health Statistics* 2:1-33, 1979.

MUTHÉN, B. Latent variable structural equation modeling with categorical data. *Journal of Econometrics* 22:48-65, 1983.

MUTHÉN, B.; GRANT, B.; AND HASIN, D. The dimensionality of alcohol abuse and dependence: Factor analysis of DSM-III-R and proposed DSM-IV criteria in the 1988 National Health Interview Survey. *British Journal of Addiction*, in press.

PLUMMER, M., AND CLAYTON, D. Measurement error in dietary assessment: An investigation using covariance structure models. *Statistics in Medicine*, in press.

ROSNER, B.; WILLET, W.C.; AND SPIEGELMAN, D. Correction of logistic regression relative risk estimates and confidence intervals for systematic within-person measurement error. *Statistics in Medicine* 8:1051-1069, 1989.

STEFANSKI, L.A., AND CARROLL, R.J. Covariate measurement error in logistic regression. *Annals of Statistics* 13(4):1335-1351, 1985.

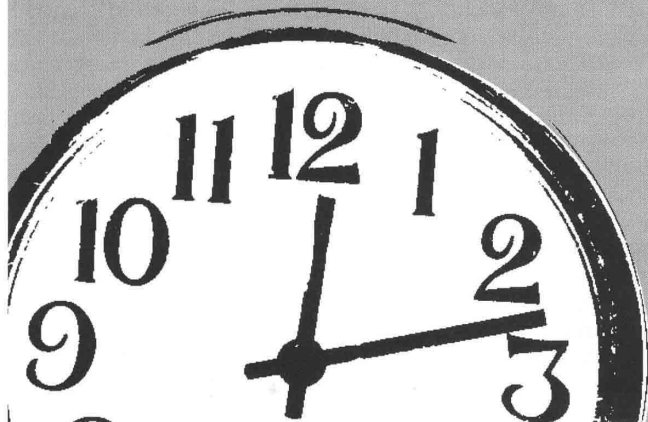
TOSTESON, T.; STEFANSKI, L.A.; AND SCHAFER, D.W. A measurement error model for binary and ordinal regression. *Statistics in Medicine* 8:1139-1147, 1989.

WILLETT, W.C.; SAMPSON, L.; STAMPFER, M.J.; ROSNER, B.; BAIN, C.; WITSCHI, J.; HENNEKENS, C.H.; AND SPEIZER, F.E. Reproducibility and validity of a semi-quantitative food frequency questionnaire. *American Journal of Epidemiology* 122:51-65, 1985.

World Health Organization. *International Classification of Diseases, Tenth Revision. Clinical Descriptions and Diagnostic Guidelines*. Geneva: the Organization, 1992.

## Alcohol Information in Minutes Not Days!

NIAAA Offers Electronic  
Resources for Anyone with  
Access to a Computer and Modem



Authoritative statistics on alcohol use and alcohol-related problems in the United States are yours in seconds through Quick Facts, the electronic bulletin board of NIAAA's Division of Biometry and Epidemiology.

Quick Facts is accessible without charge, and offers data files on alcohol consumption and patterns of use, alcohol-related morbidity and mortality, public attitudes, and other topics that meet epidemiologic research needs of the public health community.

Quick Facts also features announcements of general interest from NIAAA and other Quick Facts users. Users can receive instructions on use of the index, table of contents, and other features when connected to the system.

Join the alcohol epidemiology network today by contacting Quick Facts, c/o Cygnus Corporation, 1400 Eye Street, NW, Washington, DC 20005, 202-289-4992 (voice), 202-289-4112 (modem).

**QUICK FACTS  
For No-Cost Access  
to Alcohol Epidemiology**